Methods in Molecular Biology 2540

Springer Protocols

Christian Dahmann *Editor*

Drosophila

Methods and Protocols

Third Edition

MOREMEDIA

Humana Press

METHODS IN MOLECULAR BIOLOGY

Series Editor
John M. Walker
School of Life and Medical Sciences
University of Hertfordshire
Hatfield, Hertfordshire, UK

For further volumes:
http://www.springer.com/series/7651

For over 35 years, biological scientists have come to rely on the research protocols and methodologies in the critically acclaimed *Methods in Molecular Biology* series. The series was the first to introduce the step-by-step protocols approach that has become the standard in all biomedical protocol publishing. Each protocol is provided in readily-reproducible step-by-step fashion, opening with an introductory overview, a list of the materials and reagents needed to complete the experiment, and followed by a detailed procedure that is supported with a helpful notes section offering tips and tricks of the trade as well as troubleshooting advice. These hallmark features were introduced by series editor Dr. John Walker and constitute the key ingredient in each and every volume of the *Methods in Molecular Biology* series. Tested and trusted, comprehensive and reliable, all protocols from the series are indexed in PubMed.

Drosophila

Methods and Protocols

Third Edition

Edited by

Christian Dahmann

Institute of Genetics, Technische Universität Dresden, Dresden, Germany

Humana Press

Editor
Christian Dahmann
Institute of Genetics, Technische
Universität Dresden
Dresden, Germany

ISSN 1064-3745 ISSN 1940-6029 (electronic)
Methods in Molecular Biology
ISBN 978-1-0716-2540-8 ISBN 978-1-0716-2541-5 (eBook)
https://doi.org/10.1007/978-1-0716-2541-5

This Humana imprint is published by the registered company Springer Science+Business Media, LLC, part of Springer Nature.
The registered company address is: 1 New York Plaza, New York, NY 10004, U.S.A.

Preface

The fruit fly *Drosophila melanogaster* is a long-standing model organism in biomedical research. Research using *Drosophila* continuous to provide valuable contributions to our understanding of the genetics, development, physiology, and behavior of multi-cellular organisms and of human diseases. Scientific advance is also driven by technical and methodological inventions. This third edition of "Drosophila: Methods and Protocols" covers some of the recent methodological improvements and inventions that drosophilists have made in recent years. The book starts with a description of FlyBase, a comprehensive database of *Drosophila* genes and genomes, methods to experimentally achieve precise spatiotemporal gene expression, and methods to analyze microRNA function and single-cell transcriptome data. The next five chapters describe recent CRISPR-based methods for precise genome editing, transcriptional activation, and cell lineage tracing. Methods for studying protein function and for tagging proteins are then followed by protocols to control cell and tissue mechanics and to measure mechanical properties of tissues. The final chapters then detail methods to prepare, image, and quantitatively analyze tissues, and to study metabolism.

I thank all authors for their expert contributions to this book and hope that the book provides a valuable resource for researchers using *Drosophila* as a model organism.

Dresden, Germany ***Christian Dahmann***

Contents

Contributors

Markus Affolter • *Biozentrum der Universität Basel, Basel, Switzerland*
Gustavo Aguilar • *Biozentrum der Universität Basel, Basel, Switzerland*
Benoit Aigouy • *Aix Marseille Univ, CNRS, IBDM, Marseille, France*
Jerome Avellaneda • *Aix Marseille University, CNRS, IBDM, Turing Centre for Living Systems, Marseille, France*
Josephine Bageritz • *Centre for Organismal Studies Heidelberg (COS), Universität Heidelberg, Heidelberg, Germany*
Milena Bauer • *Biozentrum der Universität Basel, Basel, Switzerland*
Klaus Becker • *Department for Bioelectronics, FKE, Vienna University of Technology, Vienna, Austria*
Justin A. Bosch • *Department of Genetics, Blavatnik Institute, Harvard Medical School, Boston, MA, USA*
Michael Boutros • *Division Signaling and Functional Genomics, German Cancer Research Center (DKFZ) and Heidelberg University, Heidelberg, Germany*
Richard W. Carthew • *Department of Molecular Biosciences, Northwestern University, Evanston, IL, USA; NSF Simons Center for Quantitative Biology, Northwestern University, Evanston, IL, USA*
Stefano De Renzis • *European Molecular Biology Laboratory (EMBL), Developmental Biology Unit, Heidelberg, Germany*
Hans-Ulrich Dodt • *Section for Bioelectronics, Center for Brain Research, Medical University of Vienna, Vienna, Austria; Department for Bioelectronics, FKE, Vienna University of Technology, Vienna, Austria*
Xizhi Dong • *Gene Regulatory Lab, School of Medicine, Tsinghua University, Beijing, China*
Natalie A. Dye • *Mildred Scheel Nachwuchszentrum (MSNZ) P2 & Cluster of Excellence Physics of Life, Technische Universität Dresden, Dresden, Germany*
Elisabeth Fischer-Friedrich • *Biotechnology Center, Technische Universität Dresden, Dresden, Germany; Cluster of Excellence Physics of Life, Technische Universität Dresden, Dresden, Germany*
Orsolya Fölsz • *Department of Biosciences, Durham University, Durham, UK*
Jorge Garcia-Marques • *Centro Nacional de Biotecnologia, Consejo Superior de Investigaciones Cientificas, Madrid, Spain*
Jörg Großhans • *Department of Biology, Philipps University, Marburg, Germany*
Yuting Han • *Gene Regulatory Lab, School of Medicine, Tsinghua University, Beijing, China*
Helen Hertenstein • *Department of Biology, Institute of Zoology, Technische Universität Dresden, Dresden, Germany*
Thomas Hummel • *Department of Neurobiology, University of Vienna, Vienna, Austria*
Victoria K. Jenkins • *The Biological Laboratories, Harvard University, Cambridge, MA, USA*
Deqing Kong • *Department of Biology, Philipps University, Marburg, Germany*

Daniel Krueger • *European Molecular Biology Laboratory (EMBL), Developmental Biology Unit, Heidelberg, Germany; Hubrecht Institute, Royal Netherlands Academy of Arts and Sciences (KNAW), Utrecht, The Netherlands*
Aoife Larkin • *Department of Physiology, Development and Neuroscience, University of Cambridge, Cambridge, UK*
Tzumin Lee • *Janelia Research Campus, Howard Hughes Medical Institute, Ashburn, VA, USA*
Katarzyna Lepeta • *Biozentrum der Universität Basel, Basel, Switzerland*
Xiaochen Li • *Gene Regulatory Lab, School of Medicine, Tsinghua University, Beijing, China*
Yutong Li • *Gene Regulatory Lab, School of Medicine, Tsinghua University, Beijing, China*
Chun-Chieh Lin • *Department of Pathology and Laboratory Medicine, Giesel School of Medicine, Dartmouth-Hitchcock Medical Center, Lebanon, NH, USA*
Qingfei Liu • *Gene Regulatory Lab, School of Medicine, Tsinghua University, Beijing, China*
Xinyi Lu • *Gene Regulatory Lab, School of Medicine, Tsinghua University, Beijing, China*
Shinya Matsuda • *Biozentrum der Universität Basel, Basel, Switzerland*
Ellen McMullen • *Department of Molecular Biology and Genetics, Faculty of Science, University of South Bohemia, České Budějovice, Czech Republic*
Sromana Mukherjee • *Department of Biology, Indiana University, Bloomington, IN, USA*
Stephan Müller • *Department of Biology, Institute of Zoology, Technische Universität Dresden, Dresden, Germany*
Maria Jelena Musillo • *Centre for Organismal Studies Heidelberg (COS), Universität Heidelberg, Heidelberg, Germany*
Jian-Quan Ni • *Gene Regulatory Lab, School of Medicine, Tsinghua University, Beijing, China*
Kevin G. Nyberg • *Department of Molecular Biosciences, Northwestern University, Evanston, IL, USA*
Achim Paululat • *University of Osnabrück, Department of Zoology and Developmental Biology, Osnabrück, Germany*
Marko Pende • *Mount Desert Island Biological Laboratory (MDIBL), Salisbury Cove, ME, USA; Section for Bioelectronics, Center for Brain Research, Medical University of Vienna, Vienna, Austria*
Norbert Perrimon • *Department of Genetics, Blavatnik Institute, Harvard Medical School, Boston, MA, USA; Howard Hughes Medical Institute, Chevy Chase, MD, USA*
Fillip Port • *Division Signaling and Functional Genomics, German Cancer Research Center (DKFZ) and Heidelberg University, Heidelberg, Germany*
Christopher J. Potter • *The Solomon H. Snyder Department of Neuroscience, Johns Hopkins University School of Medicine, Baltimore, MD, USA*
Benjamin Prud'homme • *Aix Marseille Univ, CNRS, IBDM, Marseille, France*
Olympia-Ekaterini Psathaki • *University of Osnabrück, Department of Zoology and Developmental Biology, Osnabrück, Germany; University of Osnabrück, Center of Cellular Nanoanalytics, Integrated Bioimaging Facility Osnabrück (iBiOs), Osnabrück, Germany*
Yuhao Qiu • *Gene Regulatory Lab, School of Medicine, Tsinghua University, Beijing, China*
Olena Riabinina • *Department of Biosciences, Durham University, Durham, UK*

Saiedeh Saghafi • *Department for Bioelectronics, FKE, Vienna University of Technology, Vienna, Austria*
Pablo Sanchez Bosch • *Stanford University School of Medicine, Stanford, CA, USA*
Karla Yanín Guerra Santillán • *Institute of Genetics, Technische Universität Dresden, Dresden, Germany; Biotechnology Center, Technische Universität Dresden, Dresden, Germany; Cluster of Excellence Physics of Life, Technische Universität Dresden, Dresden, Germany*
Stefanie Schirmeier • *Department of Biology, Institute of Zoology, Technische Universität Dresden, Dresden, Germany*
Frank Schnorrer • *Aix Marseille University, CNRS, IBDM, Turing Centre for Living Systems, Marseille, France*
Xu Si • *Gene Regulatory Lab, School of Medicine, Tsinghua University, Beijing, China*
Nicholas Sokol • *Department of Biology, Indiana University, Bloomington, IN, USA*
Darya Task • *Department of Biology, Johns Hopkins University, Baltimore, MD, USA*
Jim Thurmond • *Department of Biology, Indiana University, Bloomington, IN, USA*
Uwe Töpfer • *Institute of Genetics, Technische Universität Dresden, Dresden, Germany*
Daiki Umetsu • *Graduate School of Life Sciences, Tohoku University, Sendai, Japan*
M. Alessandra Vigano • *Biozentrum der Universität Basel, Basel, Switzerland*
Schayan Yousefian • *Berlin Institute of Health (BIH) at Charité – Universitätsmedizin Berlin, Berlin, Germany; Charité-Universitätsmedizin, Berlin, Germany; Berlin Institute for Medical Systems Biology, Max Delbrück Center for Molecular Medicine in the Helmholtz Association, Berlin, Germany*

Chapter 1

Using FlyBase: A Database of *Drosophila* Genes and Genetics

Victoria K. Jenkins, Aoife Larkin, Jim Thurmond, and The FlyBase Consortium

Abstract

Since 1992, FlyBase has provided a freely available online database of information about the model organism *Drosophila melanogaster*. Data in FlyBase is curated manually from research papers as well as computationally from a variety of relevant sources, to serve as an information hub that enables and accelerates research discovery. This chapter aims to give users new to the database an overview of the layout and types of data available, as well as introducing some tools with which to access the data. More experienced users will find useful information about recent improvements and descriptions to enable more efficient navigation of the database.

Key words FlyBase, *Drosophila*, Fly, Database, Genome, Gene

1 Introduction

FlyBase is the online genomic resource for the fruit fly *Drosophila melanogaster*. As a Model Organism Database (MOD), FlyBase is responsible for the annotation and maintenance of the genome and curating data generated by the *Drosophila* research community. This includes information from published papers as well as datasets, analysis tools, correspondence sent directly to FlyBase, pre-publication data repositories, and more.

As the importance and scope of biological inquiry using *Drosophila melanogaster* grew over the twentieth century, the need to standardize and collect knowledge first in books and then in an online database [1] became evident. The modern-day FlyBase Consortium is an international, grant-supported project bringing together researchers at four main sites, plus a variety of

The members of the FlyBase Consortium are listed in the Acknowledgements.

Christian Dahmann (ed.), *Drosophila: Methods and Protocols*,
Methods in Molecular Biology, vol. 2540, https://doi.org/10.1007/978-1-0716-2541-5_1,

collaborators at other genomic and biomedical databases and projects. FlyBase is proud to be a founding member of the Alliance for Genome Resources [2], a multi-model organism coalition providing access to comparative genomics at https://www.alliancegenome.org. At FlyBase, we continually work to improve the FAIRness (Findability, Accessibility, Interoperability, Reusability) of our data and tools [3]. A recent change in our scope is to focus our curation efforts on *Drosophila melanogaster* only. All other *Drosophila* species' genomes are hosted and maintained by NCBI (https://www.ncbi.nlm.nih.gov/genome/annotation_euk/all).

This chapter is designed to help all users of FlyBase gain a better understanding of the wide range of data, tools, and information that FlyBase provides. While most readers are familiar with the aspects of FlyBase that they use regularly, they may not be aware of other features that could assist their research and enrich their participation in the *Drosophila* research community. By walking through a variety of FlyBase reports, beginning with the Gene Report, we hope to provide both overall guidance to those new to FlyBase, and specific details that may help even the experienced user. For a quick glance at many of the tools and features at FlyBase, please *see* Table 1. For a more in-depth look at some of FlyBase's features that is beyond the scope of this chapter, please see our recent publication in Genetics [4].

2 Overview

The FlyBase home page is designed for quick access to the most frequently used tools and utilities on the site. A panel of button icons at the top are one-click links to workhorse destinations within FlyBase, like the JBrowse genome viewer, BLAST, and the Batch Download tool. The sidebar has icon links to external resources of interest to *Drosophila* researchers and is headed by the FlyBase Twitter feed. A commentary panel at the bottom of the page displays a rotating, frequently updated selection of links to brief time-sensitive news articles. In the center of the page is QuickSearch, the main FlyBase search engine (*see* Subheading 3).

The NavBar (navigation menu bar) at the top of the home page appears on every page of the FlyBase site. On the right side, a search bar input box can be toggled between two methods of search: Jump to Gene/J2G and Search. When J2G is selected and a gene name or symbol is entered, clicking the Go button goes directly to the Gene Report for that gene. It will also jump straight to any inputted FlyBase data class term (*see* Subheading 3 and Table 2). This search has some flexibility (uses synonyms and case variations), but it always tries to go to a report page. When Search is selected, the

Table 1
Major interface tools at FlyBase

Category	Tool name	Function/Usage
Query tools and Portals	QuickSearch	Targeted searches on various data classes
	Vocabularies	Search or browse all controlled vocabularies used to annotate records
	Query Builder	Advanced field-by-field search for most data classes
	Interactions Browser	Explore genetic and physical interactions via static images
	ImageBrowse	Browse through *Drosophila* images by organ system, life cycle, tagma, or germ layer
	RNA-Seq Similarity	Find genes with expression similar to a query gene across modENCODE developmental stage, tissue, treatment, or cell line RNA-Seq data
Genomic and Map tools	BLAST	NCBI BLAST for finding nucleotide and protein sequences in *Drosophila* and dozens of related species
	JBrowse	Graphical representation of *D. melanogaster* genome features
	Cytosearch	Search for genetic objects mapped via cytology-based data
	Feature Mapper	Search for sequence-mapped features that overlap a specific region or gene
	RNA-Seq Profile	Find genes with specific patterns of expression across modENCODE developmental stage, tissue, treatment, or cell line RNA-Seq data
	RNA-Seq by Region	Evaluate the expression levels of exons, introns and/or intergenic regions from modENCODE developmental and tissue RNA-Seq data
Retrieve and Convert tools	Batch Download	Bulk download of individual report fields, FASTA or XML files in a variety of formats
	Sequence Downloader	Fetch nucleotide or peptide sequences for regions or genes
	Upload/ Validate IDs	Import lists of IDs for conversion or updating, then export or download list
	Coordinates Converter	Convert coordinates between genome releases (e.g., *D. melanogaster* R5 to R6)

field recapitulates the powerful Search FlyBase tab in the QuickSearch panel. This means such searches can be initiated from any page in the FlyBase website.

3 QuickSearch

QuickSearch (Fig. 1) is actually several different search engines, offering a variety of views into the vast FlyBase database [5]. The panel has tabs at the top; use these to search a specific type of data (e.g., Homologs, GO, References) or choose the Search FlyBase tab to see combined results from all of the more than 30 types of data indexed by FlyBase curators. Some QuickSearch tabs return

Table 2
Data classes used by FlyBase, with examples

Data Class	Description	Selected subtype(s)	Example	Total items in FlyBase
FBab	Aberration		In(2R)Cy	23,594
FBal	Allele		shg^{2}	265,453
FBba	Balancer		TM6B	613
FBbt	Anatomy		arista	16,643
FBcl	Clones	cDNA clone, genomic clone	AT22870	772,972
FBcv	Internal controlled vocabularies	allele class, phenotypic class, origin of mutation, publication descriptor	review	1101
GO, DO, PSI-MI, SO	External controlled vocabularies		pull down	58,346
FBdv	Developmental stage		larval stage	202
FBgg	Gene groups and pathways		WNTs	1585
FBgn	Gene		tj	249,933
FBig	Interaction group		Saf6, not	32,607
FBlc	Datasets	Biosample, project, reagent collection	AP-MS_RTK-Ras-ERK_Interactome	3681
FBmc	Molecular construct	Engineered plasmid	pWKO.1-H12	2814
FBms	Molecular segment	Cloned region	pBluescript_KS(+)	1286
FBpp	Polypeptide		egr-PA	46,865
FBrf	References	Paper, abstract, personal communication	Ahn, S.J., Marygold, S.J. (2021). Front. Physiol. 12: 648481.	232,001
FBsf	Sequence feature	TSS, insulator, polyA site, regulatory region	htl_htl_enhancer^{FL}	600,545
FBsn	Strains		iso-1	345
FBst	Stocks		BDSC 17222	135,103
FBtc	Cell lines		S2-RIKEN	319

(continued)

Table 2
(continued)

Data Class	Description	Selected subtype(s)	Example	Total items in FlyBase
FBte	Natural transposon	Transposable element	copia	1481
FBti	Insertion	Transposable element insertion site	P{lacW}A3-3-53	210,415
FBto	Experimental tool		Tag:Myr(Src64B)	724
FBtp	Transposon	Transgenic transposable element	P{UAS(-FRT)dpp.N}	145,249
FBtr	Transcript	RNA, mRNA, pseudogene, snRNA	Rab3-RA	72,548

The count of items in each data class is from release 2021_03. DO is Disease Ontology [13], GO is Gene Ontology [9, 10], MI is Molecular Interactions Ontology [12], and SO is Sequence Ontology [11]

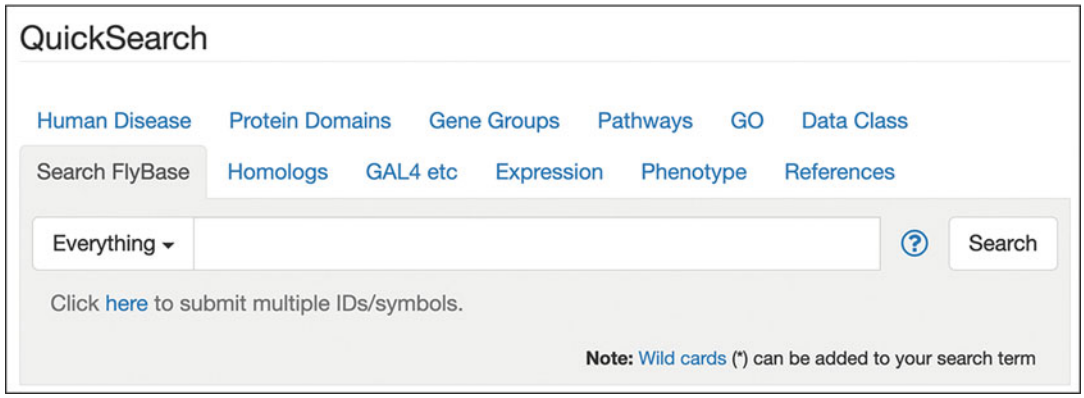


Fig. 1 The QuickSearch interface provides access to specialized searches within FlyBase

specialized result pages, such as the GO (Gene Ontology, *see* Subheading 4) tab which returns GO terms.

Most of the search tabs return results in what we call a HitList, a powerful tool all on its own. On a HitList results page, each search result item is in a panel, providing a quick overview of information on that item. Small blue or green flags to the right of some items indicate recent changes to that record; they can be moused over to provide a quick list of new data. Some items contain button links that will generate other lists, such as a Gene item having buttons for lists of related stocks, references, alleles, and more. Many results will have expandable descriptions, abstracts, or overviews. Some links connect to resources external to FlyBase; for example, Human Disease Model list items may contain links to related records at Online Mendelian Inheritance in Man (OMIM) [6] or the HGNC

[7]. Image hit items will show a thumbnail, and reference hit items may include a graphical abstract. Each data type has an associated color, used as a highlight background for the main item term and the data type "chip" in the upper right of each panel. Use these to quickly scan a long list to see approximate proportions of each type, or skip to the next item of a type of interest. New searches can be started using a search form at the top of the left sidebar.

3.1 Filtering Mixed Lists

The Search FlyBase QuickSearch tab often returns a mixed-type list composed of several data classes. Because the FlyBase database is so large, a search for a broad term can return tens or hundreds of thousands of items, across dozens of data types. To the left of the list, the Filter by Data Class panel provides a way to explore the data generated by a search request. Filtering your mixed hit list to a single data type allows it to be displayed in the Table view, and permits use of the Convert/Export/Analysis tools (all of which can be found at the top of the page). The Convert tool converts results from one type to other allowed types, for example, alleles to stocks or transgenic constructs. Export sends the results to another FlyBase tool, or exports them as a file. HitLists can also be analyzed in ways appropriate for each type of data class; for example, a list of genes can be analyzed for GO term enrichment, or a list of references analyzed for authors, sorted by frequency.

3.2 GAL4 etc QuickSearch Tab

A highly specialized QuickSearch tab is the GAL4 etc tab, which searches for GAL4 and other drivers/binary systems and reporters. Users can select one of the two main search options to limit their search by curated expression pattern (for example, by inputting an anatomy/cell type **FBbt** term such as midgut and a developmental **FBdv** term such as adult stage), or find an allele or construct that reflects the expression of a particular gene by inputting the gene symbol or ID. The search results in a HitList table that can be further refined by filtering individual columns. The GAL4 etc QuickSearch tab also contains a link to the Frequently Used GAL4 Drivers table, which lists commonly used GAL4 lines along with their expression patterns (and provides an image of the expression pattern, where available).

4 Controlled Vocabularies

FlyBase and other biological databases, like many kinds of reference databases, rely on ontologies and controlled vocabularies to organize information. An ontology, to quote Groß et al., "consist[s] of defined concepts, that are typically structured within trees or acyclic graphs where the concept nodes are interconnected by is-a, part-of and other semantic relationships" [8]. In other words, it is a network of related terms, with defined connections between certain

terms. One of the best-known biological ontologies is GO, or Gene Ontology [9, 10]. GO contains terms organized into three domains: biological processes, cellular components, and molecular functions. Each is a loose hierarchy containing terms which may have multiple parent and child terms – for example, heart formation, organelle membrane, and transcription corepressor activity. These terms can then be associated with genes via connecting functions, such as "involved_in" or "enables." By choosing or developing an ontology, curators define a controlled vocabulary (CV; the terms within the ontology), giving them a codified structure in which to categorize data.

FlyBase uses ontologies like GO, Sequence Ontology [11], Molecular Interactions ontology [12], and Disease Ontology [13], as well as internally developed controlled vocabularies, to make FlyBase data more interoperable. FlyBase data classes each have their own dedicated identifiers (IDs; *see* Table 2 for examples of each data class). All FlyBase CVs can be browsed from the Vocabularies button on the main page. On an individual term's page, the Annotations section provides HitLists of all records of a certain type (gene, allele, construct, Human Disease Model, etc.), as well as a tree view placing the term within its ontology.

5 The Gene Report

For users interested in a particular gene, the Gene Report is the best place to start (Fig. 2). The report not only focuses on gene-centric data but also provides a summary of data more fully described in other reports. Apart from the top three sections that are open by default (General Information, Genomic Location, Function), a variety of other subsections can be navigated to and opened using the right-hand Report Sections sidebar. The question mark at the top of the Report Section sidebar links to a help page listing descriptions of the contents of fields presented in a Gene Report. A variety of internal and external links are also provided throughout the report.

5.1 Naming, Classification, and Summaries

The General Information section displays gene identifiers (IDs), such as the official symbol and name of the gene, as well as the annotation symbol and FlyBase gene ID (**FBgn**). Use of the unique official gene symbol/name minimizes ambiguity when referencing genes in scientific literature. The Also Known As field lists synonyms that have been historically or extensively used in published literature; see the Synonyms and Secondary IDs subsection of the Gene Report for more details. *Drosophila* genes have the prefix **Dmel**, and foreign entities have the symbol of the origin species of the gene (e.g., **Hsap** for human genes introduced into *D. melanogaster*). The Gene Model Status field helps distinguish

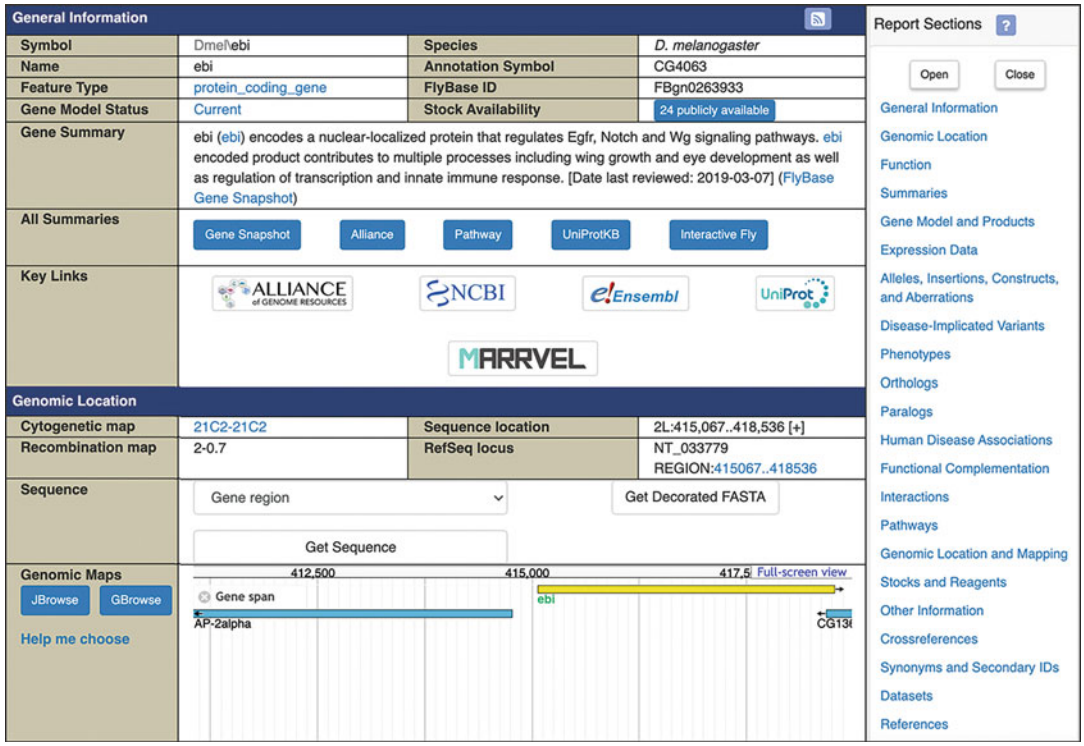


Fig. 2 Example of the top of a Gene Report, in this case *ebi*. Report sections listed at the right can be all opened or closed by the buttons there, or by the plus/minus buttons on the section's dark blue header bar in the main page. Help information for these sections is available by clicking on the question mark icon

genes with "Current" models (localized to the sequenced genome) from those whose gene model is "Incomplete" (due to sequencing gaps), "Uncertain" (supporting data is inconclusive), "Withdrawn" (new data shows a previous gene model to be unsupported), or "Unannotated" (the gene is inferred to exist from genetic data only). Certain types of gene records do not have a gene model in FlyBase, including foreign genes (e.g., Scer\GAL4, Avic\GFP). Where relevant, an Enzyme Name field will be visible, containing the systematic name(s) for an enzyme together with its Enzyme Commission (EC) number [14]. Genes that encode enzymes have been specifically reviewed to enhance our annotation of *Drosophila* data [15]. The Gene Summary field gives users a quick overview of what is known about the gene. Usually, this is a Gene Snapshot generated by FlyBase with input from users [16]. The full list of available summaries can be found further down the page in the Summaries subsection. The Stocks Availability field links to a list of publicly available stocks relating to the gene of interest, which can also be found in the Stocks and Reagents subsection (*see* Subheading 10). The Key Links field lists some links to other key bioinformatic resources; a more extensive list can be found in the External Crossreferences & Linkouts subsection.

The Genomic Location section displays the sequence location of annotated genes and links to the JBrowse [17] genome browser (*see* Subheading 9.2). Buttons in the Sequence field allow the user to download sequences associated with the gene via the Sequence Downloader tool. Other related information can be found in the Genomic Location and Detailed Mapping Data subsections.

5.2 Functionality, Orthology, and References

The Function section begins with a Gene Ontology (GO, *see* Subheading 3) summary ribbon, which gives a graphical overview of GO annotations relating to the gene of interest. The GO annotations (divided into molecular function, biological process, and cellular component) are grouped under high-level summary categories and presented as colored cells, with the depth of color reflecting how many unique terms are grouped in a particular category. Individual annotations based on experimental or predicted evidence can be examined in the related sections below the ribbon. The Function section also displays whether a gene is part of a Gene Group (*see* Subheading 11.2) or Pathway (*see* Subheading 11.3) in FlyBase. Information about the protein from UniProt [18] and InterPro [19] is shown, and details about catalytic activity (derived from GO Molecular Function annotations using EC cross-references) are shown where relevant.

Most but not all of the sections in the Gene Report will be covered in their own section in this chapter. Other sections include the Orthologs section, which lists predicted Human and Model Organism orthologs via DIOPT [20], along with other orthologs from OrthoDB [21], and the Paralogs section which lists predicted paralogs from DIOPT. The Other Information subsection provides sources for gene name or symbol changes and the (often interesting) etymology behind gene naming by authors. Finally, the References section's [22] default view uses an algorithm (which accounts for the major data types captured by FlyBase for each paper) to select up to 25 Representative Publications for the gene. Users can filter by publication type to view all publications of a specific type using the left-hand sidebar, and search for a specific publication using the filter search field.

6 Alleles and Phenotypes

Alleles and phenotypes have been essential to *Drosophila* research ever since Thomas Hunt Morgan first studied a white-eyed mutant fly [23]. Each Gene Report contains an Alleles, Insertions, Transgenic Constructs, and Aberrations subsection, which lists alleles of the gene along with any associated information. The Alleles data class in FlyBase comprises both classical alleles affecting the endogenous locus and alleles linked to transgenic constructs. Each Gene Report also contains a Phenotypes subsection, listing phenotypes

that have been associated with particular alleles of the gene. More information about phenotypes and interactions linked to a specific allele can be found on its Allele Report, along with a variety of other information. In a recent addition to FlyBase, components of alleles that act as an encoded genetic tool will link to Experimental Tool Reports (*see* Fig. 3 and Subheading 6.2), explaining and tracking their use. These subsections make use of controlled vocabularies (*see* Subheading 3), allowing users to efficiently find pertinent data or reagents.

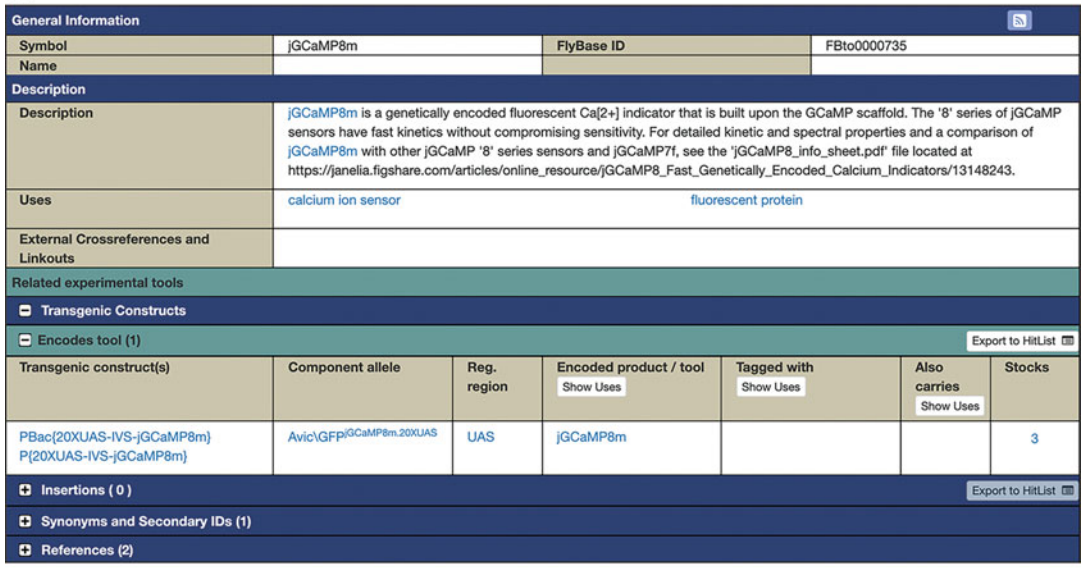
6.1 Classical Alleles, Transgenic Constructs, and Insertions

The Alleles, Insertions, Transgenic Constructs, and Aberrations section lists **FBal**, **FBti**, **FBtp**, and **FBab** data that affect the gene of interest. Each subsection lists relevant lines in a customizable table format (for an example, *see* Fig. 3b). Individual columns can be chosen for display by the Show/Hide Columns button, and a text box atop each column filters by inputted search terms. This section allows the user to easily find useful alleles or transgenic constructs, with relevant stocks listed in the Stocks column. The Classical and Insertion Alleles subsection of the Gene Report contains alleles caused by an alteration or insertion at the endogenous locus of a gene. Users can use the Mutagen column to find information (if known) about how a given allele was generated (e.g., by chemical mutagen, P-element activity, or CRISPR/Cas9 methods) or the Allele Class column to determine what kind of allele (e.g., loss of function, hypomorphic) is listed. Alleles caused by insertions are also linked to an associated Insertion Report. The Transgenic Constructs section is subdivided into those constructs containing or affecting the gene coding region (e.g., UAS lines, RNAi lines) and those constructs that contain the regulatory region of the gene (e.g., GAL4 lines, fluorescent reporters). Each construct has an associated Recombinant Construct Report (providing molecular details) as well as an associated Allele Report (for consistent curation and display of phenotypic and expression data). The customizable tables allow users to filter by information of interest, for example, the regulatory region present (e.g., UAS), and take advantage of the Experimental Tools (*see* Subheading 6.2) data class to display if the construct encodes a tool (e.g., GAL4), or is tagged (e.g., with GFP). All alleles have their own Allele Report, which contains a variety of information, including how the allele was generated (if known), curated phenotypes, interactions, and any human disease model data.

6.2 Experimental Tools

Creative *Drosophila* researchers have developed a huge number of transgenic tools with useful properties that are exploited to study biological function of a gene product or biological process. To make these complex tools more approachable, FlyBase has recently developed a custom ontology and set of **FBcv** terms to show the relationships between tools. This ontology-centric approach (for

A



General Information			
Symbol	jGCaMP8m	FlyBase ID	FBto0000735
Name			

Description

Description	jGCaMP8m is a genetically encoded fluorescent Ca[2+] indicator that is built upon the GCaMP scaffold. The '8' series of jGCaMP sensors have fast kinetics without compromising sensitivity. For detailed kinetic and spectral properties and a comparison of jGCaMP8m with other jGCaMP '8' series sensors and jGCaMP7f, see the 'jGCaMP8_info_sheet.pdf' file located at https://janelia.figshare.com/articles/online_resource/jGCaMP8_Fast_Genetically_Encoded_Calcium_Indicators/13148243.
Uses	calcium ion sensor; fluorescent protein
External Crossreferences and Linkouts	

Related experimental tools

Transgenic Constructs

Encodes tool (1) — Export to HitList

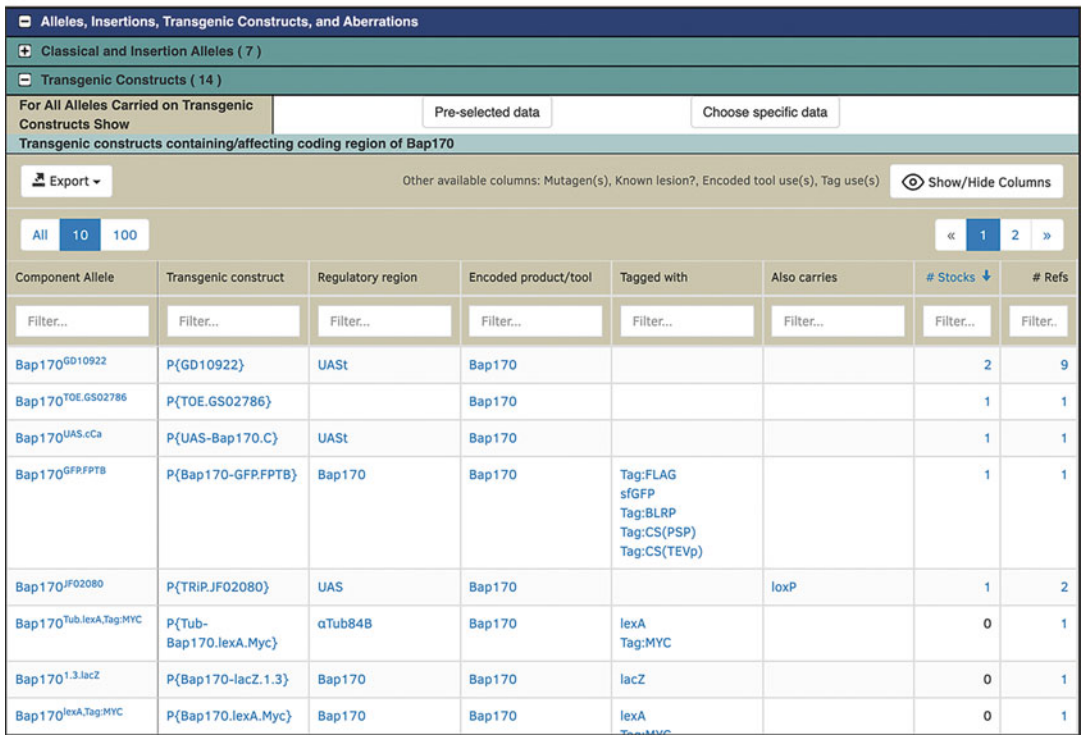
Transgenic construct(s)	Component allele	Reg. region	Encoded product / tool (Show Uses)	Tagged with (Show Uses)	Also carries (Show Uses)	Stocks
PBac{20XUAS-IVS-jGCaMP8m} P{20XUAS-IVS-jGCaMP8m}	Avic\GFP$^{jGCaMP8m.20XUAS}$	UAS	jGCaMP8m			3

Insertions (0) — Export to HitList

Synonyms and Secondary IDs (1)

References (2)

B



Alleles, Insertions, Transgenic Constructs, and Aberrations

Classical and Insertion Alleles (7)

Transgenic Constructs (14)

For All Alleles Carried on Transgenic Constructs Show: Pre-selected data | Choose specific data

Transgenic constructs containing/affecting coding region of Bap170

Export — Other available columns: Mutagen(s), Known lesion?, Encoded tool use(s), Tag use(s) — Show/Hide Columns

All 10 100 « 1 2 »

Component Allele	Transgenic construct	Regulatory region	Encoded product/tool	Tagged with	Also carries	# Stocks	# Refs
Filter...	Filter...	Filter...	Filter...	Filter...	Filter...	Filter...	Filter...
Bap170^{GD10922}	P{GD10922}	UASt	Bap170			2	9
Bap170$^{TOE.GS02786}$	P{TOE.GS02786}		Bap170			1	1
Bap170$^{UAS.cCa}$	P{UAS-Bap170.C}	UASt	Bap170			1	1
Bap170$^{GFP.FPTB}$	P{Bap170-GFP.FPTB}	Bap170	Bap170	Tag:FLAG sfGFP Tag:BLRP Tag:CS(PSP) Tag:CS(TEVp)		1	1
Bap170^{JF02080}	P{TRiP.JF02080}	UAS	Bap170		loxP	1	2
Bap170$^{Tub.lexA.Tag:MYC}$	P{Tub-Bap170.lexA.Myc}	αTub84B	Bap170	lexA Tag:MYC		0	1
Bap170$^{1.3.lacZ}$	P{Bap170-lacZ.1.3}	Bap170	Bap170	lacZ		0	1
Bap170$^{lexA.Tag:MYC}$	P{Bap170.lexA.Myc}	Bap170	Bap170	lexA		0	1

Fig. 3 Experimental Tools can be found on their own Report pages, and as alleles on Allele or Gene Reports. (**a**) Example of an Experimental Tool Report (*jGCaMP8m*). (**b**) The Allele section of a Gene Report (*Bap170*) has a customizable table to filter Experimental Tools

example, classifying all photoactivatable fluorescent protein reporters or pH sensors together) helps clarify which tools could substitute for each other when designing a multi-element genotype. In the Transgenic Constructs subsection of the Allele section, several of the columns have links to an Experimental Tool Report (**FBto**, *see* Fig. 3) to explain each component of the allele. These include tools that allow a gene product to be detected (e.g., EGFP, FLAG tag), target a gene product somewhere specific within a cell (e.g., nuclear localization signal), drive expression in a binary system (e.g., UAS, GAL4), or enable clonal or conditional expression (e.g., FLP, FRT). Reports for genetically encoded sensors are also available, allowing researchers to find reagents used to monitor changes in small molecules (e.g., GCaMP6m), pH (e.g., pHluorinE), or membrane potential (e.g., Voltron). Reports for tools that can be used to modulate neuronal activity have also been generated (e.g., CsChrimson, Kir2.1). Experimental Tool Reports list the official symbol used in FlyBase and provide a description of the tool, along with tables of transgenic constructs and insertions that carry or encode a given tool.

6.3 Phenotypes

The Phenotypes section of the Gene Report lists phenotypes that have been annotated to alleles of the gene of interest, such as decreased body size or abnormal circadian behavior. This gives an overview of anatomical and behavioral phenotypes that result from perturbation of the gene, facilitating access to more detail about specific phenotypes or alleles of interest. Phenotypic data are attached to alleles using the phenotypic class (**FBcv**) and anatomy (**FBbt**) controlled vocabularies [24, 25] and can be further refined through use of qualifier CV terms, distinguished by a "|" (pipe symbol). Genotypes requiring multiple alleles to result in a particular phenotype are indicated by the word "with" connecting alleles. All phenotypic data are curated and attributed to published papers, usually with some explanatory free text accompanying associated CV terms (included on Allele Reports). The Phenotype tab in QuickSearch allows users to search for all alleles associated with any given phenotypic class and/or anatomy CV term.

6.4 Disease Ontology Annotations

The Disease Ontology (DO) annotations, found in the Human Disease Associations section of a Gene Report, associate a DO term to either an allele (as shown by experimental evidence) or a gene (as predicted by orthology to a human gene). Models based on experimental evidence and modifiers based on experimental evidence list all DO terms that the allele was recorded as being causative of a disease or modifying another disease-causing allele. Non-allele-based models, including those caused by aberrations, are not included in DO annotations, but may be covered by a Human Disease Model (*see* Subheading 11.4). Potential Models Based on Orthology shows a human ortholog to the Gene Report's

subject, which has been associated with a DO term via OMIM. This inference is a prediction, not evidence, but it may assist researchers who are developing new fly models of human disease. DO terms' definitions and their relationships to each other can be viewed in their CV Term Report, accessible by searching for a DO term in QuickSearch or by clicking on any hyperlinked DO term.

6.5 Disease-Implicated Variants

Alleles generated to mimic variants associated with a human disease in *Drosophila* can be found in the Alleles Representing Disease-Implicated Variants section. These can be modifications of human genes expressed in *Drosophila*, as seen in Hsap\APP, or endogenous *Drosophila* genes, as seen in Dmel\crb. Alleles in the table are shown with links to the associated human variant, its Clinvar ID (https://www.ncbi.nlm.nih.gov/clinvar), the originating reference, and the associated disease. Multiple diseases can be associated with the same allele, causing there to be multiple lines per allele. To easily sort through results, columns can be filtered by the text box in the top row.

7 Expression Data

The Expression Data section starts with two ribbons which correspond to the two types of data found here: low throughput and high throughput. The first ribbon shows how frequently manually curated, low-throughput data is associated with *Drosophila* anatomy regions (**FBbt** data class); more color saturation represents greater amount of curation to the associated term. The second shows one example of high-throughput data: stage-specific RNA--seq data from modENCODE [26]. Following the high-throughput data are links to external sites with *Drosophila* expression [27–30], including images of gene expression during embryogenesis [31] for selected genes.

7.1 Low-Throughput Expression Data

FlyBase manually curates three types of low-throughput expression data: transcript-based, polypeptide-based, and reporter-based expression. Within these, data will be divided by what assay or tool was used to obtain it, such as in situ hybridization, Western blot, or a specific P-element reporter. Each line in these sections records expression seen at a specific developmental range (embryonic through adult) and tissue(s), in a particular reference, for example, "embryonic stage 11–12, embryonic myoblast, Carmena et al. 1995." Tissue/position terms may be followed by a pipe and a modifier, such as "organism | dorsal" or "muscle founder cell | subset," to provide more specificity. All stage and tissue/position terms are linked to that term's CV Term Report. The Term Report shows other expression that has been recorded for that term, as well as the term's spanning tree to see its

parent terms and child terms, if any exist. The Additional Descriptive Data section provides a summary and/or clarification of the expression data in a specific paper.

Gene expression patterns can be searched using the QuickSearch Expression tab. Like the GAL4 etc QuickSearch tab, users can search by curated expression pattern using an anatomy/cell type **FBbt** term and/or a developmental **FBdv** term. The search results in a HitList containing those genes that have been curated as having a transcript or protein expression pattern reflecting the chosen search terms. The Expression QuickSearch tab also includes links to tools to search or browse RNA-Seq expression data.

7.2 High-Throughput Expression Data

Within the Expression Data section of the Gene Report, the High-Throughput Expression Data subsection includes expression plots of high-throughput mRNA RNA-Seq data from modENCODE [32], mRNA microarray data from FlyAtlas [28], and proteome data from the Casas-Vila proteome project [33]. FlyBase has produced quantitative views of the RNA-seq and proteomic data in different stages and tissues, and views for different cell culture types and treatments of the RNA-seq data, presented as bar graphs. For the RNA-Seq data, RPKM counts (reads per kilobase per million reads) have been calculated, averaged over the exonic extent of the gene.

Several FlyBase tools are available to query calculated RPKM data. The RNA-Seq Profile tool allows retrieval of genes with RNA-Seq expression patterns and levels matching a user-defined profile; this tool can be accessed from the RNA-Seq icon link on the homepage, from the Tools menu of the NavBar, or from the QuickSearch Expression tab. A section of the QuickSearch Expression tab also provides options to search for similarly expressed genes using the modENCODE RNA-Seq datasets. The RNA-Seq by Region tool returns the average RPKM over a specified genomic region and also offers the option of a gene-specific query that returns an exon-by-exon RPKM count; this tool can be accessed from the Tools menu in the NavBar or from the High-Throughput Expression Data subsection of the Gene Report.

8 Interactions

One-to-one genetic and physical interactions between genes and gene products, respectively, are displayed in the Interactions section of a Gene Report. Both types of interactions can be explored visually using the esyN [34] network diagram. The esyN diagrams are highly manipulable; scrolling within the diagram will zoom in/out on the diagram, rather than moving the overall page up/down, and individual nodes can be dragged to new locations in the diagram. Further network diagram customization is available

by clicking on the View in Interactions Browser link. There, below the esyN diagram, a text-based list of individual interactions lists the source of the interaction observation, plus other details. The External Data for this section lists other databases that capture physical or genetic interactions, including BioGRID [35], DroID [36], and MIST [37].

8.1 Physical Interactions

Both protein–protein and protein–RNA interactions, but not DNA–protein interactions, are curated by FlyBase. Only interactions between two *D. melanogaster* gene products are recorded, with the rare exception of interactions with human gene products expressed in the fly that are disease related (for example, Hsap \APP). A list below the esyN network diagram shows, for each recorded interaction, the types of assay(s) used and the paper (s) which reported the interaction. Clicking on an individual pairwise interaction brings up its Physical Interaction Report, which provides further details on the interaction (such as each element's role in the assay, any isoform or subregion specificity, and source of the tissue assayed), and an esyN network diagram of both participants' combined interaction networks.

8.2 Genetic Interactions

Similar to physical interactions, genetic interactions are shown both as an esyN network diagram and in a list of pairwise interactions. Since the pairwise list is shown by genes rather than by alleles, there can be both an enhancing and a suppressing interaction listed for one pair of genes, but derived from different alleles. For more information about each interaction, clicking on the associated reference will list the alleles used in the paper. Selecting an allele from this list brings up its Allele Report, where detailed descriptions of genetic interactions can be found in the Interactions section.

9 Genomics Data

In March 2000, the combined sequencing and annotations efforts of FlyBase, the Berkley and European Drosophila Genome Projects (BDGP and EDGP), Celera Genomics, and other *Drosophila* experts produced Release 1 of the *Drosophila melanogaster* genome [38]. The most recent major revision to the genome is Release 6, which was published in 2014 [39]. Between major revisions, FlyBase incorporates data from publications to update the genome. These updates are identified by numbers after a period in the release number (e.g. Release 6.40, which corresponds to 2021_03), with about six updates made per year. Refinements of sequence data and annotations made by FlyBase are propagated to other sequence repositories and genome browsers. Selected archived releases are available in the Downloads menu of the NavBar, in both HTML and FTP protocols.

9.1 Gene Model Annotations in FlyBase

In FlyBase, a gene model is the set of transcripts and polypeptides produced by a gene [40]. Gene models are manually curated, based on a variety of sources, including RNA-seq data (sequences and exon junctions) and transcription start site data [26, 41]. Atypical gene models, such as genes with polycistronic or trans-spliced transcripts, require extra care during manual annotation to capture all biologically relevant isoforms [42]. In the Gene Model and Products section of each Gene report, the Comments on Gene Model section will show when the gene model was most recently reviewed by FlyBase, and any additional clarifying comments. Below that, the Sequence Ontology section will list all appropriate SO terms for that gene, such as "nuclear_gene" or "gene_with_stop_codon_readthrough." In November 2020, FlyBase updated annotations of all genes to the latest version of the Sequence Ontology, which allows for more precise classification of non-coding RNAs and some uncommon gene classes.

In a Gene Report, the Gene Model and Products section begins with an interactive view of the gene span, RNAs, and CDSs from JBrowse. Clicking on individual Transcripts or Polypeptides brings up their associated reports. These reports show the FASTA sequence, links to other products of the same gene, and links to external resources, such as GenBank [43] and UniProt. For transcripts, there are links to published cDNA clones that are consistent with the transcript. For a summary view with information about all of the transcripts and proteins from a gene, go further down the Gene Model and Products section to the Transcript Data and Polypeptide Data subsections. Information about functional domains can be found directly below the JBrowse display, where any protein domains identified by Pfam [44] and SMART [45] are shown. Elements such as known regulatory regions from REDfly [46] and mapped point mutations within the gene can be seen using the Feature Mapper tool, in the Mapped Features subsection. The Sequences Consistent with the Gene Model subsection has links to NCBI and UniProt records, and the External Data subsection links to the gene's entry in the Eukaryotic Promoter Database [47].

The *D. melanogaster* reference genome is not from a wild-type strain [48]. There are numerous genes that have been identified as carrying mutational lesions; this is indicated in controlled comments in the Comments on Gene Model section. Care must be taken when accessing sequence information for such genes: the genomic and transcript sequences represent the mutant case; for the protein sequence, FlyBase has provided a corrected sequence.

9.2 JBrowse

JBrowse (Fig. 4), a genome annotation viewer that is part of the Generic Model Organism Database (GMOD) tool suite [17], is used by FlyBase to show gene models and supporting data, such as cDNAs, ESTs (expressed sequence tags), RNA-Seq data,

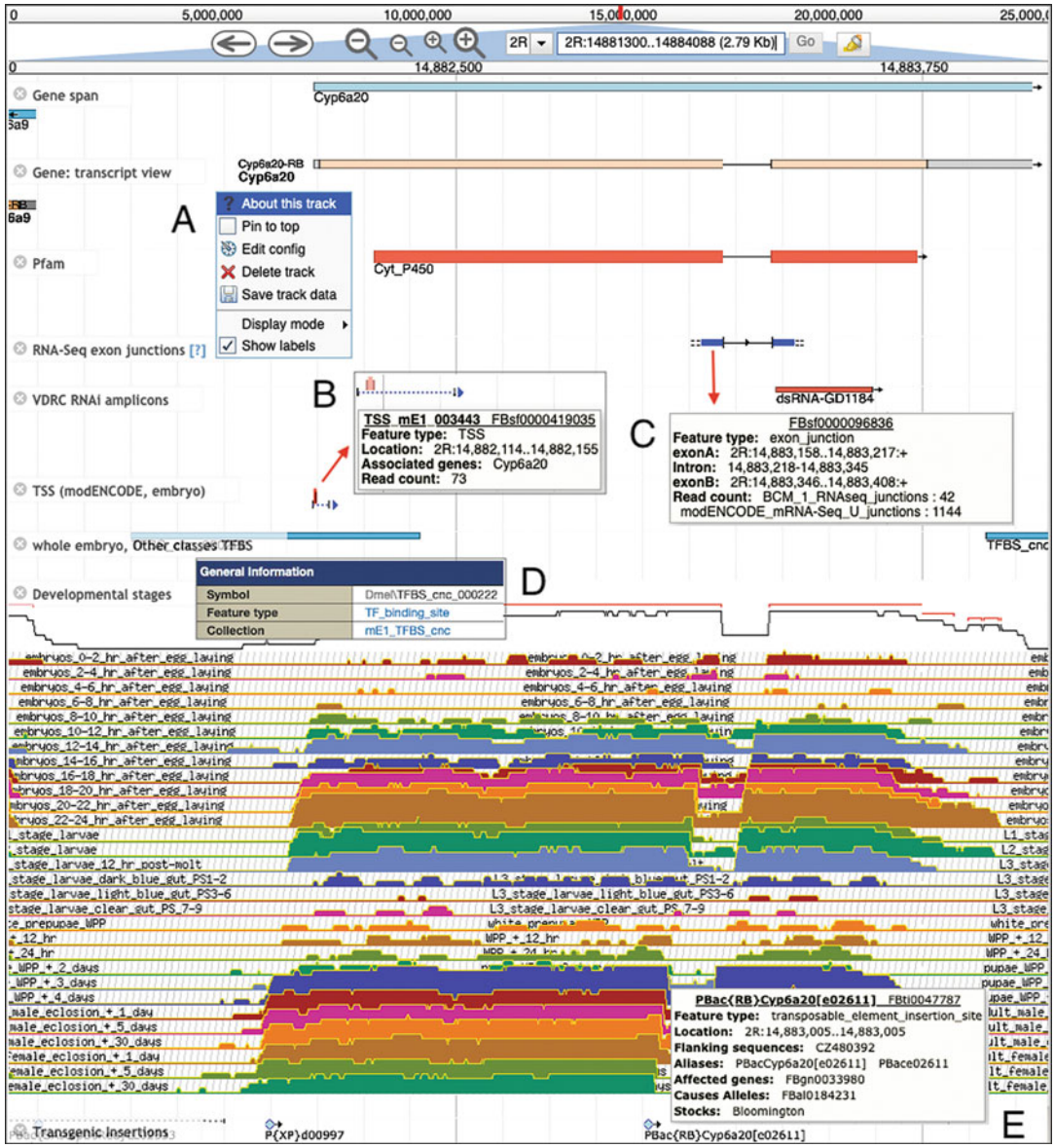


Fig. 4 The JBrowse genome browser. This view of sequence features and genomic data in the region of the Cyp6a20 gene has been customized by both using the Available Tracks left sidebar form (not shown) to turn some default tracks off and new data tracks on; and by dragging tracks to new positions vertically. Note the Developmental Stages RNA-Seq "TopoView" display indicating temporally differential expression of this gene. Most objects in JBrowse show more information on mousing over, and link to FlyBase reports or external resources when clicked. (**a**) Mousing over a track label reveals a menu toggle; the menu includes an option to download track data. The "About this track" menu option displays a pop-up with a link to the FlyBase documentation wiki entry for this track, or track group. (**b**) Some tracks show details by zooming in; this Transcription Start Site (TSS) has a bar graph with TSS distribution of reads by base. (**c**) Mousing over an RNA-Seq junction activates a pop-up with read counts; when overlapping junctions are present (not shown here), relative read counts can indicate prevalence of alternative exons. (**d**) Clicking on many objects in JBrowse links to the full FlyBase report for that feature, as shown here for a transcription factor binding site (TFBS) sequence feature. (**e**) Mousing over a Transgenic Insertion site activates a pop-up with summary information, including associated alleles and stock availability

transcription start sites, gene predictions, and aligned proteins [49]. JBrowse allows representation of many other types of sequence-based data and reagents; essentially, anything that maps to the genome can be represented on JBrowse. By using the Select Tracks option, the user can choose to view mapped genetic variants, such as mutational lesions, transgenic insertions, aberration extents, and aberration breakpoints; regions carried on transgenic constructs such as rescue fragments and RNAi reagents; microarray oligonucleotides and RNAi amplicons; or high-throughput mapping of transcription factor binding sites, insulator elements, and RNA-editing sites. When zoomed in sufficiently, the Nucleotide View track shows the nucleotide sequence by color coding, and it indicates translation starts and stops in all six forward and reverse translation frames. When zoomed in to less than 200 bp, the single letter code for nucleotides and amino acids becomes visible.

JBrowse has replaced GBrowse, and features a client-side architecture. Performance is therefore limited only by the power of the hardware running the internet browser, with little dependence on internet speed. Most users will see fast, smooth zooming and scrolling along the genome. Tracks can be dragged, opened, or closed; a useful feature as the FlyBase installation of JBrowse has well over one hundred data tracks. Track names are in the field at the top of the sidebar. Mousing over a track name in the browser display activates a drop-down menu, with display and download options.

JBrowse can be accessed from one of the button icons at the top of the homepage or via the Tools menu on the NavBar. In addition, there is a link to the appropriate genomic region in JBrowse on the reports for every localized gene and mapped sequence feature. A genomic BLAST hit obtained using the FlyBase BLAST tool includes a link to the hit in the relevant region in JBrowse. QuickSearch HitList results for several data types also link to JBrowse.

9.3 RNA-seq

Tracks of RNA-Seq expression data [32] are particularly informative when viewed in JBrowse. FlyBase JBrowse has a unique "TopoView" display for these data that shows aggregated groups of RNA-Seq tracks. RNA-Seq data tracks that are displayed using TopoView include tracks across different developmental stages, tissues, treatment conditions, and cell lines. RNA-Seq exon junction data [50], presented in a separate track on JBrowse, are extremely useful for judging alternative splicing and isoform-specific expression.

10 Reagents

There are a variety of ways to find reagents associated with particular genes or genomic regions. For a given gene, the simplest way is via the Stocks and Reagents subsection of the Gene Report; this section lists available fly stocks, genomic and cDNA clones, cell-based RNAi reagents, and antibodies used in published literature. A visual representation of sequence-based reagents (such as enhancer lines, sgRNA reagents, or RNAi lines) is best viewed using the Misc. Reagents track on JBrowse. The Transgenic Insertions Stocks in Bloomington, Kyoto track on JBrowse also provides locations of available transgenic insertion lines, while deficiency and duplication stocks can be viewed using the Aberrations track on JBrowse. The Alleles, Insertions, Transgenic Constructs, and Aberrations section of a Gene Report also allows the user to filter reagents of particular types and see associated relevant stocks in the Stocks column (*see* Subheading 6.1).

10.1 Stocks and Strains

A list of stock collections with *Drosophila* lines available to order can be found via the Links menu on the NavBar. Users interested in a particular gene will find the button in the Stock Availability field on the Gene Report the easiest way to find relevant publicly available stocks. Stock Reports (**FBst** data class) present the stock genotype and source collection, along with a stock number linked to the relevant stock center to facilitate ordering. Strain Reports (**FBsn** data class) are a good source of information about commonly used wild-type (e.g., Oregon-R), mutant (e.g., iso-1), or specifically inbred (e.g., DGRP-31, from the Drosophila Genetics Reference Panel [51]) strains. The reports contain information about the origin of the strain, along with any associated alleles or phenotypes of note.

10.2 Cell Lines

FlyBase curates stable *Drosophila* derived cell lines in the **FBtc** data class. The majority of cell lines are available from the Drosophila Genomics Resource Center (DGRC; https://dgrc.bio.indiana.edu). Information provided about lines can include the strain, developmental stage, and anatomical site of the source material, where it was originally published, and references which have used the line.

10.3 cDNAs

Most genes have a number of mapped cDNA clones, which FlyBase curates using the **FBcl** data class. This includes both large genomic clones, and gene-specific clones (both full-length cDNAs and ESTs) from GenBank, the Berkeley Drosophila Genome Project (BDGP) [52], and other sources.

10.4 Antibodies

The Antibodies section of the Gene Report provides curation of antibodies against a gene's protein product(s). This is divided into Lab Generated Antibodies (monoclonal and polyclonal antibodies curated to publications) and Commercially Available Antibodies (links to vendors, such as the Developmental Studies Hybridoma Bank, https://dshb.biology.uiowa.edu). Further description of the lab generated antibodies can be found in the Additional Polypeptide Data and Comments subsection of the Gene Report.

11 Integrated Reports

A number of reports at FlyBase aim to organize and integrate related data and resources into distinct collections. This allows association of information and metadata to sets of related data, for example, Dataset Reports for specific reagent collections, such as TRiP RNAi lines, or particular projects, such as the modENCODE transcriptome. Gene Group Reports allow easy access to related sets of genes, such as members of a gene family, or subunits of a protein complex. Pathway Reports list genes that have been experimentally shown to act within or regulate a pathway. Human Disease Model Reports are an excellent entry point for researchers interested in *D. melanogaster* models of a particular disease, as they present and integrate relevant data from numerous sources.

11.1 Large Dataset Metadata

The **FBlc** data class is used to curate a large variety of datasets and reagent collections, along with their associated metadata. This includes cDNA libraries, sets of stocks or strains, various types of next-generation sequencing projects, dsRNA reagent collections, and tissue biosamples that are sources for genomic libraries. These dataset reports represent distinct parts of a dataset or reagent collection, typically including several "biosample," "assay" (raw data), and "result" (processed data) entities, all connected to each other and grouped within a higher level "project." For example, a tissue sample may have associated raw and processed microarray data. For some **FBlc** items, such as a large raw dataset, FlyBase will not necessarily host the dataset itself, but will show links to the publication or data provider.

11.2 Gene Groups

FlyBase Gene Groups facilitate straightforward analysis of related genes, such as gene families, protein complex subunits, and other functional groupings [53, 54]. Gene Groups can be accessed by entering the name or symbol of a group or member gene in the Gene Groups tab of QuickSearch, or by following the Gene Group link on a member Gene Report. The QuickSearch Gene Groups tab also contains a link to a list of all current gene groups available in FlyBase. Gene Group Reports give an overall description of the group, along with relevant key GO terms and links to any related

gene groups. The Members subsection contains a table of the members of the gene group, with source material for membership in the group clearly attributed. This section provides buttons to view orthologs, download associated data of interest (e.g., expression data, phenotypes) using Batch Download, or to export gene group members to a HitList for further analysis. The External Data subsection lists links to equivalent gene groups in other organisms, such as WormBase [55] or HGNC [56].

11.3 Pathways

Pathway Reports (Fig. 5) facilitate access to data on pathways, providing the user with a list of manually curated genes that have been experimentally linked to particular pathways. Pathway Reports can be accessed by entering the name or symbol of a pathway or member gene in the Pathways or Gene Groups tab of QuickSearch, or by following the Pathway link on a member Gene Report. The Pathways QuickSearch tab also contains a link to a list of all current pathways available in FlyBase. Pathway Reports give an overall description of the signaling pathway and are divided into component groups (minimally: core, positive regulators, and negative regulators). The core group comprises those genes required for the activated pathway to function, whereas regulators modify activity of the core pathway members. Network diagrams are generated from physical interaction data, linking genes within the pathway. These networks can be viewed in the pathway view (distinguishes regulatory interactors) or in the functional view (nodes are colored depending on molecular function of the gene product). The Members subsection lists pathway member genes in a customizable table, enabling quick access to further information, such as references (the # Pathway Refs column lists the number of GO-curated papers that provide direct experimental evidence for a gene's role in a pathway) and reagents (e.g., antibodies, alleles, transgenic constructs). This section also provides buttons to view orthologs, download associated data of interest using Batch Download, and to export pathway members to a HitList for further analysis.

11.4 Human Disease Models

In addition to Disease Ontology curation (*see* Subheading 6.4), FlyBase curates data relevant to human health in a Human Disease Model (HDM) Report [57]. These Reports are written with the goal of making the unique features of *Drosophila* that make it such a powerful model for translational research [58] more accessible to non-Drosophilists. Thanks to the work of such organizations as GeneMatcher [59] and the Undiagnosed Disease Network [60], many more researchers are collaborating on papers including *Drosophila* work than ever before. These non-Drosophilist researchers would be aided by a human disease-centric entry to *Drosophila* genetics. Additionally, HDM Reports (which make up the **FBhh** data class) can capture non-genetic disease models that DO cannot, such as traumatic brain injury and paraquat-induced Parkinson

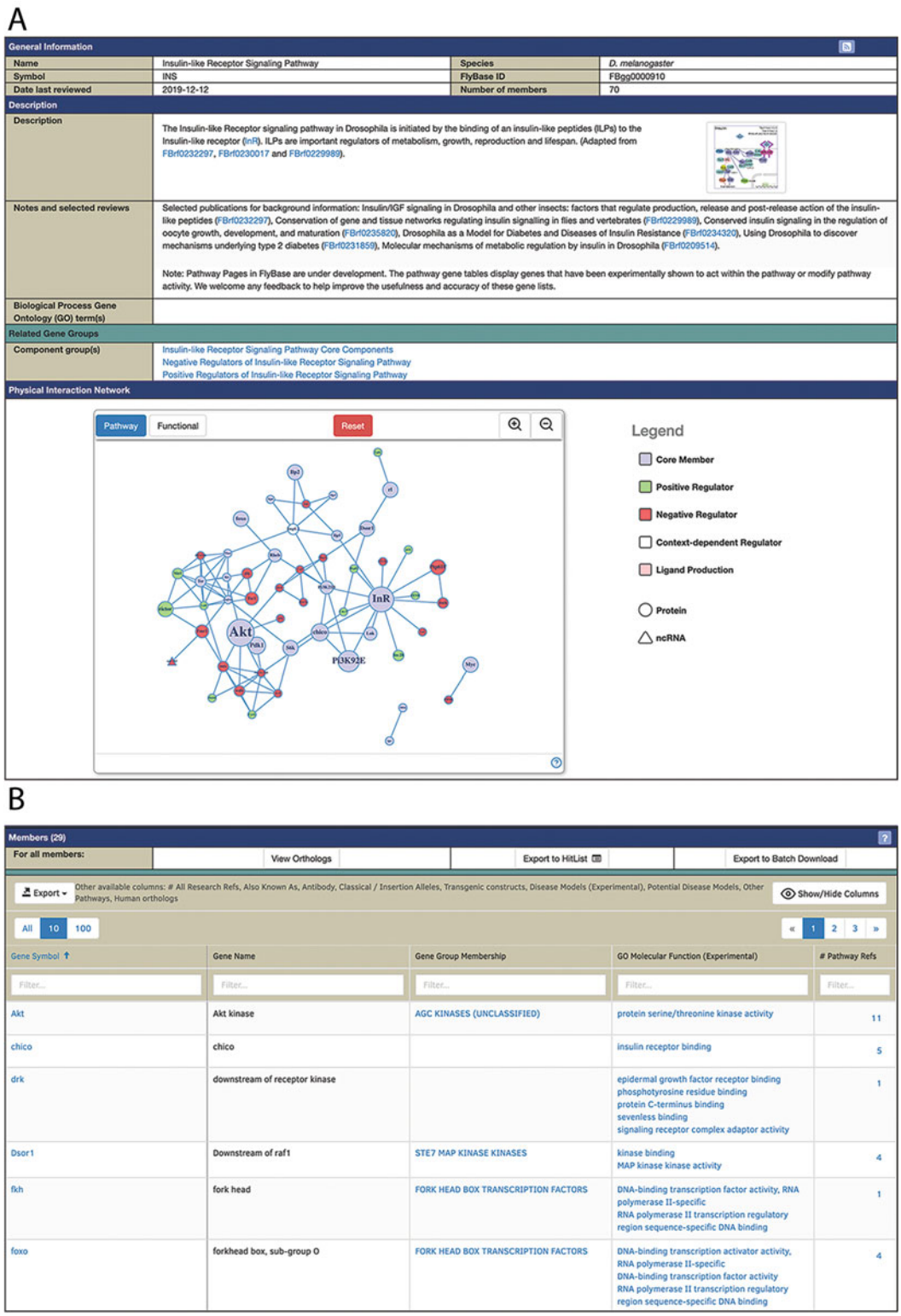


Fig. 5 Example of a Pathway Report, in this case for the Insulin-like Receptor Signaling Pathway. (**a**) Overall information about the pathway. (**b**) Core components of the signaling pathway. Pathway Reports also contain positive and negative regulators, not shown here

disease. HDM Reports connect users to external databases and websites which curate human disease data, as well as to internal FlyBase curation in Gene and Allele Reports. Diseases and their relations to each other are curated (when possible) according to their name in OMIM. Diseases are grouped either manually by FlyBase or according to Phenotypic Series groups from OMIM (all Parkinson disease subtypes, for example). This flexible report format allows us to include disease models found in the *Drosophila* literature that have not yet been included in other disease databases, which are noted by the term "postulated" in the name. A list of all HDMs can be found via a link in the Human Disease tab of QuickSearch.

HDM Reports begin with an overview of the model, including a description of the human disease phenotype and genetics, and what specific work has been performed using *Drosophila* disease models. Human-fly orthology can involve multiple paralogs in either or both organisms that are orthologous to each other, so only those that are likely to be relevant to the model are included. For the fly gene(s) in the report, relevant information from their Gene Report is provided, including DO curation and any disease-implicated variant alleles (*see* Subheadings 6.4 and 6.5). More details on the disease from a clinical viewpoint are provided in the Disease Summary section, including excerpts from OMIM and links to resources about the disease and the causal human gene(s). The Related Diseases section includes both the other members of an OMIM phenotypic series, if the subject of the HDM is in one, and/or diseases deemed relevant by FlyBase curators.

12 Bulk Data Analysis and Downloads

A list of gene, protein, or other identifiers is a common starting point for research tasks. FlyBase has tools to find and build ID lists, and to analyze lists from FlyBase or elsewhere. ID lists can be downloaded from a collection of bulk data files (*see* Table 3 for names and descriptions of precomputed files) or uploaded into FlyBase for processing using a number of tools.

12.1 Uploading and Analyzing Data

The ID Validator tool is the preferred starting point for any analysis of a list of IDs. A quick link to this tool can be found among the icon links at the top of the FlyBase home page. Lists can be pasted into a form or uploaded as a file. Instructions for formatting can be found on the ID Validator form page. Accepted ID types may change over time, but currently include FlyBase IDs, "CG" (annotation) IDs, gene symbols, PubMed IDs, and GenBank/UniProt/Swiss-Prot/TrEMBL accessions. Mixed lists are acceptable.

Table 3
Files available in the Current Release link in the Downloads drop-down menu of the NavBar

Category	File Name	Brief description
Genetic and genomic data	gene_map_table_fb_*.tsv.gz	Localization information for *Drosophila* genes
	dmel_orthologs_in_drosophila_species_fb_*.tsv.gz	Dmel genes and orthologs in sequenced *Drosophila* species
	dmel_paralogs_fb_*.tsv.gz	*D. melanogaster* paralogs
	dmel_human_orthologs_disease_fb_*.tsv.gz	Human orthologs of *D. melanogaster* genes
	gene_association.fb.gz	Gene Ontology (GO) terms assigned to Dmel genes
	gene_snapshots_fb_*.tsv.gz	Expert solicited gene snapshots visible on Gene Reports
	automated_gene_summaries.tsv.gz	Automated gene summaries as shown on Gene Reports
	gene_rpkm_report_fb_*.tsv.gz	Dmel gene expression values based on RNA-Seq
	gene_rpkm_matrix_fb_*tsv.gz	
	dmel_unique_protein_isoforms_fb_*.tsv.gz	Dmel genes and their unique protein isoforms
	allele_phenotypic_data_fb_*.tsv.gz	CV phenotypic data associated with alleles
	disease_model_annotations_fb_*.tsv.gz	Disease model data associated with alleles
	gene_genetic_interactions_fb_*.tsv.gz	Summary of Dmel gene-level genetic interactions
	allele_genetic_interactions_fb_*.tsv.gz	Allele-level genetic interactions with CV terms
	physical_interactions_mitab_fb_*.tsv.gz	Dmel gene pairs whose products physically interact, in MITAB format
	insertion_mapping_fb_*.tsv.gz	Localization information for Dmel insertions
	transposon_sequence_set.embl.txt.gz	Transposable element canonical sequences
	fu_gal4_table_fb_*.json.gz	Frequently-used GAL4 drivers
Integrated data	dataset_metadata_fb_*.tsv.gz	All dataset/collections and all associated features
	gene_group_data_fb_*.tsv.gz	All Gene Groups; relationships and members or HGNC IDs
	gene_groups_HGNC_fb_*.tsv.gz	
Reagents	genomic_clone_data_fb_*.tsv.gz	Genomic clone IDs, names, and accession numbers
	cDNA_clone_data_fb_*.tsv.gz	cDNA/EST IDs, names, library, and accession numbers
	organism_list_fb_*.tsv.gz	Data on all species in FlyBase, including abbreviations
Other data files	fbrf_pmid_pmcid_doi_fb_*.tsv.gz	All IDs for references in FlyBase that have a PMID
	stocks_FB*.tsv.gz	Genetic components and related FlyBase data about stocks

(continued)

Table 3
(continued)

Category	File Name	Brief description
Correspondence tables	fb_synonym_fb_*.tsv.gz	Symbols, names and synonyms for most features in FlyBase
	fbgn_NAseq_Uniprot_fb_*.tsv.gz	FlyBase gene IDs ⇔ nucleotide and protein accessions
	fbgn_annotation_ID_fb_*.tsv.gz	Current and secondary FBgn and annotation IDs for genes
	fbgn_fbtr_fbpp_fb_*.tsv.gz	FlyBase gene IDs ⇔ FlyBase transcript and polypeptide IDs; expanded has organism, symbol and type data
	fbgn_fbtr_fbpp_expanded_fb_*.tsv.gz	
	fbal_to_fbgn_fb_*.tsv.gz	FlyBase allele IDs ⇔ FlyBase gene IDs
	cyto-genetic-seq.tsv.gz	Dmel cytogenetic map ⇔ genetic map ⇔ genomic coord.
Ontology files	fly_anatomy.obo	Fly anatomy ontology (FBbt)
	fly_development.obo	Fly developmental stage ontology (FBdv)
	flybase_controlled_vocabulary.obo	FlyBase miscellaneous ontology (FBcv)
	flybase_stock_vocabulary.obo.gz	FlyBase stock ontology (FBsv)
	go-basic.obo	Gene ontology (GO)
	image.obo	FlyBase image ontology (FBbi)
	so-simple.obo	Sequence ontology (SO)
	doid.obo	Human disease ontology (DO)
	mi.obo	Molecular interactions ontology (MI)

An asterisk indicates where the release ID would be in a filename (e.g., YYYY_## or 2021_04), as these files are updated in every release. Other items will have their update information near the top of the file. More information on each of these files can be found in the Help menu at the top of the page

The result page displays a table with a column for submitted IDs and another for the corresponding validated ID. Record rows are color coded to indicate those IDs that were validated, those requiring some interpretation, and unrecognized IDs. Users should edit an ID list until it validates cleanly before proceeding. At this point, users can select from several dispositions for their validated list: download (only valid IDs, only unknown IDs, or a validation report) or export to one of several FlyBase tools for further analysis. Users can export the list to a FlyBase HitList (*see* Subheading 3) to access the spectrum of tools available there, or to Batch Download (*see* below in Subheading 12.2) or Query Builder [61]. If an input list was recently built using a FlyBase search tool, it should not be necessary to validate the IDs. If the list was built at FlyBase using a release earlier than the current one, validation may be necessary, as IDs occasionally change due to gene model updates or other changes.

12.2 Downloading Data

The FlyBase Batch Download tool assembles downloadable files using FlyBase report fields or FlyBase bulk data files. Batch Download can be accessed via a button link on the FlyBase home page, or from the Tools link on the NavBar. Another entry point is from a FlyBase HitList; if Batch Download is accessed this way, the query field in the form page will be pre-populated with data from the HitList. Users may also paste query lists or upload a list from a file; several kinds of identifiers are supported. To begin, users must select a data source (Report fields or precomputed file) and an output destination (browser page or file download). After these selections are set, a second form page (one for Report fields, another for precomputed files) lets the user customize their output by selecting fields or files.

FlyBase precomputes bulk data files at each release cycle and provides these files for the current release through a page accessed from the Tools → Current Release menu on the NavBar. These precomputed files contain particular slices of FlyBase data that users or collaborators have requested over the years or that are otherwise difficult to obtain in bulk (*see* Table 3). Precomputed files from previous releases are also available at the FlyBase FTP site (http://ftp.flybase.net/releases/). These files include several useful correspondence tables and ontology files used at FlyBase. Precomputed FASTA files comprise many different cuts of genomic data, including annotation categories such as small RNA classes and pseudogenes; components of gene model annotations such as exons, introns, UTRs, and predicted translations; and other genome features such as transposons and intergenic sequences.

12.3 Programmatic Access

FlyBase provides API (Application Programming Interface) access to most of its data. See https://flybase.github.io for documentation and details on how to use these API endpoints. Data are returned in JSON format in most cases (with an exception: the Chado API returns ChadoXML), and there are light restrictions on access rates. FlyBase uses many of these API endpoints internally to power its tools.

Chado is a relational database schema for managing biological data that was created by FlyBase in 2006 [62, 63]. It is currently developed and maintained by the GMOD organization (http://gmod.org/wiki/Main_Page). PostgreSQL database dumps of the FlyBase Chado database are available for every release of FlyBase via FTP: ftp://ftp.flybase.org/releases/current/psql. Instructions for installation of a FlyBase Chado database instance are here: https://flybase.github.io/docs/chado/index. A Chado instance corresponding to a 2020 FlyBase release requires about 200 GB of disk space.

13 Help Resources

Resources to aid users are collected under the Help section of the NavBar, as well as in situ on FlyBase pages via question mark icons and mouseover text (particularly in JBrowse). The Help menu links include information about how to use the website, what authors can do to ensure FlyBase can extract the most information possible out of their paper, nomenclature rules, links for new *Drosophila* investigators, and more. These documents are available as a wiki at the Help Index entry of the Help menu. At the bottom of the Help menu is Contact FlyBase, where users can send a message directly to FlyBase curators. Two offsite help resources are the video tutorials on YouTube (https://www.youtube.com/c/FlyBaseTV) and the FlyBase Twitter account (*see* Subheading 14 below). Additional updates on FlyBase and news from the *Drosophila* research community are reported in the Commentaries section of the front page, which contains a link to all archived commentaries.

14 The FlyBase Community

FlyBase aims to interact with the *Drosophila* community in a variety of ways. Users can get directly in touch with FlyBase via the Contact FlyBase page, available on the Help menu on the NavBar and linked in the footer of any FlyBase page. Authors who are planning to submit manuscripts where they name a gene are particularly encouraged to contact FlyBase before submission, to ensure that they propose a unique gene name and symbol approved by FlyBase.

Users can keep up to date with FlyBase in a variety of ways. FlyBase maintains an active Twitter profile (https://twitter.com/FlyBaseDotOrg) that is used to alert users to new data, features, and news. Regular "Tweetorials" are featured on the feed, in which a series of linked tweets visually guide users through particular FlyBase features or tools; a listing of all hashtags that can be used to find specific Tweetorials and a link to all Tweetorials are maintained on the FlyBase Help wiki (https://wiki.flybase.org/wiki/FlyBase:Tweetorial). New features are also highlighted in the Commentaries section of the homepage. Users can stay up to date by signing up to the newsletter (generally sent out with each FlyBase release), available via the Community menu on the NavBar. FlyBase representatives are also often present at *Drosophila* conferences. Presentations and pamphlets from previously attended conferences can be found via the FlyBase Presentations link in the About menu on the NavBar.

Other community resources, such as the FlyBase Community Advisory Group (FCAG) or Fast-Track Your Paper (FTYP) tool, are available via the Community menu on the NavBar. FCAG

consists of a group of volunteers who are generally contacted a handful of times per year to fill out short surveys to provide FlyBase with feedback about upcoming or existing tools and features. Anyone can join this group via the FCAG page, and we encourage all users to sign up and help shape the future of FlyBase.

The FTYP tool allows first-pass curation of publications by authors, indicating to curators the types of data in the paper and associating relevant genes with the reference [64]. Most authors will interact with the tool in response to an email from FlyBase to the corresponding author, once their publication has been fully published and added to the FlyBase bibliography. Author use of this tool is highly beneficial to FlyBase, as it facilitates curation, and benefits authors as well, by accelerating the inclusion of data from their publication into FlyBase.

During the FTYP process, authors are also encouraged to submit Gene Snapshots (short descriptions about previously uncharacterized genes that have been studied in their paper), or to highlight that their paper includes a technical advance or new type of reagent or resource that is likely to be useful for other researchers. The latter publications (along with text provided by the author which describes the technical advance) are added to the Papers with Technical Advances page, available via the Community menu in the NavBar.

15 Alliance

FlyBase is a member of the Alliance of Genome Resources (Alliance; https://www.alliancegenome.org), a consortium of multiple model organism databases (MODs) which provides an integrated platform to access model organism, human, and GO data together [2, 65]. Since its founding in 2016, in addition to FlyBase, the Alliance has brought together the Mouse Genome Database (MGI), Rat Genome Database (RGD), Saccharomyces Genome Database (SGD), WormBase, Zebrafish Information Network (ZFIN), and the GO Consortium to provide a united portal for both model organism researchers and biomedical researchers looking to leverage the power of MODs. Development work for the Alliance benefits both it and the individual MODs, making for a more FAIR environment for all.

16 The Future of FlyBase

As *Drosophila* literature, techniques, and data evolve, FlyBase is continually developing new tools and features to bring these innovations to all Drosophilists. From its founding in 1992, FlyBase has expanded from its initial core role as the repository of *Drosophila*

genetics to curating and providing access to many other kinds of data. As of the writing of this chapter, two initiatives are in progress to bring new types of data and curation to FlyBase: single-cell RNA-seq datasets and key chemicals used in *Drosophila* experiments.

16.1 Single-Cell RNA-seq

In future releases, FlyBase will make it easier to exploit data obtained from single-cell RNA-seq (scRNA-seq) experiments. *Drosophila* scRNA-seq datasets will be inventoried in FlyBase and presented in Dataset Reports, which will include data on originating tissue, how the data was generated and processed, and where raw datasets can be found. These Dataset Reports will be linked from several other types of report to make the datasets more easily discoverable. For example, a Reference Report will point to all datasets obtained in the current paper; the Gene Report will point to all datasets in which the current gene has been detected; the CV Term Report for a cell type will point to all datasets in which cells of that cell type have been identified. In addition, the Gene Report page will leverage the data generated by the Fly Cell Atlas consortium (FCA; http://flycellatlas.org) which has recently performed exhaustive scRNA-seq experiments across all tissues of the adult fly, to propose a graphical overview of all major cell types in which the current gene has been found to be expressed. FCA data will be gradually incorporated into FlyBase in various ways, including cell-type expression ribbons.

16.2 Chemical Curation

The continued prominence of flies in translational studies, in work, such as drug treatment screens, is one of the many reasons why FlyBase is beginning to curate chemicals. Although the idea of "curating chemicals used in *Drosophila* experiments" could be very wide ranging, FlyBase is choosing to focus on chemicals that are used to distinguish the experimental and control conditions, i.e., chemicals whose effects are the focus of the experiment, rather than those that are simply part of the technical process of the experiment. Chemicals will be curated as they are cataloged by EMBL-EBI's CHEBI (Chemical Entities of Biological Interest) resource [66], wherever possible, with additional information provided by PubChem [67]. This allows FlyBase to import CHEBI's ontological structure and Roles Classification, which describes common functions with biological relevance (e.g., actin polymerization inhibitor, dopaminergic agent, neurotoxin). The first wave of chemical curation will concentrate on chemicals which are used to cause or treat disease, and will be integrated with Human Disease Model Reports (*see* Subheading 11.4).

17 Concluding Remarks

Over the three decades of its existence, FlyBase has seen *Drosophila* gene information change from something one looked up in a hefty book to something to easily query and manipulate from a laptop or phone. As the display medium has shrunk, the range of what Drosophilists can find and do at FlyBase has increased dramatically, branching out from genes and genetics to the full range of data described in this chapter. Here at FlyBase, we are excited to see new innovations and advances that fly research will bring, to continue to adapt and to improve FlyBase, and to continue making *Drosophila* data FAIR for all.

Acknowledgments

Grant support for FlyBase is provided by the National Human Genome Research Institute at the U.S. National Institutes of Health (U41HG000739), the National Science Foundation (DBI-2035515, 2039324), and the British Medical Research Council (MR/N030117/1). FlyBase would especially like to thank Drosophilists around the world who support FlyBase through their website access fees.

At the time of writing, the FlyBase Consortium includes: Norbert Perrimon (PI), Nick Brown (Co-PI), Brian Calvi (Co-PI), Richard Cripps (Co-PI), Thomas Kaufman (Co-PI), Susan Russo Gelbart (PD), Julie Agapite, Giulia Antonazzo, Helen Attrill, Kris Broll, Lynn Crosby, Gil dos Santos, Kathleen Falls, Phani Garapati, Josh Goodman, Damien Goutte-Gattat, L. Sian Gramates, Victoria Jenkins, Pravija Krishna, Aoife Larkin, Ian Longden, TyAnna Lovato, Steven Marygold, Beverley Matthews, Alex McLachlan, Gillian Millburn, Arzu Ozturk-Colak, Clare Pilgrim, Jolene Seme, Victor Strelets, Christopher J. Tabone, Jim Thurmond, Vitor Trovisco, Pinglei Zhou, and Mark Zytkovicz.

References

1. Flybase Consortium (1996) FlyBase: the Drosophila database. Nucleic Acids Res 24(1): 53–56. https://doi.org/10.1093/nar/24.1.53
2. Alliance of Genome Resources Consortium (2020) Alliance of Genome Resources Portal: unified model organism research platform. Nucleic Acids Res 48(D1):D650–D658. https://doi.org/10.1093/nar/gkz813
3. Wilkinson MD, Dumontier M, Sansone SA, da Silva B, Santos LO, Prieto M, Batista D, McQuilton P, Kuhn T, Rocca-Serra P, Crosas M, Schultes E (2019) Evaluating FAIR maturity through a scalable, automated, community-governed framework. Sci Data 6(1):174. https://doi.org/10.1038/s41597-019-0184-5
4. Gramates LS, Agapite J, Attrill H, Calvi BR, Crosby M, dos Santos G, Goodman JL, Goutte-Gattat D, Jenkins VK, Kaufman T, Larkin A, Matthews B, Millburn GH, Strelets V (2022) FlyBase: a guided tour of highlighted features. Genetics 220(4):iyac035. https://doi.org/10.1093/genetics/iyac035

5. Marygold SJ, Antonazzo G, Attrill H, Costa M, Crosby MA, Dos Santos G, Goodman JL, Gramates LS, Matthews BB, Rey AJ, Thurmond J, FlyBase Consortium (2016) Exploring FlyBase data using QuickSearch. Curr Protoc Bioinformatics 56:1.31.1-1.31.23. https://doi.org/10.1002/cpbi.19
6. Amberger JS, Hamosh A (2017) Searching Online Mendelian Inheritance in Man (OMIM): a knowledgebase of human genes and genetic phenotypes. Curr Protoc Bioinformatics 58:1.2.1-1.2.12. https://doi.org/10.1002/cpbi.27
7. Braschi B, Denny P, Gray K, Jones T, Seal R, Tweedie S, Yates B, Bruford E (2019) Genenames.org: the HGNC and VGNC resources in 2019. Nucleic Acids Res 47(D1): D786–D792. https://doi.org/10.1093/nar/gky930
8. Groß A, Pruski C, Rahm E (2016) Evolution of biomedical ontologies and mappings: overview of recent approaches. Comput Struct Biotechnol J 14:333–340. https://doi.org/10.1016/j.csbj.2016.08.002
9. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet 25(1):25–29. https://doi.org/10.1038/75556
10. Gene Ontology Consortium (2021) The Gene Ontology resource: enriching a GOld mine. Nucleic Acids Res 49(D1):D325–D334. https://doi.org/10.1093/nar/gkaa1113
11. Mungall CJ, Batchelor C, Eilbeck K (2011) Evolution of the Sequence Ontology terms and relationships. J Biomed Inform 44(1): 87–93. https://doi.org/10.1016/j.jbi.2010.03.002
12. Sivade Dumousseau M, Alonso-Lopez D, Ammari M, Bradley G, Campbell NH, Ceol A, Cesareni G, Combe C, De Las Rivas J, Del-Toro N, Heimbach J, Hermjakob H, Jurisica I, Koch M, Licata L, Lovering RC, Lynn DJ, Meldal BHM, Micklem G, Panni S, Porras P, Ricard-Blum S, Roechert B, Salwinski L, Shrivastava A, Sullivan J, Thierry-Mieg N, Yehudi Y, Van Roey K, Orchard S (2018) Encompassing new use cases - level 3.0 of the HUPO-PSI format for molecular interactions. BMC Bioinformatics 19(1):134. https://doi.org/10.1186/s12859-018-2118-1
13. Bello SM, Shimoyama M, Mitraka E, Laulederkind SJF, Smith CL, Eppig JT, Schriml LM (2018) Disease ontology: improving and unifying disease annotations across species. Dis Model Mech 11(3). https://doi.org/10.1242/dmm.032839
14. McDonald AG, Tipton KF (2014) Fifty-five years of enzyme classification: advances and difficulties. FEBS J 281(2):583–592. https://doi.org/10.1111/febs.12530
15. Garapati PV, Zhang J, Rey AJ, Marygold SJ (2019) Towards comprehensive annotation of Drosophila melanogaster enzymes in FlyBase. Database (Oxford) 2019:bay144. https://doi.org/10.1093/database/bay144
16. Antonazzo G, Urbano JM, Marygold SJ, Millburn GH, Brown NH (2020) Building a pipeline to solicit expert knowledge from the community to aid gene summary curation. Database (Oxford) 2020:baz152. https://doi.org/10.1093/database/baz152
17. Buels R, Yao E, Diesh CM, Hayes RD, Munoz-Torres M, Helt G, Goodstein DM, Elsik CG, Lewis SE, Stein L, Holmes IH (2016) JBrowse: a dynamic web platform for genome visualization and analysis. Genome Biol 17:66. https://doi.org/10.1186/s13059-016-0924-1
18. UniProt Consortium (2019) UniProt: a worldwide hub of protein knowledge. Nucleic Acids Res 47(D1):D506–D515. https://doi.org/10.1093/nar/gky1049
19. Blum M, Chang HY, Chuguransky S, Grego T, Kandasaamy S, Mitchell A, Nuka G, Paysan-Lafosse T, Qureshi M, Raj S, Richardson L, Salazar GA, Williams L, Bork P, Bridge A, Gough J, Haft DH, Letunic I, Marchler-Bauer-A, Mi H, Natale DA, Necci M, Orengo CA, Pandurangan AP, Rivoire C, Sigrist CJA, Sillitoe I, Thanki N, Thomas PD, Tosatto SCE, Wu CH, Bateman A, Finn RD (2021) The InterPro protein families and domains database: 20 years on. Nucleic Acids Res 49 (D1):D344–D354. https://doi.org/10.1093/nar/gkaa977
20. Hu Y, Flockhart I, Vinayagam A, Bergwitz C, Berger B, Perrimon N, Mohr SE (2011) An integrative approach to ortholog prediction for disease-focused and other functional studies. BMC Bioinformatics 12:357. https://doi.org/10.1186/1471-2105-12-357
21. Zdobnov EM, Kuznetsov D, Tegenfeldt F, Manni M, Berkeley M, Kriventseva EV (2021) OrthoDB in 2020: evolutionary and functional annotations of orthologs. Nucleic Acids Res 49 (D1):D389–D393. https://doi.org/10.1093/nar/gkaa1009
22. Marygold SJ, Leyland PC, Seal RL, Goodman JL, Thurmond J, Strelets VB, Wilson RJ, FlyBase Consortium (2013) FlyBase: improvements to the bibliography. Nucleic Acids Res

41(Database issue):D751–D757. https://doi.org/10.1093/nar/gks1024

23. Morgan TH (1910) Sex limited inheritance in Drosophila. Science 32(812):120–122. https://doi.org/10.1126/science.32.812.120
24. Costa M, Reeve S, Grumbling G, Osumi-Sutherland D (2013) The Drosophila anatomy ontology. J Biomed Semant 4(1):32. https://doi.org/10.1186/2041-1480-4-32
25. Osumi-Sutherland D, Marygold SJ, Millburn GH, McQuilton PA, Ponting L, Stefancsik R, Falls K, Brown NH, Gkoutos GV (2013) The Drosophila phenotype ontology. J Biomed Semant 4(1):30. https://doi.org/10.1186/2041-1480-4-30
26. Graveley BR, Brooks AN, Carlson JW, Duff MO, Landolin JM, Yang L, Artieri CG, van Baren MJ, Boley N, Booth BW, Brown JB, Cherbas L, Davis CA, Dobin A, Li R, Lin W, Malone JH, Mattiuzzo NR, Miller D, Sturgill D, Tuch BB, Zaleski C, Zhang D, Blanchette M, Dudoit S, Eads B, Green RE, Hammonds A, Jiang L, Kapranov P, Langton L, Perrimon N, Sandler JE, Wan KH, Willingham A, Zhang Y, Zou Y, Andrews J, Bickel PJ, Brenner SE, Brent MR, Cherbas P, Gingeras TR, Hoskins RA, Kaufman TC, Oliver B, Celniker SE (2011) The developmental transcriptome of Drosophila melanogaster. Nature 471(7339):473–479. https://doi.org/10.1038/nature09715
27. Buchon N, Osman D, David FP, Fang HY, Boquete JP, Deplancke B, Lemaitre B (2013) Morphological and molecular characterization of adult midgut compartmentalization in Drosophila. Cell Rep 3(5):1725–1738. https://doi.org/10.1016/j.celrep.2013.04.001
28. Chintapalli VR, Wang J, Dow JA (2007) Using FlyAtlas to identify better Drosophila melanogaster models of human disease. Nat Genet 39(6):715–720. https://doi.org/10.1038/ng2049
29. Hammonds AS, Bristow CA, Fisher WW, Weiszmann R, Wu S, Hartenstein V, Kellis M, Yu B, Frise E, Celniker SE (2013) Spatial expression of transcription factors in Drosophila embryonic organ development. Genome Biol 14(12):R140. https://doi.org/10.1186/gb-2013-14-12-r140
30. Wilk R, Hu J, Blotsky D, Krause HM (2016) Diverse and pervasive subcellular distributions for both coding and long noncoding RNAs. Genes Dev 30(5):594–609. https://doi.org/10.1101/gad.276931.115
31. Kumar S, Konikoff C, Van Emden B, Busick C, Davis KT, Ji S, Wu LW, Ramos H, Brody T, Panchanathan S, Ye J, Karr TL, Gerold K, McCutchan M, Newfeld SJ (2011) FlyExpress: visual mining of spatiotemporal patterns for genes and publications in Drosophila embryogenesis. Bioinformatics 27(23):3319–3320. https://doi.org/10.1093/bioinformatics/btr567
32. Brown JB, Boley N, Eisman R, May GE, Stoiber MH, Duff MO, Booth BW, Wen J, Park S, Suzuki AM, Wan KH, Yu C, Zhang D, Carlson JW, Cherbas L, Eads BD, Miller D, Mockaitis K, Roberts J, Davis CA, Frise E, Hammonds AS, Olson S, Shenker S, Sturgill D, Samsonova AA, Weiszmann R, Robinson G, Hernandez J, Andrews J, Bickel PJ, Carninci P, Cherbas P, Gingeras TR, Hoskins RA, Kaufman TC, Lai EC, Oliver B, Perrimon N, Graveley BR, Celniker SE (2014) Diversity and dynamics of the Drosophila transcriptome. Nature 512(7515):393–399. https://doi.org/10.1038/nature12962
33. Casas-Vila N, Bluhm A, Sayols S, Dinges N, Dejung M, Altenhein T, Kappei D, Altenhein B, Roignant JY, Butter F (2017) The developmental proteome of Drosophila melanogaster. Genome Res 27(7):1273–1285. https://doi.org/10.1101/gr.213694.116
34. Bean DM, Heimbach J, Ficorella L, Micklem G, Oliver SG, Favrin G (2014) esyN: network building, sharing and publishing. PLoS One 9(9):e106035. https://doi.org/10.1371/journal.pone.0106035
35. Oughtred R, Rust J, Chang C, Breitkreutz BJ, Stark C, Willems A, Boucher L, Leung G, Kolas N, Zhang F, Dolma S, Coulombe-Huntington J, Chatr-Aryamontri A, Dolinski K, Tyers M (2021) The BioGRID database: a comprehensive biomedical resource of curated protein, genetic, and chemical interactions. Protein Sci 30(1):187–200. https://doi.org/10.1002/pro.3978
36. Murali T, Pacifico S, Yu J, Guest S, Roberts GG 3rd, Finley RL Jr (2011) DroID 2011: a comprehensive, integrated resource for protein, transcription factor, RNA and gene interactions for Drosophila. Nucleic Acids Res 39 (Database issue):D736–D743. https://doi.org/10.1093/nar/gkq1092
37. Hu Y, Vinayagam A, Nand A, Comjean A, Chung V, Hao T, Mohr SE, Perrimon N (2018) Molecular Interaction Search Tool (MIST): an integrated resource for mining gene and protein interaction data. Nucleic Acids Res 46(D1):D567–D574. https://doi.org/10.1093/nar/gkx1116
38. Misra S, Crosby MA, Mungall CJ, Matthews BB, Campbell KS, Hradecky P, Huang Y,

Kaminker JS, Millburn GH, Prochnik SE, Smith CD, Tupy JL, Whitfied EJ, Bayraktaroglu L, Berman BP, Bettencourt BR, Celniker SE, de Grey AD, Drysdale RA, Harris NL, Richter J, Russo S, Schroeder AJ, Shu SQ, Stapleton M, Yamada C, Ashburner M, Gelbart WM, Rubin GM, Lewis SE (2002) Annotation of the Drosophila melanogaster euchromatic genome: a systematic review. Genome Biol 3(12):RESEARCH0083. https://doi.org/10.1186/gb-2002-3-12-research0083

39. dos Santos G, Schroeder AJ, Goodman JL, Strelets VB, Crosby MA, Thurmond J, Emmert DB, Gelbart WM, FlyBase C (2015) FlyBase: introduction of the Drosophila melanogaster Release 6 reference genome assembly and large-scale migration of genome annotations. Nucleic Acids Res 43(Database issue): D690–D697. https://doi.org/10.1093/nar/gku1099
40. Matthews BB, Dos Santos G, Crosby MA, Emmert DB, St Pierre SE, Gramates LS, Zhou P, Schroeder AJ, Falls K, Strelets V, Russo SM, Gelbart WM, FlyBase Consortium (2015) Gene model annotations for Drosophila melanogaster: impact of high-throughput data. G3 (Bethesda) 5(8):1721–1736. https://doi.org/10.1534/g3.115.018929
41. Hoskins RA, Landolin JM, Brown JB, Sandler JE, Takahashi H, Lassmann T, Yu C, Booth BW, Zhang D, Wan KH, Yang L, Boley N, Andrews J, Kaufman TC, Graveley BR, Bickel PJ, Carninci P, Carlson JW, Celniker SE (2011) Genome-wide analysis of promoter architecture in Drosophila melanogaster. Genome Res 21(2):182–192. https://doi.org/10.1101/gr.112466.110
42. Crosby MA, Gramates LS, Dos Santos G, Matthews BB, St Pierre SE, Zhou P, Schroeder AJ, Falls K, Emmert DB, Russo SM, Gelbart WM, FlyBase Consortium (2015) Gene model annotations for Drosophila melanogaster: the rule-benders. G3 (Bethesda) 5(8):1737–1749. https://doi.org/10.1534/g3.115.018937
43. Benson DA, Cavanaugh M, Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW (2013) GenBank. Nucleic Acids Res 41(Database issue):D36–D42. https://doi.org/10.1093/nar/gks1195
44. El-Gebali S, Mistry J, Bateman A, Eddy SR, Luciani A, Potter SC, Qureshi M, Richardson LJ, Salazar GA, Smart A, Sonnhammer ELL, Hirsh L, Paladin L, Piovesan D, Tosatto SCE, Finn RD (2019) The Pfam protein families database in 2019. Nucleic Acids Res 47(D1): D427–D432. https://doi.org/10.1093/nar/gky995
45. Letunic I, Bork P (2018) 20 years of the SMART protein domain annotation resource. Nucleic Acids Res 46(D1):D493–D496. https://doi.org/10.1093/nar/gkx922
46. Rivera J, Keranen SVE, Gallo SM, Halfon MS (2019) REDfly: the transcriptional regulatory element database for Drosophila. Nucleic Acids Res 47(D1):D828–D834. https://doi.org/10.1093/nar/gky957
47. Dreos R, Ambrosini G, Groux R, Cavin Perier R, Bucher P (2017) The eukaryotic promoter database in its 30th year: focus on non-vertebrate organisms. Nucleic Acids Res 45(D1):D51–D55. https://doi.org/10.1093/nar/gkw1069
48. Hoskins RA, Carlson JW, Wan KH, Park S, Mendez I, Galle SE, Booth BW, Pfeiffer BD, George RA, Svirskas R, Krzywinski M, Schein J, Accardo MC, Damia E, Messina G, Mendez-Lago M, de Pablos B, Demakova OV, Andreyeva EN, Boldyreva LV, Marra M, Carvalho AB, Dimitri P, Villasante A, Zhimulev IF, Rubin GM, Karpen GH, Celniker SE (2015) The Release 6 reference sequence of the Drosophila melanogaster genome. Genome Res 25(3):445–458. https://doi.org/10.1101/gr.185579.114
49. dos Santos G, Schroeder AJ, Goodman JL, Strelets VB, Crosby MA, Thurmond J, Emmert DB, Gelbart WM, FlyBase Consortium (2015) FlyBase: introduction of the Drosophila melanogaster Release 6 reference genome assembly and large-scale migration of genome annotations. Nucleic Acids Res 43 (Database issue):D690–D697. https://doi.org/10.1093/nar/gku1099
50. Daines B, Wang H, Wang L, Li Y, Han Y, Emmert D, Gelbart W, Wang X, Li W, Gibbs R, Chen R (2011) The Drosophila melanogaster transcriptome by paired-end RNA sequencing. Genome Res 21(2):315–324. https://doi.org/10.1101/gr.107854.110
51. Mackay TF, Richards S, Stone EA, Barbadilla A, Ayroles JF, Zhu D, Casillas S, Han Y, Magwire MM, Cridland JM, Richardson MF, Anholt RR, Barron M, Bess C, Blankenburg KP, Carbone MA, Castellano D, Chaboub L, Duncan L, Harris Z, Javaid M, Jayaseelan JC, Jhangiani SN, Jordan KW, Lara F, Lawrence F, Lee SL, Librado P, Linheiro RS, Lyman RF, Mackey AJ, Munidasa M, Muzny DM, Nazareth L, Newsham I, Perales L, Pu LL, Qu C, Ramia M, Reid JG, Rollmann SM, Rozas J, Saada N, Turlapati L, Worley KC, Wu YQ, Yamamoto A, Zhu Y, Bergman CM, Thornton KR, Mittelman D, Gibbs RA (2012) The Drosophila melanogaster genetic reference panel. Nature

482(7384):173–178. https://doi.org/10.1038/nature10811

52. Stapleton M, Liao G, Brokstein P, Hong L, Carninci P, Shiraki T, Hayashizaki Y, Champe M, Pacleb J, Wan K, Yu C, Carlson J, George R, Celniker S, Rubin GM (2002) The Drosophila gene collection: identification of putative full-length cDNAs for 70% of D. melanogaster genes. Genome Res 12(8): 1294–1300. https://doi.org/10.1101/gr.269102
53. Attrill H, Falls K, Goodman JL, Millburn GH, Antonazzo G, Rey AJ, Marygold SJ, FlyBase Consortium (2016) FlyBase: establishing a Gene Group resource for Drosophila melanogaster. Nucleic Acids Res 44(D1): D786–D792. https://doi.org/10.1093/nar/gkv1046
54. Rey AJ, Attrill H, Marygold SJ, FlyBase Consortium (2018) Using FlyBase to find functionally related drosophila genes. Methods Mol Biol 1757:493–512. https://doi.org/10.1007/978-1-4939-7737-6_16
55. Grove C, Cain S, Chen WJ, Davis P, Harris T, Howe KL, Kishore R, Lee R, Paulini M, Raciti D, Tuli MA, Van Auken K, Williams G, WormBase Consortium (2018) Using WormBase: a genome biology resource for Caenorhabditis elegans and related nematodes. Methods Mol Biol 1757:399–470. https://doi.org/10.1007/978-1-4939-7737-6_14
56. Tweedie S, Braschi B, Gray K, Jones TEM, Seal RL, Yates B, Bruford EA (2021) Genenames.org: the HGNC and VGNC resources in 2021. Nucleic Acids Res 49(D1):D939–D946. https://doi.org/10.1093/nar/gkaa980
57. Millburn GH, Crosby MA, Gramates LS, Tweedie S, FlyBase Consortium (2016) FlyBase portals to human disease research using Drosophila models. Dis Model Mech 9(3): 245–252. https://doi.org/10.1242/dmm.023317
58. Howe DG, Blake JA, Bradford YM, Bult CJ, Calvi BR, Engel SR, Kadin JA, Kaufman TC, Kishore R, Laulederkind SJF, Lewis SE, Moxon SAT, Richardson JE, Smith C (2018) Model organism data evolving in support of translational medicine. Lab Anim (NY) 47 (10):277–289. https://doi.org/10.1038/s41684-018-0150-4
59. Sobreira N, Schiettecatte F, Valle D, Hamosh A (2015) GeneMatcher: a matching tool for connecting investigators with an interest in the same gene. Hum Mutat 36(10):928–930. https://doi.org/10.1002/humu.22844
60. Wangler MF, Yamamoto S, Chao HT, Posey JE, Westerfield M, Postlethwait J, Members of the Undiagnosed Diseases Network, Hieter P, Boycott KM, Campeau PM, Bellen HJ (2017) Model organisms facilitate rare disease diagnosis and therapeutic research. Genetics 207(1): 9–27. https://doi.org/10.1534/genetics.117.203067
61. Wilson RJ, Goodman JL, Strelets VB, FlyBase Consortium (2008) FlyBase: integration and improvements to query tools. Nucleic Acids Res 36(Database issue):D588–D593. https://doi.org/10.1093/nar/gkm930
62. Mungall CJ, Emmert DB, FlyBase Consortium (2007) A Chado case study: an ontology-based modular schema for representing genome-associated biological information. Bioinformatics 23(13):i337–i346. https://doi.org/10.1093/bioinformatics/btm189
63. Zhou P, Emmert D, Zhang P (2006) Using Chado to store genome annotation data. Curr Protoc Bioinformatics 12:9.6.1-9.6.28. https://doi.org/10.1002/0471250953.bi0906s12
64. Bunt SM, Grumbling GB, Field HI, Marygold SJ, Brown NH, Millburn GH, FlyBase Consortium (2012) Directly e-mailing authors of newly published papers encourages community curation. Database (Oxford) 2012:bas024. https://doi.org/10.1093/database/bas024
65. Alliance of Genome Resources Consortium (2019) The Alliance of genome resources: building a modern data ecosystem for model organism databases. Genetics 213(4): 1189–1196. https://doi.org/10.1534/genetics.119.302523
66. Hastings J, Owen G, Dekker A, Ennis M, Kale N, Muthukrishnan V, Turner S, Swainston N, Mendes P, Steinbeck C (2016) ChEBI in 2016: improved services and an expanding collection of metabolites. Nucleic Acids Res 44(D1):D1214–D1219. https://doi.org/10.1093/nar/gkv1031
67. Kim S, Chen J, Cheng T, Gindulyte A, He J, He S, Li Q, Shoemaker BA, Thiessen PA, Yu B, Zaslavsky L, Zhang J, Bolton EE (2021) PubChem in 2021: new data content and improved web interfaces. Nucleic Acids Res 49(D1): D1388–D1395. https://doi.org/10.1093/nar/gkaa971

Chapter 2

The Q-system: A Versatile Repressible Binary Expression System

Orsolya Fölsz, Chun-Chieh Lin, Darya Task, Olena Riabinina, and Christopher J. Potter

Abstract

Binary expression systems are useful genetic tools for experimentally labeling or manipulating the function of defined cells. The Q-system is a repressible binary expression system that consists of a transcription factor QF (and the recently improved QF2/QF2^{w}), the inhibitor QS, a *QUAS-geneX* effector, and a drug that inhibits QS (quinic acid). The Q-system can be used alone or in combination with other binary expression systems, such as GAL4/*UAS* and LexA/*LexAop*. In this review chapter, we discuss the past, present, and future of the Q-system for applications in *Drosophila* and other organisms. We discuss the in vivo application of the Q-system for transgenic labeling, the modular nature of QF that allows chimeric or split transcriptional activators to be developed, its temporal control by quinic acid, new methods to generate QF2 reagents, intersectional expression labeling, and its recent adoption into many emerging experimental species.

Key words *Drosophila*, Synthetic biology, Split-QF, Chimeric transactivators, HACK, Mosquito, *Neurospora crassa*, Bioengineering

1 Introduction

Repressible binary expression systems are versatile and powerful genetic tools in *Drosophila*. They allow flexible labeling and tissue-specific manipulations over an experimentally defined population of cells. The GAL4/*UAS* system is the most widely adopted binary expression system in *Drosophila* [1]. In many experimental situations, the GAL4 system by itself is sufficient. However, the investigation of complex systems (such as the brain) often requires the ability to manipulate two populations of cells independently. For example, an experimenter might wish to simultaneously express different effectors (e.g., RNAi or GFP) in two different cell populations (e.g., neurons or glia). Under these more demanding experimental requirements, a second repressible binary expression system

Christian Dahmann (ed.), *Drosophila: Methods and Protocols*,
Methods in Molecular Biology, vol. 2540, https://doi.org/10.1007/978-1-0716-2541-5_2,

is needed. In this chapter, we first discuss the origins of the GAL4/*UAS* system as it provided the conceptual framework for identifying the independent Q-system of binary expression. We next discuss the components of the Q-system, how these components can be utilized with GAL4/*UAS* and LexA/*LexAop* expression systems, common uses for the Q-system (intersectional expression patterns, temporal expression control using quinic acid, HACK gene conversion), and finally how the Q-system has been adapted for use in other organisms. As more species become accessible to genetic manipulations, the creative ways the Q-system is utilized in these other species can inform how it can be further implemented for use in *Drosophila*.

1.1 Origins of a Repressible Binary System: The Saccharomyces cerevisiae *GAL Gene Cluster*

The GAL4/*UAS* system originates from the GAL genes in the yeast *Saccharomyces cerevisiae*. The GAL genes are involved in the catabolism of galactose [2]. GAL4 is a transcription factor involved in the transcriptional upregulation of galactose catabolism genes in yeast. In situations where glucose is limiting, galactose can be utilized as an energy source through the action of 5 GAL genes (MELI, GAL1, GAL2, GAL7, GAL10). In yeast, GAL4 itself is negatively regulated (repressed) by GAL80, which binds to GAL4 and blocks its transcriptional activity. Under steady-state conditions, the presence of GAL80 prevents GAL4 from activating GAL genes. However, GAL80 itself can be repressed in yeast by the GAL3 protein that binds to GAL80, and whose binding to GAL80 is regulated by the presence of galactose [3]. As such, the presence of galactose in yeast leads to repression of GAL80 by GAL3, allowing GAL4 to turn on the transcription of genes involved in galactose catabolism. The GAL4/GAL80 components from yeast have been used broadly in *Drosophila*.

1.2 Neurospora crassa *qa Gene Cluster*

To adapt an independent expression system with similar properties to GAL4/*UAS*, we turned to the quinic acid (qa) repressible expression network from the bread fungus *Neurospora crassa* [4]. Extensive work, primarily by Dr. Norman Giles at the University of Georgia, characterized the qa regulatory network as being similar in design to the GAL network from yeast (Fig. 1a). Similar to the GAL system, the fungal qa regulatory network allows the catabolism of quinic acid, a small cyclohexanecarboxylic acid, in situations where glucose is scarce and quinic acid is present. Like the GAL system, the catabolite (quinic acid) induces expression of the genes needed for its catabolism (Fig. 1a). The qa gene cluster consists of 5 structural qa genes involved in the catabolism or cellular transport of quinic acid (*qa-x*, *qa-2*, *qa-3*, *qa-4*, *qa-y*) and two regulatory genes (*qa-1F* and *qa-1S*) involved in the expression of the gene cluster. The transcription factor qa-1F (abbreviated hereafter as QF) binds to DNA upstream of the *qa* genes and turns on expression of the qa gene cluster. The activity of QF is

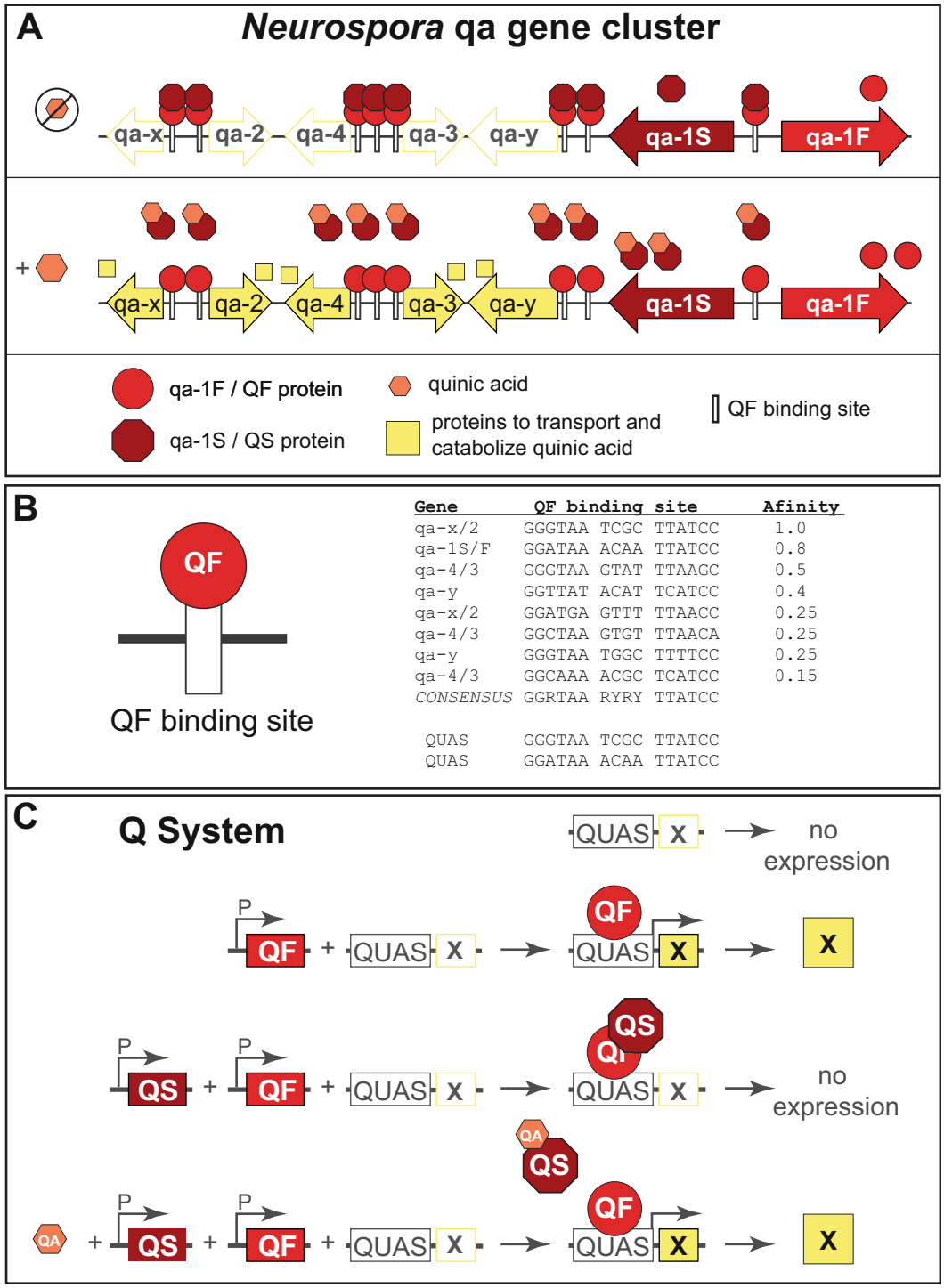


Fig. 1 Development of the Q-system from the *Neurospora qa* gene cluster. (**a**) Schematic of the quinic acid (qa) gene cluster from *Neurospora crassa*. In the absence of quinic acid, the gene cluster is silenced due to the activity of the qa-1S (QS) suppressor. In the presence of quinic acid, the QS suppressor is suppressed by quinic acid, allowing the qa-1F (QF) transcription factor to activate expression of genes involved in the efficient catabolism of qa. (**b**) Development of the *QUAS* site. The QF binding sites for the qa genes are shown, along

inhibited by qa-1S (hereafter abbreviated as QS), which binds directly to QF and blocks its transcriptional activity. In the absence of quinic acid, newly synthesized QF is bound by its repressor QS. However, in the presence of quinic acid, quinic acid binds to QS and inhibits its binding to QF. This allows QF to upregulate expression of qa genes. The regulatory genes, QF and QS, are thus analogous to GAL4 and GAL80, and represent the primary components of a repressible binary expression system.

The final component necessary to develop a repressible binary expression system for use in flies is the generation of a gene whose expression is regulated by QF. The QF DNA binding sites in the qa cluster (the QF enhancer sequence) were identified, and their binding affinity for QF was determined (Fig. 1b) [4–6]. The QF binding site is a hexameric palindrome with an internal 4 base pair spacer for a total of 20 base pairs. Note, two QF molecules bind to each single QF binding site [5]. However, for simplicity, only a single QF molecule is usually shown bound at a single site. From these *Neurospora* studies, the QF binding site consensus is defined as 5′-GGRTAA RYRY TTATCC-3′ where R is A/G and Y is C/T. Two sequences (5′-GGGTAA TCGC TTATCC-3′ and 5′-GGATAA ACAA TTATCC-3′) demonstrated the greatest affinity for QF [5]. These two sequences were chosen for use as the *Q-system Upstream Activation Sequences* (*QUAS*) incorporated into *QUAS-geneX* effectors. The number of *QUAS* sites can vary, with most *QUAS-geneX* effectors using 5 QF binding sites. The *QUAS* enhancer requires a minimal promoter to effectively induce gene expression in *Drosophila*; the *hsp70 minimal TATA* promoter is often used for this purpose.

1.3 Components of the Q-system

The Q-system, as adapted for use in flies, contains four components (Fig. 1c). The first is the *QUAS-geneX* effector (sometime referred to as a "reporter"). In the absence of the driver QF, the reporter should be silent, indicating a lack of endogenous transcription factors that can bind to the *QUAS* enhancer sequence. In the presence of QF, QF will bind *QUAS* sites, recruit transcriptional machinery to the promoter sequence, and induce expression of the effector gene. If the repressor QS is also present, QS will bind QF and inhibit its activity, effectively shutting down expression at

Fig. 1 (continued) with the affinity for QF for each site (as determined by in vitro binding assays) [5]. The *QUAS* site is derived from the highest affinity QF binding DNA sequences. The *5xQUAS* commonly used contains 4 copies of the high affinity qa-X/2 site and a single copy of the qa-1S/F binding site. (**c**) Q-system components. The *QUAS-geneX* effector is silent on its own. QF, driven by a tissue-specific promoter (P), binds to the *QUAS* enhancer and induces transcriptional expression. The activity of QF can be suppressed by the QS protein, which binds directly to QF. The QS suppressor in turn can be suppressed by the binding of quinic acid (QA). This acts to functionally induce expression from the *QUAS-geneX* effector previously silenced by QS

QUAS-geneX genes. QS in turn can be inhibited by the presence of the non-toxic drug quinic acid (QA), which binds QS and disrupts its binding to QF. QF activity is thus restored, allowing it to induce expression of *QUAS-geneX* effector genes. The Q-system is similar in design to the GAL4 system with the exception that the suppressor (QS) is directly regulated by a catabolite (QA); in the GAL4 system, the GAL80 suppressor is not directly regulated by the catabolite galactose, but instead utilizes a third protein (GAL3) for this purpose. However, GAL3 is not normally used in *Drosophila* to regulate GAL4/*UAS* system; instead, a temperature-sensitive version of GAL80 is used [7], and the expression is controlled by exposing flies to restrictive or permissive temperatures.

1.4 Modular Domains of Binary Expression Systems Enable Chimeric Transcription Factors

Repressible binary expression systems rely on the ability of GAL4 and QF2 to act as transcription factors. Transcription factor proteins contain two necessary functional domains. The DNA-binding domain (DBD), typically at the N-terminus of the protein, allows the transcription factor to bind specific enhancer DNA sequences (e.g., *UAS* or *QUAS*). The transcriptional activation domain (AD), typically at the C-terminus of the protein, recruits transcriptional machinery (RNA polymerase II and co-factors) to a promoter to begin transcription. The suppressors of transcription factors (QS and GAL80) bind directly to the activation domains to inhibit the recruitment of transcriptional machinery. A large middle domain between the DBD and AD likely allows for structural stability or modulation of transcriptional activity by external proteins (Figs. 2 and 3).

Along with GAL4/UAS and Q-system, the LexA/LexAop system has been widely adopted for use as a binary expression system in flies (Fig. 2). The LexA/LexAop system originates from the LexA repressor from *Escherichia coli* bacteria which can bind to DNA to inhibit transcription [8]. While LexA does contain a DNA-binding domain, it is not a transcription factor and does not contain a transcriptional activation domain. To be used as a binary expression system in *Drosophila*, the LexA DBD is fused with an exogenous AD, such as p65 AD [9], GAL4 AD [10], or the viral transcription factor VP16 acidic activation domain (VP16) [10]. The three binary expression systems do not cross-react and can be used simultaneously. Furthermore, given the modularity of the transcription factor protein domains, domains can be mixed and matched to generate chimeric proteins that take on the qualities of the utilized domains (Figs. 2 and 3) (see more in Subheading 2).

1.5 QF2, an Improved QF Transcription Factor for Transgenesis

The QF transcription factor from *Neurospora,* utilized in the initial publication of the Q-system [11], appeared to be toxic when broadly expressed. We performed a series of modifications to QF to address this issue [12]. As a start, the full-length QF protein was re-codonized for expression in *Drosophila* to achieve moderate

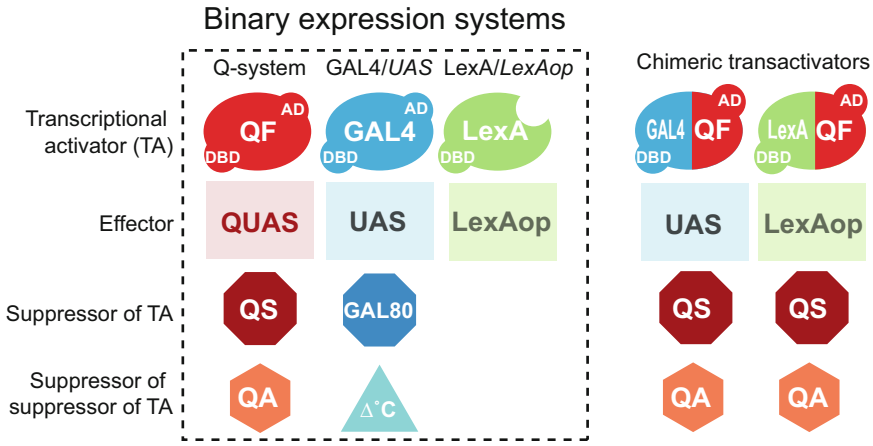


Fig. 2 Binary and chimeric expression systems. The components of the three main binary expression systems used in flies. A temperature-sensitive version of GAL80 (triangle) can be used to temporally control GAL4 [7]. Chimeric transcription factors combine the modular domains of GAL4/QF/LexA to determine which effectors or suppressors remain functional. DBD DNA-binding domain, AD activation domain

expression strength (QFrco). This also had the added benefit of removing a cryptic tracheal enhancer sequence from the original *Neurospora* QF DNA sequence. To directly address the issue of QF toxicity, a number of chimeric transcription factors were generated between GAL4, QF, and LexA to identify which QF domains, if present, led to issues of transgenic transcription factor expression when broadly expressed [12]. Chimeric constructs were generated using the strong pan-neuronal enhancer/promoters from *n-synaptobrevin* or *tubulin*. These experiments pinpointed the middle region between the QF DNA-binding domain and the QF activation domain as being a major source of toxicity when present in *Drosophila*: any construct that contained this domain failed to produce broadly expressed transgenics. Remarkably, deletion of this large domain and fusing the QF DBD with the QF AD resulted in a functional QF transcription factor that could be broadly expressed (Fig. 3). This version of QF was called QF2. QF2 retains high activity levels and is repressible by QS, similarly to the original QF [12].

The QF activation domain is more potent than the GAL4 activation domain, and while pan-neuronal or pan-tissue expression of QF2 leads to viable transgenics, the health of transgenics could still be somewhat compromised. Since the activation domain of QF depends on an acidic (negative) charge [13], this region was mutated to reduce its potency to create a "weakened" $QF2^{w}$. To this end, the last two amino acids (glutamic acid and glutamine) on the C-terminus of QF2 were replaced by four lysines (K^+ K^+ K^+ K^+) in $QF2^{w}$ (Fig. 3), which reduced its activity and increased the health of transgenic flies. Interestingly, $QF2^{w}$ is also more robustly

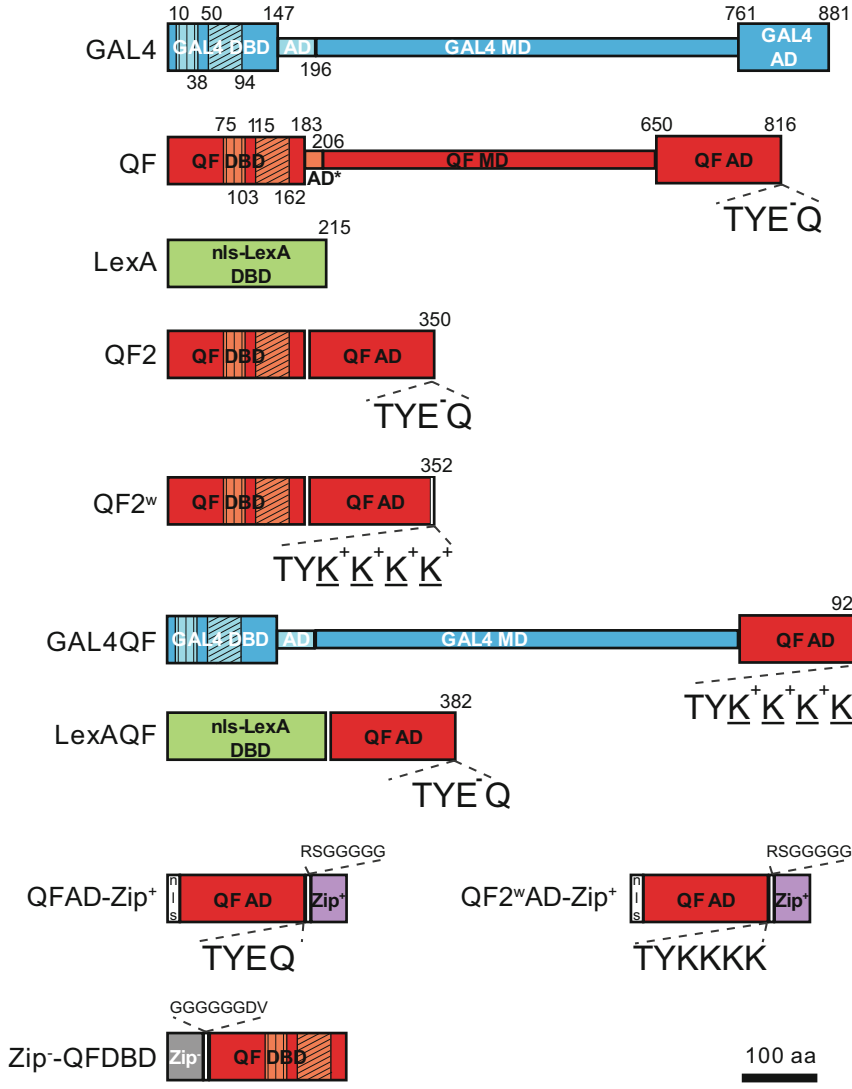


Fig. 3 Transactivator maps. Schematic representations (to scale) of original, modified, and chimeric transactivators, employed in the binary expression systems: GAL4 (blue and light blue), QF (red and light red), LexA (green). The transactivators consist of modular regions: DNA-binding domain (DBD), middle domain (MD), and activation domain (AD). Additional elements used in the split-QF system are nuclear localization domain (nls, white), Zip^+ (purple), and Zip^- (gray) leucine zippers. Vertical hatching indicates Zn_2/Cys_6 zinc finger motifs; diagonal hatchings mark dimerization domains. Numbers above and below schemes indicate amino acid position. Amino acid sequences listed above and below schematics highlight essential differences between QF2 and $QF2^w$, and indicate additional linker sequences in split-QF constructs

repressible by QS. As we noted, having QF2 expressed ubiquitously or very widely may compromise the health of fly stocks. A solution can be to have a *tubulinP-QS* transgene in the same stock (e.g., Bloomington Stock #51957), or to use the weaker $QF2^w$ version for broad expression patterns. The fact that QS alleviates the effects

of QF2 expression indicates that the activation domain of QF2 may be sequestering either non-specific targets or squelching the transcriptional machinery of a cell. The phenomenon of an exogenous transcription factor reducing the health of the fly is not unique to the QF2 activation domain – high levels of GAL4, driven by a strong promoter, can also compromise the health of a cell [14].

Expression levels of *QUAS-geneX* reporters, driven with QF2 and QF2^{w}, can be fine-tuned in a multitude of ways. A simple approach to modulate expression levels would be to change the number of *QUAS* sites. Increasing the number of *QUAS* sites from the usual 5 to 10, 15, 20, or 40 *QUAS* repeats would dramatically increase effector expression from the same driver line [9, 15]. Additional changes to the *QUAS-geneX* reporter can be used, such as including translational regulatory elements, like IVS, Syn21, and WPRE [16], or by using stronger terminator sequences, such as *SV40* or *p10* instead of *hsp70* in QF and *QUAS* constructs [16].

2 Chimeric Transactivators and Split-QF

The chimeric approach was originally used to pinpoint the source of toxicity to the middle domain of the original QF transactivator [11, 12]. Subsequently, we demonstrated that GAL4 DNA-binding and middle domains (GAL4.DBD+GAL4.MD) may be successfully fused to QF2/QF2^{w} activation domains (QF2/QF2^{w}.AD). This chimeric transactivator (GAL4QF) binds to *UAS* and can drive expression of transgenes from *UAS* reporter lines. At the same time, it is repressible by QS instead of GAL80, and may offer a useful alternative for experiments where feeding quinic acid is preferred to using temperature-sensitive GAL80. Similarly, LexA.DBD may be fused to QF2.AD or QF2^{w}.AD, to offer a QS-repressible transactivator for LexAop reporter lines (LexAQF). Both GAL4QF and LexAQF may be broadly expressed in the flies by *actin*, *tubulin*, or *synaptobrevin* promoters and efficiently repressed by QS [12]. The reverse combination of QF.DBD with GAL4.AD proved to be non-functional, although QF.DBD, similarly to GAL4.DBD [9], may be efficiently paired with p65.AD [17]. Most recently, Mao and colleagues successfully paired the rTetR DNA-binding component of the Tet expression system with the QF activation domain and used it in *C. elegans* [18]. This approach has not been implemented in *Drosophila* yet.

Instead of a direct fusion, the DBD and AD domains of a transactivator may be expressed as two separate proteins ("split") and later brought together by a linker, or zipper. Split-GAL4 [19], and later split-LexA [20] and split-QF [17, 21, 22] systems were developed to address the problem of unavailability of specific driver lines for cell labeling and manipulation experiments. For example, if a researcher wishes to silence one specific neuron in the fly, they

may not have a driver line that targets only that one neuron. Instead, they may have two driver lines that target many cells each, with only the neuron of interest in common between the two. In this ideal situation, a split system may be used to target the neuron. The enhancer/promoter sequences of one line will be used to drive the DNA-binding domain of QF2 (QF2.DBD), fused to a leucine zipper. The other driver line will drive expression of the activation domain of QF2 (QF.AD), fused to a complementary zipper sequence. Only in the neuron of interest will QF.DBD and QF.AD be co-expressed, and reconstituted into one transactivator with the help of the leucine zippers. It can then bind to the *QUAS* sequence and switch on transcription of downstream reporter genes, as usual.

The split-QF system was pioneered in *C. elegans* [21]. The implementation of split-QF in *Drosophila*, and specifically the leucine zippers used, followed the design of split-GAL4 [9] (Fig. 3). This makes split-QF fully compatible with the existing split-GAL4 and split-LexA lines, whereby DNA-binding domains of GAL4 and LexA may be successfully paired with DNA activation domains of QF2 and QF2^{w}, as described above for chimeric transactivators. The split versions of transactivators are approximately 2 times weaker than direct fusion versions but are fully repressible by QS. QA feeding recovers 25–60% of the unrepressed expression levels, depending on the life stage (larvae or adults) and transactivator version used [17].

Chimeric fusion and split transactivators work well for neuronal labeling in *Drosophila* [12, 17], zebrafish [22], and *C. elegans* [18, 21], and functional experiments, as evidenced by single cell electrophysiology recordings in *Drosophila* larvae and behavioral assays in *Drosophila* larvae and adults [17]. Most recently, split-QF was successfully used in *Aedes aegypti* mosquitoes to study olfactory receptor co-expression [23].

3 Temporal Control of Expression by Quinic Acid

The non-toxic small molecule quinic acid (QA) provides temporal control in the Q-system by relieving repression by QS. QA-mediated temporal control can be achieved with all QF variants (QF, QF2, QF2^{w}), chimeric transactivators (GAL4QF, LexAQF), and the QF activation domain (QFAD) in the split-QF system [11, 12, 17].

QA can be dissolved in water for up to ~200–300 mg/ml and adjusted to pH 7 with 5 M NaOH [17, 24]. In larvae, sufficient de-repression can be achieved by making holes in solidified fly medium, filling them with the QA solution, and growing larvae in these vials [24]. Alternatively, for both adults and larvae, up to 1.6. ml of QA solution may be added per 10 ml of fly medium,

thoroughly mixed, and allowed to dry for several hours before placing flies into the vials [17]. In adults, better ingestion can be accomplished by giving flies low-nutrient medium supplemented with a high amount of QA. For example, flies may be kept on agarose gel supplemented with 6% QA and 1% glucose [12]. Additionally, yeast paste with 3% QA solution neutralized to pH 7 with NaOH and tissue paper moistened with the same solution may be added to the vials for better results. QA feeding does not produce any known adverse effects, even during exposure for multiple generations [11], and it does not affect lifespan [25]. However, the survival of flies may be affected by low nutrition. It is thus important to balance the amount of nutrients and QA to ensure flies' survival as required for a specific experiment. De-repression can first be observed after 6 h in adult flies transferred to QA-supplemented medium following eclosion. Marked effects are observed after 24 h, while saturation is achieved in 5 days [11].

The efficacy of QA treatment depends on the dose and duration of exposure, the transactivator used, and the tissues studied. The efficacy of suppression by QS also affects the control achievable by QA. Luciferase assays on transiently transfected *Drosophila* S2 cells showed a ~23-fold re-activation of *QUAS* by high concentrations of QA compared to the control with *QUAS*, *QF,* and *5xQS* (QA has no effect on *QUAS* expression in the absence of *5xQS*) [11]. Luciferase assays on larvae raised on a QA-supplemented diet showed that expression levels with the QF variants QF2 and QF2^{w} are restored to 50–60% of unrepressed levels in the nervous system [17]. With the chimeric transactivator GAL4QF driven pan-neuronally by *synaptobrevin* promoter and enhancer regions, QA-fed larvae had expression levels of luciferase restored to 60% of unrepressed levels. QA de-repression appears to be especially efficient with the split-QF system. In QA-fed *tub-QS, nsyb-QF.DBD, nsyb-QF.AD (QF2^{w}.AD)* larvae, the expression levels after QA feeding were close to the levels in larvae lacking *tub-QS*. QA can also be used when combining components of the split-QF and split-GAL4 systems. QA feeding of larvae with QF.AD and QF2^{w}.AD activation domains used together with GAL4.DBD produced de-repression of ~20–35% of that of the unrepressed split transactivators. Quantification of QA efficiency was performed on larvae because in the adult fly brain QA works best in sensory receptor neurons and *pars intercerebralis* neurons, but appears less effective in central brain neurons [12]. This is likely due to lack of access of QA to the target tissue, possibly caused by the inability of QA to cross the glial blood–brain barrier and reach brain cells [26].

De-repression of the split QF-system by QA has been functionally validated by electrophysiological recordings and behavioral assays. Whole-cell patch-clamp recordings from aCC and RP2 motoneurons showed that feeding QA to *tub-QS,VGlut-GAL4.DBD*, *nsyb-QF2^{w}.AD* larvae partially restored depolarization and action potential count [17]. Light escape assays with larvae of the

same genotype and with ChR2 driven pan-neuronally by split-QF showed similar results. Larvae containing the QS transgene readily escaped the illuminated area, but QA feeding significantly reduced the number of escapes, due to the seizure-like overactivation of their nervous system via ChR2 [17].

QA has been used to limit the expression of a gene of interest to certain developmental stages in *Drosophila*. By feeding QA only to adults, the potential developmental effects of gene expression can be averted. For instance, QA was used to limit *SCA3polyQ78*-induced eye degeneration to adult stage with the Q-system [25]. Adult- or larval-specific induction by QA appears not to affect sleep-wake cycles in flies, suggesting that the system can be used for conditional gene expression to study circadian behaviors and sleep [27]. QA does not affect GAL4 activation of *UAS*, so it can be used together with GAL4/*UAS* system (but it may affect GAL80 repression of GAL4, as measured in S2 cell culture) [11].

Apart from *Drosophila*, QA is widely used for temporal control of gene expression in *C. elegans*. Marked de-repression appears in worms kept on plates supplemented with 7.5 mg/ml QA after 6 h and reaches its maximum after 24 h, faster than in the fly [21]. By providing QA to worms at different developmental stages, the critical period for *srb-13* chemosensory receptor signaling in sperm navigation and *vpr-1* activity in gonad development have been determined [28, 29].

The use of QA has not been widely implemented in other species with Q-system reagents. Recently, however, QA has been shown to achieve complete de-repression of the Q-system in *Saccharomyces cerevisiae* even at low concentrations (5 mg/L) [30]. To achieve suitable intracellular concentrations of the inducer, quinate permease encoded by *qa-y* was expressed. This system was used to design optogenetic circuits to dynamically regulate yeast gene expression by the simultaneous use of a GAL4-based circuit [30]. Interestingly, in transfected mammalian cells, the addition of QA causes further transcriptional repression rather than activation [11]. In zebrafish, QA was found to be toxic at functional concentrations (>0.5 mg/mL) [22]. QA/QS has not yet been implemented in mosquitoes. QS is not functional in bacteria, possibly due to the lack of necessary post-translational modifications [31].

QA provides a unique tool for temporally controlled activation of a binary system in *Drosophila* and *C. elegans*. Unlike the GAL4/*UAS* system and its temporal repression by GAL80^{ts}, QA does not require changes in temperature that might affect other aspects of fly physiology and behavior. However, the utility of QA is limited by its low availability in the fly brain and the slow onset of de-repression. Efforts to achieve faster temporal control lead to the development of a modified Q-system in *C. elegans*, where induction by another small molecule can be achieved within 2 h [32].

4 Efficient Conversion of GAL4 Transgenes to QF2 Transgenes Using Homology Assisted CRISPR Knock-in (HACK)

Generation and screening of new transactivator lines (*QF2*, *LexA*) to recapitulate existing *GAL4* expression patterns can be a tedious and time consuming process. A faster and more efficient way to generate new lines is to employ genomic conversion of *GAL4* DNA sequence at the target sites through homology-directed repair (HDR). Using the CRISPR system to create double-stranded (ds) DNA damage at the target sites [33], the original sequence of *GAL4* is disrupted, allowing integration of the desired sequences (*QF2*, *LexA*, etc.). This is called Homology Assisted CRISPR Knock-in (HACK), and the concept was first employed to convert *GAL4* enhancer trap lines into *QF2* drivers (Fig. 4) [34].

The HACK conversion of *GAL4* lines may be achieved via simple genetic crosses of HACK donor lines with *GAL4* lines and a source of Cas9 protein. There are four major components in the artificial genomic DNA of the donor: 5′ and 3′ homology arms of *GAL4*, an in-frame *T2A-QF2* coding sequence, Red Fluorescent Protein (RFP) eye marker flanked by LoxP sites (LoxP-*3xP3-mCherry*-LoxP), and U6 promoter-driven gRNAs that target *GAL4* (Fig. 4). The donor construct, marked by *3xP3-mCherry* (mCherry will be simplified as RFP hereafter), is randomly integrated into the genome by P-element insertion. In the presence of Cas9 protein driven by the *Act5C* promoter (on X) or *Vas* promoter (on X or 3rd), gRNAs bind and guide the Cas9 endonuclease to create two dsDNA breaks in the middle domain of the *GAL4* sequence. The damaged dsDNA could be repaired by either non-homologous end joining (NHEJ) or homology-directed repair (HDR). In HDR, the donor construct is utilized as a template for repair, instead of the homologous chromosome since there is no *GAL4* homology sequence in the fly genome. 5′ and 3′ *GAL4* homology arms of the donor construct hybridize with the damaged *GAL4* sequences, and the target sequence is repaired by copying the donor sequence, hence integrating the *T2A-QF2* construct into the disrupted *GAL4* sequence. The final HACKed product consists of in-frame *T2A-QF2* and LoxP-*3xP3-RFP*-LoxP flanked by 5′ and 3′ truncated *GAL4* sequences. The original promoter will drive the expression of 5′ truncated *GAL4* (non-functional) and a complete coding sequence of *QF2* (functional). Overall, the sequence of interest is thus "gene converted" to the target locus. During genetic crosses, an eye RFP marker (*3xP3-RFP*, Fig. 4) is used to monitor successful HDR-mediated insertion of the desired sequence, and can be easily removed later by Cre recombinase (BS#766, [35]) (Fig. 4).

Figure 4b, c provides detailed crossing schemes that may be employed to produce *QF2* driver lines from existing *GAL4* lines. Briefly, the target (*X-GAL4*), donor ($QF2^{G4H}$), and *Act5C-Cas9*

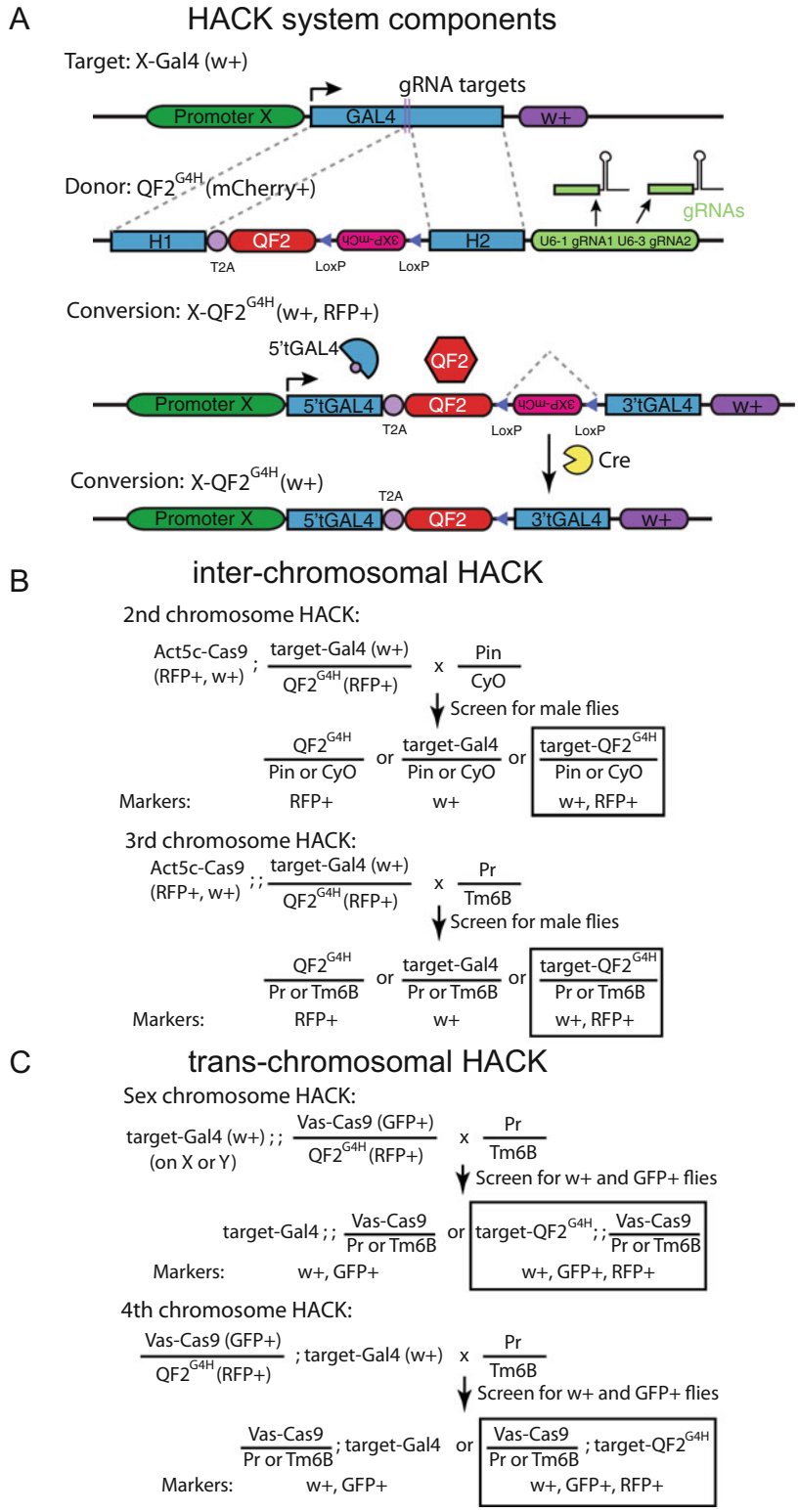


Fig. 4 HACK conversion of GAL4 lines into QF2 lines. (**a**) The target is an X promoter-driven GAL4 transgenic line (X-GAL4). The $QF2^{G4H}$ donor consists of 5′ and 3′ GAL4 homology arms (1182 and 1368 b.p., respectively)

(BS#54590, [36]) or *Vas-Cas9* (BS#51323 on X chromosome and BS#51324 on third chromosome [37]) transgenes are combined in the same animal (F_1 male, Fig. 4b, c). The male is crossed to white-eyed (*w*-) or white-eyed balancer female flies. Since *Drosophila* males do not undergo chromosomal recombination during meiosis, the F_2 male will contain either the *w*+ (*GAL4*) or RFP marker ($QF2^{G4H}$). F_2 males containing both *w*+ and *RFP* indicate successful HDR-mediated incorporation of *T2A-QF2* and floxed *3xP3-RFP* to the *GAL4* sequence (Fig. 4b, c). Due to positional effects [38], the RFP marker in the eye could demonstrate a different expression pattern/level in successfully HACKed flies compared to the original donors. The entire HACK process should take 2–3 crosses depending on the complexity and availability of combined reagents. Successfully HACKed lines are named $X\text{-}QF2^{G4H}$. This establishes a nomenclature $X\text{-}Y^{ZH}$, in which X refers to the original enhancer/promoter locus, Y refers to the introduced DNA component, and Z refers to the replaced DNA element. For instance, if *Act-GAL4* is converted to *QF2* by using the HACK method, the converted line is named $Act\text{-}QF2^{G4H}$.

Depending on the positions of the donor and target, HACK could be divided into (1) inter-chromosomal and (2) trans-chromosomal types. Inter-chromosomal HACK refers to the condition when the donor and target transgenes are on homologous chromosomes (i.e., both are on the second or third chromosomes). In this situation, cytological location could be used to reliably predict conversion efficiency. We recommend identifying two donor lines with the closest cytological locations to the desired targets for HACKing. Trans-chromosomal HACK occurs when the target is on a sex chromosome (X or Y) or fourth chromosome and the donor could be on the second or third chromosome. Taking advantage of existing *Vas-Cas9* (*GFP*+) on the third chromosome (BS#51324) and the simplicity of genetic crosses, we recommend using donors on the third chromosome. Trans-chromosomal donors with the best efficiency are usually unpredictable; thus, it is recommended to start with 3 or 4 donors at different locations. Detailed crossing schemes with selection markers are outlined in Fig. 4b, c.

Alternatively, HACK target conversion could be achieved by direct embryo injection. Flies with transgenic *GAL4* combined with a *Cas9* driver (*Act5C-Cas9* or *Vas-Cas9* on the X chromosome

Fig. 4 (continued) flanking an in-frame T2A-QF2 and a floxed 3xP3-mCherry eye marker. Downstream the 3′ GAL4 homology arm are U6:1 and U6:3 promoters driving two independent gRNAs that target GAL4 in the presence of the Cas9 protein. The successfully converted $X\text{-}QF2^{G4H}$ line contains 5′ truncated GAL4, in-frame T2A-QF2, floxed 3xP3-mCherry marker, and 3′ truncated GAL4. The 3xP3-mCherry sequence can later be removed by crossing to Cre recombinase. Of note, the two U6-gRNAs will not be converted into the target site. mCh mCherry. (**b**) and (**c**) are the detailed crossing scheme of HACK

for simplicity) are injected with donor plasmid DNA and screened for flies with double markers (*w+* and *RFP+*) after crossing to white-eyed (*w-*) or white-eyed balancer female flies. The efficiency could be slightly lower, but is useful if the target is on the sex or fourth chromosome, which requires more complex genetic screening.

Of note, the original HACK donor construct used to convert *GAL4* into *QF2* was designed for wild-type *GAL4* sequence. HACKing codon-optimized *GAL4* sequences would result in sub-optimal conversion efficiency. This applies to all other derivative constructs that used pHACK-GAL4>QF2 (Addgene#80275) as a precursor since the homology arms and gRNAs were designed according to the wild-type *GAL4* sequence [34]. Also, the HACK method utilizes *w+* as the *GAL4* driver marker. If a target *GAL4* transgene is devoid of a dominant marker or uses another fluorescent protein as a marker, the genetic and screening scheme will require modifications.

The HACK approach has been successfully expanded and used to convert *GAL4* drivers into other DNA elements, including split-*GAL4* [39], *Gal80* [39], *FLP* [40], *LexA (* [41], *Cas9* ([42]; C-C Lin, unpublished reagents), and chemically controlled *Cas9* (C-C Lin, unpublished reagents). Moreover, the concept applies beyond using *GAL4* sequence as a genetic target. HACK has been used to convert split-gene drives to full-gene drives [43] and to target specific regions of the genome for knock-ins [44].

5 Simultaneous Use of Multiple Binary Expression Systems and Intersectional Expression Utilizing the Q-system

Multiple binary expression systems can be combined to examine overlapping gene expression in different populations of cells. For example, a *QF* driver can be used to drive one effector (such as GFP), and a *LexA* or *GAL4* driver can be used to drive a second effector (like RFP) (Fig. 5a). Such co-labeling experiments have been used extensively for mapping and identifying the anatomical location and cellular/molecular identity of different neural populations (for example, [44–47]). Combining multiple binary expression systems also enables one to optogenetically activate one set of cells while simultaneously performing functional imaging from downstream partners (see [48–51]).

One of the great advantages of having multiple independent binary expression systems lies in the ability to combine them to achieve a variety of new expression patterns. These modified expression patterns can be viewed as logic gates (such as AND, NOT, or NOR), as illustrated in the schematic diagrams in Fig. 5b–d. A practical application of the *Q*-system is to use it to narrow expression patterns from many currently available *GAL4* expression lines, such as the various Janelia *GAL4* enhancer collections [9, 52–54], the Vienna Tile collection [55], InSITE [56], MiMIC [57],

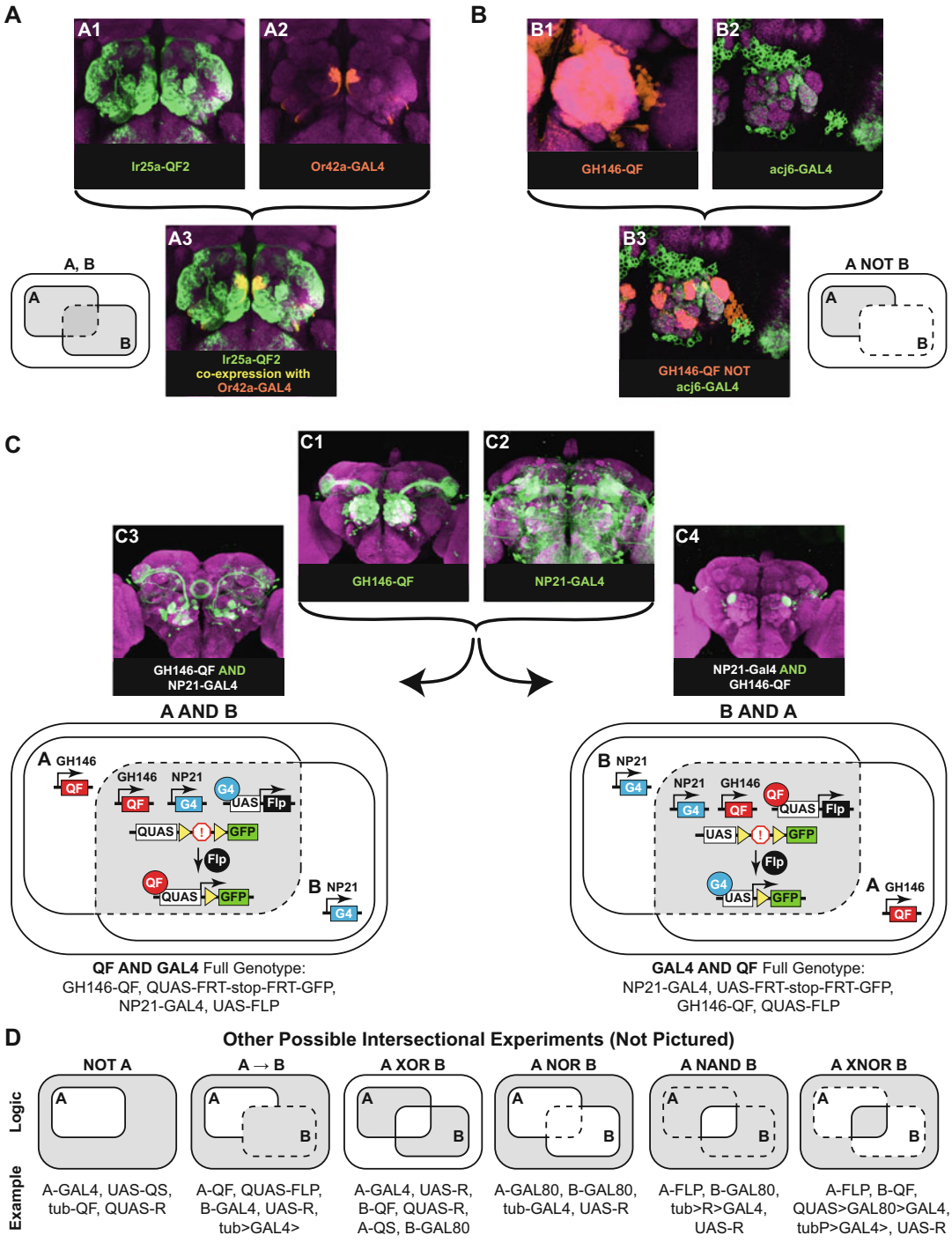


Fig. 5 Simultaneous use of multiple binary expression systems and intersectional labeling. (**a**) Simultaneous expression of two different reporters in overlapping subsets of cells with the *GAL4/UAS* and *Q*-systems (schematic bottom left). *Ir25a-QF2* (a HACK knock-in reagent, see Subheading 4 on the HACK technique) drives GFP expression in the majority of olfactory sensory neurons (A1, green). *Or42a-GAL4* drives RFP expression in one olfactory sensory neuron subtype innervating the VM7d glomerulus (A2, orange). The co-labeling experiment QF2, GAL4 reveals co-expression in this neuronal population (A3, yellow). Brains are

Trojan-MiMIC [58], and CRIMIC [59] collections. Such approaches often require FLP/FRT [60] or other recombinases [61, 62], and/or repression in specific cell subsets using *QS* or *GAL80*. One example is the NOT intersection, in which the expression pattern of one transactivator is used to limit the expression of another (Fig. 5b). This can be achieved by silencing one of the transactivators (for example, with *QS*) in cells in which both transactivators are expressed. Expression patterns can be further refined with new tools like split transactivators ([17, 19]; see Subheading 2 on *Split-QF* for more details) and Killer Zipper [63].

For many intersections, an *FRT-transcriptional stop-FRT* cassette in the reporter is essential. When such a construct is expressed with FLPase in the same cell, the FLPase permanently removes the transcriptional stop cassette from the cell and all its progeny, leading to expression of the effector downstream of the remaining FRT site (Fig. 5c). The optimal strategy for such an intersectional approach depends on the available driver, reporter, and repressor lines. One must also consider possible off-target labeling that may arise at early developmental stages. For example, the two approaches to the AND intersection could give rise to different final readouts, depending on which driver line drives initial FLPase activity, and which one drives final reporter expression. Figure 5c

Fig. 5 (continued) counterstained with anti-nc82 (magenta) in A–C. Genotype: *Ir25a-T2A-QF2, 10XQUAS-6XGFP, Or42a-GAL4, 10XUAS-IVS-mCD8::RFP*. (**b**) Using one transactivator to limit the expression pattern of another transactivator: *QF* NOT *GAL4* intersection (schematic bottom right). *GH146-QF* drives mtdTomato expression in a large proportion of olfactory projection neurons (B1, orange). *acj6-GAL4* drives GFP expression in a partly overlapping subset of projection neurons (B2, green). Using *acj6-GAL4* to additionally drive the QS repressor silences QF in those cells where GAL4 and QF expression patterns overlap. This intersection reveals those GH146+ neurons where acj6 is not expressed (B3, orange). Genotype: *GH146-QF, QUAS-mtdTomato-3XHA, acj6-GAL4, UAS-mCD8::GFP, UAS-QS*. (**c**) Limiting expression to cells that express both transactivators: the AND intersection. Expression pattern of *GH146-QF* as visualized by *QUAS-mCD8::GFP* (C1, green). Expression pattern of *NP21-GAL4* as visualized by *UAS-mCD8::GFP* (C2, green). There are two possible ways to achieve the AND intersection: *QF* AND *GAL4* (C3, bottom left diagram) versus *GAL4* AND *QF* (C4, bottom right diagram). The differences in expression between C3 and C4 arise due to the developmental timing of FLPase expression, depending on which driver line (*QF* or *GAL4*) is used to drive FLP-mediated recombination and subsequent GFP expression. In C3, *NP21-GAL4* drives *UAS-FLP*, leading to excision of the STOP cassette and GFP expression driven by *GH146-QF* (grey region of diagram). In C4, *GH146-QF* drives *QUAS-FLP*, allowing excision of the STOP cassette and subsequent *NP21-GAL4*-driven GFP expression (diagramed in grey region). C3 genotype: *GH146-QF, QUAS-FRT-stop-FRT-GFP, NP21-GAL4, UAS-FLP*. C4 genotype: *NP21-GAL4, UAS-FRT-stop-FRT-GFP, GH146-QF, QUAS-FLP*. (**d**) There are many other genetic intersections possible between binary expression systems, some of which are diagrammed here. The top row shows six additional schematics based on the intersectional logic of the experiment, while the bottom row gives one example genotype for the given intersection. In the genotypes, R stands for reporter (e.g., GFP) and $>$ stands for FRT site. A is the promoter/enhancer for one binary expression system, B for the other. "QF" could also be QF2 or QF2^{w}. It is noted that there are many other combinations possible, and this is not intended to be an exhaustive list. For additional example genotypes (including LexA and chimeric transactivators), see [12]. (Modified images and schematics in **b**, **c** reprinted with permission from [11])

shows an example of this using *GH146-QF* and *NP21-GAL4*. *GH146-QF* drives expression in many olfactory projection neurons in the adult. *NP21-GAL4* drives expression in many cells in the adult brain, including one subset of olfactory projection neurons (innervating the DA1 glomerulus). While the adult expression patterns of these two drivers suggest that an AND intersection should only label this single common glomerulus, both the QF AND GAL4 and the GAL4 AND QF intersections can label additional cells in the adult brain [11, 64]. The QF AND GAL4 intersection in particular labels four additional olfactory glomeruli (Fig. 5c3). These discrepancies between the intersections and adult co-expression are a result of broader expression of the transactivators during development, leading to excision of the FRT-stop-FRT cassette in neurons which do not have overlapping expression in the adult. This extraneous developmental expression can be mitigated or even eliminated by temporally controlling FLPase activity. For example, another chemically inducible genetic tool fuses a destabilization domain (DD) to proteins of interest [64]. The chimeric protein will be degraded unless the DD ligand is provided to the fly in food, which stabilizes the protein. Sethi and Wang [64] fused this DD to FLP and showed that restricting FLPase activity by feeding the DD ligand only to adult flies successfully removed the extraneous glomerular expression in the same *GH146-QF* AND *NP21-GAL4* intersection, recapitulating the expected adult expression pattern of only one glomerulus being labeled.

Limiting expression may be achieved in a variety of ways based on these and other tools (see Fig. 5d for more examples of logic gates and [11] for additional possible genetic combinations). Table 1 lists a selection of Q-system reagents (both plasmids and fly lines) that are a good starting place for co-expression and intersectional approaches.

Table 1
Example Q-system reagents in *Drosophila*

Component	Plasmid and/or Fly Stock #s	Purpose/function	Reference
QF2 drivers			
nsyb-QF2	Addgene# 46115; BDSC# 78338	Pan-neuronal expression	[12]
nsyb-QF2^{w}	Addgene# 46116; BDSC# 51960	Pan-neuronal expression	[12]
αtubulinP-QF2^{w}	BDSC# 51963	Pan-tissue expression under the control of the αTubulin84B enhancer+promoter	[12]

(continued)

Table 1
(continued)

Component	Plasmid and/or Fly Stock #s	Purpose/function	Reference
pGMR-QF2^{w}	BDSC# 59283	Strong expression in the eye	[12]
GH146-QF2	BDSC# 30015, 91998	Expression in olfactory projection neurons	[11]; Contributed by Lin and Potter
Orco-T2A-QF2	BDSC# 92401	CRISRP/Cas9-mediated knock-in using the HACK technique to capture Orco chemosensory receptor expression	[44]
Ir25a-T2A-QF2	BDSC# 92392	CRISPR/Cas9-mediated knock-in using the HACK technique to capture Ir25a chemosensory receptor expression	[44]
QUAS reporters			
QUAS-mCD8::GFP	Addgene# 24351, 46163; BDSC# 30001, 30002, 30003	A membrane-bound GFP reporter	[11]
10XQUAS-6XGFP	BDSC# 52263, 52264	Cytoplasmic hexameric GFP for extra-strong labeling	[15]
10XQUAS-CsChrimson. mVenus	Addgene# 163629; BDSC# 91992, 91993, 91994, 91995, 91996	A red-shifted variant of channelrhodopsin, an excitatory light-activated sodium channel. Tagged with mVenus yellow fluorescent protein	[44]
QUAS-GAL80	Addgene# 46137; BDSC# 51948, 51949, 51950	Expresses GAL80, the repressor of GAL4, under QUAS control	Contributed by Christopher Potter
QUAS-shibire^{ts}	Addgene# 24362; BDSC# 30010, 30011, 30012, 30013	Temperature-sensitive shibire protein. Inhibits synaptic transmission at 29 °C and above	[11]
QUAS-FLP	Addgene# 24357; BDSC# 30008, 30126, 30127	Expresses FLP recombinase under QUAS control	[11]
5XQUAS-GCaMP6s *5XQUAS-GCaMP6m* *5XQUAS-GCaMP6f*	BDSC# 91990; BDSC# 91988; BDSC# 91989;	Express the calcium indicator GCaMP6 with slow (s), intermediate (m), or fast (f) kinetics	Ya-Hui Chou, Academia Sinica, Taiwan

(continued)

Table 1
(continued)

Component	Plasmid and/or Fly Stock #s	Purpose/function	Reference
QUAS-TRPA1	BDSC# 91803, 91804, 91805, 91806, 91807	Expresses temperature-sensitive dTRPA1 for temporal activation of neurons	Contributed by Lin and Potter
QUAS-mtdTomato-3XHA	Addgene# 24354; BDSC# 30004, 30005	Myristoylated membrane-targeted tdTomato	[11]
10XQUAS-6XmCherry-HA	BDSC# 52269, 52270	Cytoplasmic hexameric mCherry for extra-strong labeling	[15]
QUAS-nucLacZ	Addgene# 24356; BDSC# 30007	Nuclear-localized beta-galactosidase	[11]
QUAS>stop>mCD8::GFP	Addgene# 24361; BDSC# 30134, 30135, 30136	FLP-ase-based GFP reporter	[11]
QUAS>stop>CsChrimson.mVenus	BDSC# 91999	FLP-ase-based channelrhodopsin effector	Contributed by Riabinina and Potter
QUAS>stop>shibire^{ts}	Addgene# 24363; BDSC# 30128	FLP-ase-based shibire effector	[11]
QUAS>stop>TRPA1	BDSC# 91986, 91987	FLP-ase-based TRPA1 effector	Contributed by Lin and Potter
QS repressors			
αTubulin-QS	Addgene# 24352; BDSC#: Multiple available	Expresses QS, the repressor of transactivators with the QF/QF2(w) AD, under the control of the αTubulin84B promoter	[11]
UAS-QS	Addgene# 24366; BDSC# 30033	Expresses QS, the repressor of transactivators with the QF/QF2(w) AD, under the control of UAS	[11]
Chimeric transactivators			
actin5c-GAL4QF	Addgene# 61309; BDSC# 59285	Expresses GAL4 DNA-binding domain fused to the QF activation domain ubiquitously under the control of actin5C promoter	[12]

(continued)

Table 1
(continued)

Component	Plasmid and/or Fly Stock #s	Purpose/function	Reference
nsyb-GAL4QF	Addgene# 46113; BDSC# 51945	Expresses the GAL4 DNA-binding domain fused to the QF activation domain pan-neuronally under the control of n-syb	[12]
actin5c-LexAQF	Addgene# 61311; BDSC# 62567, 62568	Expresses LexA DNA-binding domain fused to the QF activation domain ubiquitously under the control of actin5C promoter	[12]
nsyb-LexAQF	Addgene# 46123; BDSC# 51953	Expresses the LexA DNA-binding domain fused to the QF activation domain pan-neuronally under the control of n-syb	[12]
Chat-LexAQF	BDSC# 60319	Trojan-MiMIC line that expresses LexAQF in the pattern of choline acetyltransferase, targets cholinergic neurons	[58]
vGlut-LexAQF	BDSC# 60314	Trojan-MiMIC line that expresses LexAQF in the pattern of vesicular glutamate transporter, targets glutamatergic neurons	[58]
Gad1-LexAQF	BDSC# 60324	Trojan-MiMIC line that expresses LexAQF in the pattern of glutamic acid decarboxylase 1, targets GABAergic neurons.	[58]
Split-QF transactivators			
nsyb-QF2^{w}.AD-Zip^{+}	DGRC# 1478; BDSC# 81283	Pan-neuronal expression of QF2^{w}.AD	[17]
nsyb-QF2.AD-Zip^{+}	DGRC# 1479; BDSC# 81281	Pan-neuronal expression of QF2.AD	[17]
nsyb-Zip^{-}-QF.DBD	DGRC# 1480; BDSC# 81282	Pan-neuronal expression of QF. DBD	[17]
HACKed lines			
Elav-QF2^{G4H}	BDSC# 66466	Pan-neuronal QF2	[34]
Pebbled- QF2^{G4H}	BDSC# 66474	Pan-ORN, GN expression	[34]

(continued)

Table 1
(continued)

Component	Plasmid and/or Fly Stock #s	Purpose/function	Reference
GH146-QF2^{G4H}	BDSC# 66467	Olfactory projection neuron expression	[34]
TubP-QF2^{G4H}	BDSC# 66478	Pan-cellular expression	[34]
Repo-QF2^{G4H}	BDSC# 66477	Pan-glial expression	[34]
OK107-QF2^{G4H}	BDSC# 66472	Pan-mushroom body expression	[34]
OK107-GAL80^{G4H}	TBD	Pan-mushroom body expression	[39]
TH-D-DBD^{G4H}	TBD	Subset of dopaminergic neurons	[39]
Hml-Cas9^{G4H}	TBD	Larval hemocytes	[42]
OK371-Cas9^{G4H}	TBD	Glutamatergic motor neurons in third instar larva	[42]
HACK Donor lines		See https://bdsc.indiana.edu/stocks/misc/hack.html, https://potterlab.johnshopkins.edu/resources/hack-system/, and [34] for information on specific donor line selection based on cytological location of gene target	
Cas9.G4HACK Donor	BDSC# 92750, 92751	Carries a HACK donor construct for CRISPR-mediated conversion of GAL4 to Cas9	
GAL4DBD.G4HACK Donor	BDSC# Multiple available	Carries a HACK donor construct for CRISPR-mediated conversion of GAL4 to GAL4DBD	
GAL80.G4HACK	Addgene# 104874; BDSC# Multiple available	Carries a HACK donor construct for CRISPR-mediated conversion of GAL4 to GAL80	
QF2.G4HACK	Addgene# 80275, 80277, 104873; BDSC# Multiple available	Carries a HACK donor construct for CRISPR-mediated conversion of GAL4 to QF2	
QS.G4HACK	BDSC# Multiple available	Carries a HACK donor construct for CRISPR-mediated conversion of GAL4 to QS	

6 Applications Beyond *Drosophila*

Binary expression systems are widely used in species beyond *Drosophila melanogaster*, as they can provide high flexibility and amplify weak expression (Fig. 6, Table 2). The Q-system has become the primary binary system in mosquito species, especially in neurobiological studies [65, 66]. It has also been widely used in the well-established model organisms *C. elegans* and zebrafish, expanding the range of genetic tools available [21, 67, 68]. The Q-system has been implemented in mammalian cells, plants, and fungi other than *Neurospora crassa* [11, 30, 69–72]. Components of the system can efficiently control gene expression even in bacteria [31]. The following examples demonstrate how the Q-system can be adapted from *Drosophila* to virtually any species to study diverse biological questions. At the same time, applications of the Q-system in these organisms may inform on new ways it can be used in *Drosophila*.

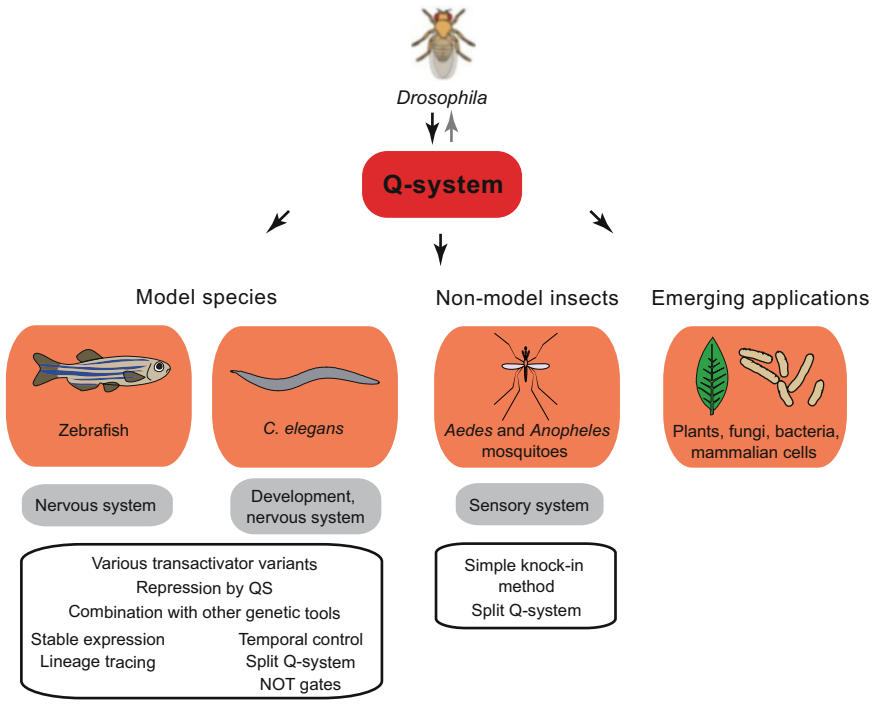


Fig. 6 Applications of the Q-system in species beyond *Drosophila melanogaster*. The Q-system has been widely used in the model systems *Danio rerio* (zebrafish) and the nematode *C. elegans*. It is the primary binary system in *Aedes aegypti* and *Anopheles gambiae* mosquitoes and has important potentials for other non-model insects. Emerging new applications include plants, fungi, mammalian cells, and bacteria. Examples for biological questions/areas widely studied using the Q-system are shown below the organisms. Boxes highlight the advantages of using the Q-system and some of the tools available in each organism. *Drosophila* was used as a guide to implement the Q-system in other species, but the novel and creative ways the system is being used in these organisms may inform on future applications in *Drosophila*

Table 2
Example Q-system reagents in species beyond *Drosophila*

Species	Type of transgene	Q-transgene	Purpose/function	Reference
An. gambiae/ coluzzii	QUAS reporter	15xQUAS-mCD8::GFP	Membrane-bound GFP	[65], available at BEI Resources
	QUAS reporter	15xQUAS-GCaMP6f	GCaMP6f calcium reporter for live imaging	[78]
	QUAS reporter	15xQUAS-AgOR2	Overexpression of *Anopheles gambiae* olfactory receptor OR2	[79]
	QF2 driver	Orco-QF2	Expresses QF2 in a subset of neurons in the pattern of the Olfactory Receptor Co-receptor Orco	[65], available at BEI Resources
	QF2 driver	Amt-QF2	Expresses QF2 in a subset of neurons in the pattern of the *An. gambiae* ammonium transporter	[80]
	QF2 driver (knock-in)	Ir76b-T2A-QF2	Expresses QF2 in a subset of neurons in the pattern of the Ionotropic Receptor Co-receptor IR76b	[87]
Ae. aegypti	QUAS reporter	15x/30xQUAS-mCD8::GFP; 15xQUAS-Syt1::tdTomato	Membrane-bound GFP and presynaptic-localized tdTomato	[66, 77, 84]
	QUAS reporter	15x-QUAS-dTomato-T2A-GCaMP6s	dTomato and GCaMP6s calcium reporter for live imaging, separated by the T2A sequence	[77, 83]
	QUAS reporter	15x/30xQUAS-CaMPARI2	CaMPARI2 photoconvertible calcium reporter for live imaging	[84]
	QUAS reporter	QUAS-CsChrimson-tdTomato	CsChrimson fused to tdTomato, for optogenetics	[85]
	QUAS reporter	15xQUAS-dTomato-T2A-TRPV1	dTomato and rat cation channel TRPV1 that can be activated by capsaicin in chemogenetic experiments, separated by the T2A sequence	[83]

(continued)

Table 2
(continued)

Species	Type of transgene	Q-transgene	Purpose/function	Reference
	QF2 driver (knock-in)	Orco- T2A-QF2; IR8a-T2A-QF2; Ir25a-T2A-QF2; IR76b-T2A-QF2; Ir7a-T2A-QF2; Ir7f-T2A-QF2	Expresses QF2 in a subset of neurons in the pattern of the Olfactory Receptors Orco/ IR8a/ IR25a/ IR76b/IR7a/IR7f	[23, 84]
	QF2 driver (knock-in)	Gr1-T2A-QF2; Gr3-T2A-QF2; Gr4-T2A-QF2	Expresses QF2 in a subset of neurons in the pattern of the CO2 receptor genes GR1/GR3 and sugar receptor GR4	[23, 83, 84]
	QF2 driver (knock-in)	ppk301-T2A-QF2	Expresses QF2 in the pattern of the DEG/ENaC ion channel ppk301	[77]
	$QF2^{w}$ driver (knock-in)	Brp-T2A-$QF2^{w}$	Expresses $QF2^{w}$ pan-neuronally in the pattern of Bruchpilot	[66]
	Split-QF driver (knock-in)	Orco-T2A-QF2.DBD; Orco-T2A-QF2.AD; Ir25a-T2A-QF2.AD	Expresses the QF2 DNA-binding domain (QF2.DBD) or activation domain (QF2.AD) in the pattern of olfactory receptors Orco/IR25a	[23]
C. elegans	QUAS reporters (extrachromosomal array and MosSCI insertion)	QUAS::GFP; QUAS::Chrimson::GFP	GFP; Chrimson fused to GFP	[21, 96]
	QUAS reporter (extrachromosomal array)	QUAS::peel-1	Sperm-specific PEEL-1 toxin	[32]
	QUAS reporter (extrachromosomal array)	QUAS::lin-3c	EGF ligand LIN-3C inducing behavioral quiescence	[32]
	QUAS reporter (extrachromosomal array)	QUAS::ChR2::GFP	Channelrhodopsin-2 fused with GFP	[99]
	QUAS reporter (extrachromosomal array)	QUAS::PH-miniSOG	Membrane-targeted blue light-activated miniature Singlet Oxygen Generator	[96]
	QUAS reporter (extrachromosomal array)	QUAS::GCaMP6::SL2::mCherry	GCaMP6 and mCherry separated by the SL2 trans-splicing sequence	[96]
	QUAS reporter (extrachromosomal array)	QUAS::unc-68ex13.1::mScarlet	Expresses mScarlet in the case of unc-68 exon 13 inclusion (splicing reporter)	[98]

(continued)

Table 2
(continued)

Species	Type of transgene	Q-transgene	Purpose/function	Reference
	QUAS reporter (extrachromosomal array)	QUAS::unc-68ex13.0::CeBFP	Expresses CeBPF in the case of unc-68 exon 13 skipping (splicing reporter)	[98]
	QF driver (extrachromosomal array)	unc-4p::QF::SL2::mCherry	Expresses QF and mCherry separated by the SL2 trans-splicing sequence in the pattern of unc-4 in A-type motor neurons (VA and DA)	[21]
	QF driver *(MosSCI insertion)*	mig-13p::QF::SL2::mCherry	Expresses QF and mCherry separated by the SL2 trans-splicing sequence in the pattern of VA12 and DA9 neurons (mig-13 promoter)	[21]
	QF driver (extrachromosomal array)	Osm-6p::QF; unc-17p::QF; gpa-10p:QF	Expresses QF in the pattern of osm-6/ unc-17/ gpa-10 promoters	[29, 95, 96]
	QF driver (extrachromosomal array)	dpy-7p::QF::SL2::mCherry	Expresses QF and mCherry separated by the SL2 trans-splicing sequence in the pattern of the gpa-10 promoter	[97]
	QF2 driver (RMCE insertion)	mec-4p::QF2	Expresses QF2 in touch receptor neurons, in the pattern of the *mec-4* promoter	[28]
	Inducible QF driver (extrachromosomal array)	unc-4p::QF::QS	Expresses QF fused to QS in the pattern of the unc-4 promoter in A-type motor neurons, when QA is present	[18]
	Inducible QF driver (extrachromosomal array)	pro-1p::QF::GR.LBD::SL::mCherry	Expresses QF fused to human glucocorticoid receptor alpha ligand-binding domain (GR.LBD) and mCherry separated by SL, in the pattern of the ubiquitous promoter pro-1, when dexamethasone is present	[32]

(continued)

Table 2
(continued)

Species	Type of transgene	Q-transgene	Purpose/function	Reference
	Inducible QF2^{w} driver (extrachromosomal array)	F25B3.3p::lox2272::mCherry::loxP::lox2272::QF2^{w}::loxP	Expresses QF2^{w} pan-neuronally in the pattern of the *F25B3.3* promoter following excision between *loxP* sites upon application of a heat shock-inducible Cre recombinase (expressed from an additional transgene)	[99]
	Split-QF.BD-DM driver (extrachromosomal array)	unc-4cp::QF.BD-DM	Expresses QF.BD-MD in the pattern of the unc-4c promoter in DA7, DA8, and DA9 neurons	[21]
	Split-QF.AD driver (extrachromosomal array)	mig-13p::QF.AD	Expresses QF.AD in the pattern of the mig-13 promoter in DA9 and VA12 neurons	[21]
	GAL4-QF driver (RMCE insertion)	mec-4p GAL4$_{SK}$-QF.AD	Expresses GAL4$_{SK}$-QF.AD in touch receptor neurons, in the pattern of the mec-4 promoter	[28]
	LexA-QF driver (RMCE insertion)	mec-4p::lexA-QF.AD	Expresses lexA-QF.AD in touch receptor neurons, in the pattern of the mec-4 promoter	[28]
	tetR-QF (tet ON) driver (RMCE insertion)	mec-4p::tetR-QF.AD	Expresses tetR-QF.AD in touch receptor neurons, in the pattern of the mec-4 promoter	[28]
	rtetR-QF (tet OFF) driver (RMCE insertion)	mec-4p::rtetR-QF.AD	Expresses rtetR-QF.AD in touch receptor neurons, in the pattern of the mec-4 promoter	[28]
	Inducible rtetR-QF (tet OFF) driver (extrachromosomal array)	rpl-28p::rtetR-QF.AD::P2A::mKate::T2A::tTS	Expresses rtetR-QF.AD, mKate and the tet-regulated transcription silencer (tTS) separated by P2A and T2A sequences under the control of the broadly active rpl-28 promoter, when tetracycline is present	[18]
	QS repressor (extrachromosomal array)	unc-4cp::QS::SL2::mCherry; unc-4p::QS::mCherry; hlh-3p::QS::SL2::mCherry	Expresses the QS repressor and mCherry under the control of the unc-4c/ unc-4/ hlh-3 promoter	[21, 96, 97]

(continued)

Table 2
(continued)

Species	Type of transgene	Q-transgene	Purpose/function	Reference
Danio rerio	QUAS reporters	QUAS: GFP; QUAS: mScarlet	GFP and mScarlet for fluorescent cell labeling	[67, 68]
	QUAS reporter	QUAS:GCaMP6s	GCaMP6s	[107]
	QUAS reporter	QUAS:nlsVenus-V2a-notch1aICD	Notch1alCD causing constitutive activation of Notch signaling	[104]
	QUAS reporter	QUAS:nVenus-V2a-SuHDN	SuHND causing inhibition of Notch signaling	[104]
	Dual QUAS and UAS reporter	QUAS:dTomato-4XnrUAS:GFP	dTomato (from QUAS) and GFP (from UAS)	[67]
	QF driver	ubb:QF; ubi:QF2^{w}	Expresses *QF2/2^{w}* ubiquitously, in the pattern of the Ubiquitin B (ubb) promoter/ ubi promoter	[67, 68]
	QF driver	mnx1:QF	Expresses QF in spinal motor neurons, the epithalamus, and pancreatic beta cells, under the mnx1 promoter	[67]
	QF driver	lfabp:QF	Expresses QF in hepatocytes, in the pattern of the lfabp promoter	[67]
	QF driver	insulin:QF	Expresses QF in pancreatic β-cells, in the pattern of the insulin promoter	[67]
	QF driver	tp1:QF	Expresses QF in cells with active Notch signaling, in the pattern of the tp1 promoter	[67]
	QF driver	Claudin15la:QF	Expresses QF in gut, in the pattern of the Claudin15la promoter	[106]
	QF2 driver	col9a2:QF2	Expresses QF2 in notochord sheath cells, in the pattern of the col9a promoter	[104]
	QF2 driver	ath5:QF2	Expresses QF2 in retinal ganglion cells, in the pattern of the ath5 promoter	[107]
	QFGAL4 driver	ubi:QFGAL4	Expresses QFGAL4 ubiquitously, in the pattern of the ubi promoter	[68]

(continued)

Table 2
(continued)

Species	Type of transgene	Q-transgene	Purpose/function	Reference
	QFGAL4 driver	acta1:QFGAL4	Expresses QFGAL4 in muscle cells, in the pattern of the acta1 promoter	[68]
	Inducible QFGAL4 driver	ubi:QFGAL4-switch	Expresses QFGAL4 ubiquitously, in the pattern of the ubi promoter, when Cre recombinase is present	[68]
N. benthamiana (some constructs also in *Glycine max*)	QF driver-reporter	35Sp-QF-NOST-QUAS-GUS-NOST	Expresses QF and β-glucuronidase under the CaMV 35 S promoter	[71]
	QF driver-reporter	Nosp/35Sp:QF:QUAS:mEmerald	Expresses QF and mEmerald under the Nos or CaMV 35 S promoter	[70]
	QF2 driver-reporter	Nosp/35Sp:QF2:QUAS:mEmerald	Expresses QF2 and mEmerald under the Nos or CaMV 35 S promoter	[70]
	$QF2^{w}$ driver-reporter	Nosp/35Sp:$QF2^{w}$:QUAS:mEmerald	Expresses $QF2^{w}$ and mEmerald under the Nos or CaMV 35 S promoter	[70]
	QF2 driver-reporter-suppressor	Nosp:QF2::35Sp:QS::QUAS:mEmerald	Expresses QF2 and mEmerald repressed by QS under the CaMV 35 S promoter	[70]
	QUAS reporter	QUAS:mScarlet-I	mScarlet-I	[70]
	QF driver-reporter and split PLP0.A2	35Sp-QF-NOST-QUAS-GUS-NOST-35Sp-PLP0.A2-NOST	Expresses QF and PLP0.A2 under the CaMV 35 S promoter	[71]
	QS repressor	35Sp-QS-NOST	Expresses QS under the CaMV 35 S promoter	[71]
	QS repressor and split PLP0.B2	35Sp-QS-NOST-35Sp-PLP0.B2-NOST	Expresses QS and PLP0.B2 under the CaMV 35 S promoter	[71]
P. chrysogenum	QUAS reporter (knock-in)	QUAS-CP-RFP	RFP under various core promoters (CP): pcbC, pcbAB, penDE, phl, nirA, or ura3	[72]
	QF.DBD:VP16.AD driver (knock-in)	gndA-QF.DBD:VP16.AD; 40Sp-QF.DBD:VP16.AD	Expresses QF-like driver and RFP	[72]
	QF.DBD:VP16.AD driver and QUAS reporter (knock-in)	40Sp-QF.DBD:VP16.AD, QUAS-Pc-pcbAB-penDE	Expresses QF-like driver and Penicillin V	[72]

(continued)

Table 2
(continued)

Species	Type of transgene	Q-transgene	Purpose/function	Reference
S. cerevisiae	QUASf reporter (CGGs replaced with AGGs)	QUASf-GFP	GFP	[30]
	QUASf reporter	QUASf- PDC1	Pyruvate decarboxylase 1	[30]
	QF2 drivers	C120p/PGK1p-QF2	Expresses QF2 under the C120 or PGK1p promoters	[30]
	QF2 driver	PGK1p-QF2-PSD	Expresses QF2 fused to a photosensitive degradation tag (PSD) under the PGK1 promoter	[30]
	QS repressor	PADH1p/ HHF2p/ TEF1p-QS	Expresses QS under the constitutive promoters ADH1, HHF2, or TEF1	[30]
	Quinate permease for QA-mediated control	PGK1p-qa-y	Expresses the QA transporter quinate permease from qa-y, under the PGK1 promoter	[30]
Mammalian cells	QF driver	CMVp-QF	Expresses QF under the human cytomegalovirus promoter (CMVp)	[11]
	QS repressor	CMVp-QS	Expresses QS under the human cytomegalovirus promoter (CMVp)	[11]
	QUAS reporter	QUAS-luc2	Mammalian codon-optimized synthetic firefly luciferase (luc2)	[11]
	QUAS reporter	QUAS-FLPo	Mammalian codon-optimized FLP recombinase	[11]
E. coli	QUAS reporter	QUAS-0-T7-ccdB	CcdB toxin. QUAS is upstream from the T7 promoter	[31]
	QUAS reporter	T7-15/10/5-QUAS-GFP	GFP. QUAS is downstream at +15, +10, or +5 nucleotides from the T7 promoter	[31]
	QUAS reporter	QUAS-15/10/5-T7-GFP	GFP. QUAS is upstream at −15, −10, or −5 nucleotides from the T7 promoter	[31]
	QF driver	T7-QF	Expresses QF under the T7 promoter	[31]

(continued)

Table 2
(continued)

Species	Type of transgene	Q-transgene	Purpose/function	Reference
	QUAS-TetO reporter	QUAS-0-T7-TetO-GFP	GFP downstream of a TetO site for TetR repressor bindings. QUAS is upstream from the T7 promoter	[31]
	QF-TetR driver	T7-LacO-QF-T7-TetO-TetR	Expresses QF in the presence of anhydrotetracycline (aTc)	[31]

6.1 Mosquitoes

Several mosquito species are important vectors of deadly diseases, such as malaria, Zika, and dengue. Binary expression systems can provide flexible genetic access to specific mosquito tissues. Although the development of genetic tools in non-model organisms is often difficult, *Drosophila* can be used as a guide in adapting new methods in mosquitoes. The GAL4 system has been introduced into the emerging mosquito model species *Aedes aegypti*, *Anopheles gambiae* (including *An. coluzzii*), and *Anopheles stephensi* [73–75]. However, the GAL4 system is poorly functional in neuronal tissues in *Aedes aegypti* [66], and its expression levels often vary in mosquito cells [76]. In contrast, the Q-system has been recently implemented in *Ae. aegypti* and *An. gambiae*, where it shows strong and consistent reporter expression [65, 66, 77, 78]. The improved QF2 and QF2^{w} trans-activators are generally well tolerated in mosquitoes [65, 66]. Therefore, the Q-system has emerged as the binary system of choice for targeted gene expression in mosquito tissues.

The Q-system has been widely used to study the mosquito chemosensory system and its involvement in human host seeking, leading to the generation of a wide range of reagents (Table 2). Two main approaches are used to generate QF2/QF2^{w} driver lines, based either on promoter fusion or CRISPR knock-in technologies. Promoter fusions were used to first introduce the Q-system into *An. gambiae* [65]. In this technique, the presumptive promoter region of the targeted gene is cloned upstream of QF2 (or its variant), and the transgene is inserted into the genome together with a dominant visible marker, such as 3xP3-ECFP/DsRed. Initial studies employed random genomic insertion using the *piggyBac* transposon system [65, 78, 79]. This method has been used to characterize *An. gambiae* olfactory sensory neurons using an enhancer region for the *odorant receptor co-receptor (Orco)*. The expression pattern of the membrane-targeted GFP reporter was validated with immunostaining, which showed overlap in over

90% of labeled cells [65]. The same method has been used for in vivo calcium imaging of Orco+ neurons [78], and for ectopic expression of an olfactory receptor (*AgOR2*) in Orco+ neurons in *An. gambiae* [79]. To avoid fitness costs associated with the random insertion locus, Ye et al. [80] employed *phiC31* site-specific integration of the promoter fusion construct to generate a driver line targeting the *An. gambiae* ammonium transporter *AgAmt*.

Driver constructs with presumed promoter sequences may fail to reflect endogenous expression patterns [44]. Intergenic sequences are much longer in mosquitoes than in the smaller *Drosophila* genome. Enhancers and regulatory elements are often further away, and tools to predict their location are largely lacking [81]. Therefore, knocking in transactivator sequences into the endogenous genomic environment using CRISPR is currently the favored approach for driver line generation.

CRISPR/Cas9-mediated homologous recombination has been used to target the C-terminal of the gene of interest and replace its stop codon with an in-frame insertion of *T2A-QF2* (or its variant) [77]. The T2A peptide causes ribosomal skipping, ensuring the separation of the protein of interest from QF2 [82]. This method led to the generation of an *Ae. aegypti* driver line for the ion channel *ppk301* (*ppk301-T2A-QF2*) after failed attempts with the cloning of *Ae. aegypti* promoters and the GAL4 system [77]. It has since been used to target diverse chemosensory co-receptors (Orco, Ir25a, Ir76b, Ir8a, and CO_2 receptor subunits Gr3 and Gr1) and tuning receptors (Ir7f, Ir7a, and Gr4) in *Ae. aegypti* [23, 83–86] and the Ir76b Ionotropic Receptor co-receptor in *An. gambiae* [87] to study host seeking and blood feeding. The generation of a pan-neuronal *Ae. aegypti* driver line (*brp-T2A-QF2*^{w}) has recently been achieved using CRISPR knock-in [66]. It offers robust expression with limited fitness costs in heterozygotes, and may provide an important tool for future investigations into the mosquito chemosensory system [81]. The line has already been used for pan-neuronal GCaMP imaging to investigate sensory discrimination between nectar and blood [83]. The split-QF system has also been adopted for *Ae. aegypti* mosquitoes. Using the *IR25a-QF2-AD* and *Orco-QF2-DBD* drivers generated by CRISPR knock-in, Younger et al. [23] found wide co-expression of the chemosensory co-receptor genes *IR25a* and *Orco* in antennal and maxillary palp olfactory sensory neurons and in axonal projections in the antennal lobe.

A main barrier for mosquito driver line development lies in the fitness costs associated with transactivator expression, especially when widely expressed. For instance, Basrur et al. [88] failed to generate healthy *fruitless-T2A-QF2/QF2w* knock-ins. Such issues can often be solved by limiting the expression of the transactivator. The generation of the *Ae. aegypti* pan-neuronal driver illustrates well the process of driver line optimization [66]. The researchers

first targeted *T2A-QF2* to the *Syt1* pan-neuronal gene, but this initial attempt caused larval lethality. Therefore, as a compromise between expression level and fitness costs, they decided to target a weaker pan-neuronal gene *brp* and swapped QF2 for the weaker $QF2^{w}$. The resulting line still resulted in strong expression but with minimal fitness costs.

Mosquito driver lines can be crossed to various reporter lines already in use to achieve unique expression patterns. As the promoters of chemosensory receptors are rather weak, multiple copies (15 and 30) of *QUAS* are often used in reporter constructs [65, 84]. Fluorophores for labeling of neural populations and GCaMP for in vivo calcium imaging are routinely used approaches in *Drosophila*, and are starting to be implemented in mosquitoes. Optogenetic experiments with a CaMPARI2 reporter [84] and chemogenetics using the *tdTomato-T2A-TRPV1* reporter have also been performed in mosquitoes [83].

Although the basic components of the Q-system are widely used in mosquitoes, complex modifications, such as intersectional expression with two binary systems or temporal control with QS/QA, have not yet been achieved. This is in part due to difficulties in maintaining multiple transgenes in mosquitoes. The use of 4 or more transgenes in the same mosquito has not been tested, and the introduction of additional transgenes is limited by the paucity of genetic markers available. The repressible Q-system may be more widely applied in insects in the future, and could be used for population control strategies against mosquitoes, other disease vectors, or agricultural pests. For instance, inducible Q-system-based male sterility systems could work in parallel with other transgenic sterility or lethality systems to counter potential resistance development [89]. With the emergence of new genetic tools in mosquitoes, such advanced modifications may soon be possible.

6.2 C. elegans

Inducible genetic expression systems are scarce in the important model *C. elegans*. Heat shock promoter (hsp)-driven DNA recombinases are widely used to achieve tissue-specific expression [90]. However, this method only provides non-reversible induction and is dependent on heat stress that might affect other aspects of cellular physiology. Increasing efforts are focused on making the use of binary expression systems easier in the nematode model, as they show great potential in targeting single neurons [91, 92]. The Q-system was the first binary system to be introduced into *C. elegans* [21]. The introduction of the GAL4-system was initially limited by reduced transcription at the low growth temperature of the worm. This issue has since been resolved by swapping the *S. cerevisiae* GAL4 sequences with those of another yeast species with higher optimal growth temperature [91]. Although all four binary systems (GAL4, QF/QF2, tetR/rtetR, and LexA) and a number of hybrid systems (tetR-QF, tetR-L-QF, $Gal4_{SK}$-QF, and

LexA-L-QF) are functional in nematodes [18, 91, 92], the Q-system is among the most strongly expressed in a single copy [92].

Driver lines in *C. elegans* can be generated via extrachromosomal arrays or single-copy integration. The latter method uses Mos1-mediated single-copy insertion or recombinase-mediated cassette exchange to achieve more stable and reproducible expression compared to extrachromosomal arrays [21, 92]. The functionality of these two transgenic approaches was validated by the labeling of A-type motor neurons using the *unc4-QF* driver [21] and touch receptor neurons using drivers for *mec-4* [92]. *C. elegans* reporters contain the *Δpes-10* minimal promoter after the *QUAS* sequence, instead of the TATA sequence used in *Drosophila* and mosquitoes [21]. The QS/QA components work well in *C. elegans*, and have been used to provide temporal control of gene expression in developmental studies [21, 28, 29]. Thanks to the high efficiency of QS-mediated repression, several studies employed a widely expressed QF driver in combination with a limiting QS transgene to drive reporter expression in cells without known specific promoters [21, 93–97]. QF is expressed under a promoter active in the cells of interest and some other cells. A QS transgene represses QF activity in these additional cells, leading to reporter expression only in the cells of interest [21]. Using this approach, optogenetic manipulations were achieved in AS motor neurons only – two QS transgenes (*pacr-5::QS::mCherry; punc-4::QS::mCherry*) restricted the expression of *unc-17p::QF* (active in all cholinergic neurons) and selective illumination ensured activation in AS motor neurons only [96]. In another application of the Q-system, fluorescent QUAS transgenes reported the inclusion or exclusion of an exon of the Ryanodine receptor *unc-68*, highlighting the tissue specificity of splice isoforms [98]. In addition, a version of the split-QF system is available in *C. elegans* [21].

Components of the Q-system have been combined with other tools to make new genetic approaches for *C. elegans*. A new ligand-gated inducer for temporal control, the glucocorticoid receptor ligand-binding domain, was added to the Q-system in nematodes [32]. Components of the system were used together with hsp-driven Cre recombinase to make a stochastic neuronal labeling tool based on Brainbow technologies [99]. This can be used to identify neuronal networks behind target behaviors via optogenetic experiments. Tet/Q, a new inducible hybrid binary system, was established based on the *C. elegans* Q-system by combining the Tetracycline-controlled transcription system with the QF activation domain [18].

6.3 Zebrafish

Although the GAL4 system is widely used in *Danio rerio* [100], progressive silencing of the *UAS* sequence by methylation of CpG islands limits its efficacy [101]. This led to the introduction of the Q-system as well as two further binary systems, LexA and TrpR

[67, 102, 103]. The Q-system is less susceptible to silencing, and its introduction into zebrafish makes intersectional labeling with two binary systems possible [22, 67]. Several stable tissue-specific driver lines were made available upon the introduction of the system into zebrafish, for instance, targeting hepatocytes (*lfabp:QF*), pancreatic β-cells (*insulin:QF*), cells with active Notch signaling (*tp1:QF*), or spinal motor neurons as well as the epithalamus and pancreatic β-cells (*mnx1:QF*) [67]. The QF, QF2, QF^{w}, and QFGAL4 transactivators have all been tested in zebrafish [22, 67, 68]. The Q-system has also been shown to work in parallel with the GAL4 system, as demonstrated by the expression of a dual reporter construct (*pBT2QUAS:dTomato-4XnrUAS:GFP*) and two driver lines, achieving distinct labeling of motor neurons and hindbrain neurons [67]. Such dual reporter constructs can save time in the zebrafish, where introducing several transgenes into a single animal takes longer. Burgess et al. [68] made important optimizations of the Q-system for zebrafish studies. Although an apparently stable *QUAS-GFP* reporter line was previously established [67], concerns about the potential silencing of *QUAS* persisted, due to its comparable CpG content to *UAS*. Burgess et al. in their study lowered the CpG content and repeat number of *QUAS*, reducing DNA methylation and leaky expression [68].

Zebrafish driver lines are commonly generated via Tol2 transposition resulting in single-copy genomic integration [67]. Additionally, gene/enhancer trapping was identified as a viable method to generate QF driver lines for cell types without known tissue-specific promoters in zebrafish [67]. Burgess et al. [68] demonstrated that driver lines can also be generated by knocking in *QFGAL4* into endogenous genes, with the help of a *pGTag-P2A-QFGAL4* construct. Using TALENs to knock in QFGAL4 to the *sox2* gene in *QUAS-GFP*^{NLS} background, the researchers visualized QFGAL4-tagged *sox2* expression in the developing nervous system. The ubiquitously expressed *ubiquitin C* gene was also tagged with *QFGAL4,* via a CRISPR knock-in method. As expected, a broad expression pattern was found in the *QUAS-GFP*^{NLS} background. In addition, Burgess et al. [68] generated transgenic lines to be used for Cre-induced lineage tracing studies.

Initial experiments established that QS can efficiently repress QF activity in zebrafish when injected as plasmids [67]. However, the addition of the QS repressor as an integrated transgene has not yet been achieved, and tests of the potential use of QA showed high health costs at functional concentrations [22]. QS mRNA injections into 1-cell stage embryos have been used to bypass the adverse early developmental effects of the misexpression of the *mespbb* gene in *col9a2*-expressing cells (*col9a2:QF2; QUAS:mespbb-p2A-GFP*) in a study on notochord segmentation [104].

The Q-system has been used in several studies on zebrafish development to overexpress genes of interest in specific tissues using stable driver lines. It was used to activate/inhibit Notch signaling specifically within the notochord sheath (*col9a2:QF2*; *QUAS:nlsVenus-V2a-notch1aICD/QUAS:nVenus-V2a-SuHDN*) in a background with a notochord segmentation marker [104]. In another study on notochord development, *QUAS:spondo-p2A-EGFP* DNA was injected into 1-cell stage *col9a2:QF2* embryos containing a notochord segmentation marker to drive overexpression of *spondo,* a mutant form of the *calymmin* gene, specifically in the notochord sheath [105]. The Q-system was also used in a study on epithelial morphogenesis, where 1-cell stage *Claudin15la:QF* embryos were injected with *pTol2-QUAS:Smtnl-GFP* DNA to visualize *Smoothelin* expression specifically in the gut [106]. Fernandes et al. [107] in their study combined transgenes expressing cytosolic GCaMP6s in retinal ganglion cells (*ath5:QF2; QUAS:GCaMP6s*) and nucleus-localized GCaMP6s in all neurons (*elavl3:nlsGCaMP6s*) to make simultaneous but distinguishable in vivo recordings of cell body layer activity and retinal ganglion cell activity in the tectal neuropil layers [107].

A recurring issue in the generation of stable zebrafish lines lies in the fitness costs associated with QF and QF2 when widely expressed [22, 68]. $QF2^{w}$ and QFGAL4 transactivators may be used instead, both of which are well tolerated even when widely expressed [68]. In fact, QFGAL4 can be ubiquitously expressed with minimal adverse effects [68]. Another consideration is the maternal deposition of *QUAS*-regulated gene expression. When studying early development, paternal transmission of the *QUAS* transgene is recommended [67].

6.4 Other Organisms

A number of recent studies have implemented the Q-system in organisms beyond widely used animal models and mosquitoes. These include plants, fungi, mammalian cells, and prokaryotes.

6.4.1 Plants

To address the paucity of well-validated expression systems in plants, two studies adopted the Q-system [70, 71]. Reis et al. [71] implemented the Q-components in transiently agroinfiltrated leaves of the model plant *Nicotiana benthamiana* (Australian tobacco). The resulting plant Q-system contains the transactivator and the reporter in the same construct driven by the *35S* constitutive promoter (*35S-QF-QUAS-GUS*), while the repressor is provided in a separate *35S-QS* construct. This system was additionally coupled with a split silencing suppressor system PLP0 [71]. Persad et al. [70] expanded and further validated the Q-system in plants, by performing in vitro assays in soybean (*Glycine max*) protoplasts and *in planta* experiments in agroinfiltrated *N. benthamiana* leaves. The transactivators QF, QF2, and QF^{w} were compared using *35S:QF/QF2/QF*^{w}*:5xQUAS:mEmerald*

constructs, and QF2 was recommended for use due to its low toxicity. As reporter signals were often low in *N. benthamiana* assays, the copy number of *QUAS* was further increased, successfully tuning expression to promoter signals of varying strength. While Reis et al. [71] reported only partial repression by QS, Persad et al. [70] achieved complete suppression using a single construct, with QF expressed under the Nos promoter and QS under the 35S promoter (*Nos:QF:2×35S:QS:5xQUAS:mEmerald*). The latter study also demonstrated that multiple reporter genes can be activated by a single QF transactivator, where QF is only expressed from one out of three constructs with distinct fluorescent reporters [70]. Although these studies demonstrate the functionality of the Q-system in plants, further efforts reducing fitness costs and optimizing codon composition may be needed before its wider adoption.

6.4.2 Fungi

Components of the Q-system have recently been introduced into two industrially relevant model species of fungi, *Penicillium chrysogenum* and *Saccharomyces cerevisiae* [30, 72]. In the *Penicillium chrysogenum* system, the QF-DBD was fused to the Herpes simplex virus VP16 activation domain and GFP with the SV40 nuclear localization signal. Expression levels were tuned by changing the promoter driving the transactivator and by increasing the copy number of *QUAS*. The functionality of the system was validated by the induction of penicillin production [72]. To introduce the Q-system into the yeast *S. cerevisiae,* Lalwani et al. [30] used the QF2 transactivator and embedded the *5xQUAS* sequence into a truncated *PCYC1* promoter. To prevent leaky cross-activation of *5xQUAS* by GAL4 in the presence of galactose, CGG sequences found in GAL4 binding sites were modified. Efficient repression with QS and tunable temporal de-repression with QA were also implemented. The system was used in optogenetic inverter and amplifier circuits, which can be combined with existing GAL4-based circuits to induce bidirectional control of gene expression by darkness and light. This was validated by acetoin production and the fine-tuning of geraniol/linalool blends involved in determining the flavor of beer [72].

6.4.3 Mammalian Cells

The Q-system has been introduced into mammalian cells including human HeLa cells, Chinese hamster ovarian cells (CHO), human embryonic kidney 293 cells, and mouse embryonic stem cells [11, 69]. In transiently transfected Hela cells, QS repressed QF expression, especially when provided in multiple copies. Interestingly, QS was not repressed but rather further activated by QA in this system [11]. Additionally, QS was found not to repress QF in transfected CHO cells [69]. To achieve temporal control, the Q-system was coupled with the LacI system in a novel genetic tool [69].

6.4.4 Bacteria

A recent paper introduced the Q-system into *E. coli* bacteria, demonstrating the functionality of eukaryotic genetic tools in prokaryotes [31]. The functionality of the *QUAS* component is strongly dependent on its placement relative to the T7 promoter driving the expression of GFP. When placed downstream of T7, it induces robust transcriptional activation. However, it results in strong transcriptional repression in an upstream position when QF expression is lacking. The underlying mechanism is unknown, but it might be because the T7 promoter is made inaccessible to the T7 RNA polymerase either by a native repressor binding *QUAS* or conformational changes induced by *QUAS* in an upstream position in the absence of QF. Taking advantage of the tight off-state of *QUAS* placed upstream of the promoter, the researchers coupled the Q-system with the TetR system to create inducible biological sensors, filters, and circuits for protein production [31].

7 Future Directions

Since its creation in 2010, the Q-system has been used in hundreds of studies on *Drosophila* and other organisms. The scientific community continues to optimize the use of the Q-system to suit specific experiments or a particular study species.

A larger collection of QF2/QF2^{w} driver lines and *QUAS* reporter lines would greatly benefit many researchers, and the HACK approach now makes QF driver creation an easy and inexpensive task. In non-melanogaster species, such as other *Drosophilids* or mosquitoes, large collections of lines would require dedicated stock centers with affordable and efficient deliveries around the world.

QF2/QF2^{w} variants were derived from QF as those that have low toxicity. However, in some cases further improvements are required to reduce the undesired effects of QF2/QF2^{w}. In the future, mutational screens of QF2/QF2^{w} could be used to determine even healthier QF variants that retain transcriptional activity. The *QUAS* sequence may be also modified to mirror changes in the QF2/QF2^{w} or may potentially be combined with *UAS* to create generic reporter constructs. The repressor QS may also be modified to produce temperature-, light-, or non-QA-drug-inducible variants.

QA delivery to the brain and potentially other tissues currently remains unresolved. Chemical modifications of the QA itself, alternative delivery methods, or expression of QA transporters in the cells of interest may all facilitate the QA delivery, and form the basis of our ongoing work.

A major future prospect lies in the application of the Q-system to non-model organisms. In insects and some other organisms CRISPR/Cas9-mediated knock-ins of Q-system components have been very successful. In the future, it may be possible to deliver the Q-system components via the REMOT system [108], bypassing labor-intensive embryo injections.

Acknowledgements

The authors thank Steve Chivasa, Tim Davies, and Jessica Mavica for insightful comments on the manuscript. OF was funded by the Laidlaw fellowship. OR was funded by the Wellcome Trust (217440/Z/19/Z) and the Royal Society (RGS\R2\192005).

References

1. Brand AH, Perrimon N (1993) Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. Development 118(2):401–415. https://doi.org/10.1101/lm.1331809
2. Johnston M (1987) A model fungal gene regulatory mechanism: the GAL genes of Saccharomyces cerevisiae. Microbiol Rev 51(4):458–476. https://doi.org/10.1128/MR.51.4.458-476.1987
3. Platt A, Reece RJ (1998) The yeast galactose genetic switch is mediated by the formation of a Gal4p–Gal80p–Gal3p complex. EMBO J 17(14):4086–4091. https://doi.org/10.1093/EMBOJ/17.14.4086
4. Giles NH, Geever RF, Asch DK, Avalos J, Case ME (1989) Organization and regulation of the qa (quinic acid) genes in Neurospora crassa and other fungi. J Hered 82:1–7
5. Baum JA, Geever R, Giles NH (1987) Expression of qa-1F activator protein: identification of upstream binding sites in the qa gene cluster and localization of the DNA-binding domain. Mol Cell Biol 7(3):1256–1266. https://doi.org/10.1128/MCB.7.3.1256-1266.1987
6. Geever RF, Huiet L, Baum JA, Tyler BM, Patel VB, Rutledge BJ et al (1989) DNA sequence, organization and regulation of the qa gene cluster of Neurospora crassa. J Mol Biol 207(1):15–34. https://doi.org/10.1016/0022-2836(89)90438-5
7. McGuire SE, Le PT, Osborn AJ, Matsumoto K, Davis RL (2003) Spatiotemporal rescue of memory dysfunction in Drosophila. Science (80-) 302(5651):1765–1768. https://doi.org/10.1126/SCIENCE.1089035
8. Lai SL, Lee T (2006) Genetic mosaic with dual binary transcriptional systems in Drosophila. Nat Neurosci 9(5):703–709. https://doi.org/10.1038/nn1681
9. Pfeiffer BD, Ngo T-TB, Hibbard KL, Murphy C, Jenett A, Truman JW et al (2010) Refinement of tools for targeted gene expression in Drosophila. Genetics 186(2):735–755. https://doi.org/10.1534/genetics.110.119917
10. Lee T, Luo L (1999) Mosaic analysis with a repressible cell marker for studies of gene function in neuronal morphogenesis. Neuron 22(3):451–461. https://doi.org/10.1016/S0896-6273(00)80701-1
11. Potter CJ, Tasic B, Russler EV, Liang L, Luo L (2010) The Q system: a repressible binary system for transgene expression, lineage tracing, and mosaic analysis. Cell 141(3):536–548. https://doi.org/10.1016/j.cell.2010.02.025
12. Riabinina O, Luginbuhl D, Marr E, Liu S, Wu MN, Luo L et al (2015) Improved and expanded Q-system reagents for genetic manipulations. Nat Methods 12(3):219
13. Gill G, Ptashne M (1987) Mutants of GAL4 protein altered in an activation function. Cell 51(1):121–126. https://doi.org/10.1016/0092-8674(87)90016-X
14. Kramer JM, Staveley BE (2003) GAL4 causes developmental defects and apoptosis when expressed in the developing eye of Drosophila melanogaster. Genet Mol Res 2(1):43–47

15. Shearin HK, MacDonald IS, Spector LP, Steven Stowers R (2014) Hexameric GFP and mCherry reporters for the Drosophila GAL4, Q, and LexA transcription systems. Genetics 196(4):951–960. https://doi.org/10.1534/genetics.113.161141
16. Pfeiffer BD, Truman JW, Rubin GM (2012) Using translational enhancers to increase transgene expression in Drosophila. Proc Natl Acad Sci 109(17):6626–6631. https://doi.org/10.1073/pnas.1204520109
17. Riabinina O, Vernon SW, Dickson BJ, Baines RA (2019) Split-QF system for fine-tuned transgene expression in Drosophila. Genetics 212(1):genetics.302034.2019. https://doi.org/10.1534/genetics.119.302034
18. Mao S, Qi Y, Zhu H, Huang X, Zou Y, Chi T (2019) A Tet/Q hybrid system for robust and versatile control of transgene expression in C. elegans. iScience 11:224–237. https://doi.org/10.1016/j.isci.2018.12.023
19. Luan H, Peabody NC, Vinson CRR, White BH (2006) Refined spatial manipulation of neuronal function by combinatorial restriction of transgene expression. Neuron 52(3):425–436. https://doi.org/10.1016/j.neuron.2006.08.028
20. Tirian L, Dickson B (2017) The VT GAL4, LexA, and split-GAL4 driver line collections for targeted expression in the Drosophila nervous system. bioRxiv. 198648. https://doi.org/10.1101/198648
21. Wei X, Potter CJ, Luo L, Shen K (2012) Controlling gene expression with the Q repressible binary expression system in Caenorhabditis elegans. Nat Methods 9(4):391
22. Ghosh A, Halpern ME (2016) Transcriptional regulation using the Q system in transgenic zebrafish. Methods Cell Biol 135:205–218. https://doi.org/10.1016/bs.mcb.2016.05.001
23. Younger MA, Herre M, Goldman OV, Lu T-C, Caballero-Vidal G, Qi Y, et al (2020) Non-canonical odor coding ensures unbreakable mosquito attraction to humans. bioRxiv. 2020.11.07.368720. https://doi.org/10.1101/2020.11.07.368720
24. Potter CJ, Luo L (2011) Using the Q system in Drosophila melanogaster. Nat Protoc 6(8):1105–1120
25. Li YX, Sibon OCM, Dijkers PF (2018) Inhibition of NF-KB in astrocytes is sufficient to delay neurodegeneration induced by proteotoxicity in neurons. J Neuroinflammation 15(1):261. https://doi.org/10.1186/s12974-018-1278-2
26. Edwards TN, Meinertzhagen IA (2010) The functional organisation of glia in the adult brain of Drosophila and other insects. Prog Neurobiol 90(4):471–497. https://doi.org/10.1016/j.pneurobio.2010.01.001
27. Li Q, Stavropoulos N (2016) Evaluation of ligand-inducible expression systems for conditional neuronal manipulations of sleep in drosophila. G3 (Bethesda) 6(10):3351–3359. https://doi.org/10.1534/g3.116.034132
28. Cottee PA, Cole T, Schultz J, Hoang HD, Vibbert J, Han SM et al (2017) The *C. elegans* VAPB homolog VPR-1 is a permissive signal for gonad development. Development 144(12):2187–2199. https://doi.org/10.1242/dev.152207
29. Hoang HD, Miller MA (2017) Chemosensory and hyperoxia circuits in C. elegans males influence sperm navigational capacity. PLoS Biol 15(6):e2002047. https://doi.org/10.1371/journal.pbio.2002047
30. Lalwani MA, Zhao EM, Wegner SA, Avalos JL (2021) The Neurospora crassa inducible Q system enables simultaneous optogenetic amplification and inversion in saccharomyces cerevisiae for bidirectional control of gene expression. ACS Synth Biol. acssynbio.1c00229. https://doi.org/10.1021/ACSSYNBIO.1C00229
31. MacDonald IC, Seamons TR, Emmons JC, Javdan SB, Deans TL (2021) Enhanced regulation of prokaryotic gene expression by a eukaryotic transcriptional activator. Nat Commun 12(1):1–10. https://doi.org/10.1038/s41467-021-24434-9
32. Monsalve GC, Yamamoto KR, Ward JD (2018) A new tool for inducible gene expression in Caenorhabditis elegans. Genetics. genetics.301705.2018. https://doi.org/10.1534/genetics.118.301705
33. Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E (2012) A programmable dual-RNA–guided DNA endonuclease in adaptive bacterial immunity. Science (80-) 337(6096):816–821
34. Lin C-C, Potter CJ (2016) Editing transgenic DNA components by inducible gene replacement in Drosophila melanogaster. Genetics 203(4):1613–1628
35. Siegal ML, Hartl DL (1996) Transgene coplacement and high efficiency site-specific recombination with the Cre/loxP system in Drosophila. Genetics 144(2):715–726.
36. Port F, Chen H-M, Lee T, Bullock SL (2014) Optimized CRISPR/Cas tools for efficient

germline and somatic genome engineering in Drosophila. Proc Natl Acad Sci 111(29): E2967–E2976. https://doi.org/10.1073/PNAS.1405500111

37. Gratz SJ, Cummings AM, Nguyen JN, Hamm DC, Donohue LK, Harrison MM et al (2013) Genome engineering of Drosophila with the CRISPR RNA-guided Cas9 nuclease. Genetics 194(4):1029–1035. https://doi.org/10.1534/GENETICS.113.152710
38. Levis R, Hazelrigg T, Rubin GM (1985) Effects of genomic position on the expression of transduced copies of the white gene of Drosophila. Science (80-) 229(4713): 558–561. https://doi.org/10.1126/SCIENCE.2992080
39. Xie T, Ho MCW, Liu Q, Horiuchi W, Lin C-C, Task D et al (2018) A genetic toolkit for dissecting dopamine circuit function in Drosophila. Cell Rep 23(2):652–665. https://doi.org/10.1016/J.CELREP.2018.03.068
40. Allen SE, Koreman GT, Sarkar A, Wang B, Wolfner MF, Han C (2021) Versatile CRISPR/Cas9-mediated mosaic analysis by gRNA-induced crossing-over for unmodified genomes. PLoS Biol 19(1):e3001061. https://doi.org/10.1371/JOURNAL.PBIO.3001061
41. Wang P, Jia Y, Liu T, Jan YN, Zhang W (2020) Visceral mechano-sensing neurons control Drosophila feeding by using piezo as a sensor. Neuron 108(4):640–650.e4. https://doi.org/10.1016/J.NEURON.2020.08.017
42. Koreman GT, Xu Y, Hu Q, Zhang Z, Allen SE, Wolfner MF et al (2021) Upgraded CRISPR/Cas9 tools for tissue-specific mutagenesis in Drosophila. Proc Natl Acad Sci 118(14). https://doi.org/10.1073/PNAS.2014255118
43. Terradas G, Buchman AB, Bennett JB, Shriner I, Marshall JM, Akbari OS et al (2021) Inherently confinable split-drive systems in Drosophila. Nat Commun 12(1): 1–12. https://doi.org/10.1038/s41467-021-21771-7
44. Task D, Lin C-C, Vulpe A, Afify A, Ballou S, Brbić M et al (2022) Chemoreceptor co-expression in *Drosophila melanogaster* olfactory neurons. eLife 11:e72599. https://doi.org/10.7554/eLife.72599
45. Chen Y, Dahanukar A (2017) Molecular and cellular organization of taste neurons in adult Drosophila pharynx. Cell Rep 21(10): 2978–2991. https://doi.org/10.1016/J.CELREP.2017.11.041
46. Lee S, Jones WD, Kim DH (2020) A cyclic nucleotide-gated channel in the brain regulates olfactory modulation through neuropeptide F in fruit fly Drosophila melanogaster. Arch Insect Biochem Physiol 103(1): e21620. https://doi.org/10.1002/arch.21620
47. Youn H, Kirkhart C, Chia J, Scott K (2018) A subset of octopaminergic neurons that promotes feeding initiation in Drosophila melanogaster. PLoS One 13(6):e0198362. https://doi.org/10.1371/JOURNAL.PONE.0198362
48. Carreira-Rosario A, Zarin AA, Clark MQ, Manning L, Fetter RD, Cardona A et al (2018) MDN brain descending neurons coordinately activate backward and inhibit forward locomotion. eLife 7. https://doi.org/10.7554/ELIFE.38554
49. Gao X, Riabinina O, Li J, Potter C, Clandinin T, Luo L (2015) A transcriptional reporter of intracellular Ca(2+) in Drosophila. Nat Neurosci 18(6):917–925. https://doi.org/10.1038/NN.4016
50. Guo F, Chen X, Rosbash M (2017) Temporal calcium profiling of specific circadian neurons in freely moving flies. Proc Natl Acad Sci 114(41):E8780–E8787. https://doi.org/10.1073/PNAS.1706608114
51. Simpson J, Looger L (2018) Functional imaging and optogenetics in Drosophila. Genetics 208(4):1291–1309. https://doi.org/10.1534/GENETICS.117.300228
52. Jenett A, Rubin G, Ngo T, Shepherd D, Murphy C, Dionne H et al (2012) A GAL4-driver line resource for Drosophila neurobiology. Cell Rep 2(4):991–1001. https://doi.org/10.1016/J.CELREP.2012.09.011
53. Li HH, Kroll JR, Lennox SM, Ogundeyi O, Jeter J, Depasquale G et al (2014) A GAL4 driver resource for developmental and behavioral studies on the larval CNS of Drosophila. Cell Rep 8(3):897–908. https://doi.org/10.1016/J.CELREP.2014.06.065
54. Pfeiffer BD, Jenett A, Hammonds AS, Ngo T-TB, Misra S, Murphy C et al (2008) Tools for neuroanatomy and neurogenetics in Drosophila. Proc Natl Acad Sci 105(28): 9715–9720. https://doi.org/10.1073/PNAS.0803697105
55. Kvon EZ, Kazmar T, Stampfel G, Yáñez-Cuna JO, Pagani M, Schernhuber K et al (2014) Genome-scale functional characterization of Drosophila developmental enhancers in vivo. Nature 512(7512):91–95. https://doi.org/10.1038/nature13395

56. Gohl DM, Silies MA, Gao XJ, Bhalerao S, Luongo FJ, Lin CC et al (2011) A versatile in vivo system for directed dissection of gene expression patterns. Nat Methods 8(3): 231–237. https://doi.org/10.1038/nmeth.1561
57. Venken KJT, Schulze KL, Haelterman NA, Pan H, He Y, Evans-Holm M et al (2011) MiMIC: a highly versatile transposon insertion resource for engineering Drosophila melanogaster genes. Nat Methods 8(9):737–747. https://doi.org/10.1038/nmeth.1662
58. Diao F, Ironfield H, Luan H, Diao F, Shropshire W, Ewer J et al (2015) Plug-and-play genetic access to drosophila cell types using exchangeable exon cassettes. Cell Rep 10(8):1410–1421. https://doi.org/10.1016/J.CELREP.2015.01.059
59. Lee PT, Zirin J, Kanca O, Lin WW, Schulze KL, Li-Kroeger D et al (2018) A gene-specific T2A-GAL4 library for drosophila. eLife 7. https://doi.org/10.7554/ELIFE.35574
60. Golic K, Lindquist S (1989) The FLP recombinase of yeast catalyzes site-specific recombination in the Drosophila genome. Cell 59(3): 499–509. https://doi.org/10.1016/0092-8674(89)90033-0
61. Bischof J, Basler K (2008) Recombinases and their use in gene activation, gene inactivation, and transgenesis. Methods Mol Biol 420: 175–195. https://doi.org/10.1007/978-1-59745-583-1_10
62. Nern A, Pfeiffer BD, Svoboda K, Rubin GM (2011) Multiple new site-specific recombinases for use in manipulating animal genomes. Proc Natl Acad Sci 108(34): 14198–14203. https://doi.org/10.1073/PNAS.1111704108
63. Dolan M, Luan H, Shropshire W, Sutcliffe B, Cocanougher B, Scott R et al (2017) Facilitating neuron-specific genetic manipulations in Drosophila melanogaster using a split GAL4 repressor. Genetics 206(2):775–784. https://doi.org/10.1534/GENETICS.116.199687
64. Sethi S, Wang JW (2017) A versatile genetic tool for post-translational control of gene expression in Drosophila melanogaster. eLife 6. https://doi.org/10.7554/eLife.30327
65. Riabinina O, Task D, Marr E, Lin C-C, Alford R, O'brochta DA et al (2016) Organization of olfactory centres in the malaria mosquito Anopheles gambiae. Nat Commun 7: 13010
66. Zhao Z, Tian D, McBride CS (2021) Development of a pan-neuronal genetic driver in Aedes aegypti mosquitoes. Cell Rep Methods 1(3):100042. https://doi.org/10.1016/J.CRMETH.2021.100042
67. Subedi A, Macurak M, Gee ST, Monge E, Goll MG, Potter CJ et al (2014) Adoption of the Q transcriptional regulatory system for zebrafish transgenesis. Methods 66(3): 433–440
68. Burgess J, Burrows JT, Sadhak R, Chiang S, Weiss A, D'Amata C et al (2020) An optimized QF-binary expression system for use in zebrafish. Dev Biol 465(2):144–156. https://doi.org/10.1016/j.ydbio.2020.07.007
69. Fitzgerald M, Gibbs C, Shimpi AA, Deans TL (2017) Adoption of the Q transcriptional system for regulating gene expression in stem cells. ACS Synth Biol 6(11):2014–2020
70. Persad R, Reuter DN, Dice LT, Nguyen M-A, Rigoulot SB, Layton JS et al (2020) The Q-system as a synthetic transcriptional regulator in plants. Front Plant Sci 11:245
71. Reis RS, Litholdo CG, Bally J, Roberts TH, Waterhouse PM (2018) A conditional silencing suppression system for transient expression. Sci Rep 8(1):1–7. https://doi.org/10.1038/s41598-018-27778-3
72. Mózsik L, Büttel Z, Bovenberg RAL, Driessen AJM, Nygård Y (2019) Synthetic control devices for gene regulation in Penicillium chrysogenum. Microb Cell Factories 18(1): 1–13. https://doi.org/10.1186/s12934-019-1253-3
73. Lynd A, Lycett GJ (2012) Development of the bi-partite Gal4-UAS system in the African malaria mosquito, Anopheles gambiae. PLoS One 7(2):e31552. https://doi.org/10.1371/JOURNAL.PONE.0031552
74. Kokoza VA, Raikhel AS (2011) Targeted gene expression in the transgenic Aedes aegypti using the binary Gal4-UAS system. Insect Biochem Mol Biol 41(8):637–644. https://doi.org/10.1016/J.IBMB.2011.04.004
75. O'Brochta DA, Pilitt KL, Harrell RA, Aluvihare C, Alford RT (2012) Gal4-based enhancer-trapping in the malaria mosquito Anopheles stephensi. G3 (Bethesda) 2(11): 1305–1315. https://doi.org/10.1534/G3.112.003582
76. Lynd A, Lycett GJ (2011) Optimization of the Gal4-UAS system in an Anopheles gambiae cell line. Insect Mol Biol 20(5):599–608. https://doi.org/10.1111/J.1365-2583.2011.01090.X
77. Matthews BJ, Younger MA, Vosshall LB (2019) The ion channel ppk301 controls freshwater egg-laying in the mosquito Aedes aegypti. eLife 8:e43963
78. Afify A, Betz JF, Riabinina O, Lahondère C, Potter CJ (2019) Commonly used

insect repellents hide human odors from Anopheles mosquitoes. Curr Biol 29(21): 3669–3680

79. Maguire SE, Afify A, Goff LA, Potter CJ (2022) Odorant-receptor-mediated regulation of chemosensory gene expression in the malaria mosquito Anopheles gambiae. Cell Reports 38(10):110494. https://doi.org/10.1016/j.celrep.2022.110494
80. Ye Z, Liu F, Sun H, Barker M, Pitts RJ, Zwiebel LJ (2020) Heterogeneous expression of the ammonium transporter AgAmt in chemosensory appendages of the malaria vector, Anopheles gambiae. Insect Biochem Mol Biol 120:103360. https://doi.org/10.1016/j.ibmb.2020.103360
81. Potter CJ (2021) Unlocking pan-neuronal expression in mosquitoes. Cell Rep Methods 1(3):100051. https://doi.org/10.1016/J.CRMETH.2021.100051
82. Diao F, White BH (2012) A novel approach for directing transgene expression in Drosophila: T2A-Gal4 in-frame fusion. Genetics 190(3):1139–1144. https://doi.org/10.1534/GENETICS.111.136291
83. Jové V, Gong Z, Hol FJH, Zhao Z, Sorrells TR, Carroll TS et al (2020) Sensory discrimination of blood and floral nectar by Aedes aegypti mosquitoes. Neuron 108(6): 1163–1180.e12. https://doi.org/10.1016/j.neuron.2020.09.019
84. Shankar S, Tauxe GM, Spikol ED, Li M, Akbari OS, Giraldo D et al (2020) Synergistic coding of human odorants in the mosquito brain. bioRxiv. 2020.11.02.365916. https://doi.org/10.1101/2020.11.02.365916
85. Sorrells TR, Pandey A, Rosas-Villegas A, Vosshall LB (2022) A persistent behavioral state enables sustained predation of humans by mosquitoes. eLife 11:e76663. https://doi.org/10.7554/eLife.76663
86. Zhao Z, Zung JL, Hinze A, Kriete AL, Iqbal A, Younger MA, et al. (2022) Mosquito brains encode unique features of human odour to drive host seeking. Nature 605: 706–712. https://doi.org/10.1038/s41586-022-04675-4
87. Ye Z, Liu F, Sun H, Baker A, Zwiebel LJ (2021) Discrete roles of the Ir76b ionotropic co-receptor impact olfaction, blood feeding, and mating in the malaria vector mosquito Anopheles coluzzii. bioRxiv. 2021.07.05.451160. https://doi.org/10.1101/2021.07.05.451160
88. Basrur NS, Elena De Obaldia M, Morita T, Herre M, von Heynitz RK, Tsitohay YN et al (2020) Fruitless mutant male mosquitoes gain attraction to human odor. eLife. https://doi.org/10.7554/eLife.63982
89. Eckermann KN, Dippel S, KaramiNejadRanjbar M, Ahmed HM, Curril IM, Wimmer EA (2014) Perspective on the combined use of an independent transgenic sexing and a multifactorial reproductive sterility system to avoid resistance development against transgenic Sterile Insect Technique approaches. BMC Genet 15(2):1–10. https://doi.org/10.1186/1471-2156-15-S2-S17
90. Hubbard JAE (2014) FLP/FRT and Cre/lox recombination technology in C. elegans. Methods 68(3):417–424. https://doi.org/10.1016/J.YMETH.2014.05.007
91. Wang H, Liu J, Gharib S, Chai CM, Schwarz EM, Pokala N et al (2016) cGAL, a temperature-robust GAL4–UAS system for Caenorhabditis elegans. Nat Methods 14(2): 145–148. https://doi.org/10.1038/nmeth.4109
92. Nonet ML (2020) Efficient transgenesis in Caenorhabditis elegans using flp recombinase-mediated cassette exchange. Genetics 215(4):903–921. https://doi.org/10.1534/genetics.120.303388
93. Schild LC, Zbinden L, Bell HW, Yu YV, Sengupta P, Goodman MB et al (2014) The balance between cytoplasmic and nuclear CaM kinase-1 signaling controls the operating range of noxious heat avoidance. Neuron 84(5):983–996. https://doi.org/10.1016/J.NEURON.2014.10.039
94. Schild LC, Glauser DA (2015) Dual color neural activation and behavior control with chrimson and CoChR in Caenorhabditis elegans. Genetics 200(4):1029–1034. https://doi.org/10.1534/genetics.115.177956
95. Jee C, Goncalves JF, LeBoeuf B, Garcia LR (2016) CRF-like receptor SEB-3 in sex-common interneurons potentiates stress handling and reproductive drive in C. elegans. Nat Commun 7(1):1–15. https://doi.org/10.1038/ncomms11957
96. Tolstenkov O, Van der Auwera P, Costa WS, Bazhanova O, Gemeinhardt TM, Bergs ACF et al (2018) Functionally asymmetric motor neurons contribute to coordinating locomotion of Caenorhabditis elegans. eLife 7. https://doi.org/10.7554/ELIFE.34997
97. Chiyoda H, Kume M, Castillo CCD, Kontani K, Spang A, Katada T et al (2021) Caenorhabditis elegans PTR/PTCHD PTR-18 promotes the clearance of extracellular hedgehog-related protein via endocytosis. PLoS Genet 17(4):e1009457. https://doi.org/10.1371/JOURNAL.PGEN.1009457

98. Marques F, Thapliyal S, Javer A, Shrestha P, Brown AEX, Glauser DA (2020) Tissue-specific isoforms of the single C. elegans Ryanodine receptor gene unc-68 control specific functions. PLoS Genet 16(10):e1009102. https://doi.org/10.1371/JOURNAL.PGEN.1009102
99. Aoki W, Matsukura H, Yamauchi Y, Yokoyama H, Hasegawa K, Shinya R et al (2018) Cellomics approach for high-throughput functional annotation of Caenorhabditis elegans neural network. Sci Rep 8(1):1–9
100. Scheer N, Campos-Ortega J (1999) Use of the Gal4-UAS technique for targeted gene expression in the zebrafish. Mech Dev 80(2):153–158. https://doi.org/10.1016/S0925-4773(98)00209-3
101. Goll MG, Anderson R, Stainier DYR, Spradling AC, Halpern ME (2009) Transcriptional silencing and reactivation in transgenic zebrafish. Genetics 182(3):747–755. https://doi.org/10.1534/genetics.109.102079
102. Suli A, Guler AD, Raible DW, Kimelman D (2014) A targeted gene expression system using the tryptophan repressor in zebrafish shows no silencing in subsequent generations. Development 141(5):1167–1174. https://doi.org/10.1242/DEV.100057
103. Emelyanov A, Parinov S (2008) Mifepristone-inducible LexPR system to drive and control gene expression in transgenic zebrafish. Dev Biol 320(1):113–121. https://doi.org/10.1016/J.YDBIO.2008.04.042
104. Wopat S, Bagwell J, Sumigray KD, Dickson AL, Huitema LFA, Poss KD et al (2018) Spine patterning is guided by segmentation of the notochord sheath. Cell Rep 22(8):2026–2038. https://doi.org/10.1016/j.celrep.2018.01.084
105. Peskin B, Henke K, Cumplido N, Treaster S, Harris MP, Bagnat M et al (2020) Notochordal signals establish phylogenetic identity of the teleost spine. Curr Biol 30(14):2805–2814.e3. https://doi.org/10.1016/j.cub.2020.05.037
106. Hachimi M, Grabowski C, Campanario S, Herranz G, Baonza G, Serrador JM et al (2021) Smoothelin-like 2 inhibits coronin-1B to stabilize the apical actin cortex during epithelial morphogenesis. Curr Biol 31(4):696–706.e9. https://doi.org/10.1016/j.cub.2020.11.010
107. Fernandes AM, Mearns DS, Donovan JC, Larsch J, Helmbrecht TO, Kölsch Y et al (2021) Neural circuitry for stimulus selection in the zebrafish visual system. Neuron 109(5):805–822.e6. https://doi.org/10.1016/j.neuron.2020.12.002
108. Chaverra-Rodriguez D, Macias VM, Hughes GL, Pujhari S, Suzuki Y, Peterson DR et al (2018) Targeted delivery of CRISPR-Cas9 ribonucleoprotein into arthropod ovaries for heritable germline gene editing. Nat Commun 9(1):3008. https://doi.org/10.1038/s41467-018-05425-9

Chapter 3

Resources and Methods for the Analysis of MicroRNA Function in *Drosophila*

Sromana Mukherjee and Nicholas Sokol

Abstract

Since the widespread discovery of microRNAs (miRNAs) 20 years ago, the *Drosophila melanogaster* model system has made important contributions to understanding the biology of this class of noncoding RNAs. These contributions are based on the amenability of this model system not only for biochemical analysis but molecular, genetic, and cell biological analyses as well. Nevertheless, while the *Drosophila* genome is now known to encode 258 miRNA precursors, the function of only a small minority of these have been well characterized. In this review, we summarize the current resources and methods that are available to study miRNA function in *Drosophila* with a particular focus on the large-scale resources that enable systematic analysis. Application of these methods will accelerate the discovery of ways that miRNAs are embedded into genetic networks that control basic features of metazoan cells.

Key words *Drosophila melanogaster*, microRNA, miRNA

1 Introduction

MicroRNAs (miRNAs) were first reported as a large, conserved class of noncoding RNAs in a trio of *Science* papers published in 2001 [1–3]. Since then, much progress has been made in identifying the complete repertoire of animal miRNAs and characterizing their biogenesis [4]. These short, 21–25 nucleotide (nt) RNAs direct a multicomponent RNA-induced silencing complex (RISC), including Argonaute1 (Ago1), to degrade or repress partially complementary mRNAs. They are derived from longer, ~70-nt hairpin RNAs, known as pre-miRNAs, that are transcribed within even longer primary transcripts, or pri-miRNAs. Mature miRNAs are released via sequential cleavage, first of the pri-miRNAs by a ribonuclease complex containing two proteins, Drosha and Pasha. The second cleavage of the pre-miRNA by the ribonuclease Dicer releases two different RNAs from the 5′ and 3′ arm of the hairpin. Through extensive cataloguing, the *Drosophila*

Christian Dahmann (ed.), *Drosophila: Methods and Protocols*,
Methods in Molecular Biology, vol. 2540, https://doi.org/10.1007/978-1-0716-2541-5_3,

melanogaster genome is now known to encode 258 pri-miRNAs and 469 mature miRNAs, a subset of which are conserved to vertebrates [5].

Due to limited resources, the function of many of these mature miRNAs in distinct cell types and tissues has remained underexplored. Initial studies of miRNA function focused on depleting core miRNA processing components, such as *dicer, drosha, pasha,* and *ago1*, which may have miRNA-independent functions in some cases [6, 7]. However, the advent of multiple genetic toolkits that allow precise manipulation of miRNA levels *in vivo* has significantly expanded miRNA functional analysis. These publicly available libraries are now being used to perform large-scale genetic screens and have revealed interesting roles of miRNAs in behavioral and developmental processes. Here we have summarized available resources and methods used for studying *in vivo* miRNA functions in *Drosophila* and discussed some of their applications.

2 Deletion Mutants

A large-scale effort in the recent past has generated a comprehensive library of 80 miRNA knockout mutants by targeted homologous recombination. In addition to previously reported mutants, this collection includes a total of 95 mutants known to delete 130 miRNAs [8]. These deletion mutants have since then been used in genetic screens to identify miRNA functions in various tissues, developmental stages, and processes of interest.

2.1 Applications

A large-scale genetic screen using a collection of 81 of these mutants that tested self-righting behavior in larvae found that ~40% of these mutants significantly affected self-righting timing and thus revealed pervasive behavioral effects in *Drosophila* larvae [9]. Several of these miRNAs acted by regulating *Abdominal-B (Abd-B)* expression.

The miRNA mutant library has also been used to screen for individual or clusters of miRNAs involved in systemic *Candida albicans* infection [10]. This study identified 6 allelic mutant backgrounds that impacted survival, pathogen numbers, and the expression of Toll-dependent transcripts, including *Drosomycin (Drs)* and *Immune-Induced Molecule 1(IM1)*.

Another large-scale screen using the same library identified miRNAs that modulate multiple aspects of male courtship behavior [11]. One particular miRNA mutant, *miR-957,* significantly increased male–male courtship and chaining behavior due to reduced inhibitory responses to animals with high levels of male-aversive pheromones.

3 miRNA Sponges

The sponge method combines the widely used GAL4/UAS system in *Drosophila* to express miRNA sponge (miR-SP) constructs in a cell or tissue-specific manner to reduce overall miRNA activity. The UAS-miRNA sponge strains consist of 20 complementary miRNA sites inserted downstream of a mCherry reporter, allowing it to be driven using cell or tissue-specific GAL4 drivers [12]. To account for dose-dependent expressions and positional effects, these constructs were double inserted at both attP2 and attP40 sites, yielding a total of 282 transgenic lines. The available transgenic miRNA sponge library is known to target 141 well-conserved miRNAs in *Drosophila*. These miRNA sponge constructs can not only be spatially regulated using tissue- or cell-specific GAL4 drivers but can also be temporally regulated in combination with the temperature-sensitive GAL80-dependent Temporal and Regional Gene Expression Targeting (TARGET) system [13], allowing for the identification of miRNA functions during adulthood.

3.1 Applications

A recent screen that exploited the advantages of miR-SP constructs revealed new roles of miRNAs in regulating multiple aspects of locomotor behavior in both *Drosophila* larvae and adults [14]. This study performed an initial screen by ubiquitously expressing miR-SP constructs and looked for changes in locomotion behavior in both 2nd instar larvae and adults using automated quantitative movement tracking system. In a secondary screen, they focused on those miR-SP lines that showed adult-specific phenotypes when combined with the temperature-sensitive Gal80 system. This enabled identification of those miRNAs that play developmental roles versus those that act specifically during adulthood. They further compared results from ubiquitous and neural-specific expressions of miR-SP constructs and found that most of the miRNAs that play important roles in locomotion behaviors act in the nervous system.

Another previous screen that also exploited the advantages of temporal regulation mediated by miR-SP constructs identified 25 well-conserved miRNAs that regulate sleep and sleep homeostasis behavior in *Drosophila* [15]. This study further investigated whether these miRNAs are tissue specific, focusing on neurons, and compared miR-SP-mediated defects with genetic ablation of cognate miRNAs. Further studies on one of the strongest hits from their screen, *let-7*, revealed both developmental and adult-specific functions particularly in mushroom body neurons.

Another set of screens using the miR-SP library focused on tissue-specific knockdown of miRNA activity and identified 20 glial-specific miRNAs that are involved in regulating circadian rhythm behavior in *Drosophila* [16]. This study followed up on two

hits from the initial screen, *miR-263b* and *miR-274*, and found that both function specifically in adult astrocytes to regulate circadian locomotor activity rhythms.

4 miRNA Overexpression/Misexpression Strains

In addition to loss-of-function mutants and miR-SP constructs, three additional libraries have been generated for overexpression or misexpression of miRNAs based on UAS/GAL4 strategies. These libraries have greatly benefited the miRNA community for studying *in vivo* functions in a tissue-specific manner. These strains can be also used to perform rescue experiments by overexpressing miRNAs in mutant backgrounds, both ubiquitously and in specific tissues.

The first library consists of 149 distinct miRNA hairpins, either expressed singly or in clusters, using random P-element insertions or site-directed øC31 integrase-based insertions [17]. The second library includes 180 evolutionary conserved miRNAs inserted at specific sites, and the third library allows overexpression of 89 individual or clustered miRNAs, thereby representing a total of 109 miRNAs [18, 19]. These libraries differ in the host reporter that is co-expressed with the miRNA as well as in the position of integration within the genome, and have collectively expanded miRNA functional analysis studies in recent years.

4.1 Applications

A group of scientists used gain-of-function screening using 160 miRNAs and identified miRNAs implicated in controlling body fat in *Drosophila* [20]. Subsequent analysis of a particular hit, *miR-969*, showed that it regulates adiposity by repressing its target *gustatory receptor 47b* (*Gr47b*) particularly in adipose tissue.

Based on a two-tiered miRNA screen using miRNA overexpression libraries, another group identified the role of *Activating transcription factor 3*(*Atf-3*) in the *Drosophila* olfactory system [21]. In a pre-screen, this group looked for miRNAs that modified *Drosophila* Odorant Receptor expression. They followed up with a bioinformatic prediction tool analysis for the positive hits and performed an RNAi-based secondary screen. They hypothesized that loss of predicted targets should mimic miRNA overexpression phenotypes and identified a previously unknown role of *Atf-3* in regulating the expression of *Orb47b*.

From another tissue-specific screen using the miRNA overexpression library, a group of scientists showed that overexpression of *miR-1* can downregulate *reversed polarity* (*repo*), which specifically acts in glial cells to transcriptionally regulate the highly conserved GABA neurotransmitter activity [22]. Loss of *repo* triggered shortened lifespan, motor activity deficits, and sensibility to seizures due to changes in the Glutamate/GABA/Glutamine cycle in *Drosophila*.

These inducible transgenic collections may have even broader applications in studying miRNA functions in disease situations in addition to examining individual miRNA phenotypes. A recent genetic screen exploited this possibility by looking for miRNAs whose misexpression could modify the effect of pro-apoptotic genes in the *Drosophila* eye [23]. This study validated a subset of the positive hits by testing for their ability to block X-ray-induced apoptosis. They further modeled a cancer-relevant setting and studied how these anti-apoptotic miRNAs promote clone survival of cells mutant for a tumor suppressor gene, *lethal giant larvae (lgl)*.

5 Target Identification and Analysis

In order to fully understand the *in vivo* functions of miRNAs in *Drosophila* tissues, it is also important to identify their biologically relevant mRNA targets. The most widely used strategies in the past to predict mRNA targets have been limited to algorithms that rely on sequence complementarity, thermodynamic stability of pairing, and conservation, as summarized in a previous review [24].

However, these bioinformatic tools are not frequently updated with new or alternative 3′UTR sequences [25]. Moreover, the high rates of false discovery of these prediction algorithms necessitate direct methods to identify putative mRNA targets. This issue can be circumvented in flies by using immunopurification of mRNAs associated with Ago1 complexes, as previously done with cell lines and embryos [26, 27]. Extensive use of these approaches can significantly improve the accuracy of identifying relevant mRNA targets. Here we have summarized the recent advances in miRNA target prediction tools in the past 5 years.

Application of crosslinking and immunoprecipitation (CLIP) of AGO-mRNA complexes followed by sequencing techniques in mouse have greatly decreased the length of search space within 3′UTR sequences to only about 30–40 nucleotides per AGO footprint [28]. However, the information generated from sequencing can sometimes be misleading due to multiple miRNA seed site matches overlapping with the AGO-binding sites. This issue can be resolved by the introduction of diagnostic events (DE) during the reverse transcription step of library preparation to accumulate 5′ upstream sequences of seed matches. Modifications in conventional CLIP techniques such as PAR-CLIP (Photoactivatable Ribonucleoside-Enhanced Crosslinking and Immunoprecipitation) and HITS-CLIP (HIgh-Throughput Sequencing of RNAs Isolated by Crosslinking Immunoprecipitation) have significantly improved *in vivo* miRNA target prediction accuracy [29–31]. A group of scientists incorporated these techniques and compared the predictive potential of single nucleotide DEs to identify functional miRNA–mRNA interactions in *Drosophila* [32].

In addition, there has also been improvement in the bioinformatics-based target prediction algorithms. One such example is the upgrade in the Combinatorial miRNA targeting (ComiR) model that is based on restricting thermodynamic binding modeling to the 3′UTR sequences of target mRNAs [33]. However, studies have shown that miRNAs can also bind within the coding regions of mRNAs [34]. An improved version of ComiR now also includes information from the coding region and is more efficient than the previous version [35].

There remain limitations in miRNA target predictions due to lack of high-throughput data of mRNA repression in *Drosophila* compared to mammals. In addition to seed pairing, there are other parameters that influence site efficacy as shown in mammals [36]. These parameters differ in flies due to shorter 3′UTR lengths and higher AU contents. A group of scientists have used new RNA sequencing data to study the effect of miRNAs in *Drosophila* cells and expanded information about regulatory sites, included 5′ untranslated region of mRNAs, and updated evolutionary analysis [4]. Based on this information, they generated a quantitative model that improves the ranking of miRNA target prediction released in the newest version of TargetScanFly, v7 (http://www.targetscan.org).

Another valuable microRNA target prediction software that complements existing toolkits is seedVicious [37]. This software not only allows prediction of canonical sites similar to other algorithms but can also identify near-target sites that have one nucleotide difference to the seed sequence. Identification of near-target sites are important in the context of population genetics and evolution studies. Studies have shown that microRNA target sites are constantly evolving due to selective pressures [38–40]. This method has been implemented to study the evolution of post-transcriptional gene regulation in *Drosophila* [41] and further opens up the possibility of identifying newer microRNA targets.

6 Hybridization and Transgenic Methods to Detect Cell-Type Expression

As described in the previous sections, *Drosophila* researchers have access to powerful resources to systematically analyze the cellular and molecular phenotypes associated with the loss of individual miRNAs. Another key question is in which cells and at what developmental stages to look for these phenotypes. While past analyses of miRNA expression patterns have indicated that, like in other systems, many fly miRNAs display cell-type specificity [42–45], a systematic understanding of the cellular expression profiles of *Drosophila* miRNAs remains lacking.

Some of the molecular biology techniques routinely used in the past to detect mature and precursor miRNA expression levels in specific tissues or whole animals include Northern Blot Analysis, Quantitative Real-Time PCRs, and miRNA Microarrays have been described in a previous version of this chapter [46]. In the section below, we have focused our attention on cell-specific detection of miRNAs.

6.1 Transcriptional Reporters

Transcriptional reporters are generated by making transgenic constructs of the promoter/enhancer region of miRNAs fused to reporters such as GFP or LacZ. These constructs are then used to report the cells which transcribe miRNAs *in vivo*. Some examples of detecting cell-specific miRNAs using transcriptional reporters include *miR-1* [47–49], *miR-124* [50], *miR-278* [51], *let-7-C* [52], *bantam* [53], *miR-279/996* [54], and *miR-958* [55].

6.2 Sensors

Another transgenic approach to successfully determine miRNA activity in specific cell types *in vivo* has been the use of miRNA sensors. These constructs are generated by fusing GFP or luciferase reporters to one or more perfectly complimentary miRNA target sequences in the 3′UTR. Downregulation of reporter activity via RNA interference-mediated destruction by miRNAs in specific cell types indicates high miRNA activity. These sensors have been used as a readout for several cell-specific miRNAs [6, 54, 56–60].

6.3 In Situ Hybridization

The most direct method to visualize miRNA expression at cellular resolution in whole animals or tissues is *in situ hybridization* (ISH) using tagged anti-sense RNA probes. The most routinely used probes in ISH are alkaline phosphatase or digoxigenin (DIG) tagged and detected using fluorescent tyramide signal amplification [61]. This technique was initially used to detect primary miRNA transcripts in *Drosophila* embryos by designing probes against regions near the miRNA hairpin precursors [42, 62]. However, primary miRNA transcripts undergo a series of processing events before they generate mature miRNAs and thus may not accurately reflect mature miRNA expression patterns *in vivo* [63]. More recent advances in generation of ISH probes have been the use of locked nucleic acid (LNA) probes. LNA probes are designed by incorporating a modified RNA substitution in which the 2′ oxygen and 4′ carbon positions are "locked" to increase hybridization stability. This method has been successful in detecting mature miRNA levels that had lower hybridization affinity due to shorter lengths. Some examples of miRNA expression detected using LNA probes are *miR-1* [49] *miR-34* [64], *let-7* [65, 66], *miR-9a* [67], *miR-980* [68], and *miR-285* [69]. A recently devised modification step of LNA probe detection can further improve signal to noise ratio, particularly for low copy number miRNAs. A group of scientists described a protocol for detection of mature miRNAs during

C. elegans oogenesis by using a DIG-based chemical development with 5-bromo-4-chloro-3-indolyl phosphate/nitro blue tetrazolium (BCIP/NBT) that significantly improved signal [70].

A more thorough understanding of miRNA and their effects on target mRNAs can only be achieved by precise quantification of their expression at single cell level rather than tissue or organism level. The recent efforts in combining single molecule fluorescence in situ hybridization (smFISH) and immunofluorescence to simultaneously detect mRNAs and proteins at single cell level can also be applied to miRNA functional studies [71, 72]. A very recent study was successfully able to detect primary transcript of the highly conserved *miR-9* and its target mRNA *rhomboid (rho)* at single-cell resolution in *Drosophila* embryos [73]. This technique can be more extensively applied to detection of miRNA and their relevant targets in future.

7 Cloning Methods to Detect Cell-Type Expression

The cloning and sequencing of miRNAs offers another approach for the analysis of miRNA expression. While discovery efforts for cataloging the complete repertoire of miRNAs focused on such analysis of whole animals or tissues, more recent efforts in flies have made advances in the cloning of miRNAs from single cell types and, in other systems, cells [74]. Such efforts offer the opportunity to generate high-resolution cell atlases that would be useful for inferring not only in which cells particular miRNAs function but also for determining how those cells or cell types are disrupted by genetic mutations. In this section, we summarize methods that have been used to characterize the cell-type-specific expression profiles of fly miRNAs.

In vitro cultured cell lines offer easy access to relatively homogenous cell types that can be used for miRNA cloning. Small RNA libraries have been generated by standard methods that employed conventional size-selection chromatography from a variety of cell lines, including predominantly embryonic hematopoietic cell lines as well as some additional lines derived from the larval wing disc, the larval central nervous system, and the adult ovary [75, 76]. The latter have been particularly useful for small RNA analysis because they harbor a population of piRNAs, enabling molecular and biochemical studies of the silencing capacity of this class of small RNAs [77, 78]. Despite the benefits of easy access and amenability for biochemical and molecular analysis, cell lines also have some limitations including that the *in vivo* correlates of many cell lines are not clear, as well as the likelihood that the process of immortalization and maintenance of cells *in vitro* alter their miRNA profiles.

MiRNA profiles that may be more relevant to endogenous expression patterns have been obtained by cloning miRNAs from cell populations sorted from tissues. For example, antibodies to *Drosophila* Prospero, Repo, and Embryonic Lethal Abnormal Vision (Elav), which are known to specifically label neuroblasts, glia, and neurons, respectively, were used to sort cell populations from two stages of *Drosophila* embryos using Fluorescent Activated Cell Sorting (FACS), and small RNAs were cloned with the NEBNext Small RNA Protocol from New England Biolabs [79]. In addition to antibodies, transgenic GAL4 driver strains have been used to drive fluorescent proteins that can be used as the basis for sorting cells from which miRNAs are cloned. Using this approach, Abruzzi et al. described methods to manually sort neurons labeled with a variety of GAL4 drivers, including *elav-Gal4* as well as the much more sparsely expressed *pigment-dispersing factor (pdf)-GAL4*, from dissected adult brains and then clone miRNAs using a modified size-selection and library generation protocol [80]. Limitations of this approach include its dependence on well-characterized antibodies or GAL4 drivers for cell selection, the possibility of contamination associated with low amounts of starting material and/or imprecise sorting methods, and the likelihood that the process of tissue disruption and cell isolation alters the miRNA profiles of sorted cells.

Some of these limitations have been addressed by a recently described method that selectively marks miRNAs in a cell-type-specific manner and uses that mark to clone them from whole tissues [81]. This approach relies on the observation that most animal miRNAs, including fly miRNAs, are not methylated at their 3′ ends, but can be methylated by the *Arabidopsis* version of the HEN1 methyltransferase. These methylated miRNAs can then be selectively cloned using a modified library generation protocol with an added chemoselection step. Using this approach, Alberti et al. cloned miRNAs from flies that expressed Arabidopsis HEN1 under UAS control in either all muscles or flight muscles and isolated 11 and 13 miRNAs, respectively, including the known muscle-specific *miR-140*. This approach is well suited to current efforts directed at the characterization of GAL4 drivers that display specific cell-type expression profiles in various tissues, including the brain and intestine [82, 83]. A complementary approach that also relies on the specificity of GAL4 drivers to isolate and clone miRNAs from intact tissue involves the cell-type expression of epitope-tagged versions of the RISC complex including, for example, Ago1. While this cell-type-specific variation has not yet been reported in flies, Ago1 has been used to isolate miRNAs from whole tissues [32], and recent protocols have been reported for the isolation and cloning of RNA cargos of epitope-tagged RNA-binding proteins [84]. A limitation of this approach is that ectopically expressed and tagged Ago1 may not display the same miRNA-binding properties as endogenous Ago1.

Given the impact that single-cell RNA sequencing (scRNA-seq) has had on the characterization of single cell identities and mRNA profiles in whole tissues, future effort will undoubtedly be focused on using this approach to profile miRNAs. Currently, the main challenge is the efficiency and accuracy of cloning miRNAs from such low input sources, since major artifacts arise from ligation bias and failure during cloning [74]. While this approach has not yet been applied to *Drosophila* cells or tissues, it holds great promise for generating detailed maps of miRNA expression at cellular resolution.

8 Concluding Summary

A variety of useful resources and methods are currently available for the analysis of miRNA function in *Drosophila*. These include collections of large numbers of fly strains that allow systematic analysis of the consequences of miRNA overexpression and loss in defined cell types across the fly lifespan. These resources will be indispensable for identifying and understanding the complete repertoire of functions of these critical regulatory factors.

References

1. Lagos-Quintana M, Rauhut R, Lendeckel W, Tuschl T (2001) Identification of novel genes coding for small expressed RNAs. Science 294(5543):853–858. https://doi.org/10.1126/science.1064921
2. Lee RC, Ambros V (2001) An extensive class of small RNAs in Caenorhabditis elegans. Science 294(5543):862–864. https://doi.org/10.1126/science.1065329
3. Lau NC, Lim LP, Weinstein EG, Bartel DP (2001) An abundant class of tiny RNAs with probable regulatory roles in Caenorhabditis elegans. Science 294(5543):858–862. https://doi.org/10.1126/science.1065062
4. Agarwal V, Subtelny AO, Thiru P, Ulitsky I, Bartel DP (2018) Predicting microRNA targeting efficacy in Drosophila. Genome Biol 19(1):152. https://doi.org/10.1186/s13059-018-1504-3
5. Kozomara A, Birgaoanu M, Griffiths-Jones S (2019) miRBase: from microRNA sequences to function. Nucleic Acids Res 47(D1):D155–D162. https://doi.org/10.1093/nar/gky1141
6. Luhur A, Chawla G, Wu YC, Li J, Sokol NS (2014) Drosha-independent DGCR8/Pasha pathway regulates neuronal morphogenesis. Proc Natl Acad Sci 111(4):1421–1426. https://doi.org/10.1073/pnas.1318445111
7. Yang J-S, Eric (2011) Alternative miRNA biogenesis pathways and the interpretation of core miRNA pathway mutants. Mol Cell 43(6):892–903. https://doi.org/10.1016/j.molcel.2011.07.024
8. Chen YW, Song S, Weng R, Verma P, Kugler JM, Buescher M, Rouam S, Cohen SM (2014) Systematic study of Drosophila microRNA functions using a collection of targeted knockout mutations. Dev Cell 31(6):784–800. https://doi.org/10.1016/j.devcel.2014.11.029
9. Picao-Osorio J, Lago-Baldaia I, Patraquim P, Alonso CR (2017) Pervasive behavioral effects of MicroRNA regulation in Drosophila. Genetics 206(3):1535–1548. https://doi.org/10.1534/genetics.116.195776
10. Atilano ML, Glittenberg M, Monteiro A, Copley RR, Ligoxygakis P (2017) MicroRNAs that contribute to coordinating the immune response in Drosophila melanogaster. Genetics 207(1):163–178. https://doi.org/10.1534/genetics.116.196584
11. Iftikhar H, Johnson NL, Marlatt ML, Carney GE (2019) The role of miRNAs in Drosophila

melanogaster male courtship behavior. Genetics 211(3):925–942. https://doi.org/10.1534/genetics.118.301901

12. Fulga TA, McNeill EM, Binari R, Yelick J, Blanche A, Booker M, Steinkraus BR, Schnall-Levin M, Zhao Y, Deluca T, Bejarano F, Han Z, Lai EC, Wall DP, Perrimon N, Van Vactor D (2015) A transgenic resource for conditional competitive inhibition of conserved Drosophila microRNAs. Nat Commun 6(1): 7279. https://doi.org/10.1038/ncomms8279
13. McGuire SE, Le PT, Osborn AJ, Matsumoto K, Davis RL (2003) Spatiotemporal rescue of memory dysfunction in Drosophila. Science 302(5651):1765–1768. https://doi.org/10.1126/science.1089035
14. Donelson NC, Dixit R, Pichardo-Casas I, Chiu EY, Ohman RT, Slawson JB, Klein M, Fulga TA, Van Vactor D, Griffith LC (2020) MicroRNAs regulate multiple aspects of locomotor behavior in Drosophila. G3 (Bethesda) 10(1): 43–55. https://doi.org/10.1534/g3.119.400793
15. Goodwin PR, Meng A, Moore J, Hobin M, Fulga TA, Van Vactor D, Griffith LC (2018) MicroRNAs regulate sleep and sleep homeostasis in Drosophila. Cell Rep 23(13):3776–3786. https://doi.org/10.1016/j.celrep.2018.05.078
16. You S, Fulga TA, Van Vactor D, Jackson FR (2018) Regulation of circadian behavior by astroglial MicroRNAs in Drosophila. Genetics 208(3):1195–1207. https://doi.org/10.1534/genetics.117.300342
17. Bejarano F, Bortolamiol-Becet D, Dai Q, Sun K, Saj A, Chou YT, Raleigh DR, Kim K, Ni JQ, Duan H, Yang JS, Fulga TA, Van Vactor D, Perrimon N, Lai EC (2012) A genome-wide transgenic resource for conditional expression of Drosophila microRNAs. Development 139(15):2821–2831. https://doi.org/10.1242/dev.079939
18. Schertel C, Rutishauser T, Forstemann K, Basler K (2012) Functional characterization of Drosophila microRNAs by a novel in vivo library. Genetics 192(4):1543–1552. https://doi.org/10.1534/genetics.112.145383
19. Szuplewski S, Kugler JM, Lim SF, Verma P, Chen YW, Cohen SM (2012) MicroRNA transgene overexpression complements deficiency-based modifier screens in Drosophila. Genetics 190(2):617–626. https://doi.org/10.1534/genetics.111.136689
20. Redmond W, Allen D, Elledge MC, Arellanes R, Redmond L, Yeahquo J, Zhang S, Youngblood M, Reiner A, Seo J (2019) Screening of microRNAs controlling body fat in Drosophila melanogaster and identification of miR-969 and its target, Gr47b. PLoS One 14(7):e0219707. https://doi.org/10.1371/journal.pone.0219707
21. Bhat S, Jones WD (2016) An accelerated miRNA-based screen implicates Atf-3 in Drosophila odorant receptor expression. Sci Rep 6: 20109. https://doi.org/10.1038/srep20109
22. Mazaud D, Kottler B, Goncalves-Pimentel C, Proelss S, Tuchler N, Deneubourg C, Yuasa Y, Diebold C, Jungbluth H, Lai EC, Hirth F, Giangrande A, Fanto M (2019) Transcriptional regulation of the glutamate/GABA/glutamine cycle in adult glia controls motor activity and seizures in Drosophila. J Neurosci 39(27): 5269–5283. https://doi.org/10.1523/JNEUROSCI.1833-18.2019
23. Bejarano F, Chang CH, Sun K, Hagen JW, Deng WM, Lai EC (2021) A comprehensive in vivo screen for anti-apoptotic miRNAs indicates broad capacities for oncogenic synergy. Dev Biol 475:10–20. https://doi.org/10.1016/j.ydbio.2021.02.010
24. Bartel DP (2009) MicroRNAs: target recognition and regulatory functions. Cell 136(2): 215–233. https://doi.org/10.1016/j.cell.2009.01.002
25. Smibert P, Miura P, Westholm JO, Shenker S, May G, Duff MO, Zhang D, Eads BD, Carlson J, Brown JB, Eisman RC, Andrews J, Kaufman T, Cherbas P, Celniker SE, Graveley BR, Lai EC (2012) Global patterns of tissue-specific alternative polyadenylation in Drosophila. Cell Rep 1(3):277–289. https://doi.org/10.1016/j.celrep.2012.01.001
26. Easow G, Teleman AA, Cohen SM (2007) Isolation of microRNA targets by miRNP immunopurification. RNA 13(8):1198–1204. https://doi.org/10.1261/rna.563707
27. Kadener S, Menet JS, Sugino K, Horwich MD, Weissbein U, Nawathean P, Vagin VV, Zamore PD, Nelson SB, Rosbash M (2009) A role for microRNAs in the Drosophila circadian clock. Genes Dev 23(18):2179–2191. https://doi.org/10.1101/gad.1819509
28. Chi SW, Zang JB, Mele A, Darnell RB (2009) Argonaute HITS-CLIP decodes microRNA-mRNA interaction maps. Nature 460(7254): 479–486. https://doi.org/10.1038/nature08170
29. Hafner M, Landthaler M, Burger L, Khorshid M, Hausser J, Berninger P, Rothballer A, Ascano M Jr, Jungkamp AC, Munschauer M, Ulrich A, Wardle GS, Dewell S, Zavolan M, Tuschl T (2010) Transcriptome-wide identification of

RNA-binding protein and microRNA target sites by PAR-CLIP. Cell 141(1):129–141. https://doi.org/10.1016/j.cell.2010.03.009

30. Zhang C, Darnell RB (2011) Mapping in vivo protein-RNA interactions at single-nucleotide resolution from HITS-CLIP data. Nat Biotechnol 29(7):607–614. https://doi.org/10.1038/nbt.1873
31. Moore MJ, Zhang C, Gantman EC, Mele A, Darnell JC, Darnell RB (2014) Mapping Argonaute and conventional RNA-binding protein interactions with RNA at single-nucleotide resolution using HITS-CLIP and CIMS analysis. Nat Protoc 9(2):263–293. https://doi.org/10.1038/nprot.2014.012
32. Wessels HH, Lebedeva S, Hirsekorn A, Wurmus R, Akalin A, Mukherjee N, Ohler U (2019) Global identification of functional microRNA-mRNA interactions in Drosophila. Nat Commun 10(1):1626. https://doi.org/10.1038/s41467-019-09586-z
33. Coronnello C, Benos PV (2013) ComiR: combinatorial microRNA target prediction tool. Nucleic Acids Res 41(Web Server issue): W159–W164. https://doi.org/10.1093/nar/gkt379
34. Brummer A, Hausser J (2014) MicroRNA binding sites in the coding region of mRNAs: extending the repertoire of post-transcriptional gene regulation. Bioessays 36(6):617–626. https://doi.org/10.1002/bies.201300104
35. Bertolazzi G, Benos PV, Tumminello M, Coronnello C (2020) An improvement of ComiR algorithm for microRNA target prediction by exploiting coding region sequences of mRNAs. BMC Bioinformatics 21(Suppl 8):201. https://doi.org/10.1186/s12859-020-3519-5
36. Grimson A, Farh KK, Johnston WK, Garrett-Engele P, Lim LP, Bartel DP (2007) MicroRNA targeting specificity in mammals: determinants beyond seed pairing. Mol Cell 27(1): 91–105. https://doi.org/10.1016/j.molcel.2007.06.017
37. Marco A (2018) SeedVicious: analysis of microRNA target and near-target sites. PLoS One 13(4):e0195532. https://doi.org/10.1371/journal.pone.0195532
38. Chen K, Rajewsky N (2006) Natural selection on human microRNA binding sites inferred from SNP data. Nat Genet 38(12): 1452–1456. https://doi.org/10.1038/ng1910
39. Saunders MA, Liang H, Li WH (2007) Human polymorphism at microRNAs and microRNA target sites. Proc Natl Acad Sci U S A 104(9): 3300–3305. https://doi.org/10.1073/pnas.0611347104
40. Marco A (2015) Selection against maternal microRNA target sites in maternal transcripts. G3 (Bethesda) 5(10):2199–2207. https://doi.org/10.1534/g3.115.019497
41. Clifton BD, Librado P, Yeh SD, Solares ES, Real DA, Jayasekera SU, Zhang W, Shi M, Park RV, Magie RD, Ma HC, Xia XQ, Marco A, Rozas J, Ranz JM (2017) Rapid functional and sequence differentiation of a tandemly repeated species-specific multigene family in Drosophila. Mol Biol Evol 34(1): 51–65. https://doi.org/10.1093/molbev/msw212
42. Aboobaker AA, Tomancak P, Patel N, Rubin GM, Lai EC (2005) Drosophila microRNAs exhibit diverse spatial expression patterns during embryonic development. Proc Natl Acad Sci U S A 102(50):18017–18022. https://doi.org/10.1073/pnas.0508823102
43. Landgraf P, Rusu M, Sheridan R, Sewer A, Iovino N, Aravin A, Pfeffer S, Rice A, Kamphorst AO, Landthaler M, Lin C, Socci ND, Hermida L, Fulci V, Chiaretti S, Foa R, Schliwka J, Fuchs U, Novosel A, Muller RU, Schermer B, Bissels U, Inman J, Phan Q, Chien M, Weir DB, Choksi R, De Vita G, Frezzetti D, Trompeter HI, Hornung V, Teng G, Hartmann G, Palkovits M, Di Lauro R, Wernet P, Macino G, Rogler CE, Nagle JW, Ju J, Papavasiliou FN, Benzing T, Lichter P, Tam W, Brownstein MJ, Bosio A, Borkhardt A, Russo JJ, Sander C, Zavolan M, Tuschl T (2007) A mammalian microRNA expression atlas based on small RNA library sequencing. Cell 129(7):1401–1414. https://doi.org/10.1016/j.cell.2007.04.040
44. Martinez NJ, Ow MC, Reece-Hoyes JS, Barrasa MI, Ambros VR, Walhout AJ (2008) Genome-scale spatiotemporal analysis of Caenorhabditis elegans microRNA promoter activity. Genome Res 18(12):2005–2015. https://doi.org/10.1101/gr.083055.108
45. Wienholds E, Kloosterman WP, Miska E, Alvarez-Saavedra E, Berezikov E, de Bruijn E, Horvitz HR, Kauppinen S, Plasterk RH (2005) MicroRNA expression in zebrafish embryonic development. Science 309(5732):310–311. https://doi.org/10.1126/science.1114519
46. Chawla G, Luhur A, Sokol N (2016) Analysis of MicroRNA function in Drosophila. Methods Mol Biol 1478:79–94. https://doi.org/10.1007/978-1-4939-6371-3_4
47. Biemar F, Zinzen R, Ronshaugen M, Sementchenko V, Manak JR, Levine MS (2005) Spatial regulation of microRNA gene expression in the Drosophila embryo. Proc Natl Acad Sci U S A 102(44):15907–15911. https://doi.org/10.1073/pnas.0507817102

48. Kwon C, Han Z, Olson EN, Srivastava D (2005) MicroRNA1 influences cardiac differentiation in Drosophila and regulates Notch signaling. Proc Natl Acad Sci U S A 102(52): 18986–18991. https://doi.org/10.1073/pnas.0509535102
49. Sokol NS, Ambros V (2005) Mesodermally expressed Drosophila microRNA-1 is regulated by Twist and is required in muscles during larval growth. Genes Dev 19(19):2343–2354. https://doi.org/10.1101/gad.1356105
50. Weng R, Cohen SM (2012) Drosophila miR-124 regulates neuroblast proliferation through its target anachronism. Development 139(8):1427–1434. https://doi.org/10.1242/dev.075143
51. Teleman AA, Maitra S, Cohen SM (2006) Drosophila lacking microRNA miR-278 are defective in energy homeostasis. Genes Dev 20(4): 417–422. https://doi.org/10.1101/gad.374406
52. Chawla G, Sokol NS (2012) Hormonal activation of let-7-C microRNAs via EcR is required for adult Drosophila melanogaster morphology and function. Development 139(10): 1788–1797. https://doi.org/10.1242/dev.077743
53. Wu YC, Lee KS, Song Y, Gehrke S, Lu B (2017) The bantam microRNA acts through Numb to exert cell growth control and feedback regulation of Notch in tumor-forming stem cells in the Drosophila brain. PLoS Genet 13(5):e1006785. https://doi.org/10.1371/journal.pgen.1006785
54. Duan H, de Navas LF, Hu F, Sun K, Mavromatakis YE, Viets K, Zhou C, Kavaler J, Johnston RJ, Tomlinson A, Lai EC (2018) The mir-279/996 cluster represses receptor tyrosine kinase signaling to determine cell fates in the Drosophila eye. Development 145(7). https://doi.org/10.1242/dev.159053
55. Mukherjee S, Paricio N, Sokol NS (2021) A stress-responsive miRNA regulates BMP signaling to maintain tissue homeostasis. Proc Natl Acad Sci U S A 118(21). https://doi.org/10.1073/pnas.2022583118
56. Brennecke J, Hipfner DR, Stark A, Russell RB, Cohen SM (2003) bantam encodes a developmentally regulated microRNA that controls cell proliferation and regulates the proapoptotic gene hid in Drosophila. Cell 113(1):25–36. https://doi.org/10.1016/s0092-8674(03)00231-9
57. Lai EC, Tam B, Rubin GM (2005) Pervasive regulation of Drosophila Notch target genes by GY-box-, Brd-box-, and K-box-class microRNAs. Genes Dev 19(9):1067–1080. https://doi.org/10.1101/gad.1291905
58. Forstemann K, Tomari Y, Du T, Vagin VV, Denli AM, Bratu DP, Klattenhoff C, Theurkauf WE, Zamore PD (2005) Normal microRNA maturation and germ-line stem cell maintenance requires Loquacious, a double-stranded RNA-binding domain protein. PLoS Biol 3(7): e236. https://doi.org/10.1371/journal.pbio.0030236
59. Kitatani Y, Tezuka A, Hasegawa E, Yanagi S, Togashi K, Tsuji M, Kondo S, Parrish JZ, Emoto K (2020) Drosophila miR-87 promotes dendrite regeneration by targeting the transcriptional repressor Tramtrack69. PLoS Genet 16(8):e1008942. https://doi.org/10.1371/journal.pgen.1008942
60. Epstein Y, Perry N, Volin M, Zohar-Fux M, Braun R, Porat-Kuperstein L, Toledano H (2017) miR-9a modulates maintenance and ageing of Drosophila germline stem cells by limiting N-cadherin expression. Nat Commun 8(1):600. https://doi.org/10.1038/s41467-017-00485-9
61. Toledano H, D'Alterio C, Loza-Coll M, Jones DL (2012) Dual fluorescence detection of protein and RNA in Drosophila tissues. Nat Protoc 7(10):1808–1817. https://doi.org/10.1038/nprot.2012.105
62. Kosman D, Mizutani CM, Lemons D, Cox WG, McGinnis W, Bier E (2004) Multiplex detection of RNA expression in Drosophila embryos. Science 305(5685):846. https://doi.org/10.1126/science.1099247
63. Thomson JM, Newman M, Parker JS, Morin-Kensicki EM, Wright T, Hammond SM (2006) Extensive post-transcriptional regulation of microRNAs and its implications for cancer. Genes Dev 20(16):2202–2207. https://doi.org/10.1101/gad.1444406
64. Soni K, Choudhary A, Patowary A, Singh AR, Bhatia S, Sivasubbu S, Chandrasekaran S, Pillai B (2013) miR-34 is maternally inherited in Drosophila melanogaster and Danio rerio. Nucleic Acids Res 41(8): 4470–4480. https://doi.org/10.1093/nar/gkt139
65. Kucherenko MM, Barth J, Fiala A, Shcherbata HR (2012) Steroid-induced microRNA let-7 acts as a spatio-temporal code for neuronal cell fate in the developing Drosophila brain. EMBO J 31(24):4511–4523. https://doi.org/10.1038/emboj.2012.298
66. Toledano H, D'Alterio C, Czech B, Levine E, Jones DL (2012) The let-7-Imp axis regulates ageing of the Drosophila testis stem-cell niche. Nature 485(7400):605–610. https://doi.org/10.1038/nature11061
67. Katti P, Thimmaya D, Madan A, Nongthomba U (2017) Overexpression of miRNA-9 generates muscle hypercontraction through

translational repression of Troponin-T in Drosophila melanogaster indirect flight muscles. G3 (Bethesda) 7(10):3521–3531. https://doi.org/10.1534/g3.117.300232

68. Kucherenko MM, Shcherbata HR (2018) Stress-dependent miR-980 regulation of Rbfox1/A2bp1 promotes ribonucleoprotein granule formation and cell survival. Nat Commun 9(1):312. https://doi.org/10.1038/s41467-017-02757-w
69. Li D, Liu Y, Pei C, Zhang P, Pan L, Xiao J, Meng S, Yuan Z, Bi X (2017) miR-285-Yki/Mask double-negative feedback loop mediates blood-brain barrier integrity in Drosophila. Proc Natl Acad Sci U S A 114(12): E2365–E2374. https://doi.org/10.1073/pnas.1613233114
70. Minogue AL, Arur S (2019) In situ hybridization for detecting mature MicroRNAs in vivo at single-cell resolution. Curr Protoc Mol Biol 127(1):e93. https://doi.org/10.1002/cpmb.93
71. Trcek T, Lionnet T, Shroff H, Lehmann R (2017) mRNA quantification using single-molecule FISH in Drosophila embryos. Nat Protoc 12(7):1326–1348. https://doi.org/10.1038/nprot.2017.030
72. Tsanov N, Samacoits A, Chouaib R, Traboulsi AM, Gostan T, Weber C, Zimmer C, Zibara K, Walter T, Peter M, Bertrand E, Mueller F (2016) smiFISH and FISH-quant - a flexible single RNA detection approach with super-resolution capability. Nucleic Acids Res 44(22):e165. https://doi.org/10.1093/nar/gkw784
73. Gallicchio L, Griffiths-Jones S, Ronshaugen M (2021) miR-9a regulates levels of both rhomboid mRNA and protein in the early Drosophila melanogaster embryo. Cold Spring Harbor Laboratory. https://doi.org/10.1101/2021.07.12.452096
74. Hucker SM, Fehlmann T, Werno C, Weidele K, Luke F, Schlenska-Lange A, Klein CA, Keller A, Kirsch S (2021) Single-cell microRNA sequencing method comparison and application to cell lines and circulating lung tumor cells. Nat Commun 12(1):4316. https://doi.org/10.1038/s41467-021-24611-w
75. Lau NC, Robine N, Martin R, Chung WJ, Niki Y, Berezikov E, Lai EC (2009) Abundant primary piRNAs, endo-siRNAs, and microRNAs in a Drosophila ovary cell line. Genome Res 19(10):1776–1785. https://doi.org/10.1101/gr.094896.109
76. Wen J, Mohammed J, Bortolamiol-Becet D, Tsai H, Robine N, Westholm JO, Ladewig E, Dai Q, Okamura K, Flynt AS, Zhang D, Andrews J, Cherbas L, Kaufman TC, Cherbas P, Siepel A, Lai EC (2014) Diversity of miRNAs, siRNAs, and piRNAs across 25 Drosophila cell lines. Genome Res 24(7): 1236–1250. https://doi.org/10.1101/gr.161554.113
77. Clark JP, Rahman R, Yang N, Yang LH, Lau NC (2017) Drosophila PAF1 modulates PIWI/piRNA silencing capacity. Curr Biol 27(17):2718–2726 e2714. https://doi.org/10.1016/j.cub.2017.07.052
78. Post C, Clark JP, Sytnikova YA, Chirn GW, Lau NC (2014) The capacity of target silencing by Drosophila PIWI and piRNAs. RNA 20(12): 1977–1986. https://doi.org/10.1261/rna.046300.114
79. Menzel P, McCorkindale AL, Stefanov SR, Zinzen RP, Meyer IM (2019) Transcriptional dynamics of microRNAs and their targets during Drosophila neurogenesis. RNA Biol 16(1): 69–81. https://doi.org/10.1080/15476286.2018.1558907
80. Abruzzi K, Chen X, Nagoshi E, Zadina A, Rosbash M (2015) RNA-seq profiling of small numbers of Drosophila neurons. Methods Enzymol 551:369–386. https://doi.org/10.1016/bs.mie.2014.10.025
81. Alberti C, Manzenreither RA, Sowemimo I, Burkard TR, Wang J, Mahofsky K, Ameres SL, Cochella L (2018) Cell-type specific sequencing of microRNAs from complex animal tissues. Nat Methods 15(4):283–289. https://doi.org/10.1038/nmeth.4610
82. Luan H, Diao F, Scott RL, White BH (2020) The *Drosophila* split Gal4 system for neural circuit mapping. Front Neural Circuits 14: 603397. https://doi.org/10.3389/fncir.2020.603397
83. Ariyapala IS, Holsopple JM, Popodi EM, Hartwick DG, Kahsai L, Cook KR, Sokol NS (2020) Identification of split-GAL4 drivers and enhancers that allow regional cell type manipulations of the Drosophila melanogaster intestine. Genetics 216(4):891–903. https://doi.org/10.1534/genetics.120.303625
84. Buddika K, Xu J, Ariyapala IS, Sokol NS (2021) I-KCKT allows dissection-free RNA profiling of adult Drosophila intestinal progenitor cells. Development 148(1). https://doi.org/10.1242/dev.196568

Chapter 4

Analysis of Single-Cell Transcriptome Data in *Drosophila*

Schayan Yousefian, Maria Jelena Musillo, and Josephine Bageritz

Abstract

The fly *Drosophila* is a versatile model organism that has led to fascinating biological discoveries. In the past few years, *Drosophila* researchers have used single-cell RNA-sequencing (scRNA-seq) to gain insights into the cellular composition, and developmental processes of various tissues and organs. Given the success of single-cell technologies a variety of computational tools and software packages were developed to enable and facilitate the analysis of scRNA-seq data. In this book chapter we want to give guidance on analyzing droplet-based scRNA-seq data from *Drosophila*. We will initially describe the preprocessing commonly done for *Drosophila*, point out possible downstream analyses, and finally highlight computational methods developed using *Drosophila* scRNA-seq data.

Key words *Drosophila*, Single-cell RNA-seq, Data analysis, Seurat, Computational methods, Droplet based, Software packages

1 Introduction

Since the first report of a single-cell RNA sequencing (scRNA-seq) technique in 2009 [1], the progress of scRNA-seq methods [2–6] has revolutionized our understanding of molecular biology and has provided deep insights into developmental processes, cell heterogeneity, and function in a variety of organisms. There are currently more than 20 commonly used scRNA-seq techniques that differ in various aspects, such as cell isolation, lysis, and reverse transcription (reviewed in [7]). Moreover, they can be distinguished based on transcript coverage (full-length, 3′-end sequencing, or 5′-end sequencing) and UMI (unique molecular identifier) usage, which allows the computational identification and removal of PCR duplicates. Nanoliter droplet-based sequencing approaches, such as Drop-seq [4] or InDrop [8], utilize a microfluidic device to separate individual cells and encapsulate them into droplets together with a barcoded mRNA-capturing bead that are sequenced in pools afterward. Although these methods only cover part of the transcript

Christian Dahmann (ed.), *Drosophila: Methods and Protocols*,
Methods in Molecular Biology, vol. 2540, https://doi.org/10.1007/978-1-0716-2541-5_4,

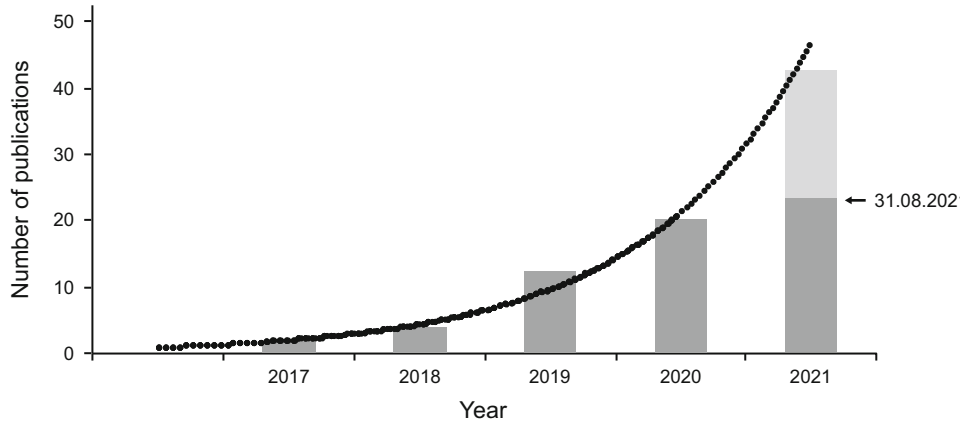


Fig. 1 Number of *Drosophila* publications (peer-reviewed and preprints) using and developing experimental and computational single-cell RNA-seq methods (status 31.08.2021). The estimated number of *Drosophila* scRNA-seq publications expected in 2021 is represented by the light gray column

(3′ end or 5′ end), their main advantages are the high throughput, comparatively low sequencing costs per cell and the usage of UMIs.

Due to these characteristics, droplet-based scRNA-seq methods have become popular and accessible to a wide range of scientists. There are also an increasing number of *Drosophila* studies in which scRNA-seq techniques have been applied (Fig. 1) to investigate, e.g., the *Drosophila* embryo, the larval blood, the larval wing discs, the larval eye discs, the adult brain, the adult ovaries and testes, the adult gut, the adult ventral nerve cord, and adult visual system (reviewed in [9, 10]). In fact, most of the studies use droplet-based scRNA-seq techniques, such as Drop-seq [4] and 10× Genomics Chromium [11], which is a commercially available single-cell sequencing platform.

Given the enormous potential of single-cell approaches, bioinformaticians and computational biologists are motivated to develop a broad range of data analysis tools. Nevertheless, it is still a relatively new field and the complexity of scRNA-seq data can pose major difficulties. Particularly challenging can be the high levels of noise [12, 13] which derive from the technical variation of single-cell experiments (e.g., low mRNA capture efficiency or cDNA amplification bias). Another major hurdles in single-cell data science are dropout events which typically occur due to the low amount of starting material (reviewed in [14, 15]).

As of end of August 2021 the computational single-cell toolbox contains 1033 different analysis methods in more than 30 categories (https://www.scrna-tools.org/; [16]), many of which have been used and developed in *Drosophila* studies (*see* Table 1). However, there is no standardized analysis pipeline, due to several reasons. One major challenge for a standardized workflow is the biological data itself, meaning that different cell types and tissues

Table 1
List of software and computational tools described in this book chapter

Software/computational tool	Description	Reference
10× Genomics Cell Ranger	Set of analysis pipelines that process Chromium single-cell raw sequencing files and generate feature-barcode count matrices	[11]
bcftools mpileup	Program that is part of the samtools repository. Can be used to derive base-substitution mutations (see [37])	[88]
Cytoscape	Open-source software to visualize experimental data, perform network analyses, and gene–gene correlations	[80]
DAVID (The Database for Annotation, Visualization and Integrated Discovery)	Web-accessible program to conduct Gene Ontology term analysis	[74, 75]
DistMap[a]	Algorithm to predict spatial gene expression patterns by using an existing reference database of in situ	[45]
Drop-seq computational cookbook	Analysis pipeline to process Drop-seq single-cell raw sequencing files and generate feature-barcode count matrices	[4]
DVEX (Drosophila Virtual Expression eXplorer)[a]	Online resource tool to explore virtual transcriptome maps of stage 6 Drosophila embryo.	[45]
GLAD (Gene list annotations for Drosophila)	An online resource containing lists of functionally related *Drosophila* genes	[73]
GOrilla (Gene Ontology enRIchment anaLysis and visuaLizAtion tool)	Web-based tool to determine and visualize enriched GO terms in ranked gene lists (e.g., spatially informative rank [97] or expression level or differential expression).	[78]
g:Profiler	Web server that can be used to perform gene set enrichment analyses on flat and ranked gene lists.	[76]
Harmony	Method to integrate scRNA-seq data by projecting cells into a shared embedding and in this way account for batch effects.	[61]
Monocle	Performs pseudotime analysis to infer cell trajectories and lineages	[81, 82]
Neural network classifier	To classify single cells in optic lobe and central brain scRNA-seq data	[32]
PANTHER (Protein Analysis Through Evolutionary Relationships)	A web-based resource to classify genes in order to facilitate high-throughput analysis	[77]
pysam	A python wrapper around the samtools package	[58]

(continued)

Table 1
(continued)

Software/computational tool	Description	Reference
Random forest machine learning classifier[a]	Model to predict transcription factors responsible for the regulation of downstream effectors	[87]
REVIGO	A web-based tool to analyze and visualize Gene Ontology term results.	[79]
RSGE (Region-specific gene enrichment)[a]	Determines a regional enrichment score for each cluster using bulk RNA-seq data from [89].	[33]
SCENIC (Single-Cell rEgulatory Network Inference and Clustering)[a]	Set of tools to infer gene regulatory networks, transcription factors, and cell types in Drosophila.	[28, 91]
ScoMap (Single-Cell Omics Mapping into spatial Axes using Pseudotime ordering)[a]	Spatial integration of single-cell omics data into virtual cells and infers enhancer-to-gene relationships	[90]
SCope[a]	Online visualization tool for scRNA-seq data	[28]
Scvelo	Toolkit for generalized RNA velocity analysis through dynamical modeling (e.g., for non-stationary populations) and for visualization of RNA velocity results	[86]
Seurat	Comprehensive R package for single-cell data analysis	[17, 34]
Slingshot	Performs pseudotime analysis to infer developmental trajectories and lineages	[83]
SRGs (Spatially restricted genes)[a]	Tool to retrieve spatially restricted genes (= genes detected in fewer cells than expected based on their average gene expression)	[43]
velocyto	Method for the analysis of expression dynamics in scRNA-seq data enabling estimations of RNA velocities of single cells by distinguishing unspliced and spliced mRNAs	[84]

[a]Tools developed using *Drosophila* scRNA-seq data

can require different filtering parameters to determine high-quality cells (Quality controls). Another reason is the choice of programming language. Although most tools are written in R or Python [16], the choice for one of those can limit the computational toolbox to methods developed in the respective language. For R users, well-documented packages are Seurat [17] and scater [18] which provide a variety of analysis tools for scRNA-seq data. In Python, Scanpy [19] and totalVI [20] offer comprehensive toolkits for single-cell data science. In this book chapter, we will put a focus on methods and tools developed in R for droplet-based scRNA-seq data preprocessing, highlight R- and Python-based downstream

analysis packages, and also include common web-based downstream analysis tools. When starting with single-cell data exploration, a few best practices and approaches should be considered [7, 21, 22]. However, as there are no standardized analysis pipelines, different tools might have to be tested, combined, and evaluated in an iterative manner.

In this book chapter, we want to give guidance on analyzing droplet-based single-cell gene expression data from *Drosophila melanogaster*. Initially, we will briefly outline the basic laboratory workflow of a *Drosophila* single-cell experiment and highlight a few key steps that are important for subsequent bioinformatic analyses. Following the generation of a count matrix, we will point out different computational methods commonly used for exploration of *Drosophila* scRNA-seq data. Notably, while other sources provide code-based files [23], online tutorials with general guidelines on scRNA-seq data analysis (https://satijalab.org/seurat/), or comprehensive Github repositories (totalVI, Scanpy, Seurat), we discuss in detail the possible settings and parameters that one can apply in particular for *Drosophila* scRNA-seq data.

2 Laboratory Workflow

In this section, we want to briefly outline the main steps of generating droplet-based single-cell libraries from *Drosophila* tissues and point out a few tips and tricks in the note section that will help in the subsequent bioinformatic analysis. You need to decide on the sample preparation protocol and the technology you want to apply, which will also impact your bioinformatic analysis strategy (reviewed in [7]).

For readers that are new to the *Drosophila* scRNA-seq field and decided to use the 10× Chromium platform, the protocol of Nguyen and colleagues analyzing *Drosophila* larval ventral cord might be a useful resource to get started in the laboratory [24] (*see* **Note 1**). After having isolated and dissociated your tissue, it is important to count your cells. This information can then later be compared to the number of cells obtained after sequencing and should allow you to estimate the number of cell doublets, an information that can be taken into account in the subsequent analysis to identify the doublets. Once you have obtained the total number of cells, you can follow the single-cell protocol.

The generated libraries can then be sequenced in bulk using next-generation sequencing (NGS) machines (*see* **Note 2**). The required sequencing depth depends on the biological question, and is influenced by many factors, including the cell type and the chosen scRNA-seq method. 10× Chromium libraries from *Drosophila* cells are typically sequenced with 20,000–200,000 mean reads per cell ([25–33], *see* **Note 3**).

3 Data Analysis

There are certain best practices to be considered when examining single-cell data [7, 21, 22]). We want to point out that there is no standardized analysis pipeline that can be applied to every cell type and dataset due to inherent species, cell type, sample preparation, and technology differences, as well as differences in the biological question. Quantifying gene expression from your raw sequencing output (text-based sequence files) is the first analysis step and returns a count matrix with following dimensions: number of barcodes x number of genes. The procedure for obtaining count matrices from droplet-based techniques is described in detail in other studies ([4, 24], computational cookbook v.1.2 (http://mccarrolllab.org/dropseq); 10× Genomics Cell Ranger (https://support.10xgenomics.com/single-cell-gene-expression/software/pipelines/latest/what-is-cell-ranger)). In this book chapter, we want to outline the preprocessing procedure (quality control, normalization, data correction, feature selection, and dimensionality reduction) for *Drosophila* scRNA-seq data, which is often done using the Seurat package in R [17, 34]. We then also highlight a few of the most commonly applied and developed computational tools for downstream analysis of *Drosophila* scRNA-seq data.

3.1 Quality Control

Starting with the generated count matrices, the first step is to select high-quality single cells. While there are many filtering criteria, selecting single *Drosophila* cells with sufficient transcript information is commonly performed using at least these two metrics: the total number of transcript counts (nCount) per cell and the total number of genes (nGene) per cell (*see* **Notes 4** and **5**).

A lower threshold is applied to remove low-quality cell barcodes (dead or dying cells with reduced mRNA content due to membrane leakage and mRNA degradation or cell barcodes with lower mRNA capture efficiency); the upper cutoff is set to discard potential cell doublets, multiplets, or to remove outliers. A good starting point is the default Seurat threshold for nGene (>200 and <2500). However, all filtering thresholds are strongly influenced by the species, cell type, technology, sample preparation, treatment, and sample batch [4, 35–39], and many *Drosophila* studies hence chose deviating thresholds. For instance, Witt and colleagues applied only a lower nCount cutoff to account for differences in the observed transcript and gene differences in the adult testis cell types [37]. While Cho and colleagues, for instance, used separate filtering criteria for each library to account for technical batch effects. The upper nCount threshold was thereby statistically calculated as the mean plus two standard deviations [36].

Another quality control (QC) criterion to select high-quality cells is to assess the fraction of counts from mitochondrial genes per cell barcode (percent.mito), which is typically an indicator for cellular stress [40]. While this percent.mito QC criteria is applied to many human and mouse scRNA-seq data with a recommended Seurat cutoff of <5%, we noted that most published *Drosophila* scRNA-seq studies did not use this filtering criteria. The *Drosophila* scRNA-seq studies that applied this metric set a threshold that ranged from <1% up to <50% ([30, 41], *see* **Note 6**).

Besides these commonly used filtering approaches, different groups have come up with novel and unique strategies for selecting high-quality cells and removing dead or broken ones. One study used the percentage of heat shock genes per cell (which are known markers for cellular stress) and the percentage of ribosomal genes (28sRNA) per cell to remove dead or dying cells [42]. Further quality control metrics are the alignment rate (percentage of mapped reads, e.g., >85% in [43]) and the library complexity (calculated as log10(genes detected)/log10(UMI counts per cell)), which can be applied to select good-quality cells with sufficient transcript information [41, 44]. For some datasets with unique scientific questions, other filtering approaches might be useful. Karaiskos and colleagues, for instance, only selected cells that express >5 genes of the Berkeley *Drosophila* Transcription Network Project (BDTNP) reference atlas [45].

Because of the inherent nature of microfluidics-based scRNA-seq methods, droplets with more than one cell cannot be avoided. However, loading a defined number of total cells allows the user to estimate the number of multiplets. For instance, with 10× Genomics technology, the multiplet rate is at ~7.6% when 16,000 cells are loaded (*see* **Note 7**). The hybrid transcriptome of multiplets can pose challenges during downstream analyses, as they potentially indicate distinct cell types or cell states that are actually not present in the data or simply do not exist. Distinct approaches to identify and remove cell doublets or multiplets have been applied to *Drosophila* scRNA-seq data. As described above, an upper nCount threshold can sometimes be set to remove those cells. Others have removed barcodes with co-expression of known marker genes of distinct cell types [30, 45, 46]. Karaiskos and colleagues additionally identified doublets as distinct populations by principal component analysis [45]. Moreover, there are tools that can infer doublets in single-cell data (Scrublet [47], DoubletFinder [48]) which have also been applied to *Drosophila* datasets [36, 49]. In case of a mixed-species experiment, a part of the multiplet population (interspecies multiplets) can be determined by assessing the transcriptome purity per cell [4, 35, 45].

Besides these cell quality controls additional filtering can be applied to the detected gene set. Often, only features are selected that are expressed in at least a certain number of cells. This metric leads to noise reduction, thus improving the overall dataset quality. In Seurat, this approach is implemented in the "CreateSeuratObject" function as "min.cells" argument. After having applied this feature selection, the reader will observe a lower total number of genes in the expression matrix. In several *Drosophila* scRNA-seq studies, the "min.cells" parameter has been set between 2 and 10 (e.g., [25, 29, 43]). To improve downstream analyses, final count matrices can also be "cleaned" for mitochondrial genes, ribosomal protein genes, genes encoding heat shock proteins, and genes from oxidative phosphorylation complexes [50]. Tools such as SoupX [51] or DecontX [52] that quantify and remove decontamination of ambient RNA have also been applied to *Drosophila* data [53]. Setting good-quality controls (for the respective scientific question asked) and proper filtering thresholds will enable the selection of high-quality, viable single-cell barcodes

3.2 Normalization

Although scRNA-seq methods have progressed enormously over the past years, all technologies still display high levels of variability and noise [13, 54]. This is mainly a result of low starting material and comparatively low efficiencies of RNA capturing and reverse transcription. Given these technical inefficiencies, it is possible for identical cells to differ in their count depth. To account for these differences, data normalization is very crucial. For count-based single-cell gene expression data, a library size normalization is commonly conducted in which the feature counts per cell are divided by the total counts of that cell. The counts are then multiplied by a scale factor and log-transformed. This type of normalization has been used in many *Drosophila* studies and was recently found to perform among the best if combined with a dimension reduction, like PCA (principal component analysis) [55]. An alternative method for UMI-based techniques called "sctransform," does not require pseudocount addition or log-transformation and has been reported to improve downstream procedures [56]. Sctransform was, for instance, applied to *Drosophila* single-cell brain data [57]. To be able to benchmark different single-cell protocols for their detected gene and transcript level, the sequencing depth should be comparable. "Pysam," a python wrapper around the samtools package (https://github.com/pysam-developers/pysam, [58]), allowed us to account for differences in sequencing depth by normalizing scRNA-seq libraries to the same average number of reads per cell based on the library

3.3 Finding Variable Features

Due to the high levels of technical noise from scRNA-seq experiments, downstream analyses such as dimensional reduction and clustering can be challenging. It is possible to overcome this by

choosing features that have high cell-to-cell variability [12]. These so-called "highly variable features" are then used as input for downstream analyses. There are several ways to determine features with varying gene expression levels. The function "FindVariableFeatures" implemented in Seurat is applied in many *Drosophila* scRNA-seq experiments. It identifies variable genes by determining the mean–variance relationship [34]. Although Seurat's method has found application in many single-cell investigations, other computational tools exist that identify variable genes using different approaches. One such method has been developed by Bageritz and colleagues and aimed at identifying spatially restricted genes (SRGs) in the larval wing imaginal disc epithelium. This method makes use of the high dropout rate in scRNA-seq data to derive SRGs, which are genes detected in fewer cells than expected

3.4 Scaling and Regressing Out Biological or Technical Effects

Prior to dimensional reduction, the count data requires scaling which is commonly performed by applying a linear transformation (in Seurat for instance). During this process, it is also possible to remove unwanted sources of variation (biological or technical noise). In few *Drosophila* studies a cell cycle score was determined and its effect removed using linear regression [30, 59]. Similarly, Hörmann et al. calculated a gender score and regressed this out [60]. Other studies have reported correcting for the effects of mitochondrial genes, another biological covariate [25, 30, 33, 42]. Zappia and colleagues have also regressed out other biological factors, including heat shock genes and genes corresponding to 28sRNA [42].

On the other side, there are also technical covariates that should be considered. To correct for differences due to library size, some studies have applied a linear regression to the number of counts [25, 30, 33]. Additionally, it is also possible to regress out the number of features

3.5 Batch Correction and Data Integration

Besides the typically high levels of noise inherent to single-cell data [13, 54], batch effects can add a further layer of variability. Batch effects can arise due to the use of different chips, due to sequencing on different lanes or because of the experimenter itself. Furthermore, the environment in which cells were harvested and prepared could impact the transcriptome [21]. It is therefore important to identify these covariates and correct them. Batch correction is usually performed on datasets from the same experiment, while data integration aims at integrating samples from multiple experiments [21]. Harmony [61], a batch correction method, has been used in few *Drosophila* studies [44, 53].

In contrast to batch correction methods, data integration tools face further challenges as they need to additionally consider differences in cellular composition of each dataset (*see* **Note 9**). Although there are several methods published for batch correction and data

integration [21], several *Drosophila* scRNA-seq experiments have applied Seurat's data integration function [34] which initially ran a canonical correlation analysis (CCA)

3.6 Dimensional Reduction and Clustering

Dimensional reduction is an important next step prior to clustering and visualization because even after filtering, scRNA-seq data can have thousands of dimensions [21], thus making downstream analyses computationally heavy and noisy. After having identified variable features (*see* Subheading 3.4), the dimensional space of the count matrix can be reduced by applying a linear dimensionality reduction algorithm. The most commonly used method in single-cell data science is PCA that summarizes the count matrix data into a new set of characteristics (called principal components, PCs). An alternative approach is the independent component analysis (ICA) that has also been applied to *Drosophila* data [31] as it was previously shown to recognize subtle variations among cells of the same class [62].

In the next step, the dimensionality of the datasets needs to be assessed, meaning how many PCs (=dimensions) actually capture the biological variability and thus should be included for clustering and visualization. Seurat provides a few options for determining the optimal number of principal components (e.g., "JackStrawPlot" provides a visualization of a statistical test that calculates the significance of your PCs, *see* **Note 10**). For the integrated Seurat datasets, different dimensionality parameters can be applied and the best-preserved local structure determined by the "LocalStruct" function (e.g., used by [63]). An alternative approach to assess a cluster's biological relevance and the number of informative PCs was conducted by Özel et al., 2021 who calculated for each cluster the Pearson correlation to log-normalized bulk transcriptomes [32].

To cluster the cells, Seurat has implemented a graph-based algorithm using a K-nearest neighbor (KNN) graph [64, 65]) followed by modularity optimization based on the Louvain algorithm which has been used in most *Drosophila* scRNA-seq studies. Finally, a non-linear dimensional reduction is performed to visualize and explore the data in a two- or three-dimensional space. Commonly used methods are t-distributed stochastic neighbor embedding

3.7 Cluster Annotation Using Marker Genes

Annotation of cell clusters can be principally done in two ways: (1) by manual curation using known marker genes, which requires background knowledge (*see* **Note 11**) or (2) reference mapping based on available reference atlases [68] to generate a joined embedding. In many *Drosophila* studies, Seurat's "FindMarkers" function is utilized, which performs differential gene expression (DGE) testing on respective identity classes. Seurat has implemented several statistical tests to assess the significance level of marker genes. Currently, it offers following methods: Wilcoxon Rank Sum test (also known as Mann–Whitney U test), likelihood ratio test

[69], ROC (Receiver operating characteristic) analysis, Student's t-test, negative binomial test, Poisson test, logistic regression, MAST [70], and DESeq2 [71]. To choose the appropriate DGE testing method, we recommend the study conducted by Wang and colleagues who have compared several DGE analysis methods [72].

The two most commonly used tests in *Drosophila* studies are the ROC analysis which returns the classification power (AUC = area under the curve) and the Wilcoxon Rank Sum test (default of Seurat's "FindMarkers" function; *see* **Note 12**).

3.8 Possible Downstream Analyses

Downstream analyses always depend on the respective biological question. In the following we want to highlight some of the possible downstream analyses that are often conducted in *Drosophila* scRNA-seq experiments. We listed the analysis tools in Table 1.

3.8.1 Gene Ontology and Network Analysis

After having identified cluster marker genes, many studies perform a GO (Gene Ontology) term enrichment analysis on these features. For *Drosophila*, several online resources and GO databases exist. In the following we want to name a few that were used in recent *Drosophila* studies: (1) Gene list annotations for *Drosophila* (GLAD) online resource [73], (2) The Database for Annotation, Visualization and Integrated Discovery (DAVID) [74, 75]), (3) g:Profiler [76], (4) Protein Analysis Through Evolutionary Relationships (PANTHER) (latest version [77]), and (5) GOrilla [78]. For analyzing and visualizing GO term results REVIGO [79] can be utilized. We and others [43, 44]) have used "Cytoscape" which is an open-source software that allows network analyses, gene–gene correlations, and visualization of experimental data [80].

3.8.2 Pseudotime Analysis and RNA Velocity

Moreover, many *Drosophila* studies conduct a pseudotime analysis to infer cell trajectories and lineages (e.g., [26, 27, 37]). The two most commonly applied methods are "Monocle" (the newest version being Monocle3 [81, 82]) and "Slingshot" [83]. In a very similar context, "velocyto" [84] has been used to determine RNA velocity and the transcriptional alterations in single cells [27, 85]. RNA velocity results can be visualized with "scvelo" [86].

3.8.3 Machine Learning Approaches

Additionally, two studies generated machine learning models using *Drosophila* scRNA-seq data. Özel and colleagues trained a neural network to classify thousands of single cells (at adult and pupal stages) and assigned them to hundreds of cell types at all stages of their development [32]. Their final model is available and can be used to annotate different datasets. A second study made use of the relationship between transcription factors and their downstream targets to generate a "random forest" machine learning network that could predict a cell's gene expression based on its transcription factor profile [87].

3.8.4 Analysis of Base-Substitution Mutations

Finally, a study on *Drosophila* testis used single-cell count data to obtain mutational information [37]. For this, they applied "bcftools mpileup" [88] by which they could identify more than 70 de novo base substitutions.

3.9 Computational Methods Developed Using Drosophila Single-Cell Data

In the following we want to highlight how *Drosophila* single-cell RNA data was used to develop novel computational methods (Table 1, highlighted with an asterisk), such as "DistMap" which was introduced by Karaiskos and colleagues [45]. Their method enables researchers to predict spatial gene expression patterns using scRNA-seq data. To do so, they generated distributed mapping scores for every sequenced cell across all tissue positions. Furthermore, they developed the *Drosophila* Virtual Expression eXplorer (DVEX) which can generate virtual in situ hybridizations (vISH) for individual and multiple genes. Bageritz and colleagues developed a method to retrieve spatial gene expression information from scRNA-seq data of the larval wing imaginal disc epithelium. This method is particularly interesting for large and undifferentiated tissues, with many stem-like cells and hence few cell types. It is based on gene expression correlations to a set of reference genes with known expression patterns. Additionally, the authors devised an algorithm that identifies a suitable set of mapping genes [43]. A tool developed by Guo X et al. [33] aims at determining the regional preference of each cluster in scRNA-seq midgut data. Their algorithm, called region-specific gene enrichment (RSGE), makes use of bulk RNA-seq data from flygut-seq [89] to determine a regional enrichment score for each cluster. Moreover, González-Blas and colleagues have developed the R package ScoMap (Single-Cell Omics Mapping into spatial Axes using Pseudotime ordering [90]) that enables spatial integration of single-cell omics data into virtual cells. This virtual environment can either resemble a specific tissue or an abstract space.

An open resource of the *Drosophila* aging brain (with more than 150,000 cells) was developed and provided by Davie and colleagues. "SCope" enables the exploration and visualization of gene expression states in an easy and user-friendly interface [28]. Additionally, based on the already existing method single-cell regulatory network inference and clustering (SCENIC [91]), Davie and colleagues developed SCENIC for the *Drosophila* genome which allows investigating gene regulatory networks in *Drosophila* [28]. To study cell–cell communication pathways in *Drosophila* scRNA-seq data Liu and colleagues developed a web-based framework, called FlyPhoneDB, which predicts receptor–ligand interactions [92].

4 Notes

1. Keep in mind that depending on the tissue of interest you might need to adapt and modify a few experimental steps. The chosen sample preparation protocol should allow recovery of high-quality cells or nuclei. Lengthy protocols, RNA degradation, or other side effects (see e.g., [93]) need to be avoided so that the recovered mRNA reflects the transcriptomic status and many high-quality cells can be retrieved for downstream analysis. Additionally, cell damage needs to be avoided to reduce the level of ambient RNA.
2. To avoid or reduce batch effects induced during sequencing we recommend to pool the different libraries that will be analyzed together (samples, replicates) and perform sequencing from this pool. If you want to pool libraries, make sure to use different index primers during library construction.
3. A loss of read information becomes apparent at each of those initial analysis steps: read QC, alignment rate, and cell selection with ambient RNA captured by empty droplets. This should be taken into account when deciding on sequencing depth for libraries. We observed losses of about 30% raw read information.
4. From our experience, the detected median gene and transcript level often reflects the cDNA library yield after the initial amplification step of the experimental library workflow. In addition, the cDNA profile of high-quality data should show a *Drosophila* specific average size, with 1300–2000 for *Drosophila* Drop-seq data. Lower average cDNA sizes can indicate RNA degradation and hence fewer high-quality cells.
5. In general, we found that filtering extreme outliers improved downstream analysis. However, we would advise to set the filtering thresholds as permissive as possible to avoid filtering out cell populations unintentionally. Decisions on the settings of a specific filtering criteria can be made based on the subsequent downstream analysis results.
6. From our experience, a $<5\%$ or up to $<20\%$ cutoff for mitochondrial genes is often enough to remove low-quality cells, which are also characterized by low nCount and/or high expression of stress response genes. A higher cutoff might be required for energy demanding tissues that can have inherently elevated mitochondrial gene expression as described for human heart or muscle tissue [94].

7. The recommended input cell concentration from 10× Chromium is based on species mixing experiments of human and mouse cells. We found that small *Drosophila* cells should be loaded about 2× more to achieve a comparable number of recovered cells and cell doublets.
8. We provide a detailed description and tutorial of the SRG tool on the Boutros laboratory repository of Github (https://github.com/boutroslab/Supplemental-Material/tree/master/Bageritz_2019). Notably, we found that the SRG tool works particularly well for pulling out genes expressed in restricted domains, so candidate genes for follow-up studies.
9. Reliable data integration is key for further downstream analysis. To assess the performance of the integration, we recommend to do the following steps: (i) perform the scRNA-seq analysis on individual datasets, (ii) assess the contribution of each experiment or replicate to the different clusters, and (iii) determine the preservation of the local structure by the Seurat "LocalStruct" function (e.g., used by [63]).
10. As also pointed out in the Seurat tutorial, we recommend exploring the dimensionality of the data with all three provided methods: JackStraw procedure for statistical analysis, the Elbow plot showing the ranked PCs based on the percentage of variance explained by each one, and visual inspection of heatmaps focusing on a principal component, in which cells and genes are sorted by their principal component scores.
11. The search database DRscDB (***Drosophila RNAi Screening Center's single-cell DataBase*** [95]) provides a valuable resource to query genes of interest in scRNA-seq studies in *Drosophila*, human and other model systems and can hence help annotate clusters.
12. Since the true distribution of single-cell data is still discussed [96], we recommend the non-parametric Wilcoxon Rank Sum test for differential gene expression analysis.

Acknowledgments

M.J.M. was supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – SFB 1324 – project number 331351713.

References

1. Tang F, Barbacioru C, Wang Y et al (2009) mRNA-Seq whole-transcriptome analysis of a single cell. Nat Methods 6:377–382. https://doi.org/10.1038/nmeth.1315
2. Islam S, Kjällquist U, Moliner A et al (2012) Highly multiplexed and strand-specific single-cell RNA 5′ end sequencing. Nat Protoc 7:813–828. https://doi.org/10.1038/nprot.2012.022

3. Ramsköld D, Luo S, Wang Y-C et al (2012) Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells. Nat Biotechnol 30:777–782. https://doi.org/10.1038/nbt.2282
4. Macosko EZ, Basu A, Satija R et al (2015) Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. Cell 161:1202–1214. https://doi.org/10.1016/j.cell.2015.05.002
5. Rosenberg AB, Roco CM, Muscat RA et al (2018) Single-cell profiling of the developing mouse brain and spinal cord with split-pool barcoding. Science 360:176–182. https://doi.org/10.1126/science.aam8999
6. Picelli S, Björklund ÅK, Faridani OR et al (2013) Smart-seq2 for sensitive full-length transcriptome profiling in single cells. Nat Methods 10:1096–1098. https://doi.org/10.1038/nmeth.2639
7. Chen G, Ning B, Shi T (2019) Single-cell RNA-seq technologies and related computational data analysis. Front Genet 10:317. https://doi.org/10.3389/fgene.2019.00317
8. Klein AM, Mazutis L, Akartuna I et al (2015) Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. Cell 161:1187–1201. https://doi.org/10.1016/j.cell.2015.04.044
9. Li H (2021) Single-cell RNA sequencing in *Drosophila*: technologies and applications. WIREs Dev Biol 10:e396. https://doi.org/10.1002/wdev.396
10. Simon F, Konstantinides N (2021) Single-cell transcriptomics in the *Drosophila* visual system: advances and perspectives on cell identity regulation, connectivity, and neuronal diversity evolution. Dev Biol 479:107–122. https://doi.org/10.1016/j.ydbio.2021.08.001
11. Zheng GXY, Terry JM, Belgrader P et al (2017) Massively parallel digital transcriptional profiling of single cells. Nat Commun 8:14049. https://doi.org/10.1038/ncomms14049
12. Brennecke P, Anders S, Kim JK et al (2013) Accounting for technical noise in single-cell RNA-seq experiments. Nat Methods 10:1093–1095. https://doi.org/10.1038/nmeth.2645
13. Grün D, Kester L, van Oudenaarden A (2014) Validation of noise models for single-cell transcriptomics. Nat Methods 11:637–640. https://doi.org/10.1038/nmeth.2930
14. Stegle O, Teichmann SA, Marioni JC (2015) Computational and analytical challenges in single-cell transcriptomics. Nat Rev Genet 16:133–145. https://doi.org/10.1038/nrg3833
15. Lähnemann D, Köster J, Szczurek E et al (2020) Eleven grand challenges in single-cell data science. Genome Biol 21:31. https://doi.org/10.1186/s13059-020-1926-6
16. Zappia L, Phipson B, Oshlack A (2018) Exploring the single-cell RNA-seq analysis landscape with the scRNA-tools database. PLoS Comput Biol 14:e1006245. https://doi.org/10.1371/journal.pcbi.1006245
17. Satija R, Farrell JA, Gennert D et al (2015) Spatial reconstruction of single-cell gene expression data. Nat Biotech 33:495–502
18. McCarthy DJ, Campbell KR, Lun ATL, Wills QF (2017) Scater: pre-processing, quality control, normalization and visualization of single-cell RNA-seq data in R. Bioinformatics 33:1179–1186. https://doi.org/10.1093/bioinformatics/btw777
19. Wolf FA, Angerer P, Theis FJ (2018) SCANPY: large-scale single-cell gene expression data analysis. Genome Biol 19:15. https://doi.org/10.1186/s13059-017-1382-0
20. Gayoso A, Steier Z, Lopez R et al (2021) Joint probabilistic modeling of single-cell multiomic data with totalVI. Nat Methods 18:272–282. https://doi.org/10.1038/s41592-020-01050-x
21. Luecken MD, Theis FJ (2019) Current best practices in single-cell RNA-seq analysis: a tutorial. Mol Syst Biol 15:e8746. https://doi.org/10.15252/msb.20188746
22. Rostom R, Svensson V, Teichmann SA, Kar G (2017) Computational approaches for interpreting scRNA-seq data. FEBS Lett 591:2213–2225. https://doi.org/10.1002/1873-3468.12684
23. Vicidomini R, Nguyen TH, Choudhury SD et al (2021) Assembly and exploration of a single cell atlas of the *Drosophila* larval ventral cord. Identification of rare cell types. Curr Protoc 1(e37). https://doi.org/10.1002/cpz1.37
24. Nguyen TH, Vicidomini R, Choudhury SD et al (2021) Single-cell RNA sequencing analysis of the *Drosophila* larval ventral cord. Curr Protoc 1:e38. https://doi.org/10.1002/cpz1.38
25. Allen AM, Neville MC, Birtles S et al (2020) A single-cell transcriptomic atlas of the adult *Drosophila* ventral nerve cord. eLife 9:e54074. https://doi.org/10.7554/eLife.54074
26. Ariss MM, Islam ABMMK, Critcher M et al (2018) Single cell RNA-sequencing identifies a metabolic aspect of apoptosis in Rbf mutant. Nat Commun 9:5024. https://doi.org/10.1038/s41467-018-07540-z

27. Cattenoz PB, Sakr R, Pavlidaki A et al (2020) Temporal specificity and heterogeneity of *Drosophila* immune cells. EMBO J 39:e104486. https://doi.org/10.15252/embj.2020104486
28. Davie K, Janssens J, Koldere D et al (2018) A single-cell transcriptome atlas of the aging *Drosophila* brain. Cell 174:982–998.e20. https://doi.org/10.1016/j.cell.2018.05.057
29. Deng M, Wang Y, Zhang L et al (2019) Single cell transcriptomic landscapes of pattern formation, proliferation and growth in *Drosophila* wing imaginal discs. Development 146: dev179754. https://doi.org/10.1242/dev.179754
30. Jevitt A, Chatterjee D, Xie G et al (2020) A single-cell atlas of adult *Drosophila* ovary identifies transcriptional programs and somatic cell lineage regulating oogenesis. PLoS Biol 18: e3000538. https://doi.org/10.1371/journal.pbio.3000538
31. Kurmangaliyev YZ, Yoo J, LoCascio SA, Zipursky SL (2019) Modular transcriptional programs separately define axon and dendrite connectivity. eLife 8:e50822. https://doi.org/10.7554/eLife.50822
32. Özel MN, Simon F, Jafari S et al (2021) Neuronal diversity and convergence in a visual system developmental atlas. Nature 589:88–95. https://doi.org/10.1038/s41586-020-2879-3
33. Guo X, Yin C, Yang F et al (2019) The cellular diversity and transcription factor code of *Drosophila* enteroendocrine cells. Cell Rep 29: 4172–4185.e5. https://doi.org/10.1016/j.celrep.2019.11.048
34. Stuart T, Butler A, Hoffman P et al (2019) Comprehensive integration of single-cell data. Cell 177:1888–1902.e21. https://doi.org/10.1016/j.cell.2019.05.031
35. Bageritz J, Krausse N, Yousefian S, et al (2021) Glyoxal as alternative fixative for single cell RNA sequencing. bioRxiv 2021.06.06.447272. https://doi.org/10.1101/2021.06.06.447272
36. Cho B, Yoon S-H, Lee D et al (2020) Single-cell transcriptome maps of myeloid blood cell lineages in *Drosophila*. Nat Commun 11:4483. https://doi.org/10.1038/s41467-020-18135-y
37. Witt E, Benjamin S, Svetec N, Zhao L (2019) Testis single-cell RNA-seq reveals the dynamics of de novo gene transcription and germline mutational bias in *Drosophila*. eLife 8:e47138. https://doi.org/10.7554/eLife.47138
38. Ziegenhain C, Vieth B, Parekh S et al (2017) Comparative analysis of single-cell RNA sequencing methods. Mol Cell 65:631–643. e4. https://doi.org/10.1016/j.molcel.2017.01.023
39. Wang X, He Y, Zhang Q et al (2021) Direct comparative analyses of 10X genomics chromium and smart-seq2. Genomics Proteomics Bioinformatics. https://doi.org/10.1016/j.gpb.2020.02.005
40. Ilicic T, Kim JK, Kolodziejczyk AA et al (2016) Classification of low quality cells from single-cell RNA-seq data. Genome Biol 17:29. https://doi.org/10.1186/s13059-016-0888-1
41. Hung R-J, Hu Y, Kirchner R et al (2020) A cell atlas of the adult *Drosophila* midgut. Proc Natl Acad Sci 117:1514. https://doi.org/10.1073/pnas.1916820117
42. Zappia MP, de Castro L, Ariss MM et al (2020) A cell atlas of adult muscle precursors uncovers early events in fibre-type divergence in *Drosophila*. EMBO Rep n/a:e49555. https://doi.org/10.15252/embr.201949555
43. Bageritz J, Willnow P, Valentini E et al (2019) Gene expression atlas of a developing tissue by single cell expression correlation analysis. Nat Methods 16:750–756. https://doi.org/10.1038/s41592-019-0492-x
44. Tattikota SG, Cho B, Liu Y et al (2020) A single-cell survey of *Drosophila* blood. eLife 9: e54818. https://doi.org/10.7554/eLife.54818
45. Karaiskos N, Wahle P, Alles J et al (2017) The *Drosophila* embryo at single cell transcriptome resolution. Science 358(6360):194–199
46. Everetts NJ, Worley MI, Yasutomi R et al (2021) Single-cell transcriptomics of the *Drosophila* wing disc reveals instructive epithelium-to-myoblast interactions. eLife 10: e61276. https://doi.org/10.7554/eLife.61276
47. Wolock SL, Lopez R, Klein AM (2019) Scrublet: computational identification of cell doublets in single-cell transcriptomic data. Cell Syst 8:281–291.e9. https://doi.org/10.1016/j.cels.2018.11.005
48. McGinnis CS, Murrow LM, Gartner ZJ (2019) DoubletFinder: doublet detection in single-cell RNA sequencing data using artificial nearest neighbors. Cell Syst 8:329–337.e4. https://doi.org/10.1016/j.cels.2019.03.003
49. Rust K, Byrnes LE, Yu KS et al (2020) A single-cell atlas and lineage analysis of the adult *Drosophila* ovary. Nat Commun 11:5628. https://doi.org/10.1038/s41467-020-19361-0
50. Kurmangaliyev YZ, Yoo J, Valdes-Aleman J et al (2020) Transcriptional programs of circuit assembly in the *Drosophila* visual system.

Neuron 108:1045–1057.e6. https://doi.org/10.1016/j.neuron.2020.10.006

51. Young MD, Behjati S (2020) SoupX removes ambient RNA contamination from droplet-based single-cell RNA sequencing data. GigaScience 9. https://doi.org/10.1093/gigascience/giaa151
52. Yang S, Corbett SE, Koga Y et al (2020) Decontamination of ambient RNA in single-cell RNA-seq with DecontX. Genome Biol 21:57. https://doi.org/10.1186/s13059-020-1950-6
53. Ghosh AC, Tattikota SG, Liu Y et al (2020) *Drosophila* PDGF/VEGF signaling from muscles to hepatocyte-like cells protects against obesity. eLife 9:e56969. https://doi.org/10.7554/eLife.56969
54. Hicks SC, Townes FW, Teng M, Irizarry RA (2018) Missing data and technical variability in single-cell RNA-sequencing experiments. Biostatistics 19:562–578. https://doi.org/10.1093/biostatistics/kxx053
55. Ahlmann-Eltze C, Huber W (2021) Transformation and preprocessing of single-cell RNA--seq data. bioRxiv 2021.06.24.449781. https://doi.org/10.1101/2021.06.24.449781
56. Hafemeister C, Satija R (2019) Normalization and variance stabilization of single-cell RNA--seq data using regularized negative binomial regression. Genome Biol 20:296. https://doi.org/10.1186/s13059-019-1874-1
57. Baker BM, Mokashi SS, Shankar V et al (2021) The *Drosophila* brain on cocaine at single-cell resolution. Genome Res. https://doi.org/10.1101/gr.268037.120
58. Li H, Handsaker B, Wysoker A et al (2009) The sequence alignment/map format and SAMtools. Bioinformatics 25:2078–2079. https://doi.org/10.1093/bioinformatics/btp352
59. Genovese S, Clément R, Gaultier C et al (2019) Coopted temporal patterning governs cellular hierarchy, heterogeneity and metabolism in *Drosophila* neuroblast tumors. eLife 8:e50375. https://doi.org/10.7554/eLife.50375
60. Hörmann N, Schilling T, Ali AH et al (2020) A combinatorial code of transcription factors specifies subtypes of visual motion-sensing neurons in *Drosophila*. Development 147. https://doi.org/10.1242/dev.186296
61. Korsunsky I, Millard N, Fan J et al (2019) Fast, sensitive and accurate integration of single-cell data with harmony. Nat Methods 16:1289–1296. https://doi.org/10.1038/s41592-019-0619-0
62. Saunders A, Macosko EZ, Wysoker A et al (2018) Molecular diversity and specializations among the cells of the adult mouse brain. Cell 174:1015–1030.e16. https://doi.org/10.1016/j.cell.2018.07.028
63. Konstantinides N, Rossi AM, Escobar A, et al (2021) A comprehensive series of temporal transcription factors in the fly visual system. bioRxiv 2021.06.13.448242. https://doi.org/10.1101/2021.06.13.448242
64. Levine JH, Simonds EF, Bendall SC et al (2015) Data-driven phenotypic dissection of AML reveals progenitor-like cells that correlate with prognosis. Cell 162:184–197. https://doi.org/10.1016/j.cell.2015.05.047
65. Xu C, Su Z (2015) Identification of cell types from single-cell transcriptomes using a novel clustering method. Bioinformatics 31:1974–1980. https://doi.org/10.1093/bioinformatics/btv088
66. van der Maaten L, Hinton G (2008) Visualizing data using t-SNE. J Mach Learn Res 9:2579–2605
67. McInnes L, Healy J, Melville J (2020) UMAP: uniform manifold approximation and projection for dimension reduction. ArXiv 180203426 Cs Stat
68. Li H, Janssens J, De Waegeneer M, et al (2021) Fly cell atlas: a single-cell transcriptomic atlas of the adult fruit fly. bioRxiv 2021.07.04.451050. https://doi.org/10.1101/2021.07.04.451050
69. McDavid A, Finak G, Chattopadyay PK et al (2013) Data exploration, quality control and testing in single-cell qPCR-based gene expression experiments. Bioinformatics 29:461–467. https://doi.org/10.1093/bioinformatics/bts714
70. Finak G, McDavid A, Yajima M et al (2015) MAST: a flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. Genome Biol 16:278. https://doi.org/10.1186/s13059-015-0844-5
71. Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 15:550. https://doi.org/10.1186/s13059-014-0550-8
72. Wang T, Li B, Nelson CE, Nabavi S (2019) Comparative analysis of differential gene expression analysis tools for single-cell RNA sequencing data. BMC Bioinformatics 20:40. https://doi.org/10.1186/s12859-019-2599-6

73. Hu Y, Comjean A, Perkins LA et al (2015) GLAD: an online database of gene list annotation for *Drosophila*. J Genomics 3: 75–81. https://doi.org/10.7150/jgen.12863
74. Huang DW, Sherman BT, Zheng X et al (2009) Extracting biological meaning from large gene lists with DAVID. Curr Protoc Bioinforma 27:13.11.1–13.11.13. https://doi.org/10.1002/0471250953.bi1311s27
75. Huang DW, Sherman BT, Lempicki RA (2009) Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res 37:1–13. https://doi.org/10.1093/nar/gkn923
76. Raudvere U, Kolberg L, Kuzmin I et al (2019) g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). Nucleic Acids Res 47: W191–W198. https://doi.org/10.1093/nar/gkz369
77. Mi H, Ebert D, Muruganujan A et al (2021) PANTHER version 16: a revised family classification, tree-based classification tool, enhancer regions and extensive API. Nucleic Acids Res 49:D394–D403. https://doi.org/10.1093/nar/gkaa1106
78. Eden E, Navon R, Steinfeld I et al (2009) GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. BMC Bioinform 10:48. https://doi.org/10.1186/1471-2105-10-48
79. Supek F, Bošnjak M, Škunca N, Šmuc T (2011) REVIGO summarizes and visualizes long lists of gene ontology terms. PLoS One 6:e21800. https://doi.org/10.1371/journal.pone.0021800
80. Su G, Morris JH, Demchak B, Bader GD (2014) Biological network exploration with Cytoscape 3. Curr Protoc Bioinform 47: 8.13.1–8.13.24. https://doi.org/10.1002/0471250953.bi0813s47
81. Trapnell C, Cacchiarelli D, Grimsby J et al (2014) The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. Nat Biotechnol 32: 381–386. https://doi.org/10.1038/nbt.2859
82. Qiu X, Mao Q, Tang Y et al (2017) Reversed graph embedding resolves complex single-cell trajectories. Nat Methods 14:979–982. https://doi.org/10.1038/nmeth.4402
83. Street K, Risso D, Fletcher RB et al (2018) Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. BMC Genomics 19:477. https://doi.org/10.1186/s12864-018-4772-0
84. La Manno G, Soldatov R, Zeisel A et al (2018) RNA velocity of single cells. Nature 560: 494–498. https://doi.org/10.1038/s41586-018-0414-6
85. Vogler G, Hum B, Tamayo M, et al (2021) Single-cell sequencing of the *Drosophila* embryonic heart and muscle cells during differentiation and maturation. bioRxiv 2021.01.15.426556. https://doi.org/10.1101/2021.01.15.426556
86. Bergen V, Lange M, Peidli S et al (2020) Generalizing RNA velocity to transient cell states through dynamical modeling. Nat Biotechnol 38:1408–1414. https://doi.org/10.1038/s41587-020-0591-3
87. Konstantinides N, Kapuralin K, Fadil C et al (2018) Phenotypic convergence: distinct transcription factors regulate common terminal features. Cell 174:622–635.e13. https://doi.org/10.1016/j.cell.2018.05.021
88. Narasimhan V, Danecek P, Scally A et al (2016) BCFtools/RoH: a hidden Markov model approach for detecting autozygosity from next-generation sequencing data. Bioinformatics 32:1749–1751. https://doi.org/10.1093/bioinformatics/btw044
89. Dutta D, Dobson AJ, Houtz PL et al (2015) Regional cell-specific transcriptome mapping reveals regulatory complexity in the adult *Drosophila* midgut. Cell Rep 12:346–358. https://doi.org/10.1016/j.celrep.2015.06.009
90. Bravo González-Blas C, Quan X-J, Duran-Romaña R et al (2020) Identification of genomic enhancers through spatial integration of single-cell transcriptomics and epigenomics. Mol Syst Biol 16:e9438. https://doi.org/10.15252/msb.20209438
91. Aibar S, González-Blas CB, Moerman T et al (2017) SCENIC: single-cell regulatory network inference and clustering. Nat Methods 14:1083–1086. https://doi.org/10.1038/nmeth.4463
92. Liu Y, Hu Y, Li JSS, et al (2021) FlyPhoneDB: an integrated web-based resource for cell-cell communication prediction in *Drosophila*. bioRxiv 2021.06.14.448430. https://doi.org/10.1101/2021.06.14.448430
93. van den Brink SC, Sage F, Vértesy Á et al (2017) Single-cell sequencing reveals dissociation-induced gene expression in tissue subpopulations. Nat Methods 14:935–936. https://doi.org/10.1038/nmeth.4437

94. Mercer TR, Neph S, Dinger ME et al (2011) The human mitochondrial transcriptome. Cell 146:645–658. https://doi.org/10.1016/j.cell.2011.06.051
95. Hu Y, Tattikota SG, Liu Y et al (2021) DRscDB: a single-cell RNA-seq resource for data mining and data comparison across species. Comput Struct Biotechnol J 19: 2018–2026. https://doi.org/10.1016/j.csbj.2021.04.021
96. Svensson V (2020) Droplet scRNA-seq is not zero-inflated. Nat Biotechnol 38:147–150. https://doi.org/10.1038/s41587-019-0379-5
97. Nitzan M, Karaiskos N, Friedman N, Rajewsky N (2019) Gene expression cartography. Nature 576:132–137. https://doi.org/10.1038/s41586-019-1773-3

Chapter 5

Prime Editing for Precise Genome Engineering in *Drosophila*

Justin A. Bosch and Norbert Perrimon

Abstract

Editing the *Drosophila* genome is incredibly useful for gene functional analysis. However, compared to gene knockouts, precise gene editing is difficult to achieve. Prime editing, a recently described CRISPR/Cas9-based technique, has the potential to make precise editing simpler and faster, and produce less errors than traditional methods. Initially described in mammalian cells, prime editing is functional in *Drosophila* somatic and germ cells. Here, we outline steps to design, generate, and express prime editing components in transgenic flies. Furthermore, we highlight a crossing scheme to produce edited fly stocks in less than 3 months.

Key words Prime editing, pegRNA, CRISPR, Genome engineering, Precise editing, *Drosophila*

1 Introduction

Prime editing is an exciting new CRISPR-based genome engineering technique to make small precise changes in the genome [1, 2]. Like many CRISPR tools, prime editing consists of two components – a prime editor (PE) protein and a prime editing guide RNA (pegRNA). Unlike wild-type CRISPR nucleases (e.g., Cas9, Cas12), PE is a fusion between a nickase form of Cas9 ($nCas9^{H840A}$) and Moloney murine leukemia virus (M-MLV) reverse transcriptase. PE2 uses an improved engineered M-MLV domain [1]. Like single guide RNAs (sgRNAs), pegRNAs contain a target-specific spacer and scaffold. pegRNAs direct PE2 to nick a protospacer target in the genome on the protospacer-adjacent motif (PAM)-containing strand. Importantly, pegRNAs have an unusual 3′ extension sequence that encodes the desired edit and is used by PE2 to reverse transcribe this edit into the nicked genomic site. As described, this is referred to as the PE2 system [1]. Nicking of the non-edited strand can bias DNA repair to increase the chances of retaining the edit [1]. This is referred to as the prime editing 3 (PE3) system, and involves co-delivery of a pegRNA and nicking sgRNA [1]. One disadvantage of the PE3 system is that

Christian Dahmann (ed.), *Drosophila: Methods and Protocols*,
Methods in Molecular Biology, vol. 2540, https://doi.org/10.1007/978-1-0716-2541-5_5,

simultaneous nicking on opposite strands by the pegRNA and sgRNA can generate double-strand breaks (DSB) and thus indels in the intervening sequence. To minimize this effect, Anzalone et al. [1] described a modification called PE3b, in which the nicking sgRNA can only bind a protospacer after the edited heteroduplex has formed.

Compared to precise editing using homology-directed repair (HDR) and CRISPR/Cas9, prime editing is associated with reduced genomic mistakes, such as insertion and deletions (indels) at on-target and off-target sites [1, 3, 4]. Compared to base editing [5], prime editing is more versatile, enabling all types of point mutations, small insertions (<100 bp), and small deletions (<100 bp) [1]. However, prime editing is currently limited to small edits, although some have adapted prime editing for larger deletions [6]. Another limitation is that prime editing efficiency varies depending on the cell type or edit [1, 7]. Due to its versatility and simplicity of use, prime editing has been quickly adopted by researchers working in diverse organisms. Therefore, it is likely that future versions of prime editing will address the issues above, such as prime editing systems with higher efficiencies or new functionalities.

Here, we outline protocols to perform prime editing in *Drosophila melanogaster* [8] (Fig. 1). First, to achieve a desired edit in the *Drosophila* genome, we outline the steps to design a pegRNA sequence, and optional nicking sgRNA sequence. Notably, this process is much more difficult than designing a traditional sgRNA, such as for gene knockout. Second, we detail bench work to molecularly clone plasmid DNA that is used to express pegRNAs and nicking sgRNAs in *Drosophila* cells. Third, we provide a guide to isolate pegRNA transgenic flies, cross these lines with PE2-expressing lines, analyze their progenies for editing, and isolate edited fly stocks.

2 Materials

2.1 Cloning pegRNA Expression Plasmids

1. Empty pegRNA expression plasmids: *pCFD3-NS* (Addgene 149545) or *pCFD5-NS* (Addgene 149546) (*see* **Note 1**).
2. BpiI with included 10× FastDigest Buffer (Fermentas, FD1014) (*see* **Note 2**).
3. FastAP (Fermentas, EF0651).
4. DNA Gel purification kit (Qiagen or similar).

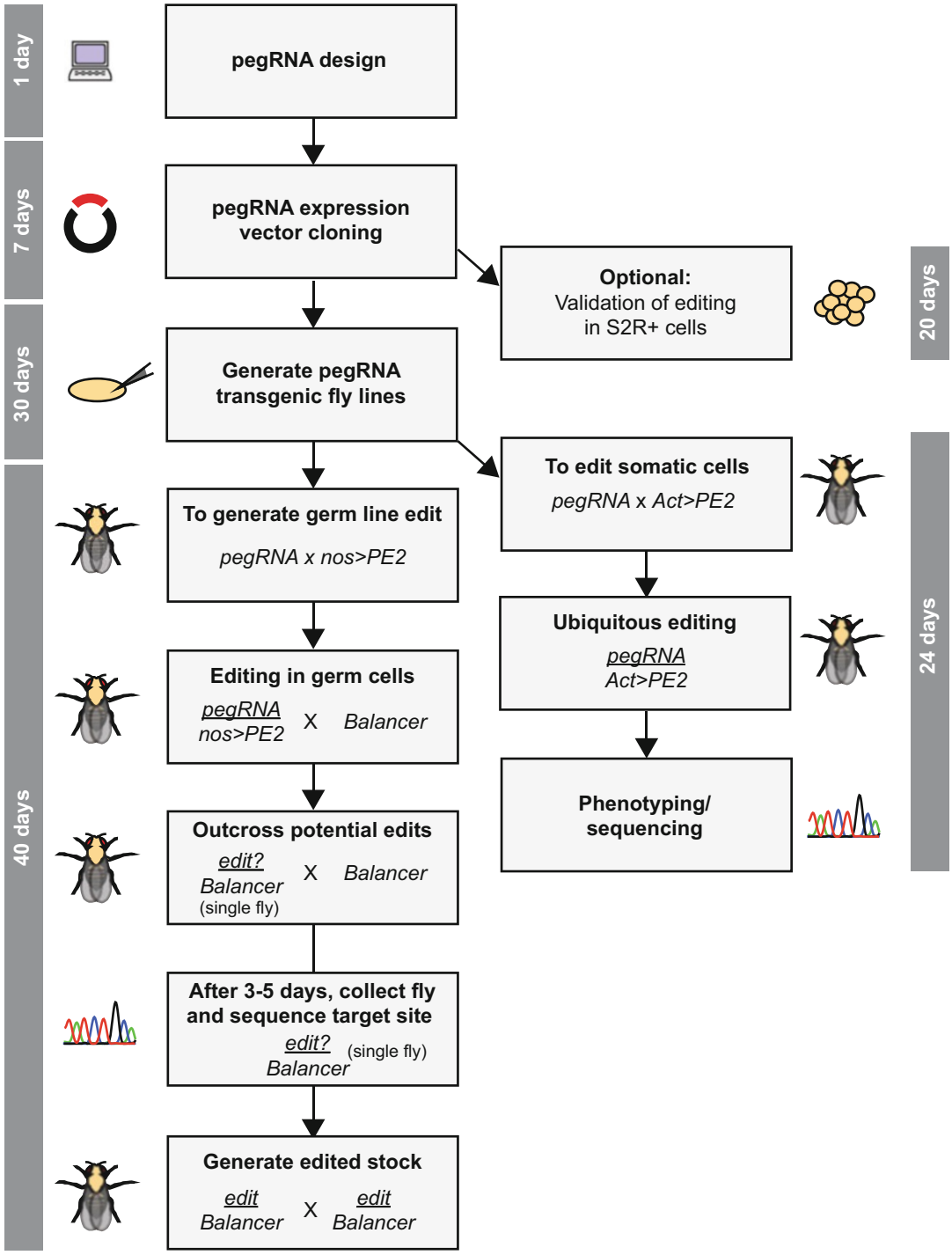


Fig. 1 Workflow to use prime editing to produce an edited fly stock by transgenic crossing. This workflow shows the steps from pegRNA design, through pegRNA molecular cloning, to transgenic crossing of pegRNA expressing lines, to generate an edited and balanced fly stock. Workflows that branch to the right (pegRNA testing in cultured S2R+ cells and editing in somatic tissues in vivo) are optional. Gray bars on left or right of the figure show the approximate timescales for sections of the workflow

5. For cloning into *pCFD3-NS* by ligation of annealed oligos:
   (a) Custom oligonucleotides (top and bottom).

```
>pegRNA_spacer_top
gtcgNNNNNNNNNNNNNNNNNNNN (Ns = 20bp spacer)
>pegRNA_spacer_bot
aaacNNNNNNNNNNNNNNNNNNNN (Ns = 20bp spacer)
>pegRNA_ext_top
gtgcNNNNNNNNNNNNNNNNNNNNNNNN (Ns = variable length extension)
>pegRNA_ext_bot
aaaaNNNNNNNNNNNNNNNNNNNNNNNN (Ns = variable length extension)
```

   (b) sgRNA scaffold oligonucleotides (top and bottom).

```
>pegRNA_scaffold_top
 gttt TAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTT
GAAAAAGTGGCACCGAGTCG
>pegRNA_scaffold_bot
 gcac CGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGC
TATTTCTAGCTCTA
```

   (c) T4 Polynucleotide Kinase (PNK) with included 10x T4 ligation buffer.
   (d) T4 DNA ligase with included 10x T4 ligation buffer.

6. For cloning into *pCFD3-NS* or *pCFD5-NS* by dsDNA overlap assembly:
   (a) Custom dsDNA fragment for *pCFD3-NS* cloning.

```
>dsDNA
AGACCTATTTTCAATTTAACGTCGNNNNNNNNNNNNNNNNNNNNGTTTTAGA
 GCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGT
 GGCACCGAGTCGGTGCNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
 NNNNNNNTTTTTTGCCTACCTGGAGCCTGAG
```

   Bold Ns = pegRNA spacer (20bp)
   Underlined Ns = pegRNA extension (variable length)
   (b) Custom dsDNA fragments for *pCFD5-NS* cloning.

```
>dsDNA fragment 1
```

CGGGTTCGATTCCCGGCCGATGCANNNNNNNNNNNNNNNNNNNNGTTTTA
GAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAA
GTGGCACCGAGTCGGTGCAACAAAGCACCAGTGGTCTAGTGGTAGAATAGT
ACCCTGCCACGGTACAGACC

Bold Ns = nicking spacer (20bp)

>dsDNA fragment 2
AACAAAGCACCAGTGGTCTAGTGGTAGAATAGTACCCTGCCACGGTACAGA
CCCGGGTTCGATTCCCGGCTGGTGCANNNNNNNNNNNNNNNNNNNNGTTT
TAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAA
AAGTGGCACCGAGTCGGTGCNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNTTTTTTGCCTACCTGGAGCCTGAG

Bold Ns = pegRNA spacer (20bp)
Underlined Ns = pegRNA extension (variable length)

(c) Gibson master mix or similar (e.g., NEBuilder HiFi DNA Assembly Master Mix).

7. Chemically competent *E. coli*.
8. Ampicillin or Carbenicillin antibiotic.
9. Agar plates and liquid media for *E. coli* growth.
10. Liquid culture tubes.
11. 37 °C shaking incubator.
12. Table-top microcentrifuge for 1.5–2 ml tubes.
13. Plasmid miniprep kit (Qiagen or similar).
14. Sequencing primer *pCFD3seqF* ACCTACTCAGCCAAGA GGC.
15. Optional colony PCR primers *pCFD3seqF* and *pCFD3seqR* ACCTACTCAGCCAAGAGGC, GCCGAGCACAATTGTCTAGAATGC.

2.2 S2R+ Cell Transfection (Optional)

1. Schneider's medium (21720-024; Thermo Fisher Scientific) with 10% fetal bovine serum and 50 U/mL penicillin-streptomycin.
2. S2R+ cells (*Drosophila* Genomics Resource Center stock #150 or similar) (*see* **Note 3**).
3. *pAC-Gal4* (Addgene 24344 or similar) (*see* **Note 4**).
4. *pUAS-PE2* (Addgene 149550) (*see* **Note 5**).
5. Transfection reagent (Effectene, 301427, Qiagen or similar).
6. Tissue culture-treated culture plates.
7. Genomic extraction reagents.

2.3 Transgenic Fly Work

1. QIAprep blue columns (Qiagen) or equivalent to repurify plasmid DNA.
2. Injection buffer (phosphate buffer): 100 μM NaPO4 pH 6.8, 5 mM KCl.
3. *PhiC31 integrase*; *attP* embryos for microinjection (e.g., from the Bloomington *Drosophila* Stock Center (BDSC), *yv nos-phiC31int*; *attP40* (BL25709), *yv nos-phiC31int*;; *attP2* (BL25710)) (can be outsourced to injection companies).
4. *yellow-*, *vermillion-* balancer stocks (e.g., *y,sc,v; Gla/CyO* BL35781, *y sc v; Dr e/TM3, Sb* BL32261).
5. For somatic editing: *w; Act-Gal4/CyO; UAS-PE2,w+ attP2 (Act>PE2)* (BL90977), or *w; UAS-PE2,w+ attP40; Tub--Gal4/TM6b* (*Tub>PE2*) (BL90974).
6. For germ line editing: *w; nos-Gal4, UAS-PE2,w+ attP40 (nos>PE2 II)* (BL91349), or *w;; nos-Gal4, UAS-PE2,w+ attP2 (nos>PE2 III)* (BL91350).
7. Balancer stocks for X, 2nd, 3rd, and 4th chromosomes available from BDSC.

2.4 Detection of Edits

1. Fly squishing buffer: 10 mM Tris–HCl, pH 8.2, 1 mM ethylenediaminetetraacetate, 25 mM NaCl, 200 μg/mL proteinase K.
2. Gene-specific forward and reverse primer to PCR amplify target site (*see* **Note 6**).
3. For Sanger sequencing: Taq polymerase.
4. For Amplicon sequencing: High fidelity polymerase (Q5, New England Biolabs; M0491L).
5. Gel purification kit (Qiagen or similar).
6. DNA fragment purification kit (QIAquick purple column or equivalent).

3 Methods

3.1 Locate and Annotate the Edit of Interest

Locate the genomic region of interest and bases to be edited (Fig. 2). Create a computer file containing the wild-type sequence, with ~1 kb flanking the edit site, and annotate the location of the desired edit (*see* **Note 7**). For example, a simple text file can be used to annotate the desired edit, such as using the following conventions from PrimeDesign [7].

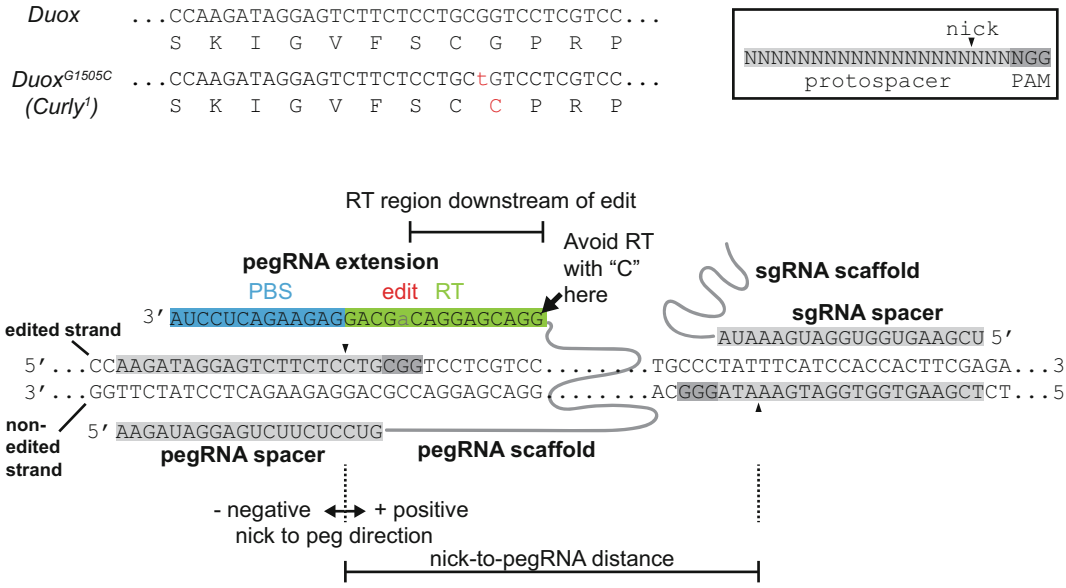


Fig. 2 Edit annotation, pegRNA design, and nicking sgRNA design. Top left shows an example annotation of the classic *Curly* mutation (*Duox*^{G1505C}). Top right shows an example of a genomic protospacer with associated NGG PAM and nick location. Bottom image shows an example pegRNA and nicking sgRNA to generate the *Curly*^{1} in a wild-type background. Overlaid on this diagram are important notes and visual aids

```
Wild-type:
...CTGGAAGGCTGGCGGCGCGTATTTGCCCATCGATC...

Substitution:
...CTGGAAGGCT(GGCG/TAAA)GCGCGTATTTGCCCATCGATC...

Insertion:
...CTGGAAGGCTGGCG(+TAAA)GCGCGTATTTGCCCATCGATC...

Deletion:
...CTGGAAGGCT(-GGCG)GCGCGTATTTGCCCATCGATC...

Combination:
...CTGGAAGGCT(-GGCG)GCGCGTAT(T/G)TGCCCATCGATC...
```

3.2 Select a Prime Editing System

Before designing pegRNAs and optional nicking sgRNAs, users will need to decide which prime editing system (e.g., PE2 vs PE3) is most appropriate (*see* **Note 8**) for their application (e.g., heritable edits vs somatic editing). PE3 results in higher editing efficiency in S2R+ cells (three edits), fly somatic cells (three edits), and fly germ cells (one edit) [8]. However, PE3 generates indels, which in some cases match or surpass the frequency of correct edits. While the PE3b system can address this issue, it has not been tested in *Drosophila*. Importantly, PE3b is limited to cases where the edit

overlaps a nicking sgRNA protospacer, and the PAM for the PE3b nicking sgRNA protospacer must be present in both the wild-type and edited sequence. When using prime editing in the germ line to make heritable edits, we recommend using the PE3 system. This is based on ~3× higher inheritance rate of an edit in *ebony* (*ebony*^{G111X}) using PE3 versus PE2 [8]. In somatic cells, indels generated by PE3 may complicate experimental results. For example, if a user intends to engineer a gain of function allele in somatic cells, PE3 will generate mosaic tissues that contain a mixture of cells with the edit and with indels that could disrupt gene function. Therefore, in general, we recommend using the PE2 system for prime editing in somatic cells. Prime editing is functional in S2R+ cells [8], but has not yet been used to make clonal S2R+ cell lines. However, considering the promising results using PE3 in mammalian cell culture systems, we recommend the use of the PE3 system to make edited S2R+ cell lines.

3.3 pegRNA/Nicking sgRNA Design

Manual design of pegRNAs and nicking sgRNAs is more complex than traditional sgRNA design. Thankfully, many free online and downloadable tools are available that simplify this process (Table 1). At the time of publication, we prefer PrimeDesign [7], but many tools have similar features (*see* **Note 9**). In nearly all cases, we recommend starting with the automatic design tools, such as those listed in Table 1. However, for those who require manual design, or want to evaluate sequences designed by online tools, we outline major design factors that influence editing efficiency below (*see* **Note 9**).

To date, there are no simple design parameters one can use to guarantee high prime editing efficiency. However, several groups have defined general design rules to help design pegRNAs and nicking sgRNAs [1, 3, 7, 9, 10] (*see* **Note 9**). As an example, we provide a hypothetical design strategy to engineer the classical *Curly* mutation (*Duox*^{G1505C}) [11] into a wild-type *Drosophila* genomic background (Fig. 2). For simplicity, we mostly refer to the rules established by Anzalone et al. and Hsu et al. [1, 7]:

1. Select a protospacer with a nick site that is 5′ relative to the intended edit site (Fig. 2) (*see* **Note 10**).
2. Select a PBS between 10 and 14 bp (Fig. 2) (*see* **Note 11**).
3. Select a RT region (containing the edit and flanking homology sequence) that is 10–20 bp for small edits (e.g., point mutations), or >20 bp for longer edits (Fig. 2) (*see* **Note 12**).
4. When using the PE3 system, select a nicking sgRNA protospacer on the non-edited strand, with the sgRNA nick site 50–100 bp away from the pegRNA nick site (Fig. 2) (*see* **Note 13**).

Table 1
pegRNA and nicking sgRNA design tools for prime editing

Tool name	References	Website
PrimeDesign	[7]	https://drugthatgene.pinellolab.partners.org/
PegIT	[19]	https://pegit.giehmlab.dk/
pegFinder	[20]	http://pegfinder.sidichenlab.org/
PnB Designer	[21]	https://fgcz-shiny.uzh.ch/PnBDesigner/
PE-Designer	[22]	http://www.rgenome.net/pe-designer/
PETAL	[23]	https://gt-scan.csiro.au/petal/
Pegassist	N/A	https://pegassist.app/
PlantPegDesigner	[10]	http://www.plantgenomeediting.net/
Prime Editing Design Tool	[24]	https://primeedit.nygenome.org/
PINE-CONE	[25]	https://github.com/xiaowanglab/PINE-CONE
Easy-Prime	[26]	http://easy-prime.cc/

3.4 Cloning pegRNA Expression Plasmids

Two empty expression plasmids are used to clone pegRNA sequences, *pCFD3-NS*, and *pCFD5-NS* (Fig. 3). *pCFD3-NS* is used to express a single pegRNA for the PE2 system. *pCFD5-NS* is used to co-express a nicking sgRNA in tandem with a pegRNA, for the PE3 system. *pCFD3-NS* and *pCFD5-NS* are derived from *pCFD3* [12] and *pCFD5* [13] and lack sgRNA scaffold sequence (NS = No Scaffold). A pegRNA cloned into *pCFD3-NS* is written as *pCFD3-PE-gene^{edit}*, and a sgRNA/pegRNA pair cloned into *pCFD5-NS* is written as *pCFD5-PE3-gene^{edit}*. *pCFD3-NS* and *pCFD5-NS* contain an *attB* sequence for site-specific integration into the fly genome and a *vermilion+* rescue transgene for transformant selection using eye color. For *pCFD3-NS* cloning, users have the option of choosing between two cloning methods – ligation of annealed oligos or insertion of double-stranded DNA (dsDNA) by Gibson cloning (or equivalent method) (*see* **Note 14**).

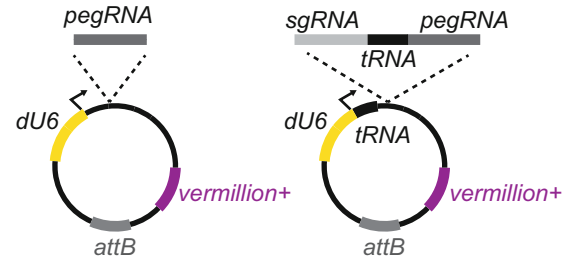
3.4.1 Cloning Annealed Oligos into pCFD3-NS

1. Design and order top and bottom oligonucleotides for the pegRNA spacer, scaffold, and pegRNA extension (Fig. 4) (*see* **Note 15**).
2. Resuspend oligos to 100 μM in H_20.
3. Phosphorylate and anneal each pair of oligos in a PCR machine (*see* **Note 16**):

   1 μl Top oligo (100 μM)

   1 μl Bottom oligo (100 μM)

   1 μl 10× T4 Ligation buffer



Name	*pCFD3-NS*	*pCFD5-NS*
Used to express	pegRNA	*pegRNA + sgRNA*
Cloning Method 1	Annealed oligos/ T4 ligase	2 dsDNAs/Gibson
Cloning Method 2	dsDNA/Gibson	
Addgene #	149545	149546
DGRC #	1528	1529

Fig. 3 pegRNA expression plasmids *pCFD3-NS* and *pCFD5-NS*. Top shows maps of *pCFD3-NS* and *pCFD5-NS*. *attB*, phiC31 integrase recombination sequence; *dU6*, *Drosophila U6:3* promoter; *vermillion+*, fly eye color transgenesis marker. Bottom table shows details distinguishing the two plasmids. DGRC Drosophila Genomics Resource Center

µl H_20

0.5 µl T4 PNK

10 µl total

PCR program: 37 °C for 30 min, 95 °C for 5 min, then ramp down to 25 °C at 5 °C/min.

4. Dilute annealed/phosphorylated oligos 1:200 in H_2O.
5. Digest/dephosphorylate *pCFD3-NS:*

   5 µg *pCFD3-NS*

   3 µl BpiI (cuts BbsI)

   3 µl FastAP

   6 µl 10× FastDigest Buffer

   Xµl H_20

   60 µl total

   Incubate at 37 °C for ~16 h.
6. Gel-purify the digested *pCFD3-NS* backbone (~6.2 kb).
7. Ligate annealed oligos into digested *pCFD3-NS* (*see* **Note 17**):

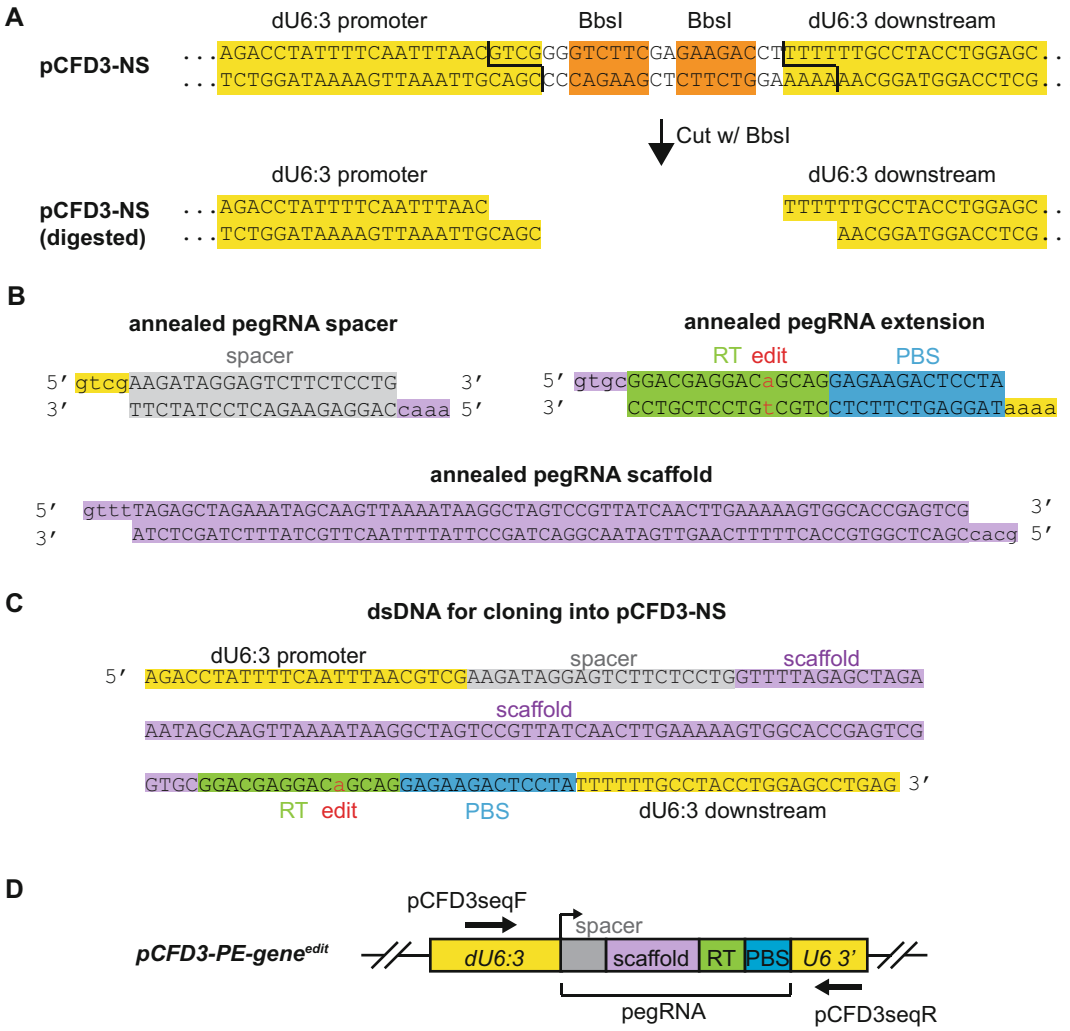


Fig. 4 pegRNA cloning into *pCFD3-NS*. (**a**) DNA sequence of *pCFD3-NS* cloning site before and after BbsI digestion. (**b**) Three pairs of annealed oligos that compose full pegRNA. (**c**) dsDNA encoding pegRNA and overlap sequence for Gibson cloning. (**d**) Final construct structure of pegRNA cloned into *pCFD3-NS*

Xμl digested *pCFD3-NS* (50 ng)
1 μl spacer diluted annealed oligo
1 μl scaffold diluted annealed oligo
1 μl 3′ extension diluted annealed oligo
1.5 μl 10× T4 Ligation Buffer
Xμl H_20
1 μl T4 DNA ligase
15 μl total
Incubate reaction at room temperature for 30 min.

8. Transform ligation into competent cells and grow colonies on LB agar ampicillin plates.
9. Culture colonies with LB + Ampicillin and sequence confirm plasmids (*see* **Note 18**).

   pCFD3seqF ACCTACTCAGCCAAGAGGC.

3.4.2 Cloning dsDNA into pCFD3-NS

1. Design and order dsDNA containing the full-length pegRNA with homology arms that match cut *pCFD3-NS* (Fig. 4) (*see* **Note 19**)
2. Digest/dephosphorylate/gel-purify *pCFD3-NS* (*see* Subheading 3.4.1 **steps 5** and **6**)
3. Mix components in a Gibson assembly reaction:

   Xμl digested *pCFD3-NS* (50 ng)

   Xμl dsDNA fragment (5 ng)

   2.5 μl Gibson master mix

   Xμl H20

   5 μl total

   Incubate reaction at 50 °C for 30 min.
4. Transform/culture bacteria and sequence plasmids (*see* Subheading 3.4.1 **steps 8** and **9**).

3.4.3 Cloning dsDNAs into pCFD5-NS

1. Design and order dsDNA fragments 1 and 2 (Fig. 5). dsDNA fragment 1 contains the nicking sgRNA, and dsDNA fragment 2 contains the pegRNA. dsDNA fragments contain homology arms that match each other and digested *pCFD5-NS* (Fig. 5) (*see* **Note 19**).
2. Digest/dephosphorylate/gel-purify *pCFD5-NS* (same as Subheading 3.4.1 **steps 5** and **6**, except using *pCFD5-NS*)
3. Mix components in a Gibson assembly reaction:

   Xμl digested *pCFD5-NS* (50 ng)

   Xμl dsDNA fragment 1 (5 ng)

   Xμl dsDNA fragment 2 (5 ng)

   2.5 μl Gibson master mix

   Xμl H20

   5 μl total

   Incubate reaction at 50 °C for 30 min.
4. Transform/culture bacteria and sequence plasmids (*see* Subheading 3.4.1 **steps 8** and **9**).

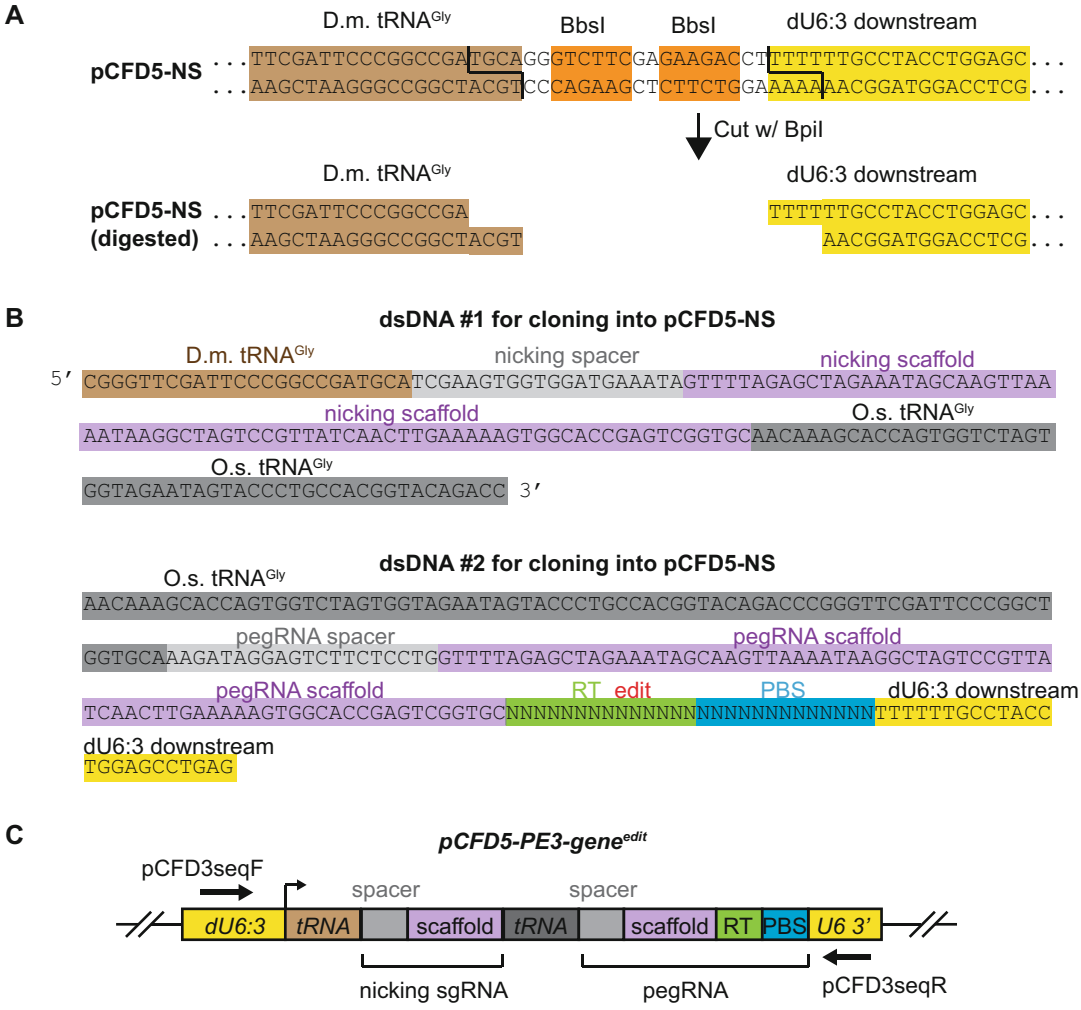


Fig. 5 pegRNA and sgRNA cloning into *pCFD5-NS*. (**a**) DNA sequence of *pCFD5-NS* cloning site before and after BbsI digestion. (**b**) Two dsDNA fragments, #1 and #2, that compose the nicking sgRNA, intervening tRNA, pegRNA, and overlap sequences for Gibson cloning. *D.m*, *Drosophila melanogaster*; *O.s.*, *Oryza sativa*. (**c**) Final construct structure of pegRNA and sgRNA cloned into *pCFD5-NS*

3.5 Validation of Editing in S2R+ Cells (Optional)

Prime editing efficiency at different target sites can be unpredictable [1, 7]. To address this issue, different groups have tested multiple pegRNA designs in cultured cells to screen for those with the highest editing efficiency. Here, we outline culture methods used in Bosch et al. 2021 [8] to deliver prime editing components to S2R+ cells.

1. Transfect the following plasmids into S2R+ cells, following the Effectene manufacturer instructions: pegRNA plasmid DNA, *pUAS-PE2*, and *pAct-Gal4* (*see* **Note 20**).
2. Grow cells at 25 °C in a humid box for 4 days (*see* **Note 21**).

3. Resuspend cells, transfer to a 1.5 ml centrifuge tube, and pellet for 10 min at 100 g.
4. Remove the supernatant, replace with 1 ml 1×PBS, gently invert the centrifuge tube, and repeat centrifugation in **step 3**.
5. Remove the supernatant, add 100 µl QuickExtract reagent, and resuspend cell pellet by pipetting up and down.
6. Incubate solution at 65 °C for 15 min and then 98 °C for 2 min.
7. Store at −20 °C.
8. *See* Subheading 3.7.3 for detection of edits by amplicon sequencing.

3.6 pegRNA Transgenic Flies and Crossing

This protocol describes prime editing by transgenic crosses between pegRNA lines and PE2 expressing flies. Embryo injection of pegRNA plasmids or synthetic pegRNAs is also possible, though with lower editing efficiency [8].

3.6.1 Isolation of pegRNA Transgenic Fly Lines

1. Re-purify plasmid DNA (*pCFD3-PE-gene^{edit}* or *pCFD5-PE3-gene^{edit}*) by column purification (QIAprep blue columns or equivalent) and elute in injection buffer at 200 ng/µl (at least 10 µl total) (*see* **Note 22**).
2. Inject plasmid DNA into at least 50 *y v nos-phiC31int; attP40* or *y v nos-phiC31int;; attP2* embryos (*see* **Note 23**).
3. Raise injected embryos at 18–25 °C.
4. Outcross injected adults (G0) to *yellow-, vermillion-* balancer stocks (e.g., *y sc v; Gla/CyO*).
5. Collect *vermillion+* G1 male progeny and repeat outcross in **step 4** (*see* **Note 24**).
6. Intercross *vermillion+/balancer* G2 male and female progeny to generate a pegRNA-expressing fly stock.

3.6.2 Prime Editing by Transgenic Crossing

For ubiquitous somatic prime editing, cross pegRNA transgenic flies to *Act>PE2* and raise at 29 °C (*see* **Note 25**). If appropriate, analyze progeny for phenotypes associated with the edit.

For germ line editing, cross pegRNA transgenic flies to *nos>PE2*. For editing genes on chromosomes X, III, and IV, use *nos>PE2* II; for editing genes on chromosomes X, II, and IV, use *nos>PE2* III.

1. Cross pegRNA line to *nos>PE2* and incubate crosses at 29 °C.
2. Every 24 h, flip crosses to a new food vial
3. Heat shock F1 progeny in food vials in a 37 °C water bath for 1 h at 24 h after egg deposition (AED), 48 h AED, 72 h AED, 96 h AED, and 120 h AED.

4. Cross ~8–10 pooled F1 male adult progeny to 10–15 females from an appropriate balancer stock (e.g., *y w;;TM3/TM6b*) (*see* **Note 26**).
5. Set up ~100 individual crosses between single F2 adult progeny and ~3–5 females of the balancer stock (*see* **Note 27**).
6. After 3–5 days of crossing, transfer single F2 adults to PCR strip tubes or a 96-well plate, freeze the tubes or plate at −20 °C, and proceed to Subheading 3.7.4 for genotyping.
7. After identifying a line with the intended edit, intercross edit/balancer males and females to generate a stable fly line.

3.7 Detection of Edits

To detect an edit, PCR is first used to amplify the target site, then the PCR fragment is sequenced by Sanger sequencing or amplicon sequencing. Using genomic sequence as reference, users should design a forward and reverse primer that flanks the edit site using programs such as Primer3 [14] or equivalent (*see* **Note 28**).

3.7.1 Primer Design

For Sanger sequencing, design primers such that the fragment size is between 200 bp and 1000 bp, ensuring that the edit is between 50 bp and 400 bp away from at least one end of the fragment.

For amplicon sequencing, design primers such that the fragment size is at least 200 bp. The maximum fragment size is determined as 10 bp less than double the read length of the sequencer. For example, Genewiz Amplicon-EZ uses 2 × 250 bp Illumina sequencing; therefore, the fragment size should be 200 bp–490 bp (*see* **Note 29**). These size ranges apply to both the wild-type and edited amplicon (*see* **Note 30**). If using a nicking sgRNA (PE3 system), ensure that both the edit site and the nicking sgRNA target site are contained in the PCR fragment.

3.7.2 Genomic DNA Extraction from S2R+ Cells or Single Flies

To collect genomic DNA from S2R+ cells, use QuickExtract and follow the manufacturer's instructions (*see* also Subheading 3.5). For collecting genomic DNA from single flies, squish a live or frozen single fly in a PCR strip tube, add 50 µl fly squishing buffer, incubate at 37 °C for 30 min and 95 °C for 2 min, and store genomic DNA at −20 °C.

3.7.3 Amplicon Sequencing of S2R+ Cells or Somatic Editing Flies

1. PCR amplify the target site using primers designed in Subheading 3.7.1 using Q5 polymerase in a 50 µl reaction.
2. Run 10 µl PCR product on a 1% agarose gel and confirm product size and lack of primer dimer.
3. Column purify (QIAquick purple column or equivalent) the remaining 40 µl PCR product and elute in 35 µl H20.
4. Submit purified DNA fragment for amplicon sequencing (e.g., Genewiz Amplicon EZ or equivalent).

5. Download next-generation sequencing (NGS) reads in .fastq format.
6. Use CRISPResso2 [15] to calculate the percent of reads with the precise edit, and the percent of reads with an indel (*see* **Note 31**).

3.7.4 Sanger Sequencing of Inherited Edits in Fly Lines

1. PCR amplify the target site using primers designed in Subheading 3.7.1 using Taq polymerase in a 20 μl reaction.
2. Run the 20 μl reaction on a 1% agarose gel, gel-purify the PCR fragment, and elute in 35 μl H20.
3. Submit purified PCR fragment for sequencing using one or both of the primers used for PCR amplification.
4. Align chromatogram trace file (e.g., .abi format) to the reference genomic sequence using DNAstar SeqMan or similar software.

4 Notes

1. *pCFD3-NS* (1528) and *pCFD5-NS* (1529) are also available from the Drosophila Genomic Resource Center.
2. In our experience, digestions using Fermentas BpiI (an isoschizomer of BbsI) perform better than BbsI-HF from NEB.
3. Do not use S2R+ cells that express Cas9 or a Cas9 variant, which could complex with pegRNAs or sgRNAs.
4. In Bosch et al. [8], the constitutive Gal4 plasmid (*pAct-Gal4*) used was an unpublished reagent from Y. Hiromi, National Institute of Genetics, Mishima, Japan. However, other Gal4-expressing plasmids should work in S2R+ cells. For example, *pAC-Gal4* #24344 from Addgene is the most suitable publicly available replacement. Alternatively, pMT-GAL4 #1042 from the DGRC could be also be used with copper sulfate added to the media.
5. *pUAS-PE2* (1527) is also available from the Drosophila Genomic Resource Center.
6. Partial illumina adapter sequences can be added on the 5′ end of primers to save on amplicon sequencing costs. For example, append the following sequences to primers for Amplicon-EZ at Genewiz.

   Forward: acactctttccctacacgacgctcttccgatc tNNNNNNNNNNNNNNNNNNN

   Reverse: gactggagttcagacgtgtgctcttccg atctNNNNNNNNNNNNNNNNNNN

7. To help annotate and edit genomic sequence, we highly recommend using sequence annotation software, such as Snapgene, VectorNTI, and Lasergene.
8. Like other CRISPR/Cas9 approaches (e.g., base editing), prime editing is a technique that will evolve and improve over time. We highly recommend users to stay up to date on new prime editing approaches before deciding which system most appropriate for their application. For example, improved versions of prime editors may have higher editing efficiencies, altered PAM specificities, lower frequency of indels using the PE3 system, or other desirable qualities.
9. pegRNA and nicking sgRNA design rules will evolve as researchers better define factors for successful editing. It is likely that, over time, design software will be updated to integrate these new rules. In addition, it is likely that completely new design software created. Researchers should check when the software listed in Table 1 was last updated, as well as search for new software tools not included in Table 1.
10. Avoid spacer sequences with T homopolymers because these can lead to premature termination of pegRNA or sgRNA transcription.

    If no protospacers are near the edit (e.g., <20 bp from protospacer nick to edit), the RT length can be increased to compensate. Programs, such as PrimeDesign, use sliders to increase the maximum RT length. RT lengths up to 34 bp have been shown to be functional [1], so it is possible that even longer RT lengths may work, but there are potential downsides to using longer RT regions. For example, longer RT regions appear to have lower editing efficiency [9], perhaps due to unpredictable pegRNA folding. Also, longer RT regions may prevent oligo synthesis if using oligo annealing to construct pegRNA expression plasmids (*see* Subheading 3.4.1). Future PE2 variants that recognize alternative PAM sequences could increase the number of possible pegRNA protospacers [16].

    High-throughput evaluation of prime editing events has determined that protospacers with high predicted cutting scores (double-strand breaks with Cas9) are also more effective at prime editing with PE2 [9]. Therefore, it may be beneficial to choose protospacers with higher predicted efficiency scores. In contrast, prime editing inherently has low off-target effects [1, 3, 4]. Therefore, it is currently unclear if choosing pegRNA protospacers with lower predicted off-target effects will benefit editing experiments. Several existing tools can be used to predicted on-target and off-target effects of protospacers in *Drosophila*, such as the DRSC "Find CRISPR gRNA search" (www.flyrnai.org/crispr3/web/) [17] or "Target Finder" at flyCRISPR (flycrispr.org/target-finder/) [18].

11. Optimal PBS length yet cannot be determined computationally. 10–14 bp is a general starting point, but can be extended to 10–17 bp [7]. The GC content of the PBS is an important determinant of prime editing efficiency. Generally, a PBS with high GC content (>50%) should be shorter, and low GC content (<50%) should be longer [1, 7]. One helpful way to choose PBS length, other than relying on automatic prediction software (Table 1), is to select a PBS length with an optimal melting temperature. One group showed that 37.4 °C is optimal in mammalian cells [7], whereas another group suggests <35 °C [9]. 30 °C appears optimal in plants [10]. Optimal PBS melting temperature for Drosophila is unknown. Future design rules may improve prediction of optimal PBS length.
12. Follow the design rules from PrimeDesign [7] to specify the length of the homology downstream of the edit (Fig. 2): 1 bp edit – 10 bp homology downstream, 2–5 bp edit – 15 bp homology downstream, 6–10 bp edit – 20 bp homology downstream, 11–15 bp edit – 25 bp homology downstream, and >15 bp edit – 35 bp homology downstream.Prime editing efficiency is higher when a PAM-disrupting mutation (i.e., in either/both GG in the NGG motif) is encoded in the RT region [1, 9]. If editing in genomic coding sequence, use a synonymous PAM-disrupting mutation.

    Avoid pegRNA extensions that start with a "C" (*see* Fig. 2).
13. Avoid choosing a nicking protospacer that is too close to the pegRNA protospacer (e.g., <50 bp away), otherwise this may increase the frequency of indels [1]. Conversely, avoid choosing a nicking protospacer that is too far from the pegRNA protospacer (e.g., >100 bp away), otherwise editing efficiency will not be improved by the nicking sgRNA.

    Anzalone et al. [1] noted that editing efficiency is slightly higher when using nicking protospacer in the positive (3′) direction, rather than the negative (5′) direction [1] (Fig. 2).

    To use the PE3b approach, a nicking protospacer must be selected that is only present after the edit is installed. Importantly, the PE3b nicking protospacer must use a PAM that is present in both the wild-type and edited sequence.
14. Oligos are generally cheaper than dsDNA fragments, but Gibson assembly is generally faster and simpler. Cloning efficiency is roughly similar between these two methods but has not been compared extensively.
15. pegRNA and nicking sgRNA design software (e.g., PrimeDesign) may output sequences with overhangs for cloning into mammalian vectors. Simply exclude these sequences for oligo/dsDNA design.

16. Oligos can also be annealed and/or phosphorylated by a company (e.g., IDT). The annealing protocol is derived from the *pCFD3* sgRNA cloning protocol [12].
17. We recommend users also prepare a negative control reaction that omits the insert. Transforming this reaction into bacteria, in parallel to the experimental reaction, will give the number of background colonies.
18. Before culturing colonies, colony PCR can be used to identify candidate pegRNA plasmids:

    pCFD3seqF ACGTTTTATAACTTATGCCCCTAAG

    pCFD3seqR GCCGAGCACAATTGTCTAGAATGC

    For *pCFD3-NS* cloning:

    Uncut backbone = 490 bp

    Correct insert = ~638 bp (depends on pegRNA length)

    For *pCFD5-NS* cloning:

    Uncut backbone = 587 bp

    Correct insert = ~846 bp (depends on pegRNA length).
19. If needed, dsDNA homology arms can be extended longer (~100 bp each). This can help decrease sequence complexity when ordering dsDNA from companies.
20. We transfect in 24-well plates (500 μl culture volume per well), using no more than 500 ng plasmid per well. Make sure to include a negative, non-transfected control. To visualize transfection efficiency, a UAS-GFP plasmid can be included. Transfect each plasmid in a 1:1:1 ng ratio.
21. A humid box can be made using water-soaked paper towels inside a sealable container larger than the tissue culture plates.
22. Miniprepped plasmid DNA can be toxic to injected embryos; therefore, we use a second column purification to remove contaminants. If embryo toxicity is too high, users can purify plasmid DNA using a Qiagen midiprep kit or equivalent.
23. We observe higher integration efficiencies using *yv nos-phiC31int; attP40* compared to *yv nos-phiC31int;; attP2.*
24. Optional: Sequence confirm pegRNA transgenic fly lines by single fly squishing (*see* Subheading 3.7.2), PCR amplification using primers pCFD3seqF/R, and sequencing using the same primers.
25. *Tub>PE2* results in slightly higher editing efficiency [8], but *Tub>PE2* is a less healthy stock. Progeny raised at 29 °C will result in slightly higher editing efficiency than progeny raised at 25 °C [8].

26. We recommend using a balancer line that is *yellow- white-* to allow easy selection against the three transgenes present in F1 flies. *nos-Gal4* and *UAS-PE2* are marked by *white+*. *attP40* or *attP2*, which contains the *U6-pegRNA* transgene, is marked by *yellow+*.
27. For edits on chromosome X that might be lethal in males, use F2 females for crosses (e.g., edit?/FM7 female × FM7/Y male).
28. It is important to identify primer pairs that result in robust PCR amplification with minimal primer dimer; therefore, we recommend testing at least two pairs for each target site.
29. Illumina adapter sequences can be added to the forward and reverse primers to facilitate downstream library assembly. For example, these adapters are used by Genewiz Amplicon-EZ and reduce the price of amplicon sequencing:

```
>forward
acactctttccctacacgacgctcttccgatctNNNNNNNNNNNNNNNNNNNN
>reverse
gactggagttcagacgtgtgctcttccgatctNNNNNNNNNNNNNNNNNNNN
```

30. If the edit is an insertion, then the edited PCR fragment will be longer than the wild-type PCR fragment. It is important that the predicted length of the edited PCR fragment also conform to the fragment size rules.
31. To calculate the percent of reads with the precise edit, we used the following parameters: "–prime_editing_pegRNA_spacer_seq," "–prime_editing_pegRNA_extension_seq," "–prime_editing_pegRNA_scaffold_sequence," "–ignore_substitutions," and "–discard_indel_reads." The precise editing frequency was calculated from "CRISPResso_quantification_of_editing_frequency.txt" for the "prime-edited" amplicon, by dividing the number of reads found under these headers – "unmodified"/"reads aligned all amplicons." To determine the percent of reads with indels we ran CRISPResso2 with standard settings and the "–ignore_substitutions" parameter. The indel frequency was calculated from "CRISPResso_quantification_of_editing_frequency.txt" as the number reads modified/number reads_aligned. When using the PE3 system, we recommend specifying a quantification window "-qwc" that encompasses the region between the pegRNA and nicking sgRNA that spans the −6 position relative to the pegRNA PAM to the −6 position relative to the sgRNA PAM.

Acknowledgments

The authors thank Gabriel Birchak for discussions and S2R+ work, Rich Binari for general lab assistance and help with fly, the TRiP and *Drosophila* RNAi Screening Center for help generating transgenic flies, Ram Viswanatha for general discussions, Gillian Millburn for discussions on pegRNA transgene nomenclature, Cathryn Murphy for general lab assistance, Cooper Cavers for help isolating transgenic flies, Jorden Rabasco for help with molecular cloning, Andrew Anzalone for advice with synthetic pegRNAs, and Ben Ewen-Campen, Jonathan Zirin, and Thai LaGraff for cloning help. J.A.B. was supported by the Damon Runyon Foundation (DRG-2258-16) and a "Training Grant in Genetics" T32 Ruth Kirschstein-National Research Service Award institutional research training grant funded through the NIH/National Institute of General Medical Sciences (T32GM007748). This work was also supported by NIH Grant P41GM132087. N.P. is an Investigator of the Howard Hughes Medical Institute.

References

1. Anzalone AV, Randolph PB, Davis JR, Sousa AA, Koblan LW, Levy JM, Chen PJ, Wilson C, Newby GA, Raguram A, Liu DR (2019) Search-and-replace genome editing without double-strand breaks or donor DNA. Nature 576(7785):149–157. https://doi.org/10.1038/s41586-019-1711-4
2. Scholefield J, Harrison PT (2021) Prime editing – an update on the field. Gene Ther 28(7):396–401. https://doi.org/10.1038/s41434-021-00263-9
3. Kim DY, Moon SB, Ko JH, Kim YS, Kim D (2020) Unbiased investigation of specificities of prime editing systems in human cells. Nucleic Acids Res 48(18):10576–10589. https://doi.org/10.1093/nar/gkaa764
4. Jin S, Lin Q, Luo Y, Zhu Z, Liu G, Li Y, Chen K, Qiu JL, Gao C (2021) Genome-wide specificity of prime editors in plants. Nat Biotechnol 39(10):1292–1299. https://doi.org/10.1038/s41587-021-00891-x
5. Rees HA, Liu DR (2018) Base editing: precision chemistry on the genome and transcriptome of living cells. Nat Rev Genet 19(12):770–788. https://doi.org/10.1038/s41576-018-0059-1
6. Choi J, Chen W, Suiter CC, Lee C, Chardon FM, Yang W, Leith A, Daza RM, Martin B, Shendure J (2021) Precise genomic deletions using paired prime editing. bioRxiv 40(2):218–226. https://doi.org/10.1101/2020.12.30.424891
7. Hsu JY, Grunewald J, Szalay R, Shih J, Anzalone AV, Lam KC, Shen MW, Petri K, Liu DR, Joung JK, Pinello L (2021) PrimeDesign software for rapid and simplified design of prime editing guide RNAs. Nat Commun 12(1):1034. https://doi.org/10.1038/s41467-021-21337-7
8. Bosch JA, Birchak G, Perrimon N (2021) Precise genome engineering in Drosophila using prime editing. Proc Natl Acad Sci U S A 118(1):e2021996118. https://doi.org/10.1073/pnas.2021996118
9. Kim HK, Yu G, Park J, Min S, Lee S, Yoon S, Kim HH (2021) Predicting the efficiency of prime editing guide RNAs in human cells. Nat Biotechnol 39(2):198–206. https://doi.org/10.1038/s41587-020-0677-y
10. Lin Q, Jin S, Zong Y, Yu H, Zhu Z, Liu G, Kou L, Wang Y, Qiu JL, Li J, Gao C (2021) High-efficiency prime editing with optimized, paired pegRNAs in plants. Nat Biotechnol 39(8):923–927. https://doi.org/10.1038/s41587-021-00868-w
11. Hurd TR, Liang FX, Lehmann R (2015) Curly encodes dual oxidase, which acts with Heme peroxidase curly Su to shape the adult

drosophila wing. PLoS Genet 11(11): e1005625. https://doi.org/10.1371/journal.pgen.1005625

12. Port F, Chen HM, Lee T, Bullock SL (2014) Optimized CRISPR/Cas tools for efficient germline and somatic genome engineering in drosophila. Proc Natl Acad Sci U S A 111(29): E2967–E2976. https://doi.org/10.1073/pnas.1405500111
13. Port F, Bullock SL (2016) Augmenting CRISPR applications in drosophila with tRNA-flanked sgRNAs. Nat Methods 13(10): 852–854. https://doi.org/10.1038/nmeth.3972
14. Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, Rozen SG (2012) Primer3--new capabilities and interfaces. Nucleic Acids Res 40(15):e115. https://doi.org/10.1093/nar/gks596
15. Clement K, Rees H, Canver MC, Gehrke JM, Farouni R, Hsu JY, Cole MA, Liu DR, Joung JK, Bauer DE, Pinello L (2019) CRISPResso2 provides accurate and rapid genome editing sequence analysis. Nat Biotechnol 37(3): 224–226. https://doi.org/10.1038/s41587-019-0032-3
16. Kweon J, Yoon JK, Jang AH, Shin HR, See JE, Jang G, Kim JI, Kim Y (2021) Engineered prime editors with PAM flexibility. Mol Ther 29(6):2001–2007. https://doi.org/10.1016/j.ymthe.2021.02.022
17. Housden BE, Valvezan AJ, Kelley C, Sopko R, Hu Y, Roesel C, Lin S, Buckner M, Tao R, Yilmazel B, Mohr SE, Manning BD, Perrimon N (2015) Identification of potential drug targets for tuberous sclerosis complex by synthetic screens combining CRISPR-based knockouts with RNAi. Sci Signal 8(393):rs9. https://doi.org/10.1126/scisignal.aab3729
18. Gratz SJ, Ukken FP, Rubinstein CD, Thiede G, Donohue LK, Cummings AM, O'Connor-Giles KM (2014) Highly specific and efficient CRISPR/Cas9-catalyzed homology-directed repair in Drosophila. Genetics 196(4): 961–971. https://doi.org/10.1534/genetics.113.160713
19. Anderson MV, Haldrup J, Thomsen EA, Wolff JH, Mikkelsen JG (2021) pegIT – a web-based design tool for prime editing. Nucleic Acids Res 49(W1):W505–W509. https://doi.org/10.1093/nar/gkab427
20. Chow RD, Chen JS, Shen J, Chen S (2021) A web tool for the design of prime-editing guide RNAs. Nat Biomed Eng 5(2):190–194. https://doi.org/10.1038/s41551-020-00622-8
21. Siegner SM, Karasu ME, Schroder MS, Kontarakis Z, Corn JE (2021) PnB Designer: a web application to design prime and base editor guide RNAs for animals and plants. BMC Bioinformatics 22(1):101. https://doi.org/10.1186/s12859-021-04034-6
22. Hwang GH, Jeong YK, Habib O, Hong SA, Lim K, Kim JS, Bae S (2021) PE-Designer and PE-Analyzer: web-based design and analysis tools for CRISPR prime editing. Nucleic Acids Res 49(W1):W499–W504. https://doi.org/10.1093/nar/gkab319
23. Adikusuma F, Lushington C, Arudkumar J, Godahewa G, Chey YCJ, Gierus L, Pilitz S, Reti D, Wilson LOW, Bauer DC, Thomas PQ (2021) Optimized nickase- and nuclease-based prime editing in human and mouse cells. Nucleic Acids Res 49(18):10785–10795
24. Morris JA, Rahman JA, Guo X, Sanjana NE (2020) Automated design of CRISPR prime editors for thousands of human pathogenic variants. bioRxiv. https://doi.org/10.1101/2020.05.07.083444
25. Standage-Beier K, Tekel SJ, Brafman DA, Wang X (2021) Prime editing guide RNA design automation using PINE-CONE. ACS Synth Biol 10(2):422–427. https://doi.org/10.1021/acssynbio.0c00445
26. Li Y, Chen J, Tsai SQ, Cheng Y (2021) Easy-prime: a machine learning-based prime editor design tool. Genome Biol 22(1):235. https://doi.org/10.1186/s13059-021-02458-0

Chapter 6

CRISPR-/Cas9-Mediated Precise and Efficient Genome Editing in *Drosophila*

Kevin G. Nyberg and Richard W. Carthew

Abstract

The CRISPR/Cas9 system provides the means to make precise and purposeful modifications to the genome via homology-directed repair (HDR). In *Drosophila*, a wide variety of tools provide flexibility to achieve these ends. Here, we detail a method to generate precise genome edits via HDR that is efficient and broadly applicable to any *Drosophila* stock or species. sgRNAs are first tested for their cleavage efficiency by injecting embryos with Cas9/sgRNA ribonucleoproteins using commercially available Cas9 protein. Using an empirically validated sgRNA, HDR is performed using a donor repair plasmid that carries two transformation markers. A fluorescent eye marker that can be seamlessly removed using PiggyBac transposase marks integration of the repair sequence. A counter-selection marker that produces small rough eyes via RNAi against *eyes absent* is used to screen against imprecise HDR events. Altogether, the enhancements implemented in this method expand the ease and scope of achieving precise CRISPR/Cas9 genome edits in *Drosophila*.

Key words *Drosophila*, CRISPR, Genome editing, Homology-directed repair, Germline

1 Introduction

The discovery and application of the CRISPR/Cas9 system has revolutionized the biological sciences, offering researchers unprecedented abilities to manipulate the genome in versatile ways [1]. In *Drosophila*, CRISPR/Cas9 is widely used to induce targeted double-stranded breaks (DSBs) and generate both imprecise INDELs via non-homologous end joining (NHEJ) and precisely engineered edits via homology-directed repair (HDR) [2–11].

A wealth of tools has been developed that provide *Drosophila* researchers flexibility in the design and construction of their edited flies. Widely available and efficient germline-specific transgenic Cas9 fly lines are popular, but Cas9 has also been introduced into the fly by injecting either expression plasmids or Cas9 mRNA/protein directly into pre-cellularized embryos [2–10, 12]. To pro-

Christian Dahmann (ed.), *Drosophila: Methods and Protocols*,
Methods in Molecular Biology, vol. 2540, https://doi.org/10.1007/978-1-0716-2541-5_6,

duce a heritable modification of the germline genome, Cas9 is complexed with a chimeric single-guide RNA (sgRNA), which fuses the targeting crRNA and the scaffolding tracrRNA [13, 14]. sgRNAs are often introduced via expression plasmids but can also be transcribed in vitro and co-injected with Cas9 protein as a ribonucleoprotein (RNP) complex [3–5, 8, 12]. For editing via HDR, both single-stranded oligonucleotides and double-stranded plasmids have been used as donor repair templates [3, 7, 9].

Still, achieving successful genome editing using CRISPR/Cas9 in *Drosophila* is not always straightforward, especially when using an HDR approach. The germline Cas9 lines reliably induce DSBs, but they also constrain the genetic backgrounds of the G0 founder flies. This may complicate experiments that rely on maintaining a specific genetic background, especially studies of quantitative polygenic traits. It also limits editing to the few *Drosophila* species with such transgenic lines. sgRNAs vary widely in cleavage efficiency, and computational predictions of sgRNA efficiency derived from studies in vertebrate systems do not always successfully predict efficiency of sgRNAs in *Drosophila* embryos [12, 15]. A poor sgRNA will doom a CRISPR/Cas9 effort from the start and will set a project back several months. Finally, because precise integration of the donor template via HDR is a relatively rare event and imprecise HDR is not uncommon, multiple visible transformation markers are advised to avoid laborious screening via PCR and sequencing. Fluorescent eye markers are widely used as markers for successful HDR integration, while a mini-*white* gene included in the donor plasmid backbone has been used to screen against imprecise HDR events [16]. The use of mini-*white* as a counter-selection marker, however, necessitates performing the CRISPR/Cas9 edits in a *white* mutant background.

With these challenges in mind, we detail here a widely applicable protocol for generating precise *Drosophila* genome edits via HDR with relative ease and efficiency [12] (Fig. 1). It also affords the ability to perform edits in the genomes of any *Drosophila* stock, including non-*melanogaster Drosophila* species. The protocol uses commercially available Cas9 protein complexed with in vitro transcribed sgRNA, which together can reliably produce DSBs. Injection of Cas9/sgRNA RNPs into embryos enables the researcher to rapidly determine the cleavage efficiency of candidate sgRNAs. This ensures that only highly efficient sgRNAs are then used for HDR-mediated editing, and greatly increases the probability of generating an expedient and successful HDR edit. Last, we have combined an eye-specific DsRed selection marker for HDR events with a counter-selection marker against those events that are imprecise [12]. The counter-selection marker is broadly applicable to all *Drosophila* stocks and species, since it produces a small rough adult eye by triggering RNAi against the *eyes absent* (*eya*) gene. The

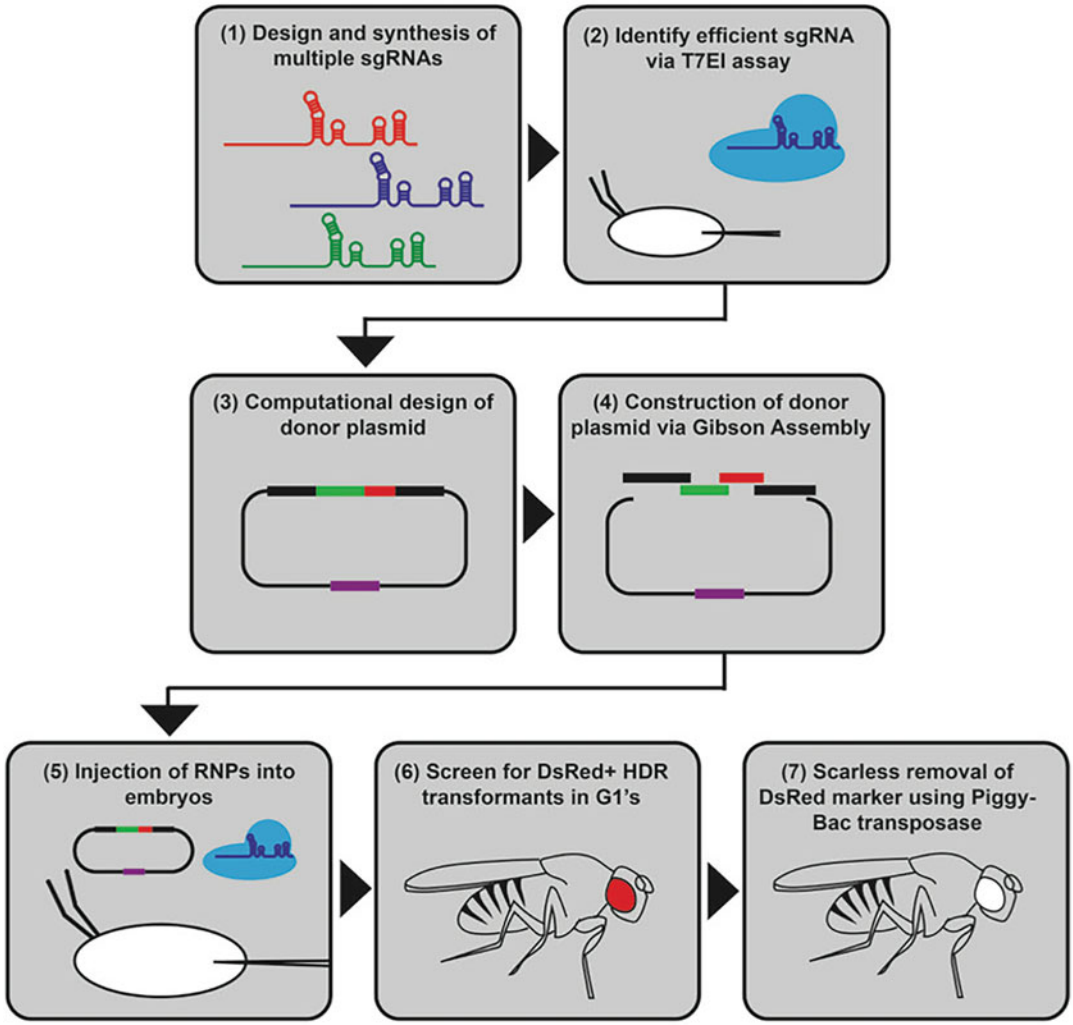


Fig. 1 Workflow to produce precise genome edits in *Drosophila* using CRISPR/Cas9. Multiple sgRNAs are designed, and Cas9/sgRNA RNPs are screened in *Drosophila* embryos to ensure efficient sgRNA cleavage activity in vivo. A donor repair plasmid is then computationally designed and constructed via Gibson Assembly. The Cas9/sgRNA RNP and donor repair plasmid are then injected together into *Drosophila* embryos and screened for successful transformation via HDR using a DsRed eye marker. This marker is then removed using a PiggyBac transposase without leaving scars in the genome sequence

DsRed marker gene is then excised from the genome by PiggyBac-mediated transposition, leaving the engineered genome edit intact and scarless [16]. Taken together, these enhancements greatly expand the ease and scope of performing precise CRISPR/Cas9 genome editing in *Drosophila*.

2 Materials

Prepare all solutions using ultrapure nuclease-free water (prepared by purifying deionized water, to attain a sensitivity of 18 MΩ-cm at 25 °C) and analytical grade reagents.

2.1 Synthesis of sgRNAs

1. sgRNA PCR Primers: sgRNA_R primer — 5′-AAAAGCACC GACTCGGTGCC-3′ and sgRNA_F primer — 5′-TTAATAC GACTCACTATAGG-[sgRNA target sequence without PAM]-GTTTTAGAGCTAGAAATAG-3′ (*See* **Note 1**). Dissolve each oligo to a final 10 μM concentration in water. Store at −20 °C.
2. Phusion High-Fidelity DNA Polymerase: (New England BioLabs).
3. dNTPs: Pool 100 mM stocks of dATP, dCTP, dGTP, dTTP, and dilute to a 10 mM final concentration of each nucleoside triphosphate.
4. pU6-BbsI-chiRNA plasmid DNA, 30 ng/μL stock in water. The plasmid can be obtained from Addgene (catalog #45946) [3].
5. Qiaquick PCR Purification Kit (Qiagen).
6. MEGAscript T7 Transcription Kit (ThermoFisher).
7. RiboLock RNase Inhibitor (ThermoFisher).
8. Monarch RNA Cleanup Kit (50 μg): (New England BioLabs).

2.2 In Vivo Assay for Cleavage and NHEJ

1. Squish buffer: 10 mM Tris-HCl (pH 8.0), 1 mM EDTA, 25 mM NaCl. Store at room temperature. Immediately before using, add 1 μL of 20 mg/mL proteinase K to 100 μL of the buffer [17].
2. Cas9/RNA RNPs: Mix 1.19 μL of 10 mg/mL Cas9-NLS protein (Integrated DNA Technologies catalog #1081058) with 2.36 μg in vitro-transcribed sgRNA, 0.38 μL 2 M KCl, and water to a final volume of 5 μL. Incubate at room temperature for 10 min. Centrifuge in a microfuge at maximum speed for 10 min at room temperature. Transfer 4 μL of supernatant into new tube to be loaded into injection needles. Store at room temperature. Prepare RNPs fresh each day.
3. Taq DNA Polymerase (non-proofreading is generally sufficient).
4. 0.5× Tris-borate-EDTA (TBE) buffer – 45 mM Tris base, 45 mM boric acid, 1 mM EDTA.
5. Ethidium bromide – 10 mg/mL stock solution.
6. T7 Endonuclease I (New England BioLabs).
7. Electrophoresis grade agarose.

2.3 Construction of Donor Plasmid

1. pBS-GMR-*eya*(shRNA) plasmid DNA (Drosophila Genomics Resource Center (catalog #1518) or Addgene (catalog #157991)) [12]. The plasmid is 3845 bp in length. It carries a short hairpin RNAi agent against *eyes absent* (*eya*) mRNA transcripts. Its transcription is driven by the eye-specific GMR enhancer. If the plasmid integrates into the *Drosophila* genome, it results in small eyes and can be used in any line with normal eye morphology.
2. pScarlessHD-DsRed plasmid DNA: (Addgene) (catalog #64703). In the plasmid, the 3xP3-DsRed cassette is flanked by PiggyBac inverted repeats that can be used to precisely excise the entire 3xP3-DsRed gene after successful integration of the modification via HDR in the *Drosophila* genome [16].
3. gBlocks gene fragments (Integrated DNA Technologies).
4. EcoRV-HF restriction enzyme (New England BioLabs).
5. Monarch DNA Gel Extraction Kit (New England BioLabs).
6. Qiaquick PCR Purification Kit (Qiagen).
7. NEBuilder HiFi DNA Master Mix (New England BioLabs).
8. Electrocompetent *E. coli* (e.g., strain DH5α).

2.4 Purification of the Donor Plasmid

1. HiSpeed Plasmid Midi Kit (Qiagen).
2. EndoFree Plasmid Mega Kit Buffer ER and Buffer QN (Qiagen). These are used to remove endotoxins and thereby reduce toxicity in injected embryos.
3. Isopropanol, ACS grade.
4. Ethanol, absolute (200 proof), molecular biology grade.
5. 3 M Sodium acetate, pH 5.2.
6. Nuclease-free water, molecular biology grade.

2.5 Cas9-Mediated Homologous DNA Repair

1. RNP/Donor Plasmid Solution: Mix 1.19 μL Cas9-NLS protein (Integrated DNA Technologies #1081058, 10 mg/mL stock) with 2.36 μg in vitro-transcribed sgRNA, 0.6 pmoles donor plasmid DNA, 0.38 μL 2 M KCl, and water to a final volume of 5 μL. Incubate at room temperature for 10 min. Centrifuge in a microfuge at maximum speed for 10 min at room temperature. Transfer 4 μL of supernatant into a new tube to be loaded into injection needles. Store at room temperature. Prepare fresh each day.

2.6 Excision of the DsRed Marker

1. Bloomington Drosophila Stock Center stock #32070 contains a PiggyBac transposase transgene under control of the *α-tubulin* promoter and tightly linked to a 3XP3-CFP transgenic marker [18]. This is located on chromosome 2. The stock

also contains 3rd chromosome balancers (MKRS/TM6B,Tb), facilitating tracking of the 3rd chromosome independent of the PiggyBac transposase.

2. Plasmid atub-pBac-K10 DNA (Drosophila Genomic Resources Center (#1155)) [19]. This P-element vector contains the *αTub84B* promoter fused to the transposase open reading frame from the PiggyBac transposable element. The 3′UTR is derived from the *fs(1)K10* gene, lending germline stability to the transposase transcript.
3. Injection buffer: 0.1 mM sodium phosphate pH 7.8, 5 mM KCl.

3 Methods

Carry out all procedures at room temperature unless otherwise specified.

3.1 Design and Synthesis of sgRNAs

1. Candidate sgRNA target should be identified using the fly-CRISPR Optimal Target Finder Web Portal (http://targetfinder.flycrispr.neuro.brown.edu). High stringency filtering is sufficient, and only NGG PAM sites should be utilized. Potential off-target sites should be minimized; zero predicted off-target sites is ideal. A sgRNA target length of 20 nucleotides works well. There is no need to have a G or GG at the 5′ end of the sgRNA (*see* **Note 2**).
2. Select two to three candidate sgRNA targets from **step 1** whose cleavage sites are within 30 bp of the desired editing site (*see* **Note 3**). Their close proximity to the editing site is to minimize the possibility of homologous recombination between the cleavage site and the editing site. Placement of the editing site within the sgRNA target sequence is ideal since it will inactivate the sgRNA target site in the donor plasmid (*see* **Note 4**).
3. Validate the target sequence for these selected sgRNAs in the *Drosophila melanogaster* stock or the stock of another species that you plan to edit. This should be done by Sanger sequencing the sgRNA targets from the stock's genomic DNA (*see* **Note 5**). Modify the sgRNA sequences to synthesize according to the observed target sequences in the stock of interest.
4. To make a DNA template for in vitro transcription of one candidate sgRNA, perform a 50 μL PCR reaction using a proofreading DNA polymerase (e.g., New England Biolabs Phusion HF DNA polymerase). The reaction should contain 1× Phusion HF buffer (with 1.5 mM $MgCl_2$), 0.2 mM dNTPs, 25 pmoles sgRNA_F oligo, 25 pmoles sgRNA_R oligo, 30 ng pU6-BbsI-chiRNA plasmid DNA, and 1 unit Phusion HF polymerase [3]. Perform 35 cycles of touchdown PCR, starting

with an annealing temperature of 60 °C and dropping 0.5 °C/cycle until a final annealing temperature of 50 °C is reached (*see* **Note 6**).

5. Purify the PCR product using standard column purification (e.g., Qiaquick PCR Purification Kit).
6. Add 300 ng of the purified PCR product to a 20 μL MEGAscript in vitro transcription reaction supplemented with 0.5 μL (20 units) of Ribolock RNase inhibitor. Incubate at 37 °C overnight in a thermocycler with a heated lid.
7. Purify the RNA product using an RNA cleanup kit and elute the RNA in 20 μL nuclease-free water (*see* **Note 7**).
8. After measurement of RNA concentration, aliquot and store at −80 °C. Successful reactions should yield concentrations of 2–4 μg/μL RNA.

3.2 Assay for Cleavage Efficiency of Synthesized sgRNAs

1. Prepare egg-laying cages with the *Drosophila* stock that you plan to edit. Let the adults acclimate to the cage for 3 days at room temperature with frequent plate changes.
2. Several hours before injections, flip flies onto fresh plates with yeast paste to allow females to lay any held embryos. Allow them to lay for at least 1 h, and then repeat for a total of two pre-clears.
3. Prepare fresh Cas9/sgRNA RNPs for injection. Since 2 or 3 candidate sgRNAs will be assayed in parallel, make separate RNP preparations for each. Fill independent needles with each RNP preparation.
4. Collect embryos over repeating 30–50 min time windows for RNP injections. From a single round of embryo collection, line up and inject 35–40 embryos for each RNP preparation. Multiple RNPs can usually be injected from a single round of embryo collection. A standard volume of ~100 pL should be injected into each embryo (*see* **Note 8**).
5. Also mock-inject 35–40 embryos as a control (*see* **Note 9**).
6. With a needle, rupture any embryos that were skipped during injection due to their advanced age or other defects. Incubate embryos in a humid chamber at 25 °C for 24 h (*see* **Note 10**).
7. Using a pipet tip, pick up either individual L1 larvae or embryos that are clearly late stage. Larvae are preferable to pick for the assay. You want to pick a total of 8–10 individuals to assay for each RNP treatment
8. For each individual embryo or larva, using the pipet tip you picked it up with, grind it in 20 μL squish buffer in a PCR tube. Pipette the mix several times (*see* **Note 11**). Incubate at 37 °C for 30 min. Incubate at 95 °C for 2 min. Store the genomic DNA prep at 4 °C.

9. Design PCR primers to amplify the genome region in which putative cleavage and NHEJ events are predicted to have occurred (Fig. 1). The PCR amplicon should be between 700 and 1200 bp long, and the sgRNA target site should be located near the center of the amplicon. If the sgRNAs being tested are spread too far apart, you might need to use different amplicons to assay them all.
10. For each pair of designed primers, amplify the genome region from two or three individual mock-injected larvae or embryos in separate PCR reactions. For each 50 μL reaction, add 3–5 μL of genomic DNA, standard Taq DNA polymerase, and the PCR primers designed in **step 9**. After the PCR reactions are complete, take 10 μL of each reaction and perform the following denaturation and reannealing steps in a PCR machine: 95 °C – 3 min, 94 °C – 1 min, 93 °C – 1 min, and 92 °C – 1 min; continue downward in 1 degree increments to 4 °C – 1 min, 6 °C – 10 s, 8 °C – 10 s, 10 °C – 10 s; and 12 °C – hold. To 10 μL reannealed DNA, add 2 μL 10 × NEBuffer 2, 0.2 μL New England Biolabs T7 Endonuclease I (T7EI), and 7.8 μL water. Incubate at 37 °C for 60 min. Electrophorese in a 2% (w/v) agarose gel: 10 μL of undigested PCR product side by side with 20 μL of the T7EI reaction. Ethidium bromide (0.5 μg/mL) and 0.5 × TBE should be used in the gel and buffer system to increase sensitivity to see faint digestion products. The PCR reaction should generate a robust amplicon, and T7EI treatment should not generate digestion products that overlap in size with the predicted fragments from sgRNA targeted NHEJ events (Fig. 2) (*see* **Note 12**).
11. If the quality control test in **step 10** passes, then the assays can proceed as follows (Fig. 1). For each genomic DNA prep of mock-injected and RNP-injected animals, perform a 50 μL PCR reaction using Taq polymerase, the PCR primers designed in **step 9**, and 3–5 μL genomic DNA.
12. Take 10 μL of each PCR reaction and perform the following denaturation and reannealing steps in a PCR machine: 95 °C – 3 min, 94 °C – 1 min, 93 °C – 1 min, and 92 °C – 1 min; continue downward in 1 degree increments to 4 °C – 1 min, 6 °C – 10 s, 8 °C – 10 s, 10 °C – 10 s; and 12 °C – hold.
13. To 10 μL reannealed DNA, add 2 μL 10 × NEBuffer 2, 0.2 μL New England Biolabs T7 Endonuclease I (T7EI), and 7.8 μL water. Incubate at 37 °C for 60 min.
14. Electrophorese reaction products on a 2% (w/v) agarose gel. For each reaction, run 10 μL of the undigested PCR amplicon side by side with 20 μL of the T7EI reaction. Ethidium bromide (0.5 μg/mL) and 0.5 × TBE should be used in the gel and buffer system to increase sensitivity to see faint digestion

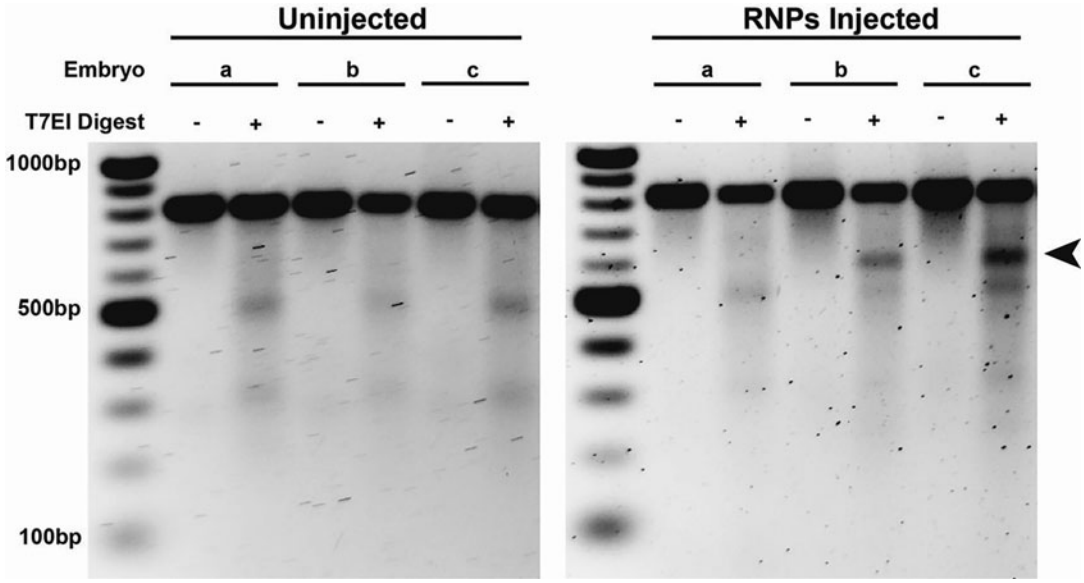


Fig. 2 In vivo screening of sgRNA cleavage activity. PCR products (826bp in size) of an sgRNA target site in the lncRNA *CR45715* gene from both uninjected and RNP-injected embryos are digested by T7EI. Three representative embryos (from eight total) are shown for each. Predicted T7EI cleavage products are 596 and 230 bp. The 596-bp product (arrowhead) is clearly visible in two RNP-injected embryos, though the 230-bp product is likely too faint to visualize. Note that T7EI produces digestion products at ~500 bp and ~300 bp even in uninjected embryos, but these are easily distinguished from the sgRNA-induced cleavage products. This sgRNA was subsequently used to generate multiple edits in the region via HDR

products. Samples with cleavage products at expected sizes from RNP-injected animals that are not present in mock-injected controls and undigested controls are indicative of sgRNA-guided cleavage (Fig. 2) (*see* **Note 13**).

3.3 Design of the Donor Plasmid

1. Once a sgRNA has been found that efficiently cleaves DNA in vivo, as determined by the T7EI assay described in Subheading 3.2, the donor plasmid can then be designed (Fig. 1). The first step is to computationally design the donor plasmid sequence using an informatics tool such as Benchling (https://www.benchling.com). It is critical that you first design the plasmid sequence you want to create before worrying about how to construct it.
2. The donor plasmid consists of a backbone plasmid with insertion of *Drosophila* genomic DNA comprising the region being edited (Fig. 3). Somewhere within the genomic DNA is inserted a synthetic positive transformation marker. The backbone plasmid itself needs no design; it is the plasmid pBS-GMR-*eya*(shRNA) [12]. The other parts are inserted altogether at the *EcoRV* site of pBS-GMR-*eya*(shRNA) within the multi-cloning region (*see* **Note 14**).

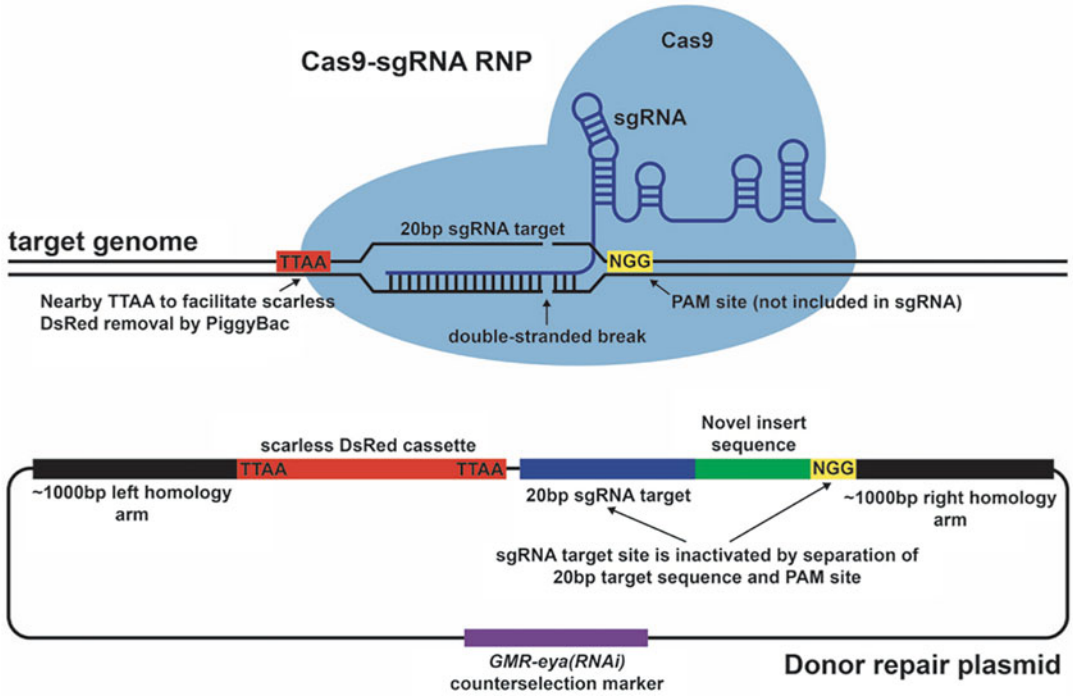


Fig. 3 Design of the sgRNA and donor repair plasmid. Important design considerations are shown at and around the sgRNA target site. The NGG PAM site is adjacent to the sgRNA target site in the genome, but is not included in the sgRNA sequence itself. A nearby TTAA sequence in the genome (ideally within 30bp of double-stranded break site) serves as site of insertion for the scarless DsRed cassette. This TTAA site can also be located within the sgRNA target sequence or within the novel insert sequence itself. The sgRNA target site in the donor repair plasmid is inactivated by placing the novel insert sequence between the 20bp sgRNA target site and the PAM site. This can also be achieved by mutating the PAM site or the very 3′ end of the sgRNA target site

3. From Subheadings 3.1 and 3.2, you have already designed and selected two key sites in the genomic DNA: the sgRNA target sequence and the site where the genome editing will occur (Fig. 3). The third key site to select is the place where the positive selection marker will be inserted into the genomic DNA. The positive transformation marker is a 3xP3-DsRed gene flanked on either side by PiggyBac inverted repeats (Fig. 3) [16]. These allow for PiggyBac transposase-mediated excision of the marker, enabling you to cleanly excise the entire marker gene after successful integration of the genome edit (Fig. 1). After excision, the remaining genome sequence will be reduced to a single TTAA site. Therefore, the 3xP3-DsRed marker must be placed immediately next to a native 5′-TTAA-3′ sequence that can be on either DNA strand (Fig. 3). The TTAA sequence can be located in the *Drosophila* genome but if so, it should be located as close to the sgRNA cleavage site as possible. Being located within 30 bp of the cleavage site is the

best (ideally within the target sequence itself if it has a TTAA motif on either strand). Alternatively, the TTAA motif to be used can be in any novel sequence being introduced by the genome edit, i.e., a TTAA motif in GFP sequence or some other sequence being inserted. Find all candidate TTAA motifs on both DNA strands that fulfill these criteria. Select the best site among these candidates.

4. Design the 3xP3-DsRed marker so it is located either on the 5′ or 3′ side of the native TTAA motif selected in **step 3**. The marker is 1691 bp in length, and must have a TTAA sequence flanking it on both sides. Thus, on one side will be the TTAA that is native, and the other side will be a TTAA that is synthetic (*see* **Note 15**).
5. There is one essential element to design of the donor plasmid; its own sgRNA target sequence must be disrupted to prevent Cas9 from cleaving the donor plasmid in vivo (Fig. 3). If possible, you should inactivate the sgRNA cleavage site in the donor plasmid either by inserting the genome editing site or the positive transformation marker into the sgRNA target sequence. If this is not possible, then you must mutate at least one basepair in the PAM site or one of the most 3′ base-pairs in the sgRNA target sequence of the donor plasmid.
6. Design the donor plasmid sequence so that there is ~1000 bp of *Drosophila* genomic DNA sequence on either side of the sgRNA cleavage site (*see* **Note 16**). These are called the left and right homology arms, and their purpose is to facilitate the double crossover events between donor plasmid and *Drosophila* chromosome after RNP-mediated cleavage of the chromosome DNA (Fig. 3).
7. Design the donor plasmid sequence to contain the precise changes in the *Drosophila* genome you want to achieve. These will be located at the editing site in the genomic DNA, either at the base of the left or right homology arm, or in the sgRNA target sequence itself (Fig. 3).

3.4 Construction of the Donor Plasmid

1. Once the donor plasmid sequence is computationally designed, Gibson assembly can be used to assemble the necessary DNA fragments into the donor plasmid [20]. The assembly typically involves five fragments of DNA: (1) pBS-GMR-*eya*(shRNA) linearized at its *EcoRV* site, (2) a PCR amplicon of the positive transformation marker, (3) left homology arm, (4) right homology arm, and (5) genome edited fragment. In most cases, the backbone plasmid is generated via EcoRV digest, and the homology arms are PCR amplified from genomic DNA from the same *Drosophila* strain or species to be used for injections. The positive transformation marker is typically

generated via PCR from pScarlessHD-DsRed plasmid DNA. The genome edited fragment can be generated either via PCR or de novo synthesis (e.g., gBlocks from Integrated DNA Technologies).

2. Digest 5–10 μg of the backbone plasmid pBS-GMR-*eya*(-shRNA) with EcoRV-HF at 37 °C for 15 min. Digestion products should be electrophoresed on a 1% (w/v) agarose gel, using multiple lanes to accommodate the large volume of the digestion reaction. Linearized plasmid DNA should be quickly and carefully excised from the gel, minimizing exposure to UV light. Purify the DNA using a Monarch DNA Gel Extraction Kit to avoid contamination of Gibson assembly reactions with trace intact circular plasmid (*see* **Note 14**).
3. Design oligo primers to amplify fragments by PCR, ensuring that the oligos also contain sequences that produce 30-bp overlap regions with adjacent fragments in the assembly. For junctions of PCR fragments with linearized plasmid or synthetic fragments, the 30-bp overlap sequence can be added entirely to the 5′ end of the PCR oligo. Alternatively, for junctions of adjacent PCR fragments, the 30-bp overlap can be split between them, resulting in less than 30 bp of additional overlap sequence added to the 5′ end of each PCR oligo. Use the NEBuilder tool (http://nebuilder.neb.com) with the following build settings: Product Kit - NEBuilder HiFi DNA Assembly Master Mix; Minimum Overlap – 30 nt; Circularize – Yes; PCR Polymerase/Kit – Phusion High-Fidelity DNA Polymerase (HF Buffer); PCR Primer Conc – 500 nM; Min. Primer Length – 18 (*see* **Notes 17–20**).
4. If possible, make the genome edited fragment by gBlocks DNA synthesis (Integrated DNA Technologies). Ensure that the gBlocks fragment also contains sequences that allow for Gibson assembly; its ends must allow for there to be a 30-bp overlap in sequence with adjacent fragments (*see* **Notes 17** and **19**). Use the NEBuilder tool (http://nebuilder.neb.com) with the following build settings: Product Kit – NEBuilder HiFi DNA Assembly Master Mix; Minimum Overlap – 30 nt; Circularize – Yes; PCR Polymerase/Kit – Phusion High-Fidelity DNA Polymerase (HF Buffer); PCR Primer Conc – 500 nM; Min. Primer Length – 18. Any synthesized DNA fragment should be briefly centrifuged and resuspended in molecular grade water to a final concentration of 10 μg/mL. Incubate at 50 °C for 15 min to facilitate resuspension.
5. To generate the PCR fragment with the positive transformation marker, perform a 50 μL PCR reaction using New England Biolabs Phusion HF Polymerase and 30–50 ng of pScarlessHD-DsRed plasmid DNA as template. Touchdown

PCR is recommended to reduce non-specific amplicons. The entire reaction should be electrophoresed on a 1% agarose gel, and the desired product should be gel extracted as in **step 2** to avoid contamination of the Gibson assembly reaction with plasmid. Similarly, gel extraction should be performed on any other PCR reaction that uses plasmid as template.

6. To generate PCR fragments of homology arms, perform a 50 μL touchdown PCR reaction for each homology arm using New England Biolabs Phusion HF Polymerase and 50 ng of genomic DNA from the same *Drosophila* strain or species that will be used for injections. If multiple amplicons are generated, then purify via agarose gel extraction as in **step 2**. Otherwise, standard column purification is sufficient (e.g., Qiagen Qiaquick PCR cleanup).
7. Determine the concentration of all DNA fragments using a fluorometer (e.g., Qubit) or spectrophotometer (e.g., Nanodrop). For a five-piece Gibson assembly reaction, fragments should be added in equimolar amounts, with total DNA content of the reaction not exceeding 0.5 pmol. A quantity of 0.08–0.1 pmol per fragment works well. The combined volume of DNA fragments should be 10 μL or less.
8. Mix all of the DNA fragments together. The combined volume should be less than 10 μL. Add water to a final volume of 10 μL.
9. Mix the DNA with 10 μL NEBuilder HiFi DNA Master Mix. Mix well.
10. Incubate at 50 °C for 1 h in a thermocycler with heated lid.
11. Transform into competent *E. coli* (*see* **Note 21**).
12. Individual colonies can be picked and screened via PCR for successful assembly across junctions. Confirm correct assembly of the entire inserted region via Sanger sequencing. Polymorphisms in non-coding regions of homology arms are not uncommon, but ensure that there are no disabling mutations in the scarless DsRed cassette or coding regions in the homology arms.

3.5 Purification of the Donor Plasmid

1. Pellet 50 mL of an overnight LB culture at 6000 × g for 15 min at 4 °C. Decant supernatant.
2. Using components from the Qiagen HiSpeed Plasmid Midi Kit, resuspend the pellet in 6 mL Buffer P1 plus RNase A by vortexing. Add 6 mL Buffer P2 and mix well by inverting 4–6 times. Incubate at RT for 5 min.
3. During incubation, screw the cap onto the outlet nozzle of the QIAfilter Cartridge. Place the cartridge into a fresh 50-mL conical tube. Add 6 mL prechilled Buffer P3 to lysate and

mix well by inverting 4–6 times. Pour the lysate into the QIAfilter Cartridge and incubate for 10 min. Remove the cap, insert the plunger, and filter the solution through the syringe filter into the 50-mL conical tube.

4. Add 1 mL buffer ER from the EndoFree Plasmid Mega Kit (Qiagen) to the filtered solution and incubate on ice for 30 min (*see* **Note 22**). During incubation, equilibrate a HiSpeed Tip with 4 mL Buffer QBT.
5. Apply the plasmid solution to the QBT-equilibrated HiSpeed Tip and allow to flow through.
6. Wash the HiSpeed Tip 2 × 10 mL with Buffer QC.
7. Place the HiSpeed Tip over a fresh 50-mL conical tube and elute by applying 5 mL Buffer QN from the EndoFree Plasmid Mega Kit (*see* **Note 22**). Add 3.5 mL isopropanol to the eluted solution. Mix by inverting and incubate for 5 min.
8. During incubation, remove the plunger from a 20-mL syringe and attach the QIAprecipitator Module from the HiSpeed Plasmid Midi kit onto the outlet nozzle. Place the QIAprecipitator over a fresh 50-mL conical tube. Transfer the plasmid eluate into the syringe and insert the plunger. Filter the mixture through using constant pressure.
9. Remove the QIAprecipitator from the syringe and pull out the plunger. Reattach the QIAprecipitator and add 2 mL 70% (v/v) ethanol to the syringe. Insert the plunger and push it through.
10. Remove the QIAprecipitator from the syringe and pull out the plunger. Attach the QIAprecipitator again and insert the plunger. Dry the membrane by pressing air through the QIAprecipitator. Repeat this step several times.
11. Dry the outlet nozzle of the QIAprecipitator with a Kimwipe. Remove the plunger from a new 5-mL syringe, attach the QIAprecipitator and hold the outlet over a fresh 1.5-mL Eppendorf tube. Add 1 mL Buffer TE from the HiSpeed Plasmid Midi kit to the syringe. Insert the plunger and elute the DNA into the collection tube using constant pressure.
12. Remove the QIAprecipitator from the 5-mL syringe and pull out the plunger. Reattach the QIAprecipitator to the syringe. Transfer the eluate from **step 9** to the 5-mL syringe and elute for a second time into the same 1.5-mL tube.
13. Estimate the volume of the DNA solution and add 1/10 volume of 3 M sodium acetate pH 5.2. Mix well. Add 3 volumes of molecular grade absolute ethanol. Incubate at −80 °C for 30 min.

14. Centrifuge in a microfuge at maximum speed for 15 min at 4 °C. Split into multiple 1.5-mL tubes if necessary. Remove supernatant and wash the pellet twice in 800 μL of ice-cold 70% (v/v) ethanol. Centrifuge at 4 °C for 5 min at maximum speed after each wash.
15. After the final wash, remove supernatant and allow to air-dry 5–10 min. Resuspend the pellet in 40 μL nuclease-free water. Measure DNA concentration using a NanoDrop or Qubit. Final concentration should be 240 nM or higher.

3.6 Cas9-Mediated Homologous DNA Repair

1. Prepare egg-laying cages with the *Drosophila* stock that you plan to edit. Let the adults acclimate to the cage for 3 days at room temperature with frequent plate changes.
2. Several hours before injections, flip flies onto fresh plates with yeast paste to allow females to lay any held embryos. Allow them to lay for at least 1 h, and then repeat for a total of two pre-clears.
3. Prepare fresh Cas9/sgRNA RNPs in the presence of the donor plasmid DNA as described in Subheading 2.5. Fill injection needles with the solution.
4. Collect embryos over repeating 30–50 min time windows for injections. Over multiple rounds of embryo collection, line up and inject 300–350 embryos. This number is typically sufficient to produce one or more germline transformants. A standard volume of ~100 pL should be injected into each embryo (*see* **Note 8**).
5. If you injected dechorionated embryos, then leave them under halocarbon oil and place the coverslip in a humid chamber. If you injected embryos with the chorion intact, then remove as much oil as possible and place coverslip with injected G0 embryos in a standard food vial. Keep the vial in a humid chamber at 25 °C overnight.
6. Once G0 adults eclose, they should be individually crossed to virgin females or males from an appropriate line. Using a *white* mutant stock for both injections and the G0 crosses is preferable in order to maintain a consistent genetic background and to expedite screening of G1 adults for the positive transformation marker phenotype, which is DsRed fluorescence in the compound eyes (*see* **Note 23**).
7. Screen all G1 offspring for two adult eye phenotypes (Fig. 4). One is for DsRed fluorescence and the second is for a small rough eye (*eya*) phenotype (*see* **Note 24**). It is possible that a fly might have none, one, or both phenotypes. Those that have both eye phenotypes have the entire donor plasmid DNA integrated into the genome. This happens when only a single crossover event occurs (imprecise HDR). Such flies should be

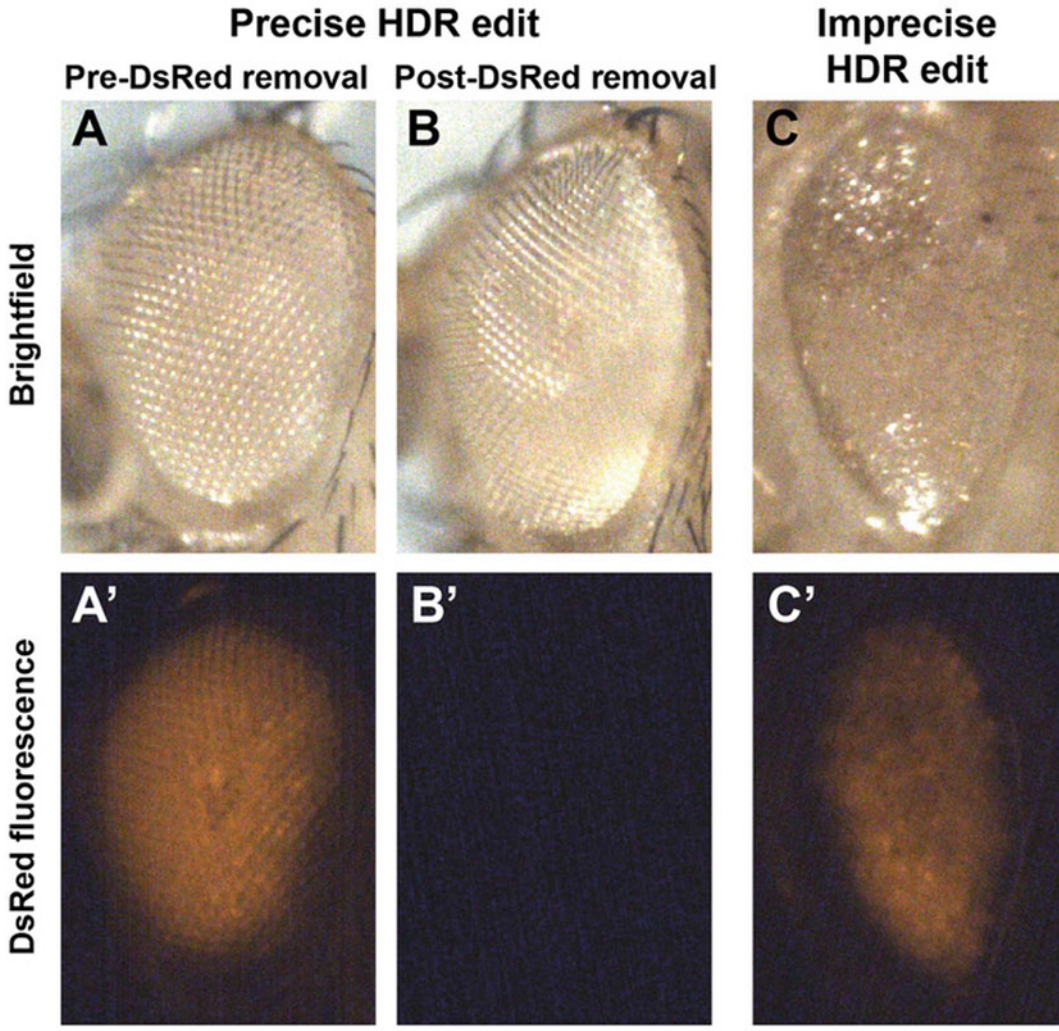


Fig. 4 Identification of precise HDR edits by eye morphology and fluorescence. Successful HDR edits using this protocol are evidenced by flies with wild-type eye morphology and DsRed fluorescence (A, A′). The fluorescent DsRed eye marker is then removed from the genome using a PiggyBac transposase, leaving non-fluorescent eyes with wild-type morphology (B, B′). An imprecise HDR edit that integrates the entire donor repair plasmid into the genome is evidenced by the DsRed eye marker and expression of the *GMR>eya(RNAi)* transgene, which produces a rough small eye phenotype (C, C′)

discarded. Likewise, flies with small eyes but no DsRed fluorescence should be discarded. Only keep G1 adults with normal sized eyes that are DsRed fluorescent (*see* **Note 25**). Such adults should be individually crossed to an appropriate balancer stock.

8. Once lines are established and stable, verification of the anticipated editing/modification must be done by PCR analysis and Sanger sequencing of that region of the genome. Imprecise

edits and spurious recombinations do occur. PCR amplify and sequence the entire region using primers that anneal outside of the homology arms coupled with internal primers.

3.7 Excision of the DsRed Marker with PiggyBac Transposase

1. Cross females from the stable DsRed+ lines to males from the stock containing the transgenes *α-tubulin*-PiggyBac transposase and 3XP3-CFP [18].
2. Select males in the F1 generation that are both DsRed positive and CFP positive (*see* **Note 26**). The PiggyBac transposase is only weakly efficient, so DsRed fluorescence will still be visible, albeit mosaic in F1 flies. Cross the F1 males to 10–20 females carrying an appropriate balancer.
3. If the DsRed marker was on an autosome, select F2 flies that have the appropriate balancer chromosome and are both DsRed-minus and CFP-minus (Fig. 4). Cross these as single-pair crosses to an appropriate balancer stock to make a balanced stock of each F2 line. If the DsRed marker was on the X chromosome, select female F2 flies that are both DsRed-minus and CFP-minus. Cross these as single-pair crosses to males from an appropriate balancer line to make a balanced stock of each F2 line.
4. If the genome editing has been performed on a species other than *D. melanogaster*, it will be necessary to inject the DsRed+ lines with the plasmid atub-pBac-K10, which expresses the PiggyBac transposase gene under *α-tubulin* promoter control [19]. Injections should be performed using a concentration of 0.6 mg/mL plasmid DNA dissolved in 0.1 mM sodium phosphate pH 7.8 + 5 mM KCl. Cross individual G0 adults to an appropriate strain and screen G1 adult offspring for the absence of DsRed eye fluorescence (*see* **Note 27**).
5. To ensure that the genome edit is still present after excision and that the excision was scarless, verify via PCR analysis and Sanger sequencing of that region of the genome.

4 Notes

1. For the forward primer sgRNA_F, the sequence encoding the 20-nucleotide sgRNA target should be sandwiched in the middle of the oligo. Note that the PAM site must not be included in the sequence that will make the sgRNA. The sequence at the 5′ end of the oligo enables T7 RNA polymerase to initiate transcription. The 19-nucleotide sequence at the 3′ end of the oligo enables it to anneal to the pU6-BbsI-chiRNA plasmid template for PCR. The complementary site on the pU6-BbsI-chiRNA plasmid corresponds to the 5′ end of the sgRNA

scaffold at position 590 of the plasmid (within crRNA sequence). The reverse primer sgRNA_R is an oligo complementary to a sequence in the pU6-BbsI-chiRNA plasmid that corresponds to the 3′ end of the sgRNA scaffold at position 669 of the plasmid (within tracrRNA sequence).

2. A G or GG at the 5′ end of the sgRNA target sequence is not necessary for in vitro transcription using T7 RNA polymerase since the GG necessary for T7 transcription is built into the forward primer.
3. The reason to select and synthesize 2–3 sgRNAs is that not every sgRNA will cleave DNA efficiently in vivo. It is necessary to assay several candidate sgRNAs for cleavage efficiency, and then choose the most efficient one for the genome editing procedure.
4. A distance of 30 bp or less is optimal, though longer distances may still work. Often, the editing site is very constrained, i.e., a GFP fusion at the end of a gene's ORF. If you have some flexibility as to where to place the site of editing, then this provides more opportunity to find a good sgRNA since you can move the edit close to or within the sgRNA target sequence.
5. Since sequence polymorphisms are prevalent across the genome of various stocks, the *Drosophila* reference genome sequence should only be taken as a guide, and the stock of interest should be sequence verified.
6. Verify successful PCR amplification using agarose gel electrophoresis. The product should be ~120 bp in length.
7. Successful in vitro transcription should yield >40 μg of RNA, a sample of which should produce a discrete band after agarose gel electrophoresis.
8. Injections can be performed on pre-cellularized embryos without dechorionation using Gompel and Schröder's method [21].
9. Mock injection can involve handling the embryos as if they were going to be injected but then not injecting them with a needle. Alternatively, one can mock-inject by injecting buffer only into the embryos.
10. If you injected dechorionated embryos, then leave them under halocarbon oil and place the coverslip in a humid chamber. If you injected embryos with the chorion intact, then remove as much of the halocarbon oil as possible from the coverslip. Place the coverslip with embryos on an egg-laying plate, and place that in a humid chamber.

11. It is easy to lose the animal, so check under a dissecting scope to ensure that the animal is still inside the PCR tube and is ruptured. This is easier with larvae.
12. This step is a quality control step. It is to verify that you can obtain a single robust PCR product from genomic DNA, and that T7EI treatment of the PCR amplicon does not generate digestion products that overlap in size with predicted cleavage products generated by sgRNA-mediated NHEJ events. Even PCR amplicons from mock-injected DNA can produce multiple digestion products after T7EI treatment. T7EI is a DNA resolvase that naturally recognizes and resolves Holliday junctions [22]. It also cleaves DNA at the site of single-stranded bubbles. However, resolvases also cleave DNA if there are long tracts of A residues, meaning that they will generate background digestion products from DNA amplicons if there are A-tracts or other unusual structures present [23]. If you find that T7EI generates such overlapping digestion products from a PCR amplicon, design new PCR primers to position the predicted site of NHEJ such that its digestion products do not overlap in size with the background products.
13. The digestion products may be very faint in the gel staining. This is due to the mosaicism of NHEJ events occurring in somatic cells of the embryo. Not all somatic cells will undergo a NHEJ event that generates DNA mismatches, and only a small number might do so. If a sgRNA is efficient at cleavage, the digestion products from this assay will be more abundant than with other sgRNAs, and it is not unusual for >50% of individuals to have detectable cleavage products.
14. The plasmid is linearized via restriction enzyme digestion before Gibson assembly. pBS-GMR-*eya*(shRNA) can be linearized with *EcoRV* (recognition sequence: GATATC), which is located in the multi-cloning site of pBluescript. Since *EcoRV* digestion generates blunt ended fragments, no nucleotides will be removed by the 5′->3′ exonuclease activity of the Gibson assembly, and thus, the assembled insert should be placed right at the cut site.
15. It is important that the 5′-TTAA-3′ repeats are direct and not inverted. Since the 3xP3-DsRed cassette can be inserted in sense or antisense orientation, it gives you a lot of flexibility as to which native TTAA motif to select as one flanking repeat. The motif can be on one strand or the other strand.
16. Lengths of homology arms can be slightly increased or reduced to provide ideal sequences for Gibson assembly (e.g., moderate GC content and non-repetitive sequence).
17. Try to alleviate flagged issues raised by NEBuilder if possible, though not all issues can be resolved. For example, you cannot change the ends of the linear backbone plasmid, even if they are

not ideal for Gibson assembly. Ends of homology arms can be slightly altered to improve Gibson overlap regions, and a synthesized gBlocks fragment can be altered to do the same. Overlap regions can be slightly altered via junction properties in NEBuilder.

18. You should limit the maximum primer length to 60 nt. This not a build setting, but necessary for a standard Integrated DNA Technologies oligo order.
19. Avoid placing repetitive regions into Gibson overlap regions. For the positive transformation marker, the PiggyBac inverted repeats at both ends of the cassette are identical and thus should not be used as Gibson overlap regions. If synthesizing a flanking fragment via gBlocks, one potential workaround is to extend the gBlocks fragment through the adjacent inverted repeat of the 3xP3-DsRed cassette and place the Gibson overlap region deeper into a non-repetitive region of the cassette. This is also applicable for flanking fragments synthesized by PCR. We have verified that the following sequences within the 3xP3-DsRed cassette can be used as Gibson overlap regions. PiggyBac left (5′) region: 5′-GTCGTTATAGTTCAAAATCA GTGACACTTA-3′; PiggyBac right (3′) region: 5′-AGATAA TCATGCGTAAAATTGACGCATGTG-3′.
20. Once all primers are designed, verify that they all will bind in your computationally assembled donor plasmid.
21. Since even successful Gibson assembly reactions can produce a small number of colonies, it is important to use *E. coli* with as high transformation efficiency as possible. Electrocompetent *E. coli* typically have higher efficiency than chemically competent *E. coli*. A successful reaction will produce one to several hundred colonies. Performing a negative control reaction in parallel is useful to distinguish a successful low-yield reaction from non-specific colonies. Negative control reactions typically contain NEBuilder HiFi DNA Master Mix and only the backbone plasmid and scarless DsRed cassette fragments, as these are most likely to introduce contaminants from trace amounts of uncut plasmid.
22. You should remove bacterial endotoxin contaminants in the plasmid DNA prep by using two reagents from the EndoFree Plasmid Mega Kit (Buffer ER and Buffer QN). This is to reduce toxicity in injected embryos.
23. The 3xP3-DsRed marker gene is used to screen for positive integration of the intended modification [16]. Note that DsRed fluorescence in the compound eye is strong if the adult is white eyed, but fluorescence is only visible in a small number of ommatidia if the adult has a wild-type eye color, making DsRed fluorescence difficult though not impossible to observe.

24. The GMR>*eya*(RNAi) gene is a negative selection marker contained in the plasmid backbone of pBS-GMR-*eya*(shRNA) [12]. If the entire donor plasmid inserts into the *Drosophila* genome by a single crossover event, it results in transformant animals with abnormally small compound eyes due to the inhibition of *eya* expression [24]. Thus, one selects for animals that do not have this small eye phenotype. If you are editing a *Drosophila* species that is not *melanogaster*, check to see if the *eya* shRNA made from pBS-GMR-*eya*(shRNA) is perfectly complementary to the *eya* gene in that species. If not, you can use site-directed mutagenesis to modify the pBS-GMR-*eya*(shRNA) sequence for perfect shRNA complementarity.
25. You might obtain multiple positive G1 adults from the same G0 parent. These may or may not be independent genome modifications. However, you can be confident that positive G1 offspring from different G0 parents will have independent genome edits. Establish stable stocks of G1 lines that come from all unique G0 parents.
26. Excision of 3XP3-DsRed typically occurs 10% of the time or less, so make sure the number of F1 males you mate is large enough to produce hundreds of F2 progeny to screen.
27. Since the PiggyBac transposase plasmid requires active P-element transposase to integrate into an injected embryo's genome, there should be no retention of the transposase gene in G1 adults.

Acknowledgements

Financial support was provided from the NIH (F32GM122349, K. G.N.; R35GM118144, R.W.C.).

References

1. Komor AC, Badran AH, Liu DR (2017) CRISPR-based technologies for the manipulation of eukaryotic genomes. Cell 168(1–2): 20–36
2. Bassett AR, Tibbit C, Ponting CP, Liu JL (2013) Highly efficient targeted mutagenesis of *Drosophila* with the CRISPR/Cas9 system. Cell Rep 4(1):220–228
3. Gratz SJ, Cummings AM, Nguyen JN, Hamm DC, Donohue LK, Harrison MM, Wildonger J, O'Connor-Giles KM (2013) Genome engineering of *Drosophila* with the CRISPR RNA-guided Cas9 nuclease. Genetics 194(4):1029–1035
4. Kondo S, Ueda R (2013) Highly improved gene targeting by germline-specific Cas9 expression in *Drosophila*. Genetics 195(3): 715–721
5. Ren X, Sun J, Housden BE, Hu Y, Roesel C, Lin S, Liu LP, Yang Z, Mao D, Sun L et al (2013) Optimized gene editing technology for *Drosophila melanogaster* using germ line-specific Cas9. Proc Natl Acad Sci U S A 110(47):19012–19017
6. Yu Z, Ren M, Wang Z, Zhang B, Rong YS, Jiao R, Gao G (2013) Highly efficient genome modifications mediated by CRISPR/Cas9 in *Drosophila*. Genetics 195(1):289–291

7. Gratz SJ, Ukken FP, Rubinstein CD, Thiede G, Donohue LK, Cummings AM, O'Connor-Giles KM (2014) Highly specific and efficient CRISPR/Cas9-catalyzed homology-directed repair in *Drosophila*. Genetics 196(4):961–971
8. Lee JS, Kwak SJ, Kim J, Kim AK, Noh HM, Kim JS, Yu K (2014) RNA-guided genome editing in *Drosophila* with the purified Cas9 protein. G3 (Bethesda) 4(7):1291–1295
9. Port F, Chen HM, Lee T, Bullock SL (2014) Optimized CRISPR/Cas tools for efficient germline and somatic genome engineering in *Drosophila*. Proc Natl Acad Sci U S A 111(29): E2967–E2976
10. Sebo ZL, Lee HB, Peng Y, Guo Y (2014) A simplified and efficient germline-specific CRISPR/Cas9 system for *Drosophila* genomic engineering. Fly (Austin) 8(1):52–57
11. Xue Z, Ren M, Wu M, Dai J, Rong YS, Gao G (2014) Efficient gene knock-out and knock-in with transgenic Cas9 in *Drosophila*. G3 (Bethesda) 4(5):925–929
12. Nyberg KG, Nguyen JQ, Kwon YJ, Blythe S, Beitel GJ, Carthew R (2020) A pipeline for precise and efficient genome editing by sgRNA-Cas9 RNPs in *Drosophila*. Fly (Austin) 14(1–4):34–48
13. Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science 337(6096):816–821
14. Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, Hsu PD, Wu X, Jiang W, Marraffini LA et al (2013) Multiplex genome engineering using CRISPR/Cas systems. Science 339(6121):819–823
15. Haeussler M, Schonig K, Eckert H, Eschstruth A, Mianne J, Renaud JB, Schneider-Maunoury S, Shkumatava A, Teboul L, Kent J et al (2016) Evaluation of off-target and on-target scoring algorithms and integration into the guide RNA selection tool CRISPOR. Genome Biol 17(1):148
16. Gratz SJ, O'Connor-Giles KM (2021) Scarless gene editing. https://flycrispr.org/scarless-gene-editing/. Accessed 26 Aug 2021.
17. Gloor GB, Engels WR (1992) Single fly DNA preps for PCR. Dros Info Ser 71:148–149
18. Horn C, Offen N, Nystedt S, Hacker U, Wimmer EA (2003) piggyBac-based insertional mutagenesis and enhancer detection as a tool for functional insect genomics. Genetics 163(2):647–661
19. Thibault ST, Singer MA, Miyazaki WY, Milash B, Dompe NA, Singh CM, Buchholz R, Demsky M, Fawcett R, Francis-Lang HL et al (2004) A complementary transposon tool kit for *Drosophila melanogaster* using P and piggyBac. Nat Genet 36(3): 283–287
20. Gibson DG, Young L, Chuang RY, Venter JC, Hutchison CA 3rd, Smith HO (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat Methods 6(5): 343–345
21. Gompel N, Schröder EA (2021) *Drosophila* germline transformation. http://gompel.org/wp-content/uploads/2021/04/Drosophila-transformation-with-chorion.pdf. Accessed 26 Aug 2021.
22. Babon JJ, McKenzie M, Cotton RG (2003) The use of resolvases T4 endonuclease VII and T7 endonuclease I in mutation detection. Mol Biotechnol 23(1):73–81
23. Bhattacharyya A, Murchie AI, von Kitzing E, Diekmann S, Kemper B, Lilley DM (1991) Model for the interaction of DNA junctions and resolving enzymes. J Mol Biol 221(4): 1191–1207
24. Silver SJ, Rebay I (2005) Signaling circuitries in development: insights from the retinal determination gene network. Development 132(1): 3–13

Chapter 7

Tissue-Specific CRISPR-Cas9 Screening in *Drosophila*

Fillip Port and Michael Boutros

Abstract

Over the last century research in *Drosophila* has resulted in many fundamental contributions to our understanding of the biology of multicellular organisms. Many of these breakthroughs have been based on the identification of novel gene functions in large-scale genetic screens. However, conventional forward-genetic screens have been limited by the random nature of mutagenesis and difficulties in mapping causal mutations, while reverse-genetic RNAi screens suffer from incomplete knockdown of gene expression. Recently developed large-scale CRISPR-Cas9 libraries promise to address these limitations by allowing the induction of targeted mutations in genes with spatial and temporal control. Here, we provide a guide for tissue-specific CRISPR screening in *Drosophila*, including the characterization of Gal4 UAS-Cas9 lines, selection of sgRNA libraries, and various quality control measures. We also discuss confounding factors that can give rise to false-positive and false-negative results in such experiments and suggest strategies on how to detect and avoid them. Conditional CRISPR screening represents an exciting new approach for functional genomics in vivo and is set to further expand our knowledge of the molecular underpinning of development, homeostasis, and disease.

Key words CRISPR-Cas9, Genome editing, Drosophila, sgRNA libraries, Screening

1 Introduction

Genetic screens aim to identify novel gene functions by testing many genes in parallel using unbiased approaches. In *Drosophila* such screens have been particularly successful and have identified many of the genes that control development, behavior, and disease in multicellular animals [1–3]. Central to genetic screening is a scalable method to reliably perturb gene expression, ideally including the capacity to abrogate gene function. Furthermore, large-scale screening can be facilitated by reverse-genetic approaches, which use reagents that are rationally designed to target specific genes, thereby allowing to test a large number of genes with limited resources.

Traditional screening methods in *Drosophila* have a number of limitations. Mutagenesis screens using chemicals, radiation, or

Christian Dahmann (ed.), *Drosophila: Methods and Protocols*,
Methods in Molecular Biology, vol. 2540, https://doi.org/10.1007/978-1-0716-2541-5_7,

transposable elements have the capacity to knockout gene function, but only do so relatively infrequently, thereby making comprehensive screens labor and resource intensive. RNAi screens use rationally designed reagents that minimize the number of experiments needed to test a large number of genes, but only knockdown gene expression, thereby missing phenotypes due to residual gene expression.

CRISPR-Cas9 gene editing is a highly scalable method for functional genomics [4]. It makes use of the RNA-guided endonuclease Cas9, which is directed to the genomic target site by a single guide RNA (sgRNA) [5]. The Cas9/sgRNA complex initially binds DNA through a protospacer adjacent motif (PAM), which in the case of *Streptococcus pyogenes* Cas9 is NGG, and represents the only restriction on genomic target space. Otherwise, target specificity is exclusively encoded in the 5′ 20 nucleotides of the sgRNA. Since it is relatively straightforward to generate new sgRNAs, this method lends itself to the generation of large resources for functional genomic screening. Binding of the Cas9 ribonucleoprotein (RNP) particle and base pairing of the sgRNA with the target sequence activates the endonuclease activity of Cas9, resulting in a cut of both DNA strands 3 base pairs upstream of the PAM. The DNA double-strand break is recognized by the endogenous DNA repair machinery and repaired by either homology-directed repair (HDR) or non-homologous end joining (NHEJ). While the former pathway results in the repair of the lesion with high fidelity, NHEJ is an error-prone repair pathway and frequently leads to small insertions and deletions (indels) at the target site (Fig. 1).

CRISPR screening typically makes use of Cas9-induced indels, which in the coding sequence of protein-coding genes can abrogate gene function. However, in many cases only indels that represent out-of-frame mutations have a strong functional impact, while in-frame indels are often silent. Moreover, in some species it has been shown that in some instances even the effect of out-of-frame indels can be attenuated by alternative splicing or genetic compensation [6–8]. As a result the efficiency of CRISPR mutagenesis is not only limited by Cas9/sgRNA activity but also by the spectrum and position of the induced mutations. Several strategies have been developed to increase the fraction of bi-alleleic knockout cells, including algorithms to predict sgRNA activity [9], bioinformatic prediction of target sites where in-frame indels are likely to disrupt gene function [10, 11], and sgRNA multiplexing to induce several mutations or larger deletions in a gene [12, 13].

CRISPR genome engineering has been quickly adopted and combined with the sophisticated genetic tools available in the *Drosophila* model system. Early work has mostly focused on establishing methods to edit the genome in *Drosophila* germ cells, allowing the generation of heritable genome edits [14–17]. More recently, several labs have also developed techniques to acutely induce knockouts in genes in somatic tissues with spatial and

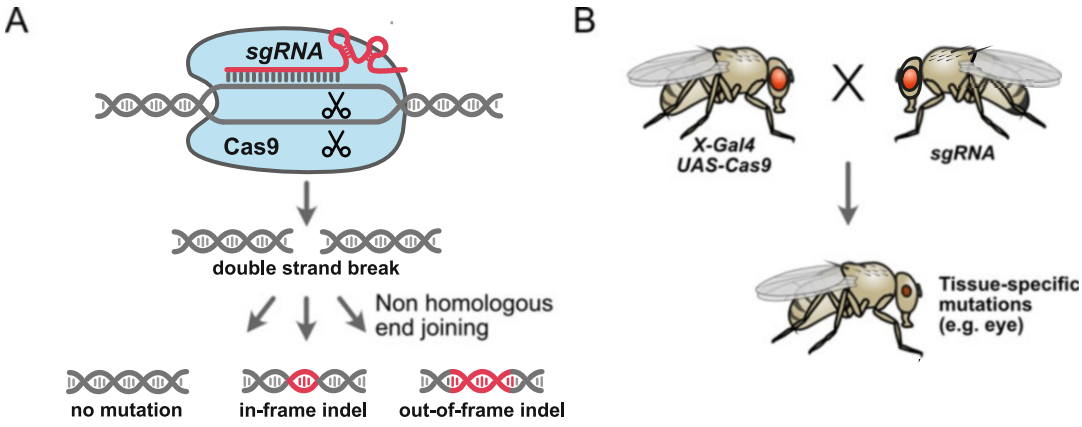


Fig. 1 CRISPR-Cas9 for gene targeting in *Drosophila*. (**a**) In CRISPR-Cas9 gene editing a small sgRNA guides the endonuclease Cas9 to the genomic target site. Binding activates the enzymatic activity, which creates a DNA double-strand break. This is recognized by the endogenous DNA repair machinery. DNA repair can result in the restoration of the original sequence, which can be cut again by Cas9, or the induction of mutations, often small insertions and deletions (indels), which can be in- or out-of-frame. (**b**) Conditional CRISPR mutagenesis in *Drosophila* involves crossing Gal4 UAS-Cas9 transgenic flies to transgenic sgRNA lines. Offspring expresses CRISPR components in Gal4-expressing cells, leading to mutations at the target locus and subsequent phenotypes

temporal control [18–21]. This is achieved by expressing CRISPR components under the control of regulatory elements from genes with restricted expression patterns, either by cloning the Cas9 coding sequence downstream of such regulatory elements or by expressing Cas9 components under control of the Gal4/UAS system. These are then combined with transgenic sgRNAs, either expressed ubiquitously or via Gal4/UAS, through a genetic cross.

Several labs are currently in the process of generating large-scale sgRNA libraries to facilitate unbiased screening for novel gene functions in *Drosophila* [19, 20, 22]. These use in part different designs, including the promoters that drive sgRNA expression and the number of sgRNAs per gene. Some libraries have already been used to mutagenize many genes either ubiquitously or in selected tissues and have been found to be highly effective, often performing significantly better than previous technologies [19, 20]. Here we provide a practical guide for performing conditional CRISPR screening in *Drosophila* and discuss strategies for quality control.

2 Materials

2.1 Transgenic Drosophila Strains Expressing Cas9

Lines that allow optimizing Cas9 expression levels independent from the strength of the Gal4 driver are available from the Vienna Drosophila Stock Center at https://stockcenter.vdrc.at [19]. UAS-uCas9 lines have the VDRC IDs 340000-340007. The HD_CFD Tools collection at VDRC also contains a number

of stocks in which UAS-uCas9 transgenes are already combined with popular Gal4 drivers. In addition, a line (VDRC ID 340008) is available that allows induction of Cas9 in negatively marked clones through FLP/FRT recombination [19]. A number of different UAS-Cas9 lines (UAS-Cas9.C, UAS-Cas9.P, UAS-Cas9.P2), as well as lines already combined with Gal4 drivers, are also available from the Bloomington Drosophila Stock Center (BDSC) at https://bdsc.indiana.edu/ [22]. Users should characterize Gal4 UAS-Cas9 lines for their intended purpose as described in Subheading 3.1.

2.2 Transgenic sgRNA Lines

Several large-scale sgRNA collections, as well as smaller collections from individual laboratories, are available from public stock centers. The Vienna Drosophila Stock Center distributes the Heidelberg CRISPR Fly Design (HD_CFD) library [19]. The Bloomington Drosophila Stock Center distributes the TRiP-KO (as well as TRiP-OE for CRISPR activation) library and the WKO library, as well as a number of smaller collections and individual lines [20–23]. Fly Stocks at the National Institute of Genetics (https://shigen.nig.ac.jp/fly/nigfly/) distributes a collection of sgRNA lines originally intended for germline mutagenesis [16]. These resources differ in their design and performance in various applications and users should consult Subheading 3.2 and Fig. 2 to choose the lines best suited for their experiments.

2.3 Fly Strains to Mark Cells with Active CRISPR-Cas9

A number of fly lines that allow to mark cells that underwent CRISPR mutagenesis have been described and are available upon request from the respective laboratories [24–26]. Note that these tools typically do not mark all cells that have been edited by CRISPR.

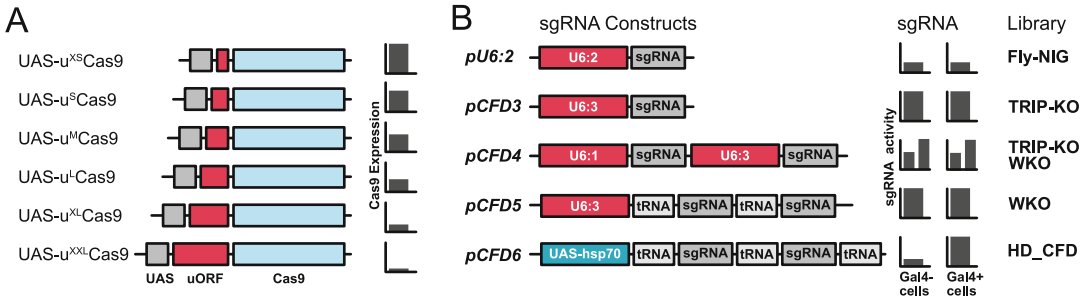


Fig. 2 Cas9 and sgRNA expression constructs for efficient CRISPR mutagenesis in *Drosophila*. (**a**) The UAS-uCas9 series comprises a number of UAS-Cas9 constructs with upstream open reading frames (uORF) in between the Cas9 coding sequence and the UAS-hsp70 promoter. Translation of the downstream Cas9 ORF is inversely correlated with the length of the uORF. Choosing the appropriate vector allows titration of Cas9 expression to optimal levels. (**b**) sgRNA expression vectors frequently used in *Drosophila* CRISPR experiments. Plasmids differ in their use of promoters for sgRNA expression and whether they allow sgRNA multiplexing or not. Transgenes with different U6 promoters have been shown to mediate gene editing with different efficiency [18]

2.4 Plasmids

There are a number of sgRNA expression plasmids available from Addgene (https://www.addgene.org/) that allow generation of transgenic sgRNA lines. We recommend use of *pCFD5* (Addgene 112645) for ubiquitous expression of sgRNAs with high efficiency and *pCFD6* (Addgene 73915) for expression of sgRNAs under control of Gal4, which leads to tighter restriction of mutagenesis to Gal4-expressing cells [12].

In case users wish to generate UAS-uCas9 lines with insertion at different genomic loci the plasmids are also available from Addgene (Plasmids 127382 – 127387).

2.5 Antibody Staining

To profile cells and tissues expressing CRISPR-Cas9 components we recommend the following antibodies: Rabbit anti-phospho-Histone 3 (Millipore, Cat# 06-570) to stain mitotic cells, rabbit anti-cleaved *Drosophila* Dcp-1 (Millipore, Cat# AB3623) to stain cells undergoing apoptosis, and mouse anti-Fasciclin III (DSHB, Clone 7G10) to visualize cell morphology.

3 Methods

3.1 Generation and Characterization of Gal4 UAS-Cas9 Fly Lines

To induce mutations only in selected target tissues and at specific stages during development the expression of Cas9 is typically controlled via the Gal4/UAS system (*see* **Note 1**). A number of Gal4 UAS-Cas9 stocks are publicly available from the Vienna Drosophila Resource Center (VDRC) and the Bloomington Drosophila Stock Center (BDSC) (*see* Subheading 2.1 and **Note 2**). However, not all of these stocks contain the latest Cas9 constructs and often users will be required to generate their own lines. To this end UAS-Cas9 transgenes are combined with a Gal4 driver of interest by conventional genetic crosses. A number of different UAS-Cas9 constructs have been described [12, 18, 19]. We recommend use of the UAS-uCas9 series, as it allows to fine tune Cas9 expression levels (as detailed in Subheading 3.1.1). It can be desirable to recombine the Gal4 driver and UAS-Cas9 transgene on the same chromosome to simplify subsequent crosses (*see* **Note 3**). Eventually stable stocks are generated that harbor the Gal4 driver and the Cas9 transgene, which can then be crossed to transgenic sgRNA flies.

1. Check if suitable Gal4 UAS-Cas9 lines are already available from the VDRC or BDSC. Order lines that are available and proceed to **step 3.1.1**.
2. If no suitable lines are readily available, order individual UAS-Cas9 transgenic lines. We recommend use of the UAS-uCas9 series developed by our lab, which is available from the VDRC (*see* **Note 2**). We typically use UAS-u^{S}Cas9 for weak Gal4 drivers, UAS-u^{M}Cas9 for medium Gal4 drivers

and UAS-u^{L}Cas9 for very strong Gal4 lines. However, we recommend to establish the optimal line empirically by testing several UAS-uCas9 transgenes in parallel.

3. Cross UAS-uCas9 lines to the Gal4 driver.
4. Select Gal4 UAS-uCas9 transgenic flies (use virgin females in case both transgenes are recombined on the same chromosome) and cross to a suitable balancer line.
5. Select balanced Gal4 UAS-uCas9 transgenic flies and set up a stable transgenic stock.

3.1.1 Testing for Cas9-Mediated Toxicity

Cas9 protein at high expression levels is toxic to flies. It is unclear what the mechanism behind this toxicity is, but it does not require the ability to cut DNA, as high levels of nuclease-dead versions of the enzyme are also detrimental [18]. Toxicity can be so severe that it leads to lethality of the animal. At intermediate expression levels effects can be more subtle, such as an increase in the number of apoptotic cells, an influence on cell shape or the physiological response to external stimuli. Fortunately, relatively low levels of Cas9 expression are sufficient for highly efficient gene editing and are well tolerated in flies [19]. It is therefore important to optimize Cas9 expression levels to balance activity and toxicity.

Expression levels of Cas9 from UAS transgenes will depend on a number of factors: the strength of the Gal4 driver, the sequence of the UAS vector, and the insertion site of the transgene in the genome. We have recently developed a series of UAS-Cas9 transgenes that uses upstream open reading frames (uORFs) between the UAS regulatory region and the Cas9 coding sequence to predictably modulate the levels of Cas9 translation (UAS-uCas9 series [19]). With these transgenes it is possible to titrate levels of Cas9 independent of the strength of the Gal4 line and ensure an optimal balance between high gene editing efficiency and low toxicity.

1. Verify that the chromosome harboring the uCas9 transgene in your Gal4 UAS-uCas9 line is homozygous viable (provided it is not recombined with a recessive lethal Gal4 driver). All UAS-uCas9 transgenes are inserted in attP landing sites that are homozygous viable (*see* **Note 4**). Homozygous strains facilitate downstream work and indicate that Cas9 is not expressed in amounts that impair viability.
2. Compare the viability, proliferation, and morphology of Cas9-expressing cells in the Gal4 UAS-uCas9 strain with a line expressing only Gal4. Viability can be tested by performing anti-cleaved Dcp-1 staining, which marks cells undergoing apoptosis. Proliferating cells can be identified by an anti-phospho Histone 3 staining. Cell morphology can be monitored by staining cells with antibodies against cell surface markers (e.g., FasIII).

3.1.2 Testing for On-Target Efficiency of Gal4 UAS-uCas9 Lines

The aim of the previous step was to select a line with low enough Cas9 expression to not cause any adverse effects. However, for the success of conditional gene editing it is important that Cas9 is present in sufficient amounts to efficiently mediate DNA double-strand breaks at the target locus. It is therefore also important to test, if the selected fly strains can mediate high gene editing activity.

There are multiple ways to assay Cas9-mediated mutagenesis at the target locus. The most direct and quantitative way to do this is via PCR amplification of the target locus coupled with high-throughput sequencing of the amplicons. However, such assays can be time consuming and expensive to do and are difficult to interpret when the Gal4 driver is only active in a small fraction of cells. Instead, we propose to test Cas9 activity through the induction of relevant phenotypes. While less sensitive and not directly reflecting Cas9 activity (*see* Subheading 3.5 for a detailed discussion), this assay has the most practical relevance for downstream applications.

1. Order several sgRNA lines targeting genes in which mutations are known to result in a clearly recognizable phenotype in the target tissue. If no such genes are known, sgRNAs targeting cell essential genes can be used and are expected to result in the death of the target cells.
2. Cross transgenic sgRNA flies to Gal4 UAS-uCas9 flies and incubate crosses at 25 °C. Include a Gal4 UAS-Cas9 line with a Cas9 construct resulting in high expression levels (e.g., UAS-Cas9.P2).
3. Record and compare the induced phenotype.
4. The ideal Gal4 UAS-uCas9 line will induce phenotypes that are comparable to the ones induced by high-expression Cas9 constructs, but not result in non-specific phenotypes in the absence of sgRNAs.

3.1.3 Analysis of Spatial Mutagenesis Patterns

Expression of CRISPR components under the Gal4/UAS system aims to restrict mutagenesis in time and space, thereby allowing for a detailed analysis of gene function in the context of a multicellular organism. In many cases researchers will already have Gal4 lines in the lab, which they know to express at a given developmental stage in a desired pattern. However, using such lines to control expression of Cas9 will in some cases lead to mutations also in other cells and tissues. This is mainly due to two reasons: First, transgenes downstream of UAS promoters can be expressed at very low levels also in the absence of Gal4 (so-called leaky expression). Such low level expression can be sufficient for Cas9-mediated mutagenesis. Second, CRISPR gene editing leads to permanent mutations that

are passed on during mitosis to both daughter cells. As a result the animal will acquire mutations in all cell lineages that expressed Gal4 at any stage during development. Of note, both of these factors play much less of a role when Gal4 lines are used to drive expression of a fluorescent protein, a method that is typically used to establish the expression pattern of a given Gal4 line. For this reason, mutagenesis patterns with CRISPR are sometimes substantially different from expectation. It is therefore advisable to experimentally test the spatial distribution of mutagenesis, in particular, if strict spatial control is an important factor for the success of the experiment.

The easiest way to visualize gene editing activity in whole animals is by using a fluorescent reporter that marks cells with Cas9-induced mutations. Several labs have developed such systems [24–26]. Upon successful gene targeting the induced mutations shift the reading frame or remove or inactivate an inhibitory sequence of the fluorophore coding sequence, thereby rendering the gene functional and the cell fluorescent. The advantage is that such systems allow visualization of gene edited cells with minimal work and without prior knowledge of the mutagenesis pattern. However, typically not all editing events result in reporter activity (*see* **Note 5**). Hence, the pattern of labeled cells is variable between different animals and an underestimate of the true gene editing pattern.

It is important to note that the spatial and temporal control of CRISPR activity does not only depend on the Gal4 line and UAS-Cas9 construct but is also significantly influenced by the sgRNA expression vector. While ubiquitous sgRNA expression from U6 promoters favors ubiquitous mutagenesis, sgRNA expression from UAS plasmids leads to CRISPR activity that is largely dependent on Gal4 (*see* below for a detailed discussion). It is therefore imperative that CRISPR activity reporters are induced with sgRNAs expressed from the same plasmid as the sgRNAs used in later experiments.

1. Order reporter lines for CRISPR activity. The sgRNA targeting the reporter must be encoded in the same expression vector as will be used later on.
2. Cross reporter flies to the Gal4 UAS-uCas9 line to be tested and incubate under the same conditions as to be used in later experiments.
3. Visualize fluorescent cells at different levels (e.g., whole body and dissected tissues). Fluorescent cells indicate they have undergone CRISPR mutagenesis. Make sure to analyze several animals and aggregate results, as stochastically some edited cells will not activate the reporter (*see* **Note 5**).

3.2 Selection of Transgenic sgRNA Lines

sgRNAs determine at which genomic locus Cas9 will cut the DNA, leading to mutations after imprecise DNA repair. Tissue-specific CRISPR screens use collections of transgenic sgRNAs stably integrated into the *Drosophila* genome. Several labs have created such large-scale collections and several thousand of them are now available from public *Drosophila* stock centers [19, 20, 22]. Should these libraries not yet contain lines targeting genes of interest users can generate new sgRNA lines themselves, which is relatively straightforward and has been previously described [12] (*see* **Note 6**).

sgRNAs are generally designed to target unique sequences in the genome with few if any highly homologous sequences elsewhere, thereby reducing the likelihood of off-target mutagenesis. Furthermore, target sites are chosen such that sgRNAs direct Cas9 to the coding sequence of protein-coding genes (*see* **Note 7**). Mutations that change the reading frame have a high chance to disrupt gene function, but functional in-frame mutations can also occur and reduce the phenotypic penetrance of CRISPR mutagenesis (*see* below).

The different sgRNA resources differ mainly in the choice of promoter used for sgRNA expression and the number of sgRNAs targeting each gene (Fig. 2). Commonly used promoters include RNA polymerase III promoters of the U6 ribosomal RNAs and the UAS-hsp70 promoter present in most UAS vectors. There are three U6 genes in *Drosophila* and early work has shown that gene editing efficiency is affected by which promoter is used [18]. The U6:3 promoter mediates editing with high efficiency, while U6:2 is significantly less active and U6:1 has intermediate activity. All U6 promoters mediate ubiquitous expression of sgRNAs. In contrast the UAS-hsp70 promoter cassette is dependent on Gal4 and can therefore be expressed in a tissue-specific manner. However, UAS-hsp70 is a RNA polymerase II promoter and results in the transcription of mRNAs, which are capped, polyadenylated and exported into the cytoplasm, which is not compatible with sgRNA activity [12, 27]. To overcome this problem sgRNAs expressed from UAS vectors are flanked by tRNAs, which mediate excision of mature sgRNAs by the endogenous tRNA processing machinery [12, 13].

The different expression strategies have complementary strengths and weaknesses. U6 promoters, and in particular U6:3, reliably mediate high CRISPR-Cas9 activity, but in our experience frequently also mediate mutagenesis in cells that do not express Gal4, due to leaky Cas9 expression. In contrast, sgRNAs expressed from UAS-hsp70 are largely dependent on Gal4 (*see* **Note 8**) and in combination with UAS-Cas9 constructs result in much tighter restriction of mutagenesis to Gal4-expressing cells [12]. However, sgRNA levels can become limiting when these constructs are combined with weak Gal4 drivers [20].

The second aspect that differentiates publicly available sgRNA transgenes is the number of sgRNAs encoded in each line. Currently, almost all lines encode either one or two sgRNAs targeting the same gene. It has been shown that multiplexing sgRNAs leads to higher mutagenesis efficiency and more penetrant phenotypes [12]. This is because additional sgRNAs can compensate for sgRNAs with low efficiency and inducing independent mutations in the same gene increases the chances of abrogating gene function and can lead to the induction of larger deletions. However, the use of multiple sgRNAs also increases the risk of off-target mutagenesis and phenotypes caused by a general response to DNA damage.

3.2.1 The Heidelberg CRISPR Fly Design Library (Boutros Lab)

The lines of the Heidelberg CRISPR Fly Design (HD_CFD) library encode two sgRNAs in the *pCFD6* expression vector and have been described in Port et al., 2020 [19]. *pCFD6* is a UAS vector and mutagenesis with this library has been observed to be well restricted to Gal4-expressing cells. Each sgRNA pair targets sites preferentially located in the 5′ half of the coding sequence and usually spaced apart by approximately 500 bp. Characterization of a random sample of sgRNA lines indicated that the large majority of sgRNA lines is active and most lines generate indels at both target sites. Deletions between target sites also occur, but are less frequent. Phenotypic analysis of a large number of HD_CFD lines further supported the high efficiency of this library, with 80–90% of lines giving rise to the expected phenotype.

The HD_CFD library contains mostly sgRNAs targeting transcription factors, kinases, and phosphatases, as well as a number of genes with human homologs implicated in disease. There are currently about 2000 HD_CFD lines available from the VDRC.

3.2.2 The TRIP Knockout Lines (Perrimon Lab)

The Transgenic RNAi Project (TRiP) is producing sgRNA lines for gene activation (TRiP-OE) and gene knockouts (TRiP-KO) and these resources are described in Zirin et al., 2020 [22]. Since this chapter focuses on CRISPR loss-of-function screens, we will only describe the TRiP-KO library. This library currently consists of over 2000 transgenic lines, the majority of which encodes single sgRNAs in *pCFD3*. This vector expresses sgRNAs under the control of the strong, ubiquitous U6:3 promoter. A smaller proportion of lines express two sgRNAs per gene encoded in *pCFD4*, which uses two different U6 promoters to drive expression of each sgRNA. The TRiP has announced that production of TRiP-KO lines has been shifted to using the *pCFD6* vector to improve the Gal4 dependency of mutagenesis [22]. The lines of the TRiP sgRNA libraries are available via the Bloomington Drosophila Stock Center (BDSC).

3.2.3 The Weizmann Knockout Project Lines (Schuldiner Lab)

sgRNA lines of the Weizmann knockout project encode two sgRNAs per gene and over 300 lines are currently available from the Bloomington Drosophila Stock Center. This resource is described in Meltzer et al. 2019 [20]. sgRNAs are expressed from either *pCFD4* or *pCFD5*, which both mediate ubiquitous expression of sgRNAs. An evaluation of these tools in the nervous system suggested that sgRNAs expressed from *pCFD5* have higher efficiency [20]. This might be due to the fact that in this vector both sgRNAs are expressed from the most efficient U6:3 promoter and are flanked by tRNAs. Targets covered by the WKO library are enriched for genes implicated in neuronal development, but also encompass genes involved in other biological processes. The efficiency of this library has been successfully demonstrated through a screen in the *Drosophila* mushroom body [20].

3.2.4 The NIG sgRNA Lines (Kondo and Ueda Lab)

The fly facility at the National Institute of Genetics in Japan has created a large collection of transgenic sgRNA strains mainly targeting genes on the second chromosome. The experimental strategy has been previously described [16], but so far no report of a larger-scale screen with these lines has been published. The lines are mainly intended to create germline alleles. They can also be combined with a conditional Cas9 source for somatic mutagenesis, but genetic mosaicism is expected to be particularly pronounced, since they encode a single sgRNA expressed under the weakest U6 promoter.

3.2.5 Other Publicly Available sgRNA Collections

At the time of writing the Bloomington Drosophila Stock Center also distributed 99 sgRNA lines generated by the Rhagu lab targeting genes implicated in the regulation of phosphoinositides [23]. These encode two sgRNAs per gene expressed from a U6:2 promoter. The target sites are located at the beginning and end of the coding sequence of each gene. Deletions between the target sites are therefore deleting the entire gene, creating null mutations. However, in cases where such deletions do not occur, individual indels at the downstream target site have a relatively low chance of disrupting gene function.

In addition, the BDSC also distributes 54 sgRNA lines generated by the Han lab [21, 26]. These encode single sgRNAs expressed from the U6:3 promoter.

3.3 Controls and Experimental Conditions

The goal of CRISPR screening is to reveal phenotypes induced by mutations in the target gene. However, a number of other factors have the potential to induce phenotypes independent of on-target mutagenesis. Such factors include toxicity associated with excessive amounts of Cas9 (as discussed above) and the cellular response to the DNA damage that is caused by Cas9. The goal of negative controls is to detect such effects:

3.3.1 Negative Controls

1. Order sgRNA lines targeting genes not expressed in the target tissue or known to not be involved in the process of interest. We do not recommend the use of "non-targeting" sgRNAs, as they do not induce DNA damage and, therefore, miss an often important contribution to non-specific phenotypes.
2. Cross sgRNA flies to Gal4 UAS-uCas9 flies and incubate under the same conditions as envisioned for the screen. Raise in parallel non-transgenic flies and animals that only harbor sgRNA transgenes or Gal4 UAS-uCas9 transgenes.
3. Compare the phenotype of animals expressing an active CRISPR system to animals harboring only individual components or are not transgenic.

3.3.2 Positive Controls

Positive controls will confirm or refute that a method, here conditional CRISPR-Cas9 mutagenesis, is suitable to uncover genes that play a role in the process of interest. To this end, mutagenesis is performed with sgRNAs targeting genes already known to be involved and where the phenotype upon conditional loss of function is established. Occasionally, when performing particularly innovative screens, no such prior knowledge might be available. In such cases, it is advisable to perform mutagenesis with sgRNAs targeting genes that are essential to cell survival or other recognizable phenotypes, to establish that efficient mutagenesis can be performed under the chosen conditions:

1. Order sgRNA lines targeting genes known to be involved in the process of interest and where a phenotype of gene knockout is known or can be inferred.
2. Cross sgRNA lines to Gal4 UAS-uCas9 lines and incubate under the same conditions as planned for the screen.
3. Compare the phenotype of animals expressing the positive-control sgRNAs to animals expressing negative-control sgRNAs.

3.3.3 Experimental Conditions

A relatively low number of sgRNA lines that are typically used as positive and negative controls make it feasible to test different experimental conditions to find the optimal set-up for a screen. Parameters that should be tested include the number of flies in each vial, whether vials need to be flipped at a certain time point and the temperature at which the crosses are incubated. Furthermore, the exact time point at which the phenotype is recorded needs to be established, with a longer time between induction of CRISPR mutagenesis and analysis favoring stronger phenotypes. It is also advisable to test the variation between replicates of the same cross and to establish if differential phenotypes are observed in male and female animals.

3.3.4 Pilot Screen

Once the experimental conditions are established, a pilot screen should be performed to further optimize the screening workflow. While experiments with positive and negative controls will give information about the feasibility of a screen, only a pilot screen with a larger number of sgRNAs will reveal the breath and variability of phenotypes and highlight potential problems that might arise once experiments are done at large scale:

1. Order a large selection of sgRNA lines. These will typically be several tens to a few hundred lines. Include lines targeting genes that you know or suspect might be involved in the process of interest, some that are highly unlikely to be involved and a number of genes that represent a random selection of the genes you plan to test in a future screen.
2. Cross sgRNA lines to Gal4 UAS-uCas9 lines and incubate under the same conditions as planned for the screen.
3. Establish a reliable system for setting up and tracking each cross during the screen, recording of the phenotype, and analysis of the data.

3.4 Confounding Factors

While conditional CRISPR mutagenesis is now well established and has been found to often yield better results than previous technologies, there remain some factors that can occasionally lead one astray. These can be broadly classified as false-negative and false-positive results.

3.4.1 Effects That Can Give Rise to False-Negative Results

Low Mutagenesis Efficiency

In some cases, animals expressing Cas9 and a gene-specific sgRNA will not or only infrequently acquire mutations at the target locus. The most likely cause is inactive sgRNAs. It is long established that sgRNAs vary in their potency to guide and activate Cas9. While there is a large literature on sgRNA activity prediction, the resulting algorithms have often been found to be poor predictors of sgRNA activity in *Drosophila* [28]. This is likely due to the fact that such models are typically based on data from transient assays performed in mammalian cells, which differ in key parameters from transgenic CRISPR systems in vivo. Furthermore, genomic variation, rather than sgRNA activity per se, can be the basis of low on-target mutagenesis, due to polymorphisms at the target site. As a result, it is currently not possible to reliably predict sgRNAs libraries consisting of only highly active sgRNAs.

Another factor that has been shown to limit CRISPR mutagenesis activity is chromatin context. Target sites in heterochromatic regions of the genome are less accessible to Cas9, although some mutagenesis has been detected also at such sites. Since chromatin is dynamic and differs between cell types and states it is expected that activity of the same sgRNA will vary depending on where and when it is used.

mRNA and Protein Stability

After successful bi-allelic CRISPR mutagenesis only mutant copies of mRNA are transcribed. However, existing functional copies of mRNA will remain for some time and continue to be translated. The translated protein will also persist in the cell and perform its function until it is degraded. As a result at what point mutant phenotypes can be observed will depend on the level of preexisting mRNA and protein and their decay kinetics. This can be problematic in early stages of development or for experimental designs in which the CRISPR system is only activated shortly before phenotypes are recorded. Furthermore, actively dividing cells are likely to acquire phenotypes faster, as preexisting mRNA and protein is rapidly diluted in the growing cell population.

Silent Mutations and Genetic Compensation

Conditional CRISPR mutagenesis is based on error-prone DSB repair by non-homologous end joining, which typically leads to a variety of mutations at the target site, most prominently small indels. Often not all such mutations impact gene function. For example, in-frame mutations are often silent, as many proteins tolerate the deletion or insertion of a few amino acids at many positions. Furthermore, research in other organisms has established that even the impact of out-of-frame mutations can sometimes be suppressed, by mechanisms, such as alternative splicing, translational initiation at downstream start codons and upregulation of genes with overlapping functions [6, 7].

Genetic Mosaics

While sgRNA lines that completely fail to mediate on-target mutagenesis are rare, so are lines that lead to bi-alleleic knockout of the target gene in all cells. In the majority of cases CRISPR mutagenesis with one or two sgRNAs leads to genetic mosaics, in which cells that have been independently edited by Cas9 carry different mutations (Fig. 3). Some of these cells carry silent mutations or non-mutated alleles. This heterogeneity can attenuate phenotypes or lead to significant variation in observed phenotypes.

3.4.2 Effects That Can Give Rise to False-Positive Results

Off-Target Mutagenesis

The natural function of CRISPR-Cas9 as an adaptive immune system targeting quickly evolving pathogens means that it is unlikely to have evolved absolute target specificity. It is therefore not surprising that Cas9 can tolerate a certain degree of mismatches between the target and the sgRNA protospacer. In the context of genome editing this can lead to unwanted off-target effects, where Cas9 induces DSBs at loci elsewhere in the genome with incomplete homology to the sgRNA. It has been shown that the number of mismatches that can be tolerated are sgRNA specific and not easily predicted [29]. Unfortunately, a genome-wide assessment of off-target effects across many sgRNAs in *Drosophila* is currently lacking. However, results from large-scale experiments utilizing many sgRNAs suggest that off-target effects are not so common that they would lead to a high number of false-positive results [19, 20]. Nevertheless, it is imperative to exclude a causal role of off-target mutations in follow-up experiments (*see* below).

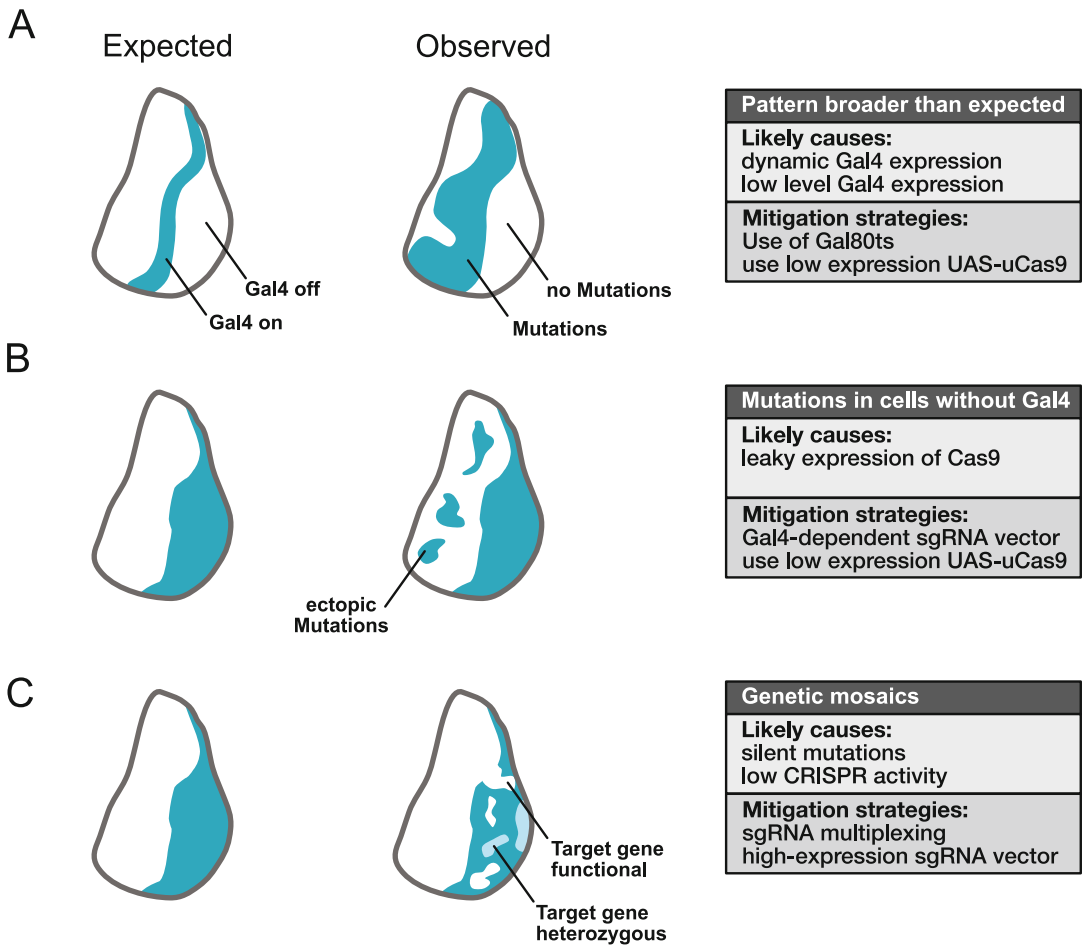


Fig. 3 Potential limitations of conditional CRISPR experiments and strategies to avoid them. (**a**) CRISPR mutagenesis can in some cases be observed in broader pattern than the known Gal4 expression domain. This can be due to broad Gal4 expression in early development and Gal4 expression below the detection limit of other methods. To mitigate such effects Cas9 lines without excessive expression levels should be used. Gal4 expression can be further suppressed with temperature-dependent Gal80. (**b**) Mutations can be observed in cells that do not express Gal4. Ectopic mutations are the result of low level "leaky" expression of Cas9 in combination with sufficient amounts of sgRNA. To mitigate this effect tRNA-flanked sgRNAs can be expressed from the UAS vector *pCFD6*, which substantially improves Gal4 dependency of mutagenesis. (**c**) Conditional CRISPR mutagenesis frequently gives rise to genetic mosaics, which comprise cells with one, two, and no functional alleles of the target gene. These are typically caused by the induction of functional in-frame mutations and sgRNAs with low activity. To increase the frequency of bi-allelic knockout cells sgRNAs can be multiplexed to induce multiple mutations in the same gene

Large Deletions and Genomic Rearrangements

While small indels seem to be the dominant form of mutations induced by CRISPR-Cas9, several others have been documented. These include large deletions at the on-target locus, genomic rearrangements, including translocations and inversions, and chromotypsis [30–33]. Importantly, most assays used to profile CRISPR-induced mutations will not detect such events, as they are typically

PCR-based and these genomic aberrations delete one or both primer binding sites. This is particularly problematic in samples that also contain alleles with small indels or which are not edited. Larger genomic alterations have the potential to give rise to phenotypes not related to the target gene, for example, by deleting neighboring genes or regulatory elements that can act on distantly located genes.

Loss of Heterozygosity

It has been shown in several organisms that DSBs caused by CRISPR-Cas9 can lead to mitotic recombination, when sister chromatids from homologous chromosomes are exchanged [24, 34–36]. This phenomenon is particularly common in *Drosophila*, because the homologous chromosomes pair during mitosis. The consequence of such recombination events is that the daughter cells become homozygous for the region of the chromosome that is distal to the cut site. If this region carries recessive mutations, these can give rise to phenotypes not observed in heterozygous tissue and not related to the CRISPR target gene.

3.5 Strategies to Confirm Causality of On-Target Mutations

A CRISPR screen will give a list of sgRNA lines that are associated with phenotypes of interest. However, as detailed in the previous section in some cases, such phenotypes could represent false-positive results and it is therefore crucial to test, if there exists a causal relationship between mutations in the target gene and the observed phenotype.

3.5.1 Perform Mutagenesis with Independent sgRNA Lines

In some cases, publicly available sgRNA libraries already contain independent lines targeting the same gene. If only a single line targeting the candidate gene exists, additional lines can be generated with relatively little effort. Lines targeting different positions within the same gene differ in their protospacer sequence and are therefore very unlikely to share off-target sequences with incomplete sequence homology. Therefore, phenotypes that are shared between independent lines are unlikely to arise from mutations caused by off-target cutting of Cas9. However, independent lines could still share other off-target effects, such as those caused by loss of heterozygosity or deletion of neighboring genes. If different phenotypes are obtained with lines targeting the same gene this might indicate differences in on-target knockout efficiency or the presence of off-target effects not shared between the lines.

3.5.2 Perturb Gene with Alternative Methods

To strengthen the link between the loss of function of a gene and an observed phenotype, it is often beneficial to perturb gene expression by orthogonal methods. For example, RNAi can be used to knockdown mRNA levels or nanobodies fused to degrons can mediate protein turnover of tagged proteins [37, 38]. While shared phenotypes are a strong indication for a causal relationship between gene function and the observed phenotype, differential outcomes

are more difficult to interpret. These might arise due to off-target effects of one or the other method, but often reflect the different kinetics and strength of the different perturbation methods. For example, RNAi is often limited by a substantial amount of residual mRNA expression and a failure to reproduce a phenotype with this method could simply reflect this fact.

3.5.3 Create Sequence Verified Germline Alleles

The creation of a sequence verified germline allele is an essential step during the in-depth characterization of a gene. To this end sgRNA lines can be crossed to Cas9 lines combined with a germline driver (such as nanos-Gal4). Offspring will have an active CRISPR system in the germline and can give rise to mutant progeny. Mutant alleles can be sequenced using PCR amplicons generated with primers flanking the mutation sites. Furthermore, mutant alleles can be backcrossed into a desired genetic background, which will strongly reduce the likelihood of other genetic alterations being linked to the on-target mutation. Often multiple different alleles can be generated in a single experiment and crossed together to generate trans-heterozygous animals for analysis. In cases where knockout animals are not viable, mutations can be induced directly on FRT chromosomes to combine with conditional expression of FLP recombinase to induce homozygous mutant cells in heterozygous animals.

3.5.4 Rescue the Phenotype

Rescue experiments are a powerful way to confirm the link between gene function and phenotype. If an observed phenotype is caused by the absence of a certain gene, reintroduction of that gene should rescue the phenotype. This can be done by inserting a plasmid encoding the gene into the mutant fly strain. Using UAS plasmids will allow testing where in the organism the gene needs to be expressed in order to rescue the phenotype. However, transgene expression via the Gal4/UAS system typically leads to highly unphysiological expression levels. To express a gene at physiological levels and in its natural expression pattern, exogenous DNA sequences containing all regulatory elements are required. Bacterial artificial chromosomes are one way to insert such typically large stretches of DNA into the fly genome. An alternative approach is to correct the originally created mutation though CRISPR-assisted HDR.

4 Notes

1. An alternative approach is to directly express Cas9 under the control of tissue-specific regulatory elements by encoding them in the same plasmid or by inserting the Cas9 sequence into an endogenous locus by homology-directed repair [21, 26]. The

advantage of this approach is that the Gal4/UAS system can be used in parallel for other purposes and that Cas9 expression is induced faster than via Gal4. However, this method often leads to mutations in undesired locations, as it does not allow for dual conditional control of Cas9 and sgRNA expression, and Cas9 expression levels cannot be easily tuned for an optimal balance between activity and toxicity.

2. Information about available Gal4 UAS-Cas9 stocks are available at https://stockcenter.vdrc.at/control/library_hdcfd and https://bdsc.indiana.edu/stocks/genome_editing/triptoolbox_grna.html. We recommend use of UAS-uCas9 transgenes.
3. For some applications users will need to add other transgenes in their Gal4 UAS-Cas9 stocks, such as UAS-Fluorophore to label Cas9-expressing cells (note that this is an acute label and does not cover cells that might have expressed Cas9 at earlier time points during development) and tub-Gal80ts, which can be used to inhibit Cas9 expression by temperature.
4. Sometimes chromosomes will not become homozygous not because of toxicity of the transgene but because of unrecognized recessive lethal mutations present in the original stock.
5. CRISPR-Cas9 mutagenesis can lead with some frequency to larger deletions at the target locus, which can delete the promoter or fluorophore reporter in such constructs. The result are mutant cells without fluorophore expression. Since these events occur stochastically, the best remedy is to average the expression pattern of the reporter over many animals.
6. We recommend the use of *pCFD5* for ubiquitous expression of sgRNA and *pCFD6* for Gal4-dependent sgRNA expression. Step-by-step protocols for cloning of one or several sgRNA into these vectors are available under http://www.crisprflydesign.org/ and as supplementary protocol in ref. 12.
7. For other applications, such as CRISPR activation or CRISPR interference, sgRNA is designed to target other locations (e.g., the transcriptional start site).
8. While expression of sgRNAs from the UAS vector *pCFD6* is largely dependent on the presence of Gal4, some ubiquitous expression also exists. This is likely at least in part due to the presence of sequences encoding tRNAs, which have been shown to act as RNA polymerase III promoters. As a result, some mutagenesis in non-Gal4-expressing cells is also present when sgRNAs are expressed from *pCFD6*, although much reduced in comparison to strong ubiquitous sgRNA vectors.

Acknowledgments

The authors would like to thank Jun Zhou for helpful discussions and comments on the manuscript; Claudia Strein, Mona Stricker, and other members of the CRISPR Fly Design team in the Boutros lab for generating and characterizing CRISPR tools; Erich Brunner, Ben Ewen-Campden, Stephanie Mohr, Jonathan Zirin, and Norbert Perrimon for sharing unpublished information; and Lisa Meadows, Reinhard Klug, and the VDRC for distributing fly lines. Work in the laboratory of M.B. is in part supported by the European Research Council (ERC DECODE) and the DFG (SFB/TRR186, SFB1324).

References

1. Nüsslein-Volhard C, Wieschaus E (1980) Mutations affecting segment number and polarity in Drosophila. Nature 287:795–801
2. Boutros M, Ahringer J (2008) The art and design of genetic screens: RNA interference. Nat Rev Genet 9:554–566
3. St Johnston D (2002) The art and design of genetic screens: Drosophila melanogaster. Nat Rev Genet 3:176–188
4. Pickar-Oliver A, Gersbach CA (2019) The next generation of CRISPR–Cas technologies and applications. Nat Rev Mol Cell Biol 20: 490–507
5. Jinek M et al (2012) A programmable dual-RNA–guided DNA endonuclease in adaptive bacterial immunity. Science 337:816–821
6. El-Brolosy MA et al (2019) Genetic compensation triggered by mutant mRNA degradation. Nature 568:193–197
7. Smits AH et al (2019) Biological plasticity rescues target activity in CRISPR knock outs. Nat Methods 16:1087–1093
8. Mou H et al (2017) CRISPR/Cas9-mediated genome editing induces exon skipping by alternative splicing or exon deletion. Genome Biol 18:108
9. Cui Y, Xu J, Cheng M, Liao X, Peng S (2018) Review of CRISPR/Cas9 sgRNA design tools. Interdiscip Sci 10:455–465
10. Molla KA, Yang Y (2020) Predicting CRISPR/Cas9-induced mutations for precise genome editing. Trends Biotechnol 38:136–141
11. Michlits G et al (2020) Multilayered VBC score predicts sgRNAs that efficiently generate loss-of-function alleles. Nat Methods 17:708–716
12. Port F, Bullock SL (2016) Augmenting CRISPR applications in Drosophila with tRNA-flanked sgRNAs. Nat Methods 13: 852–854
13. Xie K, Minkenberg B, Yang Y (2015) Boosting CRISPR/Cas9 multiplex editing capability with the endogenous tRNA-processing system. Proc Natl Acad Sci U S A 112:3570–3575
14. Gratz SJ et al (2013) Genome engineering of Drosophila with the CRISPR RNA-guided Cas9 nuclease. Genetics 194:1029–1035
15. Bassett AR, Tibbit C, Ponting CP, Liu J-L (2013) Highly efficient targeted mutagenesis of Drosophila with the CRISPR/Cas9 system. Cell Rep 4:220–228
16. Kondo S, Ueda R (2013) Highly improved gene targeting by germline-specific Cas9 expression in Drosophila. Genetics 195: 715–721
17. Port F, Bullock SL (2016) Creating heritable mutations in drosophila with CRISPR-Cas9. Methods Mol Biol 1478:145–160
18. Port F, Chen H-M, Lee T, Bullock SL (2014) Optimized CRISPR/Cas tools for efficient germline and somatic genome engineering in Drosophila. Proc Natl Acad Sci U S A 111: E2967–E2976
19. Port F et al (2020) A large-scale resource for tissue-specific CRISPR mutagenesis in Drosophila. elife 9:e53865
20. Meltzer H et al (2019) Tissue-specific (ts)-CRISPR as an efficient strategy for in vivo screening in Drosophila. Nat Commun 10: 2113

21. Poe AR et al (2019) Robust CRISPR/Cas9-mediated tissue-specific mutagenesis reveals gene redundancy and perdurance in drosophila. Genetics 211:459–472
22. Zirin J et al (2020) Large-scale transgenic drosophila resource collections for loss- and gain-of-function studies. Genetics 214:755–767
23. Trivedi D et al (2020) A genome engineering resource to uncover principles of cellular organization and tissue architecture by lipid signaling. elife 9:e55793
24. Brunner E et al (2019) CRISPR-induced double-strand breaks trigger recombination between homologous chromosome arms. Life Sci Alliance 2(3):e201800267
25. Garcia-Marques J et al (2019) Unlimited genetic switches for cell-type-specific manipulation. Neuron 104:227–238, e7
26. Koreman GT et al (2021) Upgraded CRISPR/Cas9 tools for tissue-specific mutagenesis in Drosophila. Proc Natl Acad Sci U S A 118: e2014255118
27. Ren X et al (2013) Optimized gene editing technology for Drosophila melanogaster using germ line-specific Cas9. Proc Natl Acad Sci U S A 110:19012–19017
28. Port F, Muschalik N, Bullock SL (2015) Systematic evaluation of Drosophila CRISPR tools reveals safe and robust alternatives to autonomous gene drives in basic research. G3 (Bethesda) 5:1493–1502
29. Ren X et al (2014) Enhanced specificity and efficiency of the CRISPR/Cas9 system with optimized sgRNA parameters in Drosophila. Cell Rep 9:1151–1162
30. Leibowitz ML et al (2021) Chromothripsis as an on-target consequence of CRISPR–Cas9 genome editing. Nat Genet 53:895–905
31. Papathanasiou S et al (2021) Whole chromosome loss and genomic instability in mouse embryos after CRISPR-Cas9 genome editing. Nat Commun 12:5855
32. Kosicki M, Tomberg K, Bradley A (2018) Repair of double-strand breaks induced by CRISPR-Cas9 leads to large deletions and complex rearrangements. Nat Biotechnol 36: 765–771
33. Shin HY et al (2017) CRISPR/Cas9 targeting events cause complex deletions and insertions at 17 sites in the mouse genome. Nat Commun 8:15464
34. Allen SE et al (2021) Versatile CRISPR/Cas9-mediated mosaic analysis by gRNA-induced crossing-over for unmodified genomes. PLoS Biol 19:e3001061
35. Sadhu MJ, Bloom JS, Day L, Kruglyak L (2016) CRISPR-directed mitotic recombination enables genetic mapping without crosses. Science 352:1113–1116
36. Alanis-Lobato G et al (2021) Frequent loss of heterozygosity in CRISPR-Cas9–edited early human embryos. Proc Natl Acad Sci U S A 118:e2004832117
37. Heigwer F, Port F, Boutros M (2018) RNA interference (RNAi) screening in drosophila. Genetics 208:853–874
38. Caussinus E, Kanca O, Affolter M (2012) Fluorescent fusion protein knockout mediated by anti-GFP nanobody. Nat Struct Mol Biol 19:117–121

Chapter 8

CRISPR-Based Transcriptional Activation in *Drosophila*

Yuting Han, Xinyi Lu, Yutong Li, Yuhao Qiu, Xizhi Dong, Xiaochen Li, Xu Si, Qingfei Liu, and Jian-Quan Ni

Abstract

Overexpression is one of the classical approaches to study pleiotropic functions of genes of interest. To achieve overexpression, we often increase the transcription by introducing genes on exogenous vectors or by using the CRISPR/dCas9-based transcriptional activation system. To date, the most efficient CRISPR/dCas9-based transcriptional activator is the Synergistic Activation Mediator (SAM) system whereby three different transcriptional activation domains are directly fused to dCas9 and MS2 phage Coat Protein (MCP), respectively, and the system in *Drosophila* is named flySAM. Here we describe the effective and convenient transcriptional activation system, flySAM, starting from vector construction, microinjection, and transgenic fly selection to the phenotypic analysis.

Key words Overexpression, CRISPR/dCas9, Transcriptional activation, flySAM, *Drosophila*

1 Introduction

Our knowledge of functional genomes has been greatly benefited from model organisms, such as *Drosophila*. Equipped with powerful genetic tools, we can systematically analyze the phenotypic consequence by either depletion or overexpression of target genes at molecular, cellular, and organismal levels [1–6]. Previously, several approaches to overexpress genes have been developed – for example, cloning of an entire gene sequence or the coding sequence (CDS) downstream of efficient promoters, such as the ubiquitously active *actin* promoter [7]. This can significantly improve the level of gene expression, but because of the ubiquitous nature of expression, it often results in lethality before adulthood or produces false-positive results due to forced expression in tissues, which are not of interest. A further approach is to knock in a transcriptional enhancer element or upstream activation sequence (UAS) in front of the gene of interest [8]. While it gives unexpected efficiency of overexpression, it is time consuming to screen for

Christian Dahmann (ed.), *Drosophila: Methods and Protocols*,
Methods in Molecular Biology, vol. 2540, https://doi.org/10.1007/978-1-0716-2541-5_8,

proper transgenic lines. Another often used approach to overexpress genes is to clone the CDS or entire gene region into a pUAS vector, which allows the transcription of the inserted gene to be conditionally controlled by Gal4 [9]. Recently, an efficient and simplified gene transcriptional activation system named flySAM has been developed, which is based on the CRISPR/dCas9 (**C**lustered **R**egularly **I**nterspaced **S**hort **P**alindromic **R**epeats/dead-**C**RISPR-**AS**sociated nuclease) system [10].

The CRISPR/Cas9 system has been widely used in gene editing [3]. By introduced single amino acid mutations within the HNH and RuvC domains of Cas9, the mutant Cas9 loses endonuclease activity [11]. This nuclease dead Cas9 is named dCas9 and still remains the ability to be guided by sgRNA on specific loci. If the dCas9 is fused with a transcriptional activator, it can be recruited by sgRNA to the sequence of interest to activate downstream transcription [12].

Transcriptional activation can occur through fusing dCas9 with transcription factors. The often-used transcription activators include VP64, p65, and HSF1. VP64 is the tetramer of VP16, a viral protein, which plays a key role in triggering the initiation of gene transcription [13]. p65, one of the members of NF-κB family of transcription factors, is a well-known transcription activator [14]. HSF1 is rapidly induced by various stress and then activates the transcription through binding the promoter region of target genes [15]. The activation domain of these transcriptional activators can be sequentially fused to dCas9. So far, one of the most efficient transcriptional activation systems, the Synergistic Activation Mediator (SAM) system, has been developed by creating dCas9 and MS2 phage Coat Protein (MCP) fusion protein with additional aptamers in sgRNA [16].

The SAM system in *Drosophila* includes three regulatory cassettes to control the transcription of dCas9-VP64, MCP-p65-HSF1, and sgRNA [10]. To improve the efficiency, and to simplify the cloning procedure, the three cassettes are built into a single vector, and the resulting vector is named flySAM. For this vector, *vermilion* is selected as transgenic selectable marker, as the *white* gene has a deleterious behavior; it has an attB site, allowing this vector to site-specifically integrate into *Drosophila* genome through phiC31. Moreover, both dCas-VP64 and MCP-p65-HSF1 are controlled by a single UAS basal promoter sequence and function independently through T2A mediated division. Finally, the sgRNA that has two binding motifs and is driven by efficient U6b promoter.

flySAM is an ideal transcriptional activation system, because it is more efficient than other dCas9-based transcriptional activation systems. Moreover, the severity of phenotypes from flySAM is not only comparable to the traditional Gal4/UAS overexpression system but also superior to it at many aspects. First, for flySAM, the

researcher only needs to synthesize and clone a 20 bp targeting sequence, which greatly reduces the cost and time. Second, flySAM can target and modulate multiple genes simultaneously, allowing researchers to interrogate the roles of proteins with redundant functions or protein complexes involved in different biological processes. This is particularly useful for targeting multiple genes, as traditional methods require complicated genetic recombination. Third, unlike other dCas9-based transcriptional activation systems, flySAM can generate mosaic cellular clones with only one genetic cross, which provide a powerful genetic analysis tool to precisely analyze the behavior of the marked cells. Last and most importantly, for genes with multiple transcripts, one flySAM line can activate all the transcripts.

Having established flySAM as superior to other combinations, the parameters of the sgRNA have been systematically analyzed [17]: the most efficient sgRNAs normally accumulate in the region around -200 bp upstream of the transcription start site (TSS); the high GC content within the sgRNA targeting sequence often associates with efficiency; sgRNAs targeting to the non-template strand of genomic DNA is optimal. The high performance and relative ease in manipulation, together with these criteria for sgRNA design, will greatly facilitate the application of flySAM to benefit the *Drosophila* community [18]. In this chapter, we describe the flySAM system, including the construction of the flySAM plasmid, plasmid microinjection, and the selection of transgenic fly lines.

2 Materials

2.1 Fly Stocks

All *Drosophila* lines used to construct flySAM transgenic lines (Table 1) are stored at the Tsinghua Fly Center (THFC) and cultured on standard food under standard laboratory conditions (25 °C, 60% humidity).

2.2 Plasmid Construction Reagents

1. ddH_2O.
2. Annealing buffer (1×): 10 mM Tris–HCl, pH 7.5, 1 mM EDTA, 0.1 M NaCl.
3. Restriction endonucleases: *Bbs* I, *Nhe* I, *Spe* I, CIP.
4. Backbone plasmid (the flySAM without sgRNA).
5. DNA Ligase: T4 DNA ligase.
6. Trans5α Chemically Competent Cells.
7. GO Taq.
8. TAE buffer (50×): Dissolve 242 g Tris–HCl and 18.612 g EDTA in 800 ml ddH_2O. Then add 57.1 ml acetic acid. Make up to 1 L with ddH_2O.

Table 1
Fly stocks

THFC number	Genotype	Comment
TB16	*y,sc,v,nanos-integrase*; attP40	This line is used for injection. *Nanos-integrase* and attP40 site locate at chromosome X and II respectively.
TB18	*y,sc,v,nanos-integrase*;;attP2	This line is used for injection. *Nanos-integrase* and attP2 site locate at chromosome X and III respectively.
TB23	*y,sc,v;Gla,Bc/CyO*	This line is used for transgenic selection, carrying *CyO* on chromosome II.
TB77	*y,sc,v*	This line is used for transgenic selection to discard *nanos-integrase*.
TB139	*y,sc,v;;Dr,e/TM3, Sb*	This line is used for transgenic selection, carrying *TM3,Sb* on chromosome III.

9. TAE buffer (1×): Dilute 20 ml 50× TAE with 980 ml ddH_2O.
10. 1.2% Agarose Gel: Add 1.2 g agarose in 100 ml TAE buffer (1×). Heat the solution to dissolve agarose. After cool down to 60 °C, add 5 μl 5 mg/ml ethidium bromide (EB). Mix thoroughly and then pour into the casting tray.
11. 100 mg/ml ampicillin solution: Dissolve 1 g ampicillin in 9 ml ddH_2O; then add ddH_2O up to 10 ml. Aliquot it into 1.5-ml tubes after filtering through 0.22-μm Millipore filters. Store at −20 °C.
12. LB Liquid Medium: Add 25 g mixture (40% Tryptone, 20% yeast extract, and 40% NaCl) in 900 ml ddH_2O; add ddH_2O up to final volume of 1 L. Autoclave at 121 °C for 15 min. After autoclaving, add 1 ml 100 mg/ml ampicillin. Mix thoroughly and keep at 4 °C.
13. AxyPrep Plasmid MiniPrep Kit.
14. AxyPrep DNA Gel Extraction Kit.
15. LB Solid Medium: Add 25 g mixture (40% Tryptone, 20% yeast extract, and 40% NaCl) and 10 g agar in 900 ml ddH_2O, and then make up to 1 L with ddH_2O. After autoclaving at 121 °C for 15 min, add 1 ml 100 mg/ml ampicillin at about 60 °C, then pour the solution into 90 mm petri dishes, around 20 ml per dish. Store at 4 °C.
16. Primers:

    Forward primer (5'-ttcg NNNNNNNNNNNNNNNNNNN N-3′).

    Reverse primer (5′-aacc MMMMMMMMMMMMMMMMM MMM-3′).

    Forward Primer F1 (5'-TCAACAAACGAACAATAGGACAC-3′).

Reverse Primer R1 (5'-AAAAAAGCACCGACTCGGTG-3′).

Forward Primer F2 (5'-CACACTTATTACGTGGCCAG-3′).

Reverse Primer R2 (5'-TCCCACACCTCCCACTGAAC-3′).

2.3 Injection Reagents

1. 0.2 M, pH 6.8 sodium phosphate buffer: Mix 49 ml 0.2 M Na_2HPO_4 solution and 51 ml 0.2 M NaH_2PO_4 solution.
2. 0.5 M KCl: Dissolve 23.8 g KCl with 800 ml ddH_2O and then add ddH_2O up to 1 L.
3. Injection Buffer (1×): Mix 0.5 ml 0.2 M, pH 6.8 sodium phosphate buffer with 10 ml 0.5 M KCl and make up to 1 L with sterile ddH_2O. Filter through 0.22-μm Millipore filters. Aliquot it into a 50-ml centrifuge tube and store at −20 °C before injection.
4. PurePlasmid Mini Kit.

3 Methods

3.1 Backbone Plasmid (the flySAM Vector without sgRNA) Amplification

1. Add 0.1 μl of approximately 150 ng/μl backbone plasmid (Fig. 1) into 10 μl Trans5α Chemically Competent Cells (*see* **Note 1**). After standard transformation procedure, plate it with 200 μl tips on LB solid 90 mm plate containing 100 μg/ml ampicillin. Incubate the plate at 37 °C overnight.
2. Select a single clone and inoculate into LB liquid medium containing 100 μg/ml ampicillin. Incubate it in shaker overnight at 37 °C, 200 r/min.
3. Collect the bacteria pellet by centrifugation. Then extract the plasmid by standard procedure using AxyPrep Plasmid Mini-Prep Kit. If it is eluted with 100 μl ddH_2O, the concentration shall be around 150 ng/μl.

3.2 Backbone Preparation (Linearization by Enzyme Digestion)

1. Use *Bbs* I to linearize the backbone plasmid at 37 °C for 1.5 h (Table 2).
2. Run the digestion product by using 1.2% agarose gel. After 180 V electrophoresis for 20 min, two bands will be observed under UV light, one is 700 bp and another is 14 kb (Fig. 2). Cut and collect the 14 kb large band and then extract the DNA with AxyPrep DNA Gel Extraction Kit following standard procedure. Elute DNA fragment with 30 μl ddH_2O. The concentration should be around 40 ng/μl.

3.3 Oligo Design

1. Select a proper sgRNA targeting sequence, the 20 nt length sequence upstream NGG PAM (*see* **Note 2**).
2. To facilitate the ligation, add the overhang sequence 5′-ttcg-3′ into forward primer, e.g., 5'-ttcgNNNNNNNNNNNNNNNNNNN-3′, and add 5′-aaac-3′ to the reverse primer, e.g.,

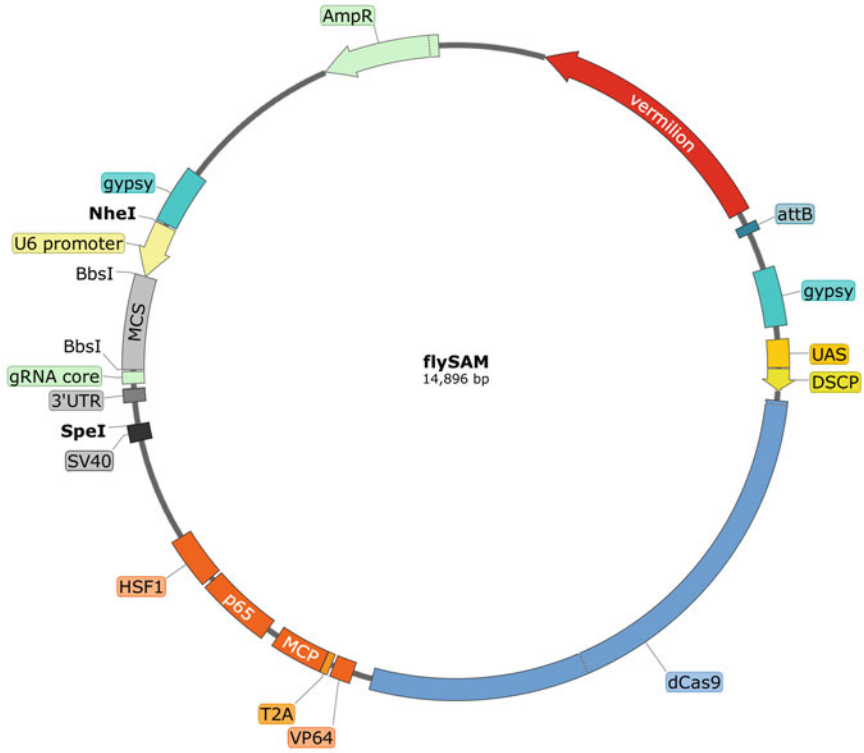


Fig. 1 The map of flySAM without sgRNA. dCas9 and MCP activators are under the control of UAS and *Drosophila* synthetic core promoter (DSCP). sgRNA is controlled by U6 promoter. The attB sequence and *vermilion*$^{+}$ gene are included in this plasmid

Table 2
Enzyme digestion solution

Content	Volume
Bbs I-HF (20,000 units)	1 μl
10× NEB Cutsmart buffer	5 μl
Backbone plasmid	2.5 μg
Add ddH_2O to	50 μl

5′-aaacMMMMMMMMMMMMMMMMMMMM-3′, which are complementary to the sticky ends of the linearized backbone digested by *Bbs* I.

3. Set up annealing to get a double-stranded DNA fragment (Table 3). Denature the primers in annealing buffer with PCR Thermal Cycler at 95 °C for 5 min and then turn off the power to slowly cool the temperature down.

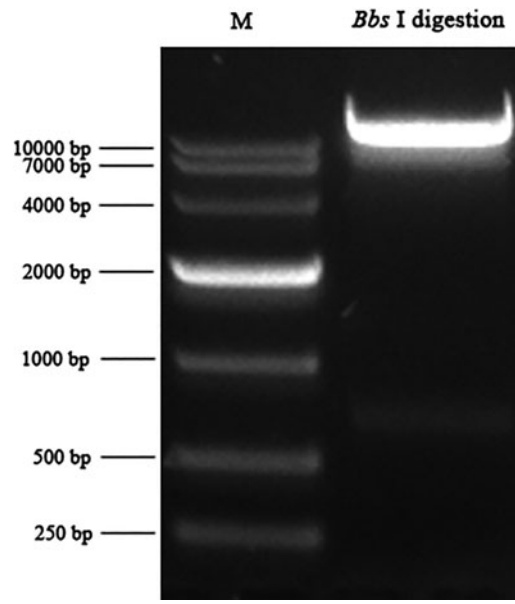


Fig. 2 The image taken from flySAM digested by *Bbs* I. M is DNA marker. The fragment with a size over 10 kb shall be extracted

Table 3
Annealing solution

Content	Volume
Forward Primer (20 μM)	5 μl
Reverse Primer (20 μM)	5 μl
Annealing buffer	40 μl
Total	50 μl

3.4 Ligation and Transformation

1. Ligate above annealed DNA primers with linearized backbone at 16 °C for 1 h (Table 4).
2. Add 5 μl of the ligation product into 10 μl Trans5α Chemically Competent Cells. After transformation with standard procedure, plate it on the LB solid plate containing 100 μg/ml ampicillin. Incubate the plate at 37 °C overnight.
3. Select the correct bacterial clone by PCR (Tables 5 and 6). The size of the PCR product from the correct clone should be about 300 bp.
4. Inoculate the bacterial correct clone in 5 ml LB liquid medium with 100 μg/ml ampicillin in a 15-ml centrifuge tube, and incubate in shaker overnight at 37 °C, 200 r/min.

Table 4
Ligation solution

Content	Volume
Annealed sgRNA product	4 μl
Linearized backbone	30 ng
T4 ligase	0.5 μl
10× T4 ligase buffer	1 μl
ddH_2O	Add to 10 μl

Table 5
PCR solution

Content	Volume
2× Go Taq Mix	5 μl
Forward Primer F1 (5'-TCAACAAACGAACAATAGGACAC-3′)	0.1 μl
Reverse Primer R1 (5'-AAAAAAGCACCGACTCGGTG-3′)	0.1 μl
ddH_2O	4.8 μl

Table 6
Thermal cycling condition

Temperature	Time	
95 °C	3 min	
95 °C	30 s	30 cycles
55 °C	30 s	
72 °C	30 s	
72 °C	5 min	
4 °C	∞	

5. Collect the bacteria pellet and miniprep plasmid with AxyPrep Plasmids Miniprep Kit. Elute the plasmid with 100 μl ddH_2O and the concentration shall be around 150 ng/μl.
6. Before injection, the flySAM plasmid is further confirmed through DNA sequencing by using the Forward Primer F1 (5'-TCAACAAACGAACAATAGGACAC-3′) (*see* **Note 3**).

Table 7
Enzyme digestion solution

Content	Volume
Spe I-HF (20,000 units)	0.5 μl
10× NEB Cutsmart buffer	5 μl
Plasmid	2.5 μg
ddH_2O	Add to 50 μl

Table 8
Enzyme digestion solution

Content	Volume
Nhe I-HF (20,000 units)	0.5 μl
Spe I-HF (20,000 units)	0.5 μl
10× NEB Cutsmart buffer	5 μl
Plasmid	2.5 μg
ddH_2O	Add to 50 μl

3.5 Cloning Multiple sgRNA into flySAM System

1. First clone individual sgRNA targeting sequence into flySAM vector, respectively, using above approaches.
2. Backbone preparation. Use *Spe* I to linearize flySAM plasmid carrying single sgRNA at 37 °C for 1.5 h (Table 7). Add 0.1 μl CIP at room temperature for 10 min (*see* **Note 4**). Then purify the digested product by AxyPrep DNA Gel Extraction Kit. Elute it with 30 μl ddH_2O and the concentration will be around 40 ng/μl.
3. The inserted fragment containing sgRNA preparation. Use *Nhe* I and *Spe* I to digest the flySAM plasmid carrying another sgRNA at 37 °C for 1.5 h (Table 8). Run the digestion product by using 1.2% agarose gel. After electrophoresis, the smaller band, which has a size of approximately 800 bp, will be extracted by AxyPrep DNA Gel Extraction Kit. Elute it with 30 μl ddH_2O and the DNA concentration will be about 5 ng/μl.
4. Ligate the linearized backbone from procedure 2 and the fragment carrying sgRNA from procedure 3 with T4 DNA ligase at 16 °C for 1.5 h (Table 9).
5. Add 5 μl of the ligation product into 10 μl Trans5α *E. coli* Chemically Competent Cells. After transformation, plate it on an LB solid plate containing 100 μg/ml ampicillin. Incubate the plate at 37 °C overnight.

Table 9
Ligation solution

Content	Volume
Fragment	10 ng
Backbone	30 ng
T4 ligase	0.5 μl
10× T4 ligase buffer	1 μl
ddH_2O	Add to 10 μl

Table 10
PCR solution

Content	Volume
2× Go Taq Mix	5 μl
Forward Primer F2 (5'-CACACTTATTACGTGGCCAG-3′)	0.1 μl
Reverse Primer R2 (5'-TCCCACACCTCCCACTGAAC-3′)	0.1 μl
ddH_2O	4.8 μl

6. Select the correct bacterial clone by PCR (Tables 10 and 11). The size of PCR product from correct clone should be about 1.8 kb if there are two sgRNAs.
7. Inoculate the correct bacterial clone in LB liquid medium containing 100 μg/ml ampicillin. Incubate overnight at 37 °C, 200 r/min.
8. Collect the bacteria pellet and miniprep plasmid with AxyPrep Plasmids Miniprep Kit. Elute the plasmid with 100 μl ddH_2O and the concentration shall be around 150 ng/μl.
9. Further confirm the correct clone through DNA sequencing by using the Reverse Primer R2 (5'- TCCCACACCTCCCACT GAAC-3′).
10. To clone additional sgRNAs, follow above procedures.

3.6 FlySAM Plasmid Purification and Microinjection

1. For microinjection, the plasmid requires to be purified by Pure-Plasmid Mini Kit and dissolved with injection buffer. The plasmid concentration is around 150 ng/μl.
2. Collect the embryos and inject the flySAM vector (*see* **Note 5**).
3. After microinjection, the injected embryos will be put in vial with standard food for development at 25 °C.

Table 11
Thermal cycling condition

Temperature	Time	
95 °C	3 min	
95 °C	30 s	30 cycles
55 °C	30 s	
72 °C	120 s	
72 °C	5 min	
4 °C	∞	

3.7 Transgenic Selection (for Detailed Procedure, see *Fig. 3)*

1. After eclosion, the adult flies are named G0. Select males and cross each of them with 3–5 virgins of *y,sc,v*;; (TB77). Their offspring is G1.
2. Transgenic selection based on eye color. Select G1 males with dark red eyes (Fig. 4) and cross with virgin flies carrying balancer (*see* **Note 6**). Their offspring is named G2.
3. After G2 flies eclosion, select male and virgin siblings that have both dark red eye balanced with *CyO* or *TM3,Sb*. Set up cross for them; then select homozygous transgenic flies from their offspring G3.
4. Keep the transgenic flies, which are ready for phenotypic analysis.

4 Notes

1. The sequence of flySAM vector without sgRNA.

CACCTAAATTGTAAGCGTTAATATTTTGTTAAAATTCG
CGTTAAATTTTTGTTAAATCAGCTCATTTTTTAACCAA
TAGGCCGAAATCGGCAAAATCCCTTATAAATCAAAA
GAATAGACCGAGATAGGGTTGAGTGTTGTTCCAG
TTTGGAACAAGAGTCCACTATTAAAGAACGTGG
ACTCCAACGTCAAAGGGCGAAAAACCGTCTATC
AGGGCGATGGCCCACTACGTGAACCATCACCCTAAT
CAAGTTTTTTGGGGTCGAGGTGCCGTAAAGCAC
TAAATCGGAACCCTAAAGGGAGCCCCCGATTTAGA
GCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAA
GGAAGGGAAGAAAGCGAAAGGAGCGGGCGCTA

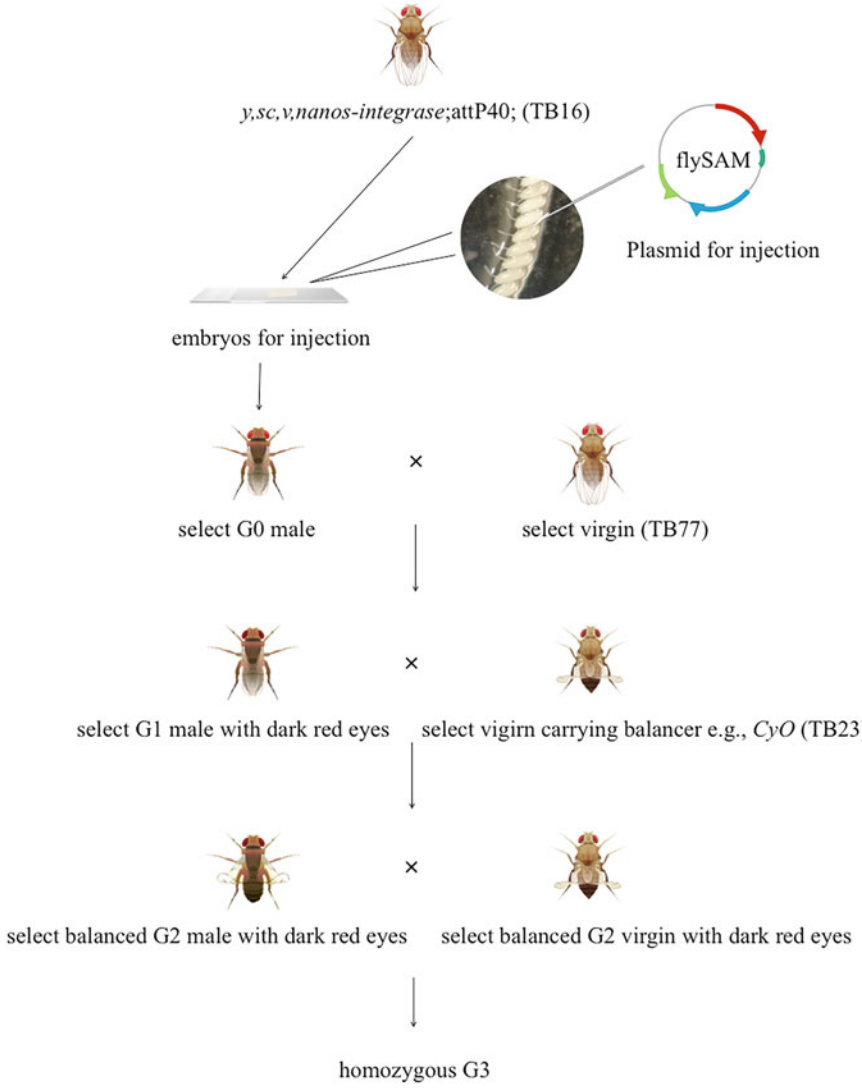


Fig. 3 Workflow for flySAM transgenic line construction (e.g., flySAM plasmid integrating into Chromosome II)

GGGCGCTGGCAAGTGTAGCGGTCACGCTGCGCG
TAACCACCACACCCGCCGCGCTTAATGCGCCGCTA
CAGGGCGCGTCCCATTCGCCATTCAGGCTGCG
CAACTGTTGGGAAGGGCGATCGGTGCGGGCCTCTTC
GCTATTACGCCAGCTGGCGAAAGGGGGATGTGCT
GCAAGGCGATTAAGTTGGGTAACGCCAGGGTTTT
CCCAGTCACGACGTTGTAAAACGACGGCCAGT
GAATTGTAATACGACTCACTATAGGGCGAATTGGG
TACAAGCTTATTTATTTTGTTATGTTATATGTATTATA
TGTCAGACATAAAGAAAAGGAACACATCAAATGTGA
TAACAAAGACTAAACAAGTAATTTTATTACACC
AAAACGACAAAACAGTAGGCAGAACAAACAACGCA

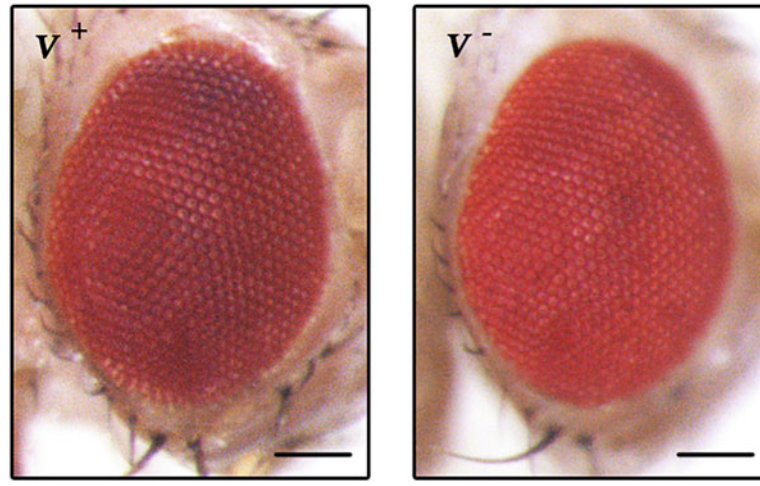


Fig. 4 Selectable marker for transgenic fly selection. Transgenic flies have dark red eyes (left); while flies without flySAM integration have red eyes (right). (Scale bars: 250 μm)

TAGCCAAACATTGACGAATTGGATACCCTGCCGATTG
TCAGACACTTTTGTTGATCAGTTTCTTGCGAA
TGGTCTCGTCCAGCGGTGGAATCGCCTCGCGGGGAA
TCAGAAAAGTGGACAGATTGAACAGATCCAGAAA
CACCTTGTACCGATCACTGAAACCAAAAAAAAAC
AAAGGGAGAACAGTTTGAGTTCATTGATCCCCGATA
TAATCACATCTGCGATGATCACCTGAGAGTGGA
GCGCAGATATTGATAACCAGACGAGCCACCAGTG
CCCAACTGTTGCGATCCAATCATGCGTTGCACCATG
ATCACGTGATTGTCTGCGGCGGGAATAGAAAGTA
TTTGGTTAGGAAAACCAGTCTTAAACATAAGATAT
ATTTATAAAAGAGTATCAAAGAATGCAATACTTAC
ATCTCCACTTGGTTATTAACGAGTCGATGTCCATGAG
CAGGGTGAGCAACTGGTGTGGTTGGCTGAA
CCTGGGTTCATCCCTATAGAAGGTGATCATGA
TGGCTCCCTGAAGGGCACGATGGCTAAACCGGCG
ATCCCCACGACGCACCAGTGCATCGTGCA
CTGCCGGATCAAAGATGGAGCGATACA
CCTCGCGTCGCTTCTCAATGTCCATGAGGCGGTA
GTTTTTCGCCTTCTCCACGGGCTCCTCCATGGCGCT
CTGTACCTGCGCCTCCAGGAATCGATCGACGCTCTCC
TGAAACTTGGCCCAGAAGTTGAAGCCA
CTCTCCTCCAGTCCGGGCGTCCTCTCCAGCCATCGC
TGCACTAGCTCCAGTAGCGAGGGATCTTTCTCC
GAGTTGCGAATCGAGTTCCGCGCCTCCTCGTCGCT
AAAGACATCCGAGTACTTCTGGTTGTATCTCACCCGC
TGCTCTGTCAGAACTCCCAGCTTGTTCTCGAT
CAAACGGAACTGCAGCGACTGAAAACCAGAT
GCGGGTGCCAGGTACTTGCGGAAGTCCATGAAG

TCTAGCGGGGTCATGGTCTCCAGAATGGGCACTT
GGTCCACCAGGAGCTGTACAAAGGAAGTTATAAAGG
ATTTTGGTAAGAGATTCAGAAAGCACTCACTTTTA
GAATCAGAACCACTCGGTTCAGTCGCTTGACAA
TCTCCAGCGTCTTGGTTTCATCGATGACCTCTGCAT
CCAACATGTCTCGTATGGAGTCGAACTCAAAGAT
GATCTGCTTGAACCAAAGCTCGTAGGCTGTGGCGAA
GGTACTTAAATGCCATTGAGTGTTGTCATCAAAGTTG
TAAACCTACTCACCCTGGTGCGTGATGATGAACAGAT
GCTCATCGTGCACGGGTCGCTTGTCCTCCTCGGACA
GCATACACTGGGCATCCAGCAGTTTGTCCAGCATCA
GATACTCTCCATAGATTTTGCCCACTTCCGTGGTTAAT
GGCACCGCCGAATCATCGTGATCGTTTCTGTATGGG
TTTGAATTGAATCGCAGAACTGAAGATCGATT
GGCATTCCTGGACAGCACGTGCTGGTGCTCACCCG
TTTCCTGCATAGGGACAGCTCATGGTGCACAGCTC
AGATCAGATCGTGACTCCTCGACCGGCGGATGCT
GGCGAACTGATCTCCGCCAGCGGACCGGAGATGA
GACCCCAGCGAACCGATAACAGAGCGAGAGA
GCTCCAGTTCCGACTGATTGCACAGTCGGTGATCT
GGGCGATGGGCACTGCCAGATAGGCTGGGAATTAT
CAATCACTTGAGGTGAAAGTGCGGCGCACACAAAT
CCAAGCTTGATATCATCGATCTCGACGCTGCAT
CCAACGCGTTGGGAGCTCTCCGGATCAATTCGGCT
TCACGTACCGTCGACGATGTAGGTCACGGTCT
CGAAGCCGCGGTGCGGGTGCCAGGGCGTG
CCCTTGGGCTCCCCGGGCGCGTACTCCACCTCACCC
ATCTGGTCCATCATGATGAACGGGTCGAGGT
GGCGGTAGTTGATCCCGGCGAACGCGCGGCG
CACCGGGAAGCCCTCGCCCTCGAAACCGCTGGGC
GCGGTGGTCACGGTGAGCACGGGACGTGCGACGG
CGTCGGCTGGTGCGGATACGCGGGGCAGCGTCAG
CGGGTTCTCGACGGTCACGGCGGGCATGTCGACAAG
CCGAATTGATCCACTAGAAGGCCTAATTCGGTACAC
TATGCTAGTATAACTTCGTATAATGTATGCTATACGAA
GTTATGCTAGTTGGCCACGTAATAAGTGTGCGTTGAA
TTTATTCGCAAAAACATTGCATATTTTCGGCAAAG
TAAAATTTTGTTGCATACCTTATCAAAAAATAAGT
GCTGCATACTTTTTAGAGAAACCAAATAATTTTTTATT
GCATACCCGTTTTTAATAAAATACATTGCATACCCT
CTTTTAATAAAAAATATTGCATACTTTGACGAAAC
AAATTTTCGTTGCATACCCAATAAAAGATTATTATATT
GCATACCCGTTTTTAATAAAATACATTGCATACCCT
CTTTTAATAAAGAATATTGCATACGTTGACGAAA
CAAATTTTCGTTGCATACCCAATAAAAGATTATTATA
TTGCATACCTTTTCTTGCCATACCATTTAGCCGATCAA
TTCTGCTCGGCAACAGTATATTTGTGGTGTGCCAACC
AACAACACTAGCATAACTTCGTATAATGTATGCTATAC
GAAGTTATGAGCTCGCTCGGGTAATCGCTTATCCTC

GGGTAATCGCTTATCCTTAAGCTGCAGGTCGGAGTAC
TGTCCTCCGAGCGGAGTACTGTCCTCCGAGCGGAGT
ACTGTCCTCCGAGCGGAGTACTGTCCTCCGAGCGGA
GTACTGTCCTCCGAGCGGAGACTCCCGCGGTC
GGAGTACTGTCCTCCGAGCGGAGTACTGTCCTCCG
AGCGGAGTACTGTCCTCCGAGCGGAGTACTGTCC
TCCGAGCGGAGTACTGTCCTCCGAGCGGAGACTC
GTCGACGAGCTCGCCCGGGGATCGAGCGCAGC
GGTATAAAAGGGCGCGGGGTGGCTGAGAGCATCAG
TTGTGAATGAATGTTCGAGCCGAGCAGACGT
GCCGCTGCCTTCGTTAATATCCTTTGAATAAGCCAA
CTTTGAATCACAAGACGCATACCAAAGTCTAGAGG
TACCGCCACCATGGCCCCAAAGAAGAAGC
GGAAGGTCGGTATCCACGGTGTCCCAGCAGCCATGG
ACAAGAAGTACTCCATTGGGCTCGCTATCGGCAC
AAACAGCGTCGGCTGGGCCGTCATTACGGACGAGTA
CAAGGTGCCGAGCAAAAAATTCAAAGTTCT
GGGCAATACCGATCGCCACAGCATAAAGAAGAACCT
CATTGGCGCCCTCCTGTTCGACTCCGGGGAGA
CGGCCGAAGCCACGCGGCTCAAAAGAACAGCA
CGGCGCAGATATACCCGCAGAAAGAATCGGA
TCTGCTACCTGCAGGAGATCTTTAGTAATGAGA
TGGCTAAGGTGGATGACTCTTTCTTCCATAGGCT
GGAGGAGTCCTTTTTGGTGGAGGAGGATAAAAAGCA
CGAGCGCCACCCAATCTTTGGCAATATCGTGGACG
AGGTGGCGTACCATGAAAAGTACCCAACCATATATCA
TCTGAGGAAGAAGCTTGTAGACAGTACTGAT
AAGGCTGACTTGCGGTTGATCTATCTCGCGCT
GGCGCATATGATCAAATTTCGGGGACACTTCCTCATC
GAGGGGGACCTGAACCCAGACAACAGCGATGTCGA
CAAACTCTTTATCCAACTGGTTCAGACTTACAATC
AGCTTTTCGAAGAGAACCCGATCAACGCATCCGGA
GTTGACGCCAAAGCAATCCTGAGCGCTA
GGCTGTCCAAATCCCGGCGGCTCGAAAACCTCA
TCGCACAGCTCCCTGGGGAGAAGAAGAACGGCC
TGTTTGGTAATCTTATCGCCCTGTCACTCGGGCTGA
CCCCCAACTTTAAATCTAACTTCGACCTGGCCGAA
GATGCCAAGCTTCAACTGAGCAAAGACACCTACGA
TGATGATCTCGACAATCTGCTGGCCCAGATCGGCGA
CCAGTACGCAGACCTTTTTTTGGCGGCAAAGAA
CCTGTCAGACGCCATTCTGCTGAGTGATATTCTGCGA
GTGAACACGGAGATCACCAAAGCTCCGCTGA
GCGCTAGTATGATCAAGCGCTATGATGAGCACCACC
AAGACTTGACTTTGCTGAAGGCCCTTGTCAGACAG
CAACTGCCTGAGAAGTACAAGGAAATTTTCTTCGAT
CAGTCTAAAAATGGCTACGCCGGATACATT
GACGGCGGAGCAAGCCAGGAGGAATTTTACAAATTT
ATTAAGCCCATCTTGGAAAAAATGGACGGCACC
GAGGAGCTGCTGGTAAAGCTTAACAGAGAAGATCT

GTTGCGCAAACAGCGCACTTTCGACAATGGAAGCAT
CCCCCACCAGATTCACCTGGGCGAACTGCACGCTAT
ACTCAGGCGGCAAGAGGATTTCTACCCCTTTTT
GAAAGATAACAGGGAAAAGATTGAGAAAATCCTCA
CATTTCGGATACCCTACTATGTAGGCCCCCTCG
CCCGGGGAAATTCCAGATTCGCGTGGATGACT
CGCAAATCAGAAGAGACCATCACTCCCTGGAA
CTTCGAGGAAGTCGTGGATAAGGGGGCCTCT
GCCCAGTCCTTCATCGAAAGGATGACTAACTTTGAT
AAAAATCTGCCTAACGAAAAGGTGCTTCCTAAAC
ACTCTCTGCTGTACGAGTACTTCACAGTTTATAAC
GAGCTCACCAAGGTCAAATACGTCACAGAAGGGA
TGAGAAAGCCAGCATTCCTGTCTGGAGAGCAGAA
GAAAGCTATCGTGGACCTCCTCTTCAAGACGAA
CCGGAAAGTTACCGTGAAACAGCTCAAAGAGGA
C T A T T T C A A A A A G A T T G A A T G T T T C G A
CTCTGTTGAAATCAGCGGAGTGGAGGATCGCTTCAA
CGCATCCCTGGGAACGTATCACGATCTCCTGAAAA
TCATTAAAGACAAGGACTTCCTGGACAATGAGGA
GAACGAGGACATTCTTGAGGACATTGTCCTCA
CCCTTACGTTGTTTGAAGATAGGGAGATGATTGAA
GAACGCTTGAAAACTTACGCTCATCTCTTCGACGAC
AAAGTCATGAAACAGCTCAAGAGGCGCCGATATAC
AGGATGGGGGCGGCTGTCAAGAAAACTGATCAAT
GGGATCCGAGACAAGCAGAGTGGAAAGACAATCC
TGGATTTTCTTAAGTCCGATGGATTTGCCAACC
GGAACTTCATGCAGTTGATCCATGATGACTCTCTC
ACCTTTAAGGAGGACATCCAGAAAGCACAAGTTTC
TGGCCAGGGGGACAGTCTGCACGAGCACATCGC
TAATCTTGCAGGTAGCCCAGCTATCAAAAAGGGAA
TACTGCAGACCGTTAAGGTCGTGGATGAACTCGTC
AAAGTAATGGGAAGGCATAAGCCCGAGAATATC
GTTATCGAGATGGCCCGAGAGAACCAAACTACCC
AGAAGGGACAGAAGAACAGTAGGGAAAGGATGAA
GAGGATTGAAGAGGGTATAAAAGAACTGGGG
TCCCAAATCCTTAAGGAACACCCAGTTGAAAACA
CCCAGCTTCAGAATGAGAAGCTCTACCTGTACTACC
TGCAGAACGGCAGGGACATGTACGTGGATCAGG
AACTGGACATCAATCGGCTCTCCGACTACGACGTGG
ATGCTATCGTGCCCCAGTCTTTTCTCAAAGATGATTC
TATTGATAATAAAGTGTTGACAAGATCCGATAAAAA
TAGAGGGAAGAGTGATAACGTCCCCTCAGAAGAAG
TTGTCAAGAAAATGAAAAATTATTGGCGGCAGCTG
CTGAACGCCAAACTGATCACACAACGGAAGTTCGA
TAATCTGACTAAGGCTGAACGAGGTGGCCTGTCTG
AGTTGGATAAAGCAGGCTTCATCAAAAGGCAGCTTG
TTGAGACACGCCAGATCACCAAGCACGTGG
CCCAAATTCTCGATTCACGCATGAACACCAAGTACG
ATGAAAATGACAAACTGATTCGAGAGGTGAAAG

TTATTACTCTGAAGTCTAAGCTGGTCTCAGATTTCAG
AAAGGACTTTCAGTTTTATAAGGTGAGAGAGATCAA
CAATTACCACCATGCGCATGATGCCTACCTGAATGCA
GTGGTAGGCACTGCACTTATCAAAAAATATCCCAA
GCTTGAATCTGAATTTGTTTACGGAGACTATAAAGT
GTACGATGTTAGGAAAATGATCGCAAAGTCTGAG
CAGGAAATAGGCAAGGCCACCGCTAAGTACTT
CTTTTACAGCAATATTATGAATTTTTTCAAGACCGA
GATTACACTGGCCAATGGAGAGATTCGGAAGCGA
CCACTTATCGAAACAAACGGAGAAACAGGAGAAAT
CGTGTGGGACAAGGGTAGGGATTTCGCGACAGT
CCGGAAGGTCCTGTCCATGCCGCAGGTGAACAT
CGTTAAAAAGACCGAAGTACAGACCGGAGGCTTCT
CCAAGGAAAGTATCCTCCCGAAAAGGAACAGCG
ACAAGCTGATCGCACGCAAAAAAGATTGGGA
CCCCAAGAAATACGGCGGATTCGATTCTCCTACAG
TCGCTTACAGTGTACTGGTTGTGGCCAAAGTGGA
GAAAGGGAAGTCTAAAAAACTCAAAAGCGTCAAGG
AACTGCTGGGCATCACAATCATGGAGCGAT
CAAGCTTCGAAAAAAAACCCCATCGACTTTCT
CGAGGCGAAAGGATATAAAGAGGTCAAAAAAGA
CCTCATCATTAAGCTTCCCAAGTACTCTCTCTTT
GAGCTTGAAAACGGCCGGAAACGAATGCTCGCTAGT
GCGGGCGAGCTGCAGAAAGGTAACGAGCTGGCACT
GCCCTCTAAATACGTTAATTTCTTGTATCT
GGCCAGCCACTATGAAAAGCTCAAAGGGTCT
CCCGAAGATAATGAGCAGAAGCAGCTGTTCGT
GGAACAACACAAACACTACCTTGATGAGATCAT
CGAGCAAATAAGCGAATTCTCCAAAAGAGTGATCCT
CGCCGACGCTAACCTCGATAAGGTGCTTTCTGCTT
ACAATAAGCACAGGGATAAGCCCATCAGGGAG
CAGGCAGAAAACATTATCCACTTGTTTACTCTGACC
AACTTGGGCGCGCCTGCAGCCTTCAAGTACTTCGA
CACCACCATAGACAGAAAGCGGTACACCTCT
ACAAAGGAGGTCCTGGACGCCACACTGATTCAT
CAGTCAATTACGGGGCTCTATGAAACAAGAAT
CGACCTCTCTCAGCTCGGTGGAGACAAGCGTCCT
GCTGCTACTAAGAAAGCTGGTCAAGCTAAGAAAAA
GAAACAATTCGGAGGAGGTGGAAGCGGAGGAGG
AGGAAGCGGAGGAGGAGGTAGCGGACCTAAG
AAAAAGAGGAAGGTGGCGGCCGCTGGTT
CCGGACGGGCTGACGCATTGGACGATTTTGATCTGG
ATATGCTGGGAAGTGACGCCCTCGATGATTTTG
ACCTTGACATGCTTGGTTCGGATGCCCTTGATG
ACTTTGACCTCGACATGCTCGGCAGTGACGCCCTTG
ATGATTTCGACCTGGACATGCTGATTAACCAATTCGG
AAGCGGAGAGGGCAGAGGAAGTCTGCTAACATG
CGGTGACGTCGAGGAGAATCCTGGACCTATGG
CTTCAAACTTTACTCAGTTCGTGCTCGTGGACAATGG

T G G G A C A G G G G A T G T G A C A G T G G
CTCCTTCTAATTTCGCTAATGGGGTGGCAGAGTGGAT
CAGCTCCAACTCACGGAGCCAGGCCTACAAGGTG
ACATGCAGCGTCAGGCAGTCTAGTGCCCAGAAGAG
AAAGTATACCATCAAGGTGGAGGTCCCCAAAGTGG
CTACCCAGACAGTGGGCGGAGTCGAACTGCCTGTCG
CCGCTTGGAGGTCCTACCTGAACATGGAGCTCAC
TATCCCAATTTTCGCTACCAATTCTGACTGTGAA
CTCATCGTGAAGGCAATGCAGGGGCTCCTCAAAG
ACGGTAATCCTATCCCTTCCGCCATCGCCGCTAACTC
AGGTATCTACAGCGCTGGAGGAGGTGGAAGCGG
AGGAGGAGGAAGCGGAGGAGGAGGTAGCGGACC
TAAGAAAAAGAGGAAGGTGGCGGCCGCTCAATTG
CCTTCAGGGCAGATCAGCAACCAGGCCCTGG
CTCTGGCCCCTAGCTCCGCTCCAGTGCTGGCCCAG
ACTATGGTGCCCTCTAGTGCTATGGTGCCTCTGG
CCCAGCCACCTGCTCCAGCCCCTGTGCTGACCCC
AGGACCACCCCAGTCACTGAGCGCTCCAGT
GCCCAAGTCTACACAGGCCGGCGAGGGGACTC
TGAGTGAAGCTCTGCTGCACCTGCAGTTCGACGC
TGATGAGGACCTGGGAGCTCTGCTGGGGAACAG
CACCGATCCCGGAGTGTTCACAGATCTGGCCTCCG
TGGACAACTCTGAGTTTCAGCAGCTGCTGAATC
AGGGCGTGTCCATGTCTCATAGTACAGCCGAACC
AATGCTGATGGAGTACCCCGAAGCCATTACCCGG
CTGGTGACCGGCAGCCAGCGGCCCCCCGACCCCGC
T C C A A C T C C C C T G G G A A C C A G C G G C C T G
CCTAATGGGCTGTCCGGAGATGAAGATTTCTCAAG
CATCGCTGATATGGACTTTAGTGCCCTGCTGTCACA
GATTTCCTCTAGTGGGCAGGGAGGAGGTGGAAGCGG
CTTCAGCGTGGACACCAGTGCCCTGCTGGACCTG
TTCAGCCCCTCGGTGACCGTGCCCGACATGAGCCTG
CCTGACCTTGACAGCAGCCTGGCCAGTATCCAAGA
GCTCCTGTCTCCCCAGGAGCCCCCCAGGCCTCCCG
AGGCAGAGAACAGCAGCCCGGATTCAGGGAAGCAG
CTGGTGCACTACACAGCGCAGCCGCTGTTCCTG
CTGGACCCCGGCTCCGTGGACACCGGGAGCAACG
ACCTGCCGGTGCTGTTTGAGCTGGGAGAGGG
CTCCTACTTCTCCGAAGGGGACGGCTTCGCCGAGG
ACCCCACCATCTCCCTGCTGACAGGCTCGGAGCC
TCCCAAAGCCAAGGACCCCACTGTCTCCTAGC
CTAGGACTAGAGCAAACTAGAATTGTTGGCATCAGG
TAGGCATCACACACGATTAACAACCCCTAAAAATAC
ACTTTGAAAATATTGAAAATATGTTTTTGTATACA
TTTTTGATATTTTCAAATAATACGCAGTTATAAAACTC
ATTAGCTAACCCATTTTTTCTTTGCTTATGCTTAC
A G A T T G C A A A G A A C T A G A G C C G C G G G A T C
TTTGTGAAGGAACCTTACTTCTGTGGTGTGACATAA
TTGGACAAACTACCTACAGAGATTTAAAGCTCT

AAGGTAAATATAAAATTTTTAAGTGTATAATGTGTT
AAACTACTGATTCTAATTGTTTGTGTATTTTAGATT
CCAACCTATGGAACTGATGAATGGGAGCAGTGGT
GGAATGCCTTTAATGAGGAAAACCTGTTTTGCTCA
GAAGAAATGCCATCTAGTGATGATGAGGCTACTGCT
GACTCTCAACATTCTACTCCTCCAAAAAAGAAGAG
AAAGGTAGAAGAGCCCAAGGACTTTCCTTCAGAATT
GCTAAGTTTTTTGAGTCATGCTGTGTTTAGTAAT
AGAACTCTTGCTTGCTTTGCTATTTACACCACAAAGG
AAAAAGCTGCACTGCTATACAAGAAAATTA
TGGAAAAATATTTGATGTATAGTGCCTTGACTAGAGA
TCATAATCAGCCATACCACATTTGTAGAGGTTTTACTT
GCTTTAAAAAACCTCCCACACCTCCCACTGAA
CCTGAAACATAAAATGAATGGAATTGTTGTTGTT
AACTTGTTTATTGCAGCTTATAATGGTTACAAAT
AAAGCAATAGCATCACAAATTTCACAAATAAAG
CATTTTTTTCACTGCATTCTAGTTGTGGTTTGT
CCAAACTCATCAATGTATCTTAACTAGTGATGAAAA
CAAAAACAACTGTGTTGAAAATTAAAAATTTTGTATG
CATACGCATTAAGCGAACATTAAAAAGATGTGAAAA
CATAACTATTATGTCTAAATAAACACACGTCAGATGT
ATGTACGTCAACGGAAAACCATTGTCTATATATT
ACGCTAGTAATTCTTTCCTGTAAAATAAATATAAAT
AAATATGCTTTATTGACAGAAAATTTGATGTTTTTGTA
TTTGAGTAGGAGCAATCACAGGTGAGCAAAAAAAG
CACCGACTCGGTGCCACTTGGCCCTGCAGACATGGG
TGATCCTCATGTTGGCCAAGTTGATAACGGACTAG
CCTTATTTTAACTTGCTAGGCCCTGCAGACATGGGTG
ATCCTCATGTTGGCCTAGCTCTAAAACAGG
TCTTCTCCAATTCACATCGACTGAAATCCCTGGTAAT
CCGTTTTAGAATCCATGATAATAATTTTTTGGATGATT
GGGAGCTTTTTTTTGCACGTTCAAAATTTTTT
GCAACCCCTTTTTGGAAACGAACACCACGGTA
GGCTGCGAAATGCCCATACTGTTGAGCAATTCACGTT
CATTATAAATGTCGTTCGCGGGCGCAACTGCAACT
CCGATAAATAACGCGCCCAACACCGGCATAAAGAA
TTGAAGAGAGTTTTCACTGCATACGACGATTCTGTGA
TTTGTATTCAGCCCATATCGTTTCATAGCTTCT
GCCAACCGAACGGACATTTCGAAGTACTCAGCGT
AAGTGATGTCCACCTCGATATGTGCATCTGTAAAAG
CAATTAATTGTTCCAGGAACCAGGGCGTATCTCTT
CATAGCCTTATGCAGTTGCTCTCCTCTAGTGGT
GAAGGGGGCGGCCGCGGAGCCTGCTTTTTTGT
ACAAAGTTGGCATTATAAAAAAGCATTGCTCAT
CAATTTGTTGCAACGAACAGGTCACTATCAGT
CAAAATAAAATCATTATTTGGGGCCCGAGCTTAAGA
CTGGCCGTCGTTTTACAACGTCGTGACTGGGAAAA
CATCCATGCAAGCGGCTGAATATGGGATGTTTTAT
GGGATGTTTTCTAGACTCTACGAAGACCCCGAAG

TATTGAGGAAAACATACCTATATAAATGATCAACAT
CAGGAAAGAGCAGTTGAGAATTATAAGAATT
GGCAAATGGTCTTAAGAACCCTCTGCTTAAGATTTT
CAAAATTTCCTTTAGAATCAAAGTGTCCTATTGTT
CGTTTGTTGAAAACACAGTTCGATTTATTGACTAT
AATAAATTGATAGTTTTAAATATAGAGGCACGACTAA
GAGAGTTTGGTTTTGTTTTGGTTAGGGACC
AAAAAAAGTATACATAACGAATAAAAAAGGATTTAA
GTGCAAATGGTAAAAAAGTGTCGAGTTTTTCTT
GAGTTGATTGTGCTGTAATGAGACTCTGCATTCGGC
ATTTGACTCGGCTTTTCCTACTCGTGCCGTATTT
CAGGCTGCAAGTCGAACAAGCTAGCAGAGCTCT
GGCCACGTAATAAGTGTGCGTTGAATTTATT
CGCAAAAACATTGCATATTTTCGGCAAA
GTAAAATTTTGTTGCATACCTTATCAAAAAATAAGT
GCTGCATACTTTTTAGAGAAACCAAATAATTTTTTATT
GCATACCCGTTTTTAATAAAATACATTGCATACCCT
CTTTTAATAAAAAATATTGCATACTTTGACGAAA
CAAATTTTCGTTGCATACCCAATAAAAGATTATTAT
ATTGCATACCCGTTTTTAATAAAATACATTGCATA
CCCTCTTTTAATAAAGAATATTGCATACGTTGAC
GAAACAAATTTTCGTTGCATACCCAATAAAAGATT
ATTATATTGCATACCTTTTCTTGCCATACCATTTAGCC
GATCAATTCTGCTCGGCAACAGTATATTTGTGGTGT
GCCAACCAACAACGAGCTCCAGCTTTTGTTCCCTTT
AGTGAGGGTTAATTTCGAGCTTGGCGTAATCATGGT
CATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCA
CAATTCCACACAACATACGAGCCGGAAGCATAAAGT
GTAAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCA
CATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGT
CGGGAAACCTGTCGTGCCAGCTGCATTAATGAAT
CGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATT
GGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCT
GCGCTCGGTCGTTCGGCTGCGGCGAGCGGTAT
CAGCTCACTCAAAGGCGGTAATACGGTTATCCACA
GAATCAGGGGATAACGCAGGAAAGAACATGTGAG
CAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAA
GGCCGCGTTGCTGGCGTTTTTCCATAGGCT
CCGCCCCCCTGACGAGCATCACAAAAATCGACGCT
CAAGTCAGAGGTGGCGAAACCCGACAGGACTAT
AAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCT
CGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACC
GGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGT
GGCGCTTTCTCATAGCTCACGCTGTAGGTATCT
CAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCT
GTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGC
GCCTTATCCGGTAACTATCGTCTTGAGTCCAACCC
GGTAAGACACGACTTATCGCCACTGGCAGCAGCCAC
TGGTAACAGGATTAGCAGAGCGAGGTATGT

AGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCT
AACTACGGCTACACTAGAAGGACAGTATTTGGTATCT
GCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGA
GTTGGTAGCTCTTGATCCGGCAAACAAACCA
CCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAG
CAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGA
TCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGT
GGAACGAAAACTCACGTTAAGGGATTTTGGTCATGA
GATTATCAAAAAGGATCTTCACCTAGATCCTTTT
AAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATA
TATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAA
TCAGTGAGGCACCTATCTCAGCGATCTGTCTA
TTTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTGT
AGATAACTACGATACGGGAGGGCTTACCATCT
GGCCCCAGTGCTGCAATGATACCGCGAGACCCACGC
TCACCGGCTCCAGATTTATCAGCAATAAACCAGCCA
GCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAA
CTTTATCCGCCTCCATCCAGTCTATTAATTGTT
GCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAAT
AGTTTGCGCAACGTTGTTGCCATTGCTACAGGCAT
CGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATT
CAGCTCCGGTTCCCAACGATCAAGGCGAGTTACAT
GATCCCCCATGTTGTGCAAAAAAGCGGTTAGCT
CCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGG
CCGCAGTGTTATCACTCATGGTTATGGCAGCACT
GCATAATTCTCTTACTGTCATGCCATCCGTAAGAT
GCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATT
CTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTT
GCCCGGCGTCAATACGGGATAATACCGCGCCACAT
AGCAGAACTTTAAAAGTGCTCATCATT
GGAAAACGTTCTTCGGGGCGAAAACTCTCAAGG
ATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCC
ACTCGTGCACCCAACTGATCTTCAGCATCTTTT
ACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGG
AAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGA
CACGGAAATGTTGAATACTCATACTCTTCCTTTTT
CAATATTATTGAAGCATTTATCAGGGTTATTGTCTCAT
GAGCGGATACATATTTGAATGTATTTAGAAAAATAAA
CAAATAGGGGTTCCGCGCACATTTCCCCGAAAA
GTGC.

2. The targeting sequence of sgRNA can be designed manually or through the website (*http://www.flyrnai.org/crispr/*). The targeting site of sgRNA locates nearby −200 nt upstream the target gene's TSS. It shall not include TTTT sequence. Limit complementarity to other genomic loci to less than 17 nt to avoid off target effects. To

3. If only one targeting sequence of the sgRNA needs to be cloned, the next step will switch to procedure 3.6. If multiple sgRNAs need to be cloned, the next step should go directly to procedure 3.5.
4. *Nhe* I and *Spe* I produce compatible ends. To avoid self-ligation of the linearized vector, use CIP to prevent this event.
5. If the flySAM construct is integrated into chromosome II, collect the embryos of *y,sc,v,nanos-integrase*;attP40 (TB16) for injection. On the other hand, if it is integrated into chromosome III, collect embryos of *y,sc,v,nanos-integrase*;;attP2 (TB18) for injection.
6. G1 males shall be crossed with corresponding balancer, their offspring carrying balancer will limit the unpredicted recombination. If the plasmid is integrated into attP40 on chromosome II, *y,sc,v*;*Gla,Bc/CyO* (TB23) will be used to cross with G1 males; on the other hand, *y,sc,v*;;*Dr,e/TM3,Sb* (TB139) will be chosen for chromosome III.

Acknowledgments

This work was supported by the National Key Technology Research and Development Program of the Ministry of Science and Technology of the People's Republic of China (2016YFE0113700) and the National Natural Science Foundation of China (20181300988, 20201300797).

References

1. Xu J, Ren X, Sun J, Wang X, Qiao HH, Xu BW et al (2015) A toolkit of CRISPR-based genome editing systems in Drosophila. J Genet Genomics 42(4):141–149. https://doi.org/10.1016/j.jgg.2015.02.007
2. Xu RG, Wang X, Shen D, Sun J, Qiao HH, Wang F et al (2019) Perspectives on gene expression regulation techniques in Drosophila. J Genet Genomics 46(4):213–220. https://doi.org/10.1016/j.jgg.2019.03.006
3. Ren X, Sun J, Housden BE, Hu Y, Roesel C, Lin S et al (2013) Optimized gene editing technology for Drosophila melanogaster using germ line-specific Cas9. Proc Natl Acad Sci U S A 110(47):19012–19017. https://doi.org/10.1073/pnas.1318481110
4. Qiao HH, Wang F, Xu RG, Sun J, Zhu R, Mao D et al (2018) An efficient and multiple target transgenic RNAi technique with low toxicity in Drosophila. Nat Commun 9(1):4160. https://doi.org/10.1038/s41467-018-06537-y
5. Peng P, Wang X, Shen D, Sun J, Jia Y, Xu RG et al (2019) CRISPR-Cas9 mediated genome editing in Drosophila. Bio Protoc 9(2):e3141. https://doi.org/10.21769/BioProtoc.3141
6. Wang F, Qiao HH, Xu RG, Sun J, Zhu R, Mao D et al (2019) pNP transgenic RNAi system manual in Drosophila. Bio Protoc 9(3):e3158. https://doi.org/10.21769/BioProtoc.3158
7. Huang Y, Bugg W, Bangs M, Qin G, Drescher D, Backenstose N et al (2021) Direct and pleiotropic effects of the Masou Salmon Delta-5 Desaturase transgene in F1 channel catfish (Ictalurus punctatus). Transgenic Res 30(2):185–200. https://doi.org/10.1007/s11248-021-00242-1
8. Cox DN, Chao A, Lin H (2000) Piwi encodes a nucleoplasmic factor whose activity modulates the number and division rate of germline stem cells. Development 127(3):503–514. https://doi.org/10.1242/dev.127.3.503

9. Smart AD, Pache RA, Thomsen ND, Kortemme T, Davis GW, Wells JA (2017) Engineering a light-activated caspase-3 for precise ablation of neurons in vivo. Proc Natl Acad Sci U S A 114(39):E8174–E8183. https://doi.org/10.1073/pnas.1705064114
10. Jia Y, Xu RG, Ren X, Ewen-Campen B, Rajakumar R, Zirin J et al (2018) Next-generation CRISPR/Cas9 transcriptional activation in Drosophila using flySAM. Proc Natl Acad Sci U S A 115(18):4719–4724. https://doi.org/10.1073/pnas.1800677115
11. Dominguez AA, Lim WA, Qi LS (2016) Beyond editing: repurposing CRISPR-Cas9 for precision genome regulation and interrogation. Nat Rev Mol Cell Biol 17(1):5–15. https://doi.org/10.1038/nrm.2015.2
12. Gilbert LA, Larson MH, Morsut L, Liu Z, Brar GA, Torres SE et al (2013) CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. Cell 154(2):442–451. https://doi.org/10.1016/j.cell.2013.06.044
13. Bak RO, Gomez-Ospina N, Porteus MH (2018) Gene editing on center stage. Trends Genet 34(8):600–611. https://doi.org/10.1016/j.tig.2018.05.004
14. Jimi E, Fei H, Nakatomi C (2019) NF-kappaB signaling regulates physiological and pathological chondrogenesis. Int J Mol Sci 20(24). https://doi.org/10.3390/ijms20246275
15. Gomez-Pastor R, Burchfiel ET, Thiele DJ (2018) Regulation of heat shock transcription factors and their roles in physiology and disease. Nat Rev Mol Cell Biol 19(1):4–19. https://doi.org/10.1038/nrm.2017.73
16. Konermann S, Brigham MD, Trevino AE, Joung J, Abudayyeh OO, Barcena C et al (2015) Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. Nature 517(7536):583–588. https://doi.org/10.1038/nature14136
17. Mao D, Jia Y, Peng P, Shen D, Ren X, Zhu R et al (2020) Enhanced efficiency of flySAM by optimization of sgRNA parameters in Drosophila. G3 (Bethesda) 10(12):4483–4488. https://doi.org/10.1534/g3.120.401614
18. Sun J, Wang X, Xu RG, Mao D, Shen D, Wang X et al (2021) HP1c regulates development and gut homeostasis by suppressing Notch signaling through Su(H). EMBO Rep 22(4):e51298. https://doi.org/10.15252/embr.202051298

Chapter 9

Tracing and Manipulating *Drosophila* Cell Lineages Based on CRISPR: CaSSA and CLADES

Jorge Garcia-Marques and Tzumin Lee

Abstract

Cell lineage defines the mitotic connection between cells that make up an organism. Mapping these connections in relation to cell identity offers an extraordinary insight into the mechanisms underlying normal and pathological development. The analysis of molecular determinants involved in the acquisition of cell identity requires gaining experimental access to precise parts of cell lineages. Recently, we have developed CaSSA and CLADES, a new technology based on CRISPR that allows targeting and labeling specific lineage branches. Here we discuss how to better exploit this technology for lineage studies in *Drosophila*, with an emphasis on neuronal specification.

Key words Cell lineage, CRISPR, CaSSA, CLADES, Neuronal specification, Genetic cascades, Genetic intersections

1 Introduction

The *Drosophila* central brain hosts a diverse population of neurons, emerging from ~100 neuroblasts (the neural stem cells) per hemisphere [1, 2]. Determined by molecular gradients and other specification factors, most of these neuroblasts exhibit unique identity features. Neuroblasts divide ~100 times, producing a clone of neurons (for some lineages also glia) that we refer to as neuronal lineages. There are two types of these lineages in the fly brain: type I and type II. In type I lineages, neuroblasts divide to self-renew and produce a ganglion mother cell (GMC) that, in turn, divides again and gives rise to two neurons. Due to molecular mechanisms of asymmetric cell division, each of these neurons acquires a different identity. In type II lineages, the neuroblast divides to produce another neuroblast and an intermediate neural progenitor (INP). Each INP divides approximately eight to nine times to self-renew

Jorge Garcia-Marques and Tzumin Lee contributed equally with all other contributors.

Christian Dahmann (ed.), *Drosophila: Methods and Protocols*,
Methods in Molecular Biology, vol. 2540, https://doi.org/10.1007/978-1-0716-2541-5_9,

and produce a GMC that, in turn, will divide again to produce two neurons. In some lineages, especially type II lineages, the neurogenesis ends with transformation of the neuroblast into a glial progenitor that undergoes mitosis to produce a variable number of glial cells. Together with the identity acquired by the neuroblast from spatial signals, the division of the GMC results in asymmetrical molecular inheritance that contributes to neuron diversity. Another source for neuronal diversity is the temporal change undergone by the neuroblast that produces different types of neurons over time. Tracing representative type I lineages has revealed diverse manifestation of neuronal temporal fate patterning. In the four repeated mushroom body lineages, the neuroblast only produces three main classes of neurons throughout embryonic, larval, and pupal development [3]. By contrast, the dorsal antennal lobe lineage, which ends shortly after pupal formation and produces only one viable neuron from each GMC, can yield an invariant sequence of ~40 neuronal types [4, 5].

Understanding the molecular mechanisms underlying neuronal specification requires genetic access to single neuroblast types and their progeny. This demand inspired a new generation of genetic tools based on the combination of CRISPR and single-strand annealing (SSA), a conserved mechanism for DNA repair [6]. This technology includes CaSSA [7], a tool to genetically target discrete neuroblast types, and CLADES [8], to program genetic cascades that progress throughout the neuroblast lifespan (Fig. 1).

2 CaSSA

2.1 Rationale

CaSSA is a platform to create unlimited genetic switches based on Cas9 (Ca) and SSA (Fig. 2). Essential for the CRISPR function, Cas9 is an endonuclease that cuts DNA in a region specified by a 20-bp sequence encoded in a guide RNA (gRNA) [9, 10]. SSA is a well-conserved repair mechanism evolved to repair repetitive DNA [6]. When a double-strand break (DSB) occurs in a region between two direct repeats, SSA repairs the DNA by removing one of the repeats, as well as the intervening sequence. The fact that the DNA repair outcome is predictable allows the design of genetic elements that, in their original state, are inactive and become active only after SSA occurs. These genetic switches are the basis for multiple applications.

2.2 A Simple Reporter

In the simplest implementation of CaSSA, a switchable reporter gene is generated by introducing a switch cassette in the middle of its sequence (Fig. 2a). This cassette consists of two direct repeats that are part of the reporter sequences, separated by one or multiple target sites for a specific gRNA. The presence of this switch cassette

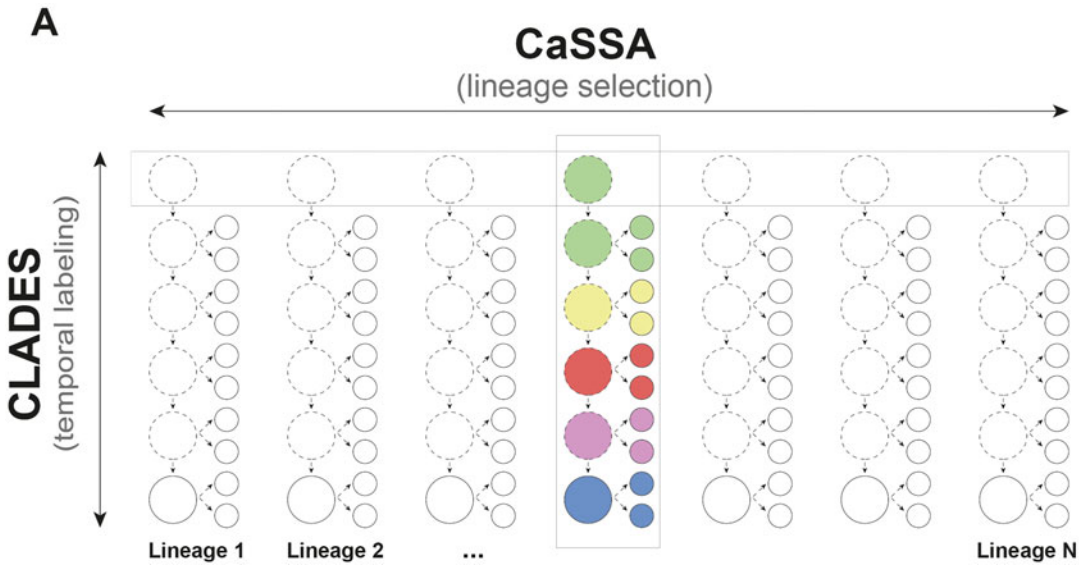


Fig. 1 Overview of the applications of CaSSA and CLADES for cell lineage studies. Whereas CaSSA targets specific cell lineages based on genetic intersections of multiple genetic drivers, CLADES triggers genetic cascades of reporter genes for birth dating

interrupts the open reading frame (ORF) of the reporter, thus impeding the translation of the second half of the reporter. The expression of Cas9 and a matching gRNA generates a DSB in the switch cassette, which is repaired by SSA. This mechanism for DNA repair acts by removing one of the repeats and the region with the target sites, resulting in the intact reporter sequence. The power of CaSSA derives from CRISPR that allows the design of a virtually unlimited number of distinct gRNAs. Different gRNAs can be used at the same time to activate independent reporters.

Activating a CaSSA reporter is a non-reversible event that permanently changes the DNA of the cell. When CaSSA is triggered in a dividing cell, all the progeny inherits that change and therefore expresses the reporter gene. This property can be used to target specific neuronal lineages by expressing a gRNA in a subset of neuroblasts. The CaSSA event is confined to neuroblasts by expressing Cas9 from deadpan (dpn) regulatory sequences [11]. The combination of Cas9 expressed exclusively in neuroblasts and a gRNA in a subset of neuroblasts labels specific neuronal lineages. We have used this strategy to target the antennal lobe lineages by using the 44F03 regulatory fragment driving the transcription of the gRNA, or the 41A10 or 14D11 fragments for mushroom body lineages [7].

The structure of a gRNA has constraints that complicate their expression from promoters that recruit the RNA polymerase II. These type II promoters are not suitable because the 5' CAP region, possible 5′/3' UTR, and the polyA tail may hinder the

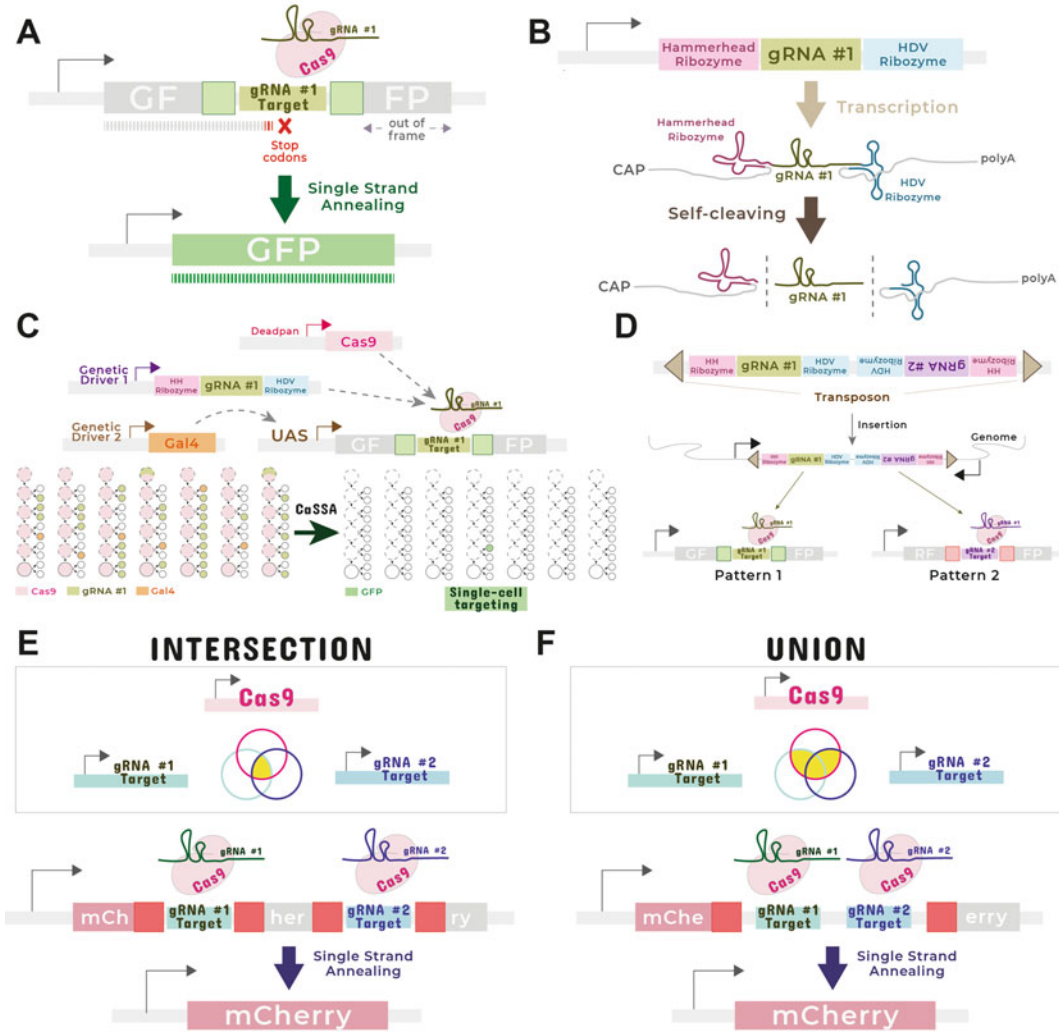


Fig. 2 Key features of CaSSA. (**a**) A basic CaSSA reporter is a reporter gene that is disrupted by a switch cassette. This cassette consists of two direct repeats that duplicate part of the reporter sequence, separated by one or multiple target sites for a specific gRNA. The switch cassette contains stop codons to prevent translation of the second half of the reporter, which is also in a different ORF. In the presence of Cas9 and the gRNA matching the target site in the reporter, cells repair this reporter by SSA. This mechanism for DNA repair removes the intervening sequence and one of the repeats, thus resulting in a reconstituted version of the reporter. (**b**) In CaSSA, gRNAs are transcribed from type II promoters by using a system of ribozymes. Each gRNA is flanked by the HH ribozyme in the 5′-end and the HDV virus ribozyme in the 3′-end. The mRNA undergoes self-cleavage catalyzed by the intrinsic nuclease activities of the ribozymes to release the desired gRNA. This gRNA is perfectly functional as it lacks the 5′-end capping, the 3′-end polyadenylation and any other accessory sequences. (**c**) CaSSA can be combined with the GAL4 system to target specific cells. In a typical experiment, CaSSA activates a reporter in a specific neuronal lineage in the presence of Cas9 and the gRNA #1. Since Cas9 is expressed under the neuroblast-specific dpn promoter and the gRNA #1 is driven by a genetic driver that is active only in a subset of lineages, the intersection results in only a couple of lineages in which the reporter is active. However, the transcription of this reporter is controlled by GAL4, resulting in a single cell expressing GFP. (**d**) CaSSA can be applied to finding genetic drivers of interest. A transposon vector containing two different gRNAs in opposite directions is inserted into the genome at random locations. Thus, the regulatory sequences flanking the insertion site control the transcription of each gRNA, capturing the

gRNA function [12]. Initially, gRNAs were typically transcribed from type III promoters. These are ubiquitous promoters that guarantee a very strong and ubiquitous transcription. Although these promoters produce uncapped, non-polyA RNAs, they lack tissue or cell specificity. An important advance for the field was the synthesis of gRNAs from type II promoters, which include most of the tissue-specific promoters described, by incorporating ribozyme sequences [12]. These self-excising RNA sequences process gRNAs so that the polyA or CAP regions are removed, producing the exact short sequence required for the gRNA function (Fig. 2b). Since type II are not as strong as type III promoters, multiple copies are often needed. CaSSA achieves effective levels of gRNA expression by using type II promoters followed by multiple copies of a RGR (ribozyme-gRNA-ribozyme) cassette, which contains the sequence of a gRNA flanked by the Hammerhead (HH) and the HDV ribozyme.

2.3 Refining Expression Patterns with GAL4

The ability to consistently target expression of transgenes to small subsets of neurons is crucial for structural-functional studies in complex nervous systems. Given one neural stem cell can produce multiple neural types, it is beneficial to target neural types rather than an entire lineage for detailed structural-functional analyses. Further, neurons do not migrate actively in the fly central brain, resulting in densely packed neuronal clones. Since discerning neuronal morphology is difficult when most of the neurons in the clone are labeled, targeting a small portion of the lineage is often desirable. However, targeting small parts of a lineage is problematic as the activation of CaSSA reporters is a non-reversible event that labels the entire neuroblast progeny.

Binary systems (i.e., GAL4, LexAp65, etc.) and the rest of genetic approaches used in fly do not rely on SSA, which makes them fully compatible with CaSSA. Collections of thousands of GAL4 [13–15] and Split-GAL4 [16, 17] fly lines have been particularly useful to reproducibly direct expression of transgenes to defined, small subsets of neurons. A GAL4 line consists of a genomic fragment controlling the expression of GAL4 that, in turn, activates the expression of genes controlled by UAS elements [18]. Further refinement is achieved with a Split-GAL4 line

Fig. 2 (continued) transcription of the sense and antisense strand at the site of insertion, and resulting in different lineages labeled with CaSSA. (**e**) In a CaSSA reporter for genetic intersection, the reporter gene is disrupted by two switch cassettes. This reporter can be reconstituted only in the presence of Cas9 and two gRNAs matching the target site in each of the switch cassettes. Cas9 and each of these gRNAs are driven by a different genetic driver. The final pattern results from the intersection between the three different genetic drivers. (**f**) The CaSSA reporter gene for genetic union relies on a single switch cassette that contains two or more target sites for different gRNAs. In the presence of Cas9 and a gRNA matching any of these target sites, the reporter gene undergoes SSA and becomes active

[19]. These lines contain two genomic fragments, each driving the expression of either the DNA-binding (DBD) or the transcription activation (AD) GAL4 domain. The expression of both domains reconstitutes a functional GAL4 protein, thus intersecting the expression pattern of both genomic fragments.

To combine CaSSA with GAL4, we have generated versions in which the reporter expression is controlled by UAS regulatory elements, restricting the expression of a CaSSA reporter to a subset of GAL4-positive neurons. This approach allowed us to target a specific neuron per hemisphere in the adult fly brain, based on the intersection between a Split-GAL4 line and a line bearing a lineage-specific driver that activates CaSSA in a few lineages (Fig. 2c).

2.4 Finding New Genetic Drivers via CaSSA-Based Gene Trapping

While we primarily utilize screened ~3-kb genomic fragments to restrict gRNA expression, CaSSA can be applied to finding new genetic drivers that are specifically expressed in the lineage(s) of interest. For that we developed a gene trapping system consisting of two gRNA cassettes placed in opposite directions in a transposon vector (Fig. 2d). Hopping this transposon by inducing the expression of transposase with heat shock can capture the expression of the flanking genes.

Previous gene trapping strategies are highly inefficient. Most of the events of transposon hopping are not informative for several reasons, including the reporter gene landing in the correct locus but on a different ORF or orientation [20]. CaSSA is an extremely efficient strategy for gene trapping, with most of the hits revealing expression patterns. CaSSA relies on the transcription of gRNAs, which makes it independent of the ORF, thus theoretically increasing the number of useful hits by 66%. Also, CaSSA uses two different gRNAs in opposite directions to avoid the problem of the transposon landing in an incorrect orientation. Each gRNA then activates a different reporter in the same brain. The only limitation of this strategy is that gene expression can only be reported in the context of lineages. Since the CaSSA event labels the entire progeny arising from a progenitor cell, the expression pattern of the reporter reflects both present and past gRNA expression.

2.5 Intersections and Unions

Gaining genetic access to a specific neuronal lineage requires targeting a particular neuroblast with precision. Due to the pleiotropism of most genes and enhancers, a single genetic driver is often unable to select a single out of the ~100 existent neuroblasts. We need to resort to genetic intersections in which the genetic targeting occurs only in cells expressing multiple genetic markers. Genetic intersections can be achieved in various ways. Whereas some approaches require both genetic drivers to be simultaneously active in the same cell (i.e., Split-Gal4), other strategies are based on the activity of the drivers in the same lineage either at the same or different developmental times [11].

CaSSA is a powerful method to achieve intersections of multiple genetic drivers, thus resulting in a more refined expression pattern. Distinct gRNAs can control multiple target sites independently to reconstitute a reporter gene only when all the switches are turned on. CaSSA reporters used for intersections are disrupted by two different switch cassettes, each with target sites for a different gRNA (Fig. 2e). The simultaneous expression of a gRNA driven by a specific genetic driver and Cas9 results in the independent reconstitution of each switch cassette. The combined activity of the three drivers, one for each gRNA and one for Cas9, reconstitutes both switch cassettes, thus activating reporter expression. While this intersection requires each gRNA and Cas9 to be present at the same time, both gRNAs do not need to be present in the same cell if they exist sometime in the same lineage. This strategy succeeded to genetically target the single type of mushroom body lineages out the ~100 total lineages [7]. Finding new genetic drivers with restricted activity could extend this specificity to other lineages.

Similar to genetic intersections, CaSSA also achieves the union of the expression pattern of two or more genetic drivers (Fig. 2f). This strategy is useful in those cases in which the genetic drivers are not fully penetrant. CaSSA reporters designed for genetic union consist of a switch cassette containing target sites for more than one gRNA. Thus, even if a genetic driver remains silent, a second genetic driver can still activate the reporter. This makes the system redundant and increases the effectiveness of the approach.

2.6 Notes

- CaSSA relies on inducing DSBs that are repaired by SSA. Researchers often ask in our presentations about the impact of DSBs on the cell and organism physiology. Despite this is a reasonable concern, we have never observed developmental defects related to the use of CaSSA, in contrast to other genetic tools including GAL4, recombinases or base-editing Cas9 variants. Every cell undergoes an average of natural 10–50 DSBs per day [21], which are compatible with their physiology. Another aspect to consider is that CaSSA-induced DSB is not accompanied by DNA lesions, in contrast to DNA-damaging chemical agents. Also, CaSSA reporters contain homologous repeats that probably make the DNA repair process easier for cells.
- The sequence of the ribozymes that process the gRNA varies depending on the sequence of the gRNA. Specific instructions to design gRNAs flanked by the HH and HDV ribozymes can be found in Gao & Zhao [12].
- When translation follows transcription, a single mRNA copy is capable to produce multiple protein units, thus amplifying the genetic response. gRNAs do not undergo this amplification

process so that a single gRNA copy is often not enough to achieve high expression and effectiveness of CaSSA. We have tested the impact of the gRNA copy number on CaSSA effectiveness, which reaches a plateau at 12 gRNA copies. Recursive cloning based on compatible restriction enzymes makes easy to assemble cassettes with multiple copies of gRNAs flanked by ribozymes. In our experience, up to 24 copies are easily assembled and remain stable in bacteria and flies.

- There are two mechanisms to disrupt the reporter gene while introducing the switch cassette: (1) incorporating stop codons in between the target sites and (2) placing the second half of the reporter out of frame. When possible, it is recommendable to design constructs that exploit both mechanisms.
- The position of the switch cassette in the reporter is important. It should be inserted in such a way that the first half of the reporter is not active without the second half. For reporters that consist of dimers, such as tdTomato, placing the switch cassette after the first monomer is not sufficient to abolish fluorescence since the first monomer remains intact.
- Designing a CaSSA construct should take into consideration the existence of potential splicing sites. These sites may result in isoforms that lack the switch cassette and produce a functional reporter gene. The strongest splicing sites should be destroyed by codon-optimizing the reporter sequence. Splicing can be predicted by using NNSPLICE [22] or similar tools.
- Although SSA acts in DNA repeats as small as 10 bp, longer repeats increase effectiveness with a plateau at 400 bp [23]. We have observed high CaSSA effectiveness with 100-bp repeats. Another way to increase effectiveness is by minimizing the distance from each repeat to the gRNA target site. Although SSA can repair DNA when repeats are as distant as hundreds of kb, the process is length dependent [24].
- The number of target sites per switch cassette is important since each target represents an opportunity for SSA to occur. If one of these sites is destroyed by non-homologous end joining (NHEJ), there are additional opportunities for SSA to recombine homologous repeats and remove all the target sites. In our constructs, we have included at least two to four sites. Adjacent sites should be oriented in opposite directions to avoid SSA happening between these sites. Since the probability of SSA depends on the distance between repeats, adjacent sites oriented in the same direction might trigger SSA more efficiently, thus delaying the desired SSA event.
- A limiting factor in the effectiveness of the intersections is the possibility of inter-target deletions destroying the reporter.

To avoid that, new reporters in which switch cassettes are separated by introns would probably reduce the probability of these deletions.

- SSA is highly dependent on proteins expressed in mitotic cells, which confers additional specificity to target progenitor cells in developmental studies [25]. However, CaSSA remains limited for targeting non-dividing cells. Screening for Cas9 fusions to proteins involved in DNA repair might enable CaSSA in non-dividing cells. These Cas9 variants would also increase the efficiency of CaSSA in dividing cells.

3 Clades

3.1 Key Concepts

Having the tools to specifically select one or few lineages and deliver genetic reagents only solves part of the problem in lineage studies in *Drosophila*. Often it is necessary to link different cell identities to birth order. Mapping how a progenitor generates different cell types at different developmental times is crucial to understand cell specification. How much of this process is due to temporal mechanisms occurring in the stem cell? Is temporal specification linked to developmental stage? To what extent the number of mitoses undergone by the stem cell or its developmental age determine cell identity? CLADES was invented to solve the problem of easily tracking birth order of neurons in a single individual. With this tool, a cell lineage can be divided into five different color windows, each labeling the neuronal progeny derived from a very specific temporal window in the life of the neuroblast (Fig. 3a).

CLADES delivers a cascade of reporter activation and inactivation that is pre-programmed. This cascade relies on switchable gRNAs that can be activated by other gRNAs, based on the same mechanism as CaSSA reporters. However, given the size constrains of the gRNA sequence, the direct repeats are only ~20 bp long. Despite various studies suggesting that the structure of the gRNA is very sensitive to minor modifications [26], we needed a thorough process of optimization to create a gRNA design in which the insertion of the switch cassette abolished gRNA function. Having gRNAs that can be activated by another gRNA allows the design of gRNA cascades in which a gRNA #1 can trigger a sequence of gRNA activation. Each of these gRNAs can then be linked to specific events of reporter activation or inactivation, resulting in coordinated reporter expression.

Both Cas9 cutting and SSA are events that are highly stochastic. Therefore, the order of reporter activation does not necessarily follow the same order of gRNA activation. The second reporter can be activated prior to the first reporter even if the gRNA#1 was activated before the gRNA#2. Strict coupling of the activation with

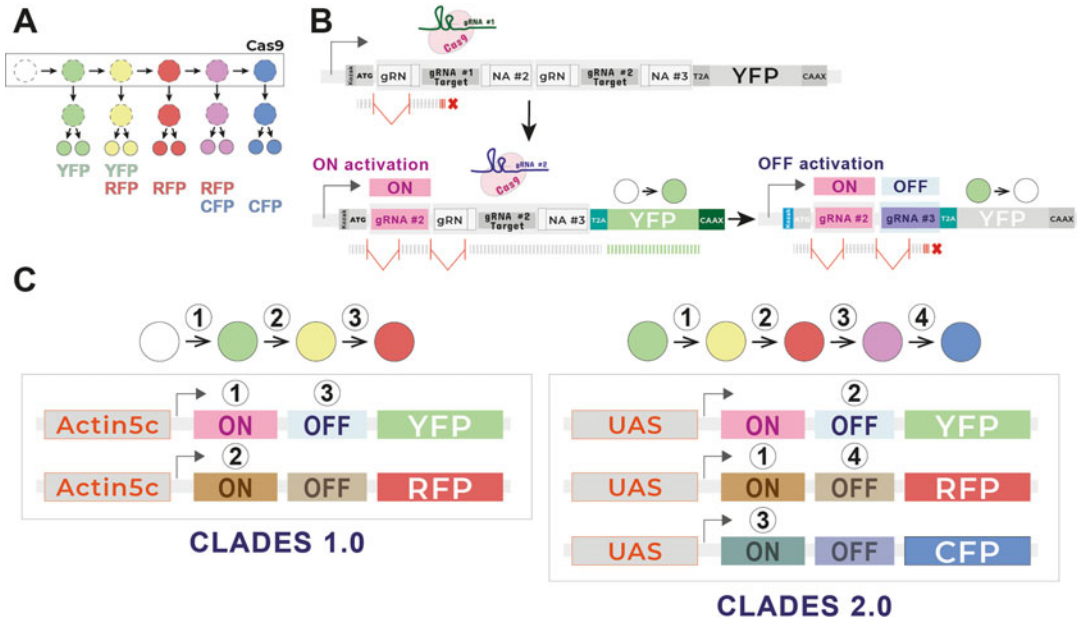


Fig. 3 Key features of CLADES. (**a**) CLADES enables triggering cascades of sequential activation and inactivation of reporter genes. Since Cas9 is exclusively expressed in the stem cell, the cascade progression results in progeny labeled in different colors. The sequence of activation and deactivation in the scheme is as follows: YFP (activation), RFP (activation), YFP (deactivation), BFP (activation), RFP (deactivation). (**b**) A CLADES construct consists of a reporter gene whose expression is controlled by two regulatory cassettes, the ON and the OFF cassette. Each cassette consists of a conditional gRNA to be activated by another gRNA by Cas9 and SSA. When Cas9 is expressed together with a matching gRNA, the gRNA in the ON cassette is activated. Simultaneously, this event brings the reporter in frame thus activating it. The expression of Cas9 and another matching gRNA (reconstituted from another conditional gRNA) activates the gRNA in the OFF cassette, which brings the reporter out of frame. For the proper function of this construct, a system of controlled splicing guarantees the reporter gene is in the correct ORF in every event. (**c**) In CLADES 1.0, the cascade consists of two different reporters driven by the ubiquitous Actin5C promoter. The sequence in the cascade is as follows: YFP, YFP and RFP, RFP. (**d**) In CLADES 2.0, the cascade consists of three reporters controlled by UAS sequences that allow reporter expression only in the presence of GAL4. Since the first reporter is pre-activated, CLADES 2.0 guarantees the lineage is labeled from the beginning. The sequence of reporter expression is as follows: YFP, YFP and RFP, RFP, RFP and CFP, CFP

the activation or inactivation of the reporter is key to ensure the desired combinations of reporters. This coupling is guaranteed by the CLADES construct, which is based on the insertion of switchable gRNA cassettes in the same ORF as the reporter (Fig. 3b).

3.2 The CLADES Construct

The CLADES construct is based on a type II promoter driving a reporter gene (Fig. 3b). Inserted in the ORF of the reporter, two switchable gRNAs control the expression of the reporter. In the initial configuration, the first switchable gRNA (the ON cassette) interrupts the ORF in such a way that only after Cas9 cuts the target region in this gRNA and SSA occurs, the switchable gRNA is reconstituted bringing the reporter in frame. The second

switchable gRNA (the OFF cassette) controls the inactivation of the reporter. An event of DNA cutting and SSA repair activates the switchable gRNA in the OFF cassette and inactivates the reporter by bringing it out of frame.

In our first attempts with this design, we realized the reporter was activated despite being out of frame. We therefore searched for splicing sites in the construct and found that the U6 promoters controlling the expression of the gRNAs had strong, predicted splicing sites that could interfere with the design, making the reporter active despite being in a different ORF. We therefore generated a construct in which all the splicing sites were removed. However, the reporter in this construct exhibited no expression, even in the control construct in which the ON cassette was pre-reconstituted. Although the reasons for the reporter expression requiring splicing are unknown, we hypothesize splicing might remove sequences that prevent translation. Finally, we generated a new construct in which splicing occurred in a controlled way (Fig. 3b), resulting in reporter activation or inactivation when the ON or the OFF cassette was reconstituted, respectively.

In summary, a CLADES construct consists of a reporter controlled by two regulatory cassettes, the ON and the OFF cassette. The reconstitution of the gRNA in each cassette results in activation or inactivation of the reporter. At the beginning of the reporter, a Kozak sequence precedes the ORF, followed by a splicing donor. This donor site defines the beginning of the first intron, which ends at the splicing acceptor located just upstream of the ON cassette. The second intron is defined by another splicing donor located after the ON cassette, followed by a splicing acceptor before the OFF sequence. No more splicing sites exist in the construct. Between the regulatory cassettes and the reporter there is a T2A signal that detaches the reporter from the small peptide generated from translating the regulatory cassettes.

CLADES has two versions. In CLADES 1.0, the Actin5C promoter controls the ubiquitous expression of the reporter (Fig. 3c). This version consists of two CLADES constructs, each with a different reporter, YFP or RFP. The sequence of gRNA activations starts with the ON cassette in the YFP construct, labeling the cells in green. The activation of the gRNA in the ON cassette of the YFP construct triggers the ON sequence in the RFP reporter, whose gRNA will trigger the OFF sequence in the YFP. Thus, the color sequence is as follows: (1) YFP, (2) YFP and RFP, and (3) RFP alone. To label the lineage from the beginning, the CLADES 2.0 version implements a pre-activated version of the first reporter (Fig. 3d). In this version the Actin5C promoter was substituted by UAS sequences that allow CLADES to be combined with GAL4. Also, a third construct encoding the CFP reporter was built to expand the number of colors with which a lineage can be traced. In CLADES 2.0, the sequence of reporters is as follows:

(1) YFP (pre-activated), (2) YFP and RFP, (3) RFP, (4) RFP and CFP, and (5) CFP. In both versions, the YFP and the CFP reporters are tagged with V5 and HA, respectively. These tags and the RFP reporter can be independently detected by immunostaining, thus enhancing the reporter signal. A CAAX sequence reveals cell morphology by directing the reporters to the plasma membrane.

3.3 Controlling Cascade Progression

In the first CLADES versions, Cas9 was driven by a neuroblast-specific dpn driver. This driver makes the cascade to progress in the neuroblast, and the progeny inherits the color corresponding to the period in which the cell was generated. But Cas9 can be driven by any other genetic driver. This opens the possibility to use CLADES as a system to report different biological events. We showed proof of principle by driving Cas9 expression with the Bam promoter, a transitory germ line promoter [27, 28]. Since Bam is only active for a short period of time between the germ stem cell and the mature germ cell, the progression of the cascade is limited. Whereas in females the cascade progresses a single step per generation, multiple steps occur in males for unknown reasons.

The continuous induction of Cas9 in cycling neuroblasts using the dpn promoter provides no control over the cascade progression. The asynchronous stochastic progression can exhaust the cascade prematurely, and not knowing the timing of actual cascade progression makes birth dating impossible. To address these limitations, we developed a Cas9 version controlled by a heat shock promoter. In the first design, due to the basal activity of the heat shock promoter, we observed cascade progression even without heat shock. Also, the fact that the heat shock promoter is expressed ubiquitously complicates lineage tracing. A ubiquitous heat shock promoter driving Cas9 would result in a majority of cells expressing the reporters since any proliferating cells could undergo cascade transitions. To overcome this issue, we fused Cas9 to a small domain of the Geminin that promotes Cas9 degradation when the cell is not in G2 through M phase [29]. In other words, only proliferating cells (for a neuronal lineage only the neuroblast and the GMC) have Cas9 activity.

A more informative approach especially in the complex brain is to combine CLADES with GAL4 drivers that are expressed in the subset of neurons of interest. This temporally inducible version of CLADES is extremely useful for birth dating as we can link the timing in which we apply each heat shock with the color transition and cell identity revealed by the GAL4 pattern.

3.4 Notes

- The design of CLADES constructs for new animal models should begin with the generation of three important controls: (1) a construct containing the ON and OFF cassettes exhibiting no reporter expression, (2) a construct with the ON cassette pre-reconstituted showing strong reporter expression, and (3) a

construct with the OFF cassette pre-reconstituted showing no expression at all. In case a strategy of controlled splicing is needed, these control constructs will guarantee splicing occurs as predicted.

- In CLADES constructs, a T2A peptide detaches the reporter from the small peptide translated from the regulatory cassettes. Some peptide signals need to be in the N-terminus of the ORF. This is the case of the myristoylation signal that targets a protein to the plasma membrane. In our designs we had to use CAAX, which is functional in the C-terminus, to label the plasma membrane.
- Despite SSA is very efficient in proliferative cells, such as neuroblasts, some cutting events will be repaired by NHEJ, resulting in the collapse of the cascade. This collapse occurs with higher probability for the activation of the reporters since only one out of three possible ORFs results in reporter expression. Reporter inactivation is more likely to occur even when the DSB is repaired by NHEJ, since two out of three possible ORFs will bring the reporter out of frame (although in that case the cascade would most likely end in that step since the gRNA would not be reconstituted properly).
- Despite we use dpn to restrict the expression of Cas9 to neuroblasts, residual Cas9 can remain in the GMC. This leads to minimal cascade progression in the GMC. In some cases, the cascade can undergo that minimal progression in the GMC but not in the neuroblast. Thus, the progeny derived from the early GMC might be labeled by a subsequent color compared to the later-derived progeny. We refer to this phenomenon as GMC events.
- In extremely rare cases, the cascade might not follow the order of reporters predicted or even three reporters might coincide in the same cell. These exceptions are probably the result of very unlikely recombination events.

4 Analyzing *Drosophila* Cell Lineages with CaSSA and CLADES

Initial efforts to analyze *Drosophila* cell lineages mostly relied on tools that could label cell clones without resolving birth order. The invention of Twin-Spot MARCM (TS-MARCM) was a major leap forward for tracing birth order in cell lineages [30]. Mostly applied to neuronal lineages, TS-MARCM exploits the highly developmental stereotypy in flies to resolve birth order by integrating data from multiple experiments and individuals. Outperforming any other tool in terms of resolution, this approach exhibits four main limitations: (1) it assumes full developmental stereotypy, overlooking any

possible variation among individuals; (2) it requires thousands of samples to assemble a single neuronal lineage in a complex brain; (3) it can only label cells in two distinct colors, limiting phenotypic analysis in perturbation studies; and (4) it does not solve the problem of gaining genetic access to specific lineages. The technology presented in this chapter can help overcome these limitations.

How can CaSSA and CLADES help your cell lineage studies? How to design your experiment based on these tools? The answers to these questions depend on the biological system and problem you are dealing with. These tools can be useful for many aspects of your research, including finding new genetic drivers for your lineage(s) of interest, targeting a specific lineage, or even triggering a cascade of effectors (e.g., transcription factors) throughout the progression of a specific cell lineage. Whereas CaSSA is susceptible to be used with any cell type, CLADES performs better in those cells directly arising from stem cells (asymmetrical lineages). For symmetrical lineages, in which each progenitor duplicates in every mitosis to produce two exact progenitor cells, Cas9 remains active in these cells, thus driving cascade progression. Over multiple mitosis, this progression results in all cells expressing the last color in the cascade.

To illustrate the power of these techniques, we here focus on a typical experiment of resolving the birth order of a group of neurons of interest. Most neurons in the fly brain derive from asymmetrical, type I, or type II lineages. As discussed above, having a driver to label the neurons of interest greatly facilitates the experiment. If such a driver labels a subset of neurons, one first need to determine if that subset arises from a single or multiple lineages. Since neurons do not migrate in the fly brain, the location of the soma can help discern the number of involved lineages. In case of doubt, one can carry out clonal analysis by mosaic labeling using CaSSA or any other tool based on recombinases. Under sparse induction, having two cells labeled with the same reporter in multiple experiments is a good indicator that these cells arise from the same progenitor.

Once it is clear that neurons belong to the same lineage, the experiment for assessing birth order is as simple as crossing a GAL4 line that labels the population of interest to the UAS-Clades 2.0 line (Fig. 4a). After applying established protocols of immunostaining and imaging [31], the color for each neuron can be revealed and annotated across multiple samples. Assigning a birth order score to each color of the cascade allows ranking the neurons based on their color scores. To limit the effect of errors related to GMC events, a minimum of 50 samples should be analyzed. Also, rare events, such as those not following the correct cascade order (for instance, three reporters expressed in the same cell), should be discarded.

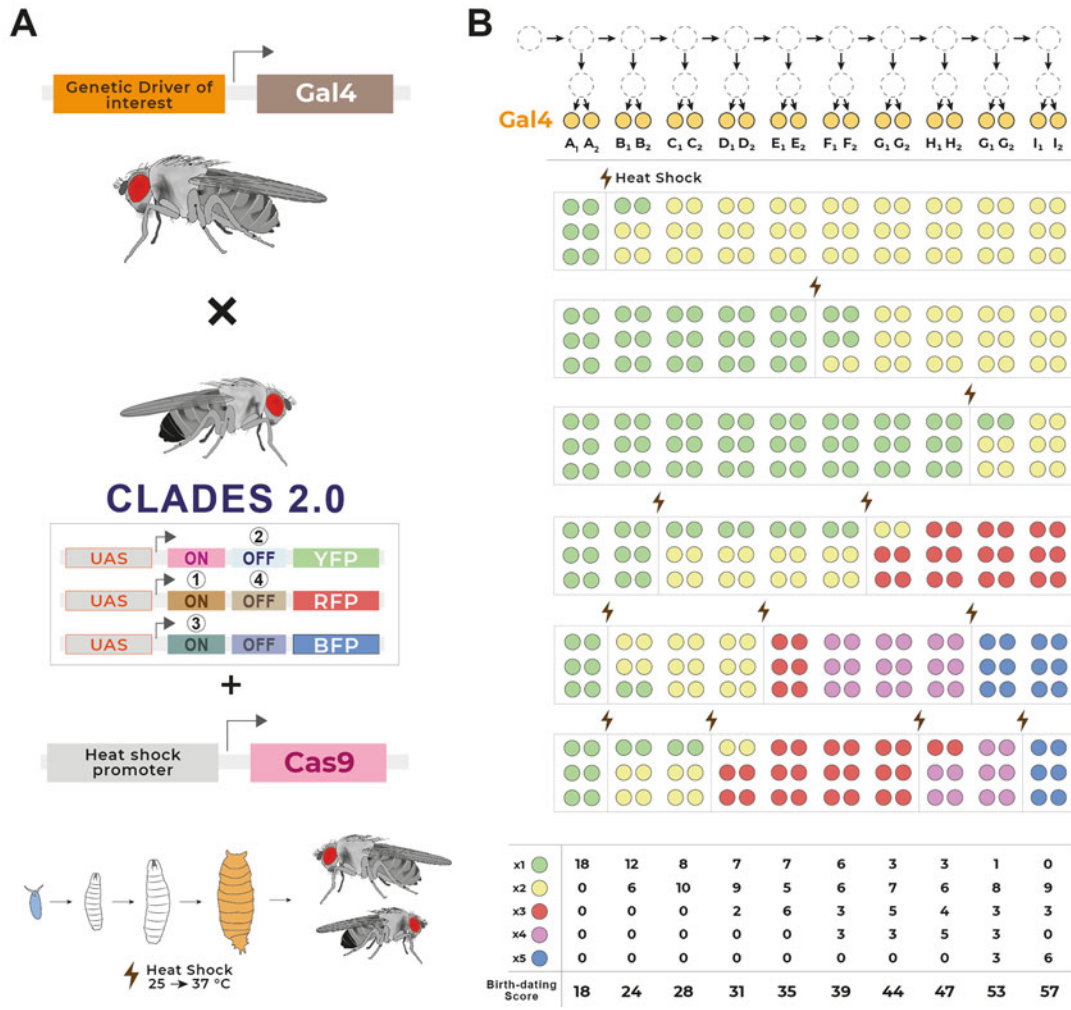


Fig. 4 A simple experiment of lineage tracing based on CLADES. (**a**) The most efficient strategy for lineage tracing based on CLADES relies on a single cross of (1) a line bearing the three CLADES 2.0 constructs and a Cas9 driven by a heat shock promoter, and (2) a GAL4 line with expression in the lineage of interest. The progeny originated from this cross should be heat shocked at different stages, one or multiple times in each experiment. (**b**) Then, the color for each of the cells comprising the lineage should be annotated in every sample. A birth-date score can be calculated by multiplying the order of the color in the cascade by the number of samples in which that cell was labeled with that color. The lineage order can then be assembled by ordering the progeny based on this score. It is noted that heat shock does not guarantee the progression of the cascade in all samples.

To determine the temporal window in which each neuron is born, the experiment ought to be repeated using the hs-Cas9 version (Fig. 4b). For this, synchronized animals need to be obtained and heat shocked at different developmental stages, following established protocols [31]. Depending on the required resolution and the specific features of the lineage of interest, the interval of synchronization should be adjusted. We typically use 2-hour windows for larvae harvesting.

Tracing a full lineage requires GAL4 to be expressed in all progeny, which is a rare phenomenon. One solution is to utilize multiple lines that label overlapping parts of the lineage. Another option would be inducing mosaic labeling of neuronal clones with lineage-specific genetic drivers controlling a CaSSA switchable Cas9 version and then combining it with CLADES. Sparse labeling is critical to distinguish neurons based on their morphology and/or the expression of molecular markers. Labeling an entire lineage complicates cell identification owing to the limited number of available colors.

To extend this approach to genetic and functional studies, one can compare the derived color scores and birth order in the wild type to elucidate whether gene perturbations modify the birth order/time, morphology, or any other neuronal feature. New versions of CLADES in which the reporter genes were substituted by effectors of binary systems (i.e., GAL4, LexAp65, etc.) would facilitate these perturbations without needing to generate new CLADES constructs. This technology would be particularly powerful since, well designed, one could study several genetic perturbations in a single lineage with few samples.

References

1. Urbach R, Schnabel R, Technau GM (2003) The pattern of neuroblast formation, mitotic domains and proneural gene expression during early brain development in Drosophila. Development 130:3589–3606
2. Yu H-H, Awasaki T, Schroeder MD, Long F, Yang JS, He Y, Ding P, Kao J-C, Wu GY-Y, Peng H et al (2013) Clonal development and organization of the adult Drosophila central brain. Curr Biol 23:633–643
3. Lee T, Lee A, Luo L (1999) Development of the Drosophila mushroom bodies: sequential generation of three distinct types of neurons from a neuroblast. Development 126: 4065–4076
4. Lai S-L, Awasaki T, Ito K, Lee T (2008) Clonal analysis of Drosophila antennal lobe neurons: diverse neuronal architectures in the lateral neuroblast lineage. Development 135: 2883–2893
5. Jefferis GSXE, Marin EC, Stocker RF, Luo L (2001) Target neuron prespecification in the olfactory map of drosophila. Nature 414: 204–208
6. Bhargava R, Onyango DO, Stark JM (2016) Regulation of single-strand annealing and its role in genome maintenance. Trends Genet 32:566–575
7. Garcia-Marques J, Yang CP, Espinosa-Medina-I, Mok K, Koyama M, Lee T (2019) Unlimited genetic switches for cell-type-specific manipulation. Neuron 104:227–238.e7
8. Garcia-Marques J, Espinosa-Medina I, Ku K, Yang C, Koyama M, Yu H, Lee T (2020) A programmable sequence of reporters for lineage analysis. Nat Neurosci 23:1618–1628
9. Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science (80) 337:816–821
10. Gasiunas G, Barrangou R, Horvath P, Siksnys V (2012) Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. Proc Natl Acad Sci 109:E2579–E2586
11. Awasaki T, Kao C-F, Lee Y-J, Yang C-P, Huang Y, Pfeiffer BD, Luan H, Jing X, Huang Y-F, He Y et al (2014) Making Drosophila lineage–restricted drivers via patterned recombination in neuroblasts. Nat Neurosci 17:631–637
12. Gao Y, Zhao Y (2014) Self-processing of ribozyme-flanked RNAs into guide RNAs in vitro and in vivo for CRISPR-mediated genome editing. J Integr Plant Biol 56: 343–349

13. Kvon EZ, Kazmar T, Stampfel G, Yáñez-Cuna JO, Pagani M, Schernhuber K, Dickson BJ, Stark A (2014) Genome-scale functional characterization of Drosophila developmental enhancers in vivo. Nature 512:91–95
14. Jenett A, Rubin GM, Ngo T-TB, Shepherd D, Murphy C, Dionne H, Pfeiffer BD, Cavallaro A, Hall D, Jeter J et al (2012) A GAL4-driver line resource for Drosophila neurobiology. Cell Rep 2:991–1001
15. Gohl DM, Silies MA, Gao XJ, Bhalerao S, Luongo FJ, Lin C-C, Potter CJ, Clandinin TR (2011) A versatile in vivo system for directed dissection of gene expression patterns. Nat Methods 8:231–237
16. Dionne H, Hibbard KL, Cavallaro A, Kao J-C, Rubin GM (2018) Genetic reagents for making Split-GAL4 lines in Drosophila. Genetics 209: 31–35
17. Tirian, L., and Dickson, B.J. (2017). The VT GAL4, LexA, and split-GAL4 driver line collections for targeted expression in the Drosophila nervous system. BioRxiv
18. Brand AH, Perrimon N (1993) Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. Development 118:401–415
19. Luan H, Peabody NC, Vinson CR, White BH (2006) Refined spatial manipulation of neuronal function by combinatorial restriction of transgene expression. Neuron 52:425–436
20. Venken KJT, Schulze KL, Haelterman NA, Pan H, He Y, Evans-Holm M, Carlson JW, Levis RW, Spradling AC, Hoskins RA et al (2011) MiMIC: a highly versatile transposon insertion resource for engineering Drosophila melanogaster genes. Nat Methods 8:737–743
21. Vilenchik MM, Knudson AG (2003) Endogenous DNA double-strand breaks: production, fidelity of repair, and induction of cancer. Proc Natl Acad Sci 100:12871–12876
22. Reese MG, Eeckman FH, Kulp D, Haussler D (1997) Improved splice site detection in genie. J Comput Biol 4:311–323
23. Sugawara N, Ira G, Haber JE (2000) DNA length dependence of the single-strand annealing pathway and the role of Saccharomyces cerevisiae RAD59 in double-Strand break repair. Mol Cell Biol 20:5300–5309
24. Vaze MB, Pellicioli A, Lee SE, Ira G, Liberi G, Arbel-Eden A, Foiani M, Haber JE (2002) Recovery from checkpoint-mediated arrest after repair of a double-Strand break requires Srs2 helicase. Mol Cell 10:373–385
25. Al-Minawi AZ, Saleh-Gohari N, Helleday T (2007) The ERCC1/XPF endonuclease is required for efficient single-strand annealing and gene conversion in mammalian cells. Nucleic Acids Res 36:1–9
26. Briner AE, Donohoue PD, Gomaa AA, Selle K, Slorach EM, Nye CH, Haurwitz RE, Beisel CL, May AP, Barrangou R (2014) Guide RNA functional modules direct Cas9 activity and orthogonality. Mol Cell 56:333–339
27. Chen D, McKearin DM (2003) A discrete transcriptional silencer in the bam gene determines asymmetric division of the drosophila germline stem cell. Development 130:1159–1170
28. White-Cooper H (2012) Tissue, cell type and stage-specific ectopic gene expression and RNAi induction in the drosophila testis. Spermatogenesis 2:11–22
29. Gutschner T, Haemmerle M, Genovese G, Draetta GF, Chin L (2016) Post-translational regulation of Cas9 during G1 enhances homology-directed repair. Cell Rep 14: 1555–1566
30. Yu H-H, Chen C-H, Shi L, Huang Y, Lee T (2009) Twin-spot MARCM to reveal the developmental origin and identity of neurons. Nat Neurosci 12:947–953
31. Lin S, Kao C-F, Yu H-H, Huang Y, Lee T (2012) Lineage analysis of drosophila lateral antennal lobe neurons reveals notch-dependent binary temporal fate decisions. PLoS Biol 10:e1001425

Chapter 10

Studying Protein Function Using Nanobodies and Other Protein Binders in *Drosophila*

Katarzyna Lepeta, Milena Bauer, Gustavo Aguilar, M. Alessandra Vigano, Shinya Matsuda, and Markus Affolter

Abstract

The direct manipulation of proteins by nanobodies and other protein binders has become an additional and valuable approach to investigate development and homeostasis in *Drosophila*. In contrast to other techniques, that indirectly interfere with proteins via their nucleic acids (CRISPR, RNAi, etc.), protein binders permit direct and acute protein manipulation. Since the first use of a nanobody in *Drosophila* a decade ago, many different applications exploiting protein binders have been introduced. Most of these applications use nanobodies against GFP to regulate GFP fusion proteins. In order to exert specific protein manipulations, protein binders are linked to domains that confer them precise biochemical functions. Here, we reflect on the use of tools based on protein binders in *Drosophila*. We describe their key features and provide an overview of the available reagents. Finally, we briefly explore the future avenues that protein binders might open up and thus further contribute to better understand development and homeostasis of multicellular organisms.

Key words Protein Binders, Nanobodies, *Drosophila*, GFP, DARPins, Development

1 Introduction

The first use of a nanobody in *Drosophila melanogaster* was reported in 2011 [1]. In this study, the authors reported deGradFP, a novel method to specifically induce degradation of a protein of interest (POI). The rationale behind this tool was to capitalize on the well-characterized ubiquitin-based degradation pathway to induce degradation of GFP fusion proteins. In order to achieve this, a nanobody recognizing GFP [2] was fused to a subunit of an E3 ubiquitin ligase complex. As it turned out, the co-expression of this synthetic "tool" and a GFP fusion protein led, in many cases, to the degradation of the latter.

Katarzyna Lepeta, Milena Bauer and Gustavo Aguilar contributed equally with all other contributors.

Christian Dahmann (ed.), *Drosophila: Methods and Protocols*,
Methods in Molecular Biology, vol. 2540, https://doi.org/10.1007/978-1-0716-2541-5_10,

This first report of the use of a nanobody to specifically and directly manipulate a protein in *Drosophila* was quickly followed by many different applications in the fly as well as in other multicellular organisms. Protein binders can be used directly to disrupt POI function or "functionalized" by fusing them to different enzymatically active domains or to scaffolds with discrete subcellular localization. This way, proteins can be degraded (deGradFP), inactivated (PAPTi), visualized (chromobodies, LlamaTag), and relocated to, or trapped in, different places (Morphotrap, GrabFP, JabbaTrap). Using the versatile and widely used GAL4 system [3] to drive the expression of functionalized protein binders, manipulation of protein function can be achieved in a spatially and temporally regulated manner. Large numbers of GAL4 lines exist, allowing to functionally manipulate proteins in discrete cell types [4]. The continuous development of GAL4-based expression strategies further adds to a more sophisticated use of protein-regulating tools.

Given the wide array of tools to manipulate gene products, the question is unavoidable: why does it help to interfere with protein function directly, rather than modifying the sequence or the expression of the corresponding gene using gene-editing approaches or RNAi? Direct manipulation of proteins allows for a number of applications that are difficult to achieve otherwise. Many (or most) proteins are contributed maternally, and in order to study their specific roles at the onset of development, their removal needs to be triggered at the protein level in late stages of oogenesis or very early in embryogenesis. In later developmental stages or in adult flies, when cell division slows down or stops, many proteins do not turn over rapidly [5], and their disappearance from specific cells upon the removal of the corresponding gene or mRNA is too slow to permit the analysis of their loss-of-function phenotype. The potential use of protein binders is not limited to the generation of loss-of-function phenotypes. Bringing a POI to a distinct cellular location might help to test existing models or to generate new models concerning the role of the POI in a given process. Similarly, specific post-translational modification of a POI in a given cell at a given time might also help to better understand the role of signaling pathways in cellular processes. In sum, the availability of protein binders with desired functionalizations, combined with the strength of the available tools for their controlled expression, allows for numerous novel applications in the *Drosophila* system.

2 Key Features of Protein Binder-Based Approaches

In this method paper, we will concentrate on the use of validated binders recognizing widely used tags. In general, three components are necessary for in vivo protein manipulation of tagged proteins: (1) A POI fused to a tag that can be recognized by an available protein binder; (2) A functionalized version of the binder allowing

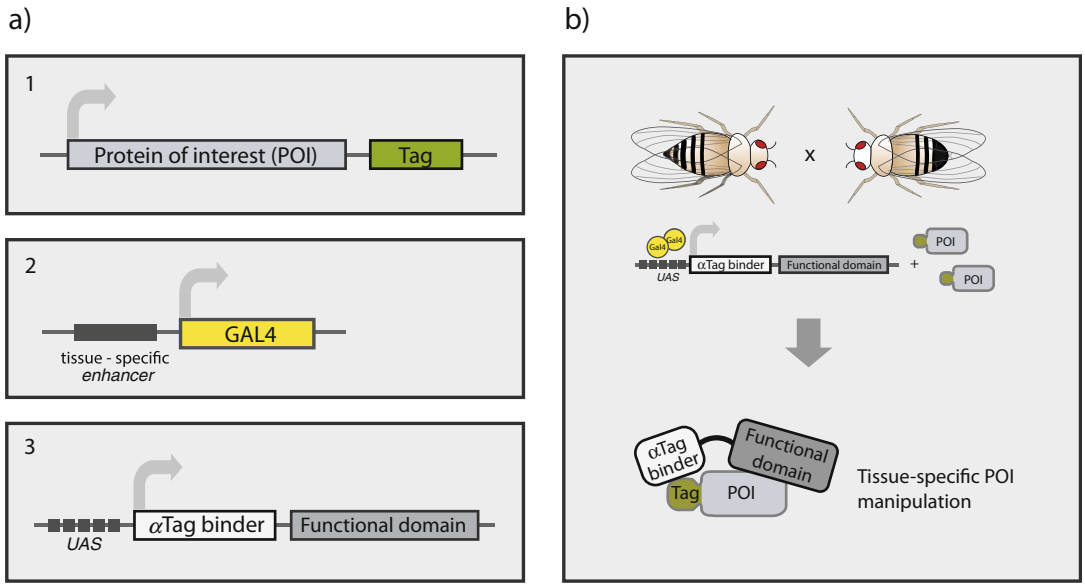


Fig. 1 Basic principles underlying direct protein manipulations in *Drosophila*. (**a**) The three main components required: (1) A POI fused to a tag that can be recognized by an available protein binder; (2) A GAL4 driver line active in a given tissue and/or at a given time; and (3) A protein binder recognizing the tag, functionalized to manipulate the POI in a predictable manner, generally expressed under UAS control, hence under the control of Gal4. (**b**) Upon crossing of flies containing all three components, the binder will tightly recognize the tag, bringing in close proximity the functional domain and the POI, resulting in the desired manipulation. It is noted that transgene distribution is arbitrary. If loss of function is sought, both genomic copies of the POI gene must be tagged

for a particular manipulation of the fusion protein, generally under UAS control; and (3) A driver GAL4 line allowing to express the functionalized binder in a given tissue and/or at a given time (Fig. 1).

2.1 Availability of a Protein Binder

Antibodies have been used as invaluable tools in basic and biomedical research, in diagnostics, as well as in disease therapy. However, their large size and complex structural build-up, with various intra- and intermolecular disulfide bridges, hinders their in vivo applications as genetically encoded synthetic tools. The discovery and engineering of much smaller, monomeric protein-based protein binders (in short, protein binders) recognizing target proteins with high affinity and specificity, such as scFvs, nanobodies, and DARPins [6–9], led to their use as genetically encoded protein binders in vivo. The logic behind the use of such binders is the following: if one were able to target or "grab" a protein in its natural environment, be it inside or outside the cell, in a specific cellular compartment, or in a specific tissue, and affect its function in a desired manner, for example, by chemically modifying, degrading, relocating, or visualizing it, it would be possible to learn more about this protein's function in vivo. Furthermore, existing models

arguing for specific functions of a given protein could be tested, or screens for discovery approaches could be designed.

There are two distinct approaches that can be chosen to directly manipulate proteins. Either one uses a validated binder recognizing a protein or peptide (such as a fluorescent protein, or a small peptide tag) which has been fused in frame with the POI, or, one alternatively, uses a protein binder recognizing specific structural features on the wild-type POI surface (Fig. 2). While bypassing the need of POI tagging, the later approach requires the generation and validation of new binders for each target POI. However, directly recognizing the POI permits the visualization and/or manipulation of specific POI's conformational states as well as POIs bearing distinct post-translational modifications. Protein binders directly recognizing the POI are prone to disrupt protein function when binding to critical domains or surfaces involved in protein–protein interactions. In *Drosophila*, few examples of such inhibitory use of protein binders have been reported. Along these lines, the Perrimon lab developed intracellular binders disrupting the Wnt pathway [10]. Similarly, binders directly disrupting the function of secreted proteins have also been described [11, 12]. The widespread availability of platforms and companies allowing for the screening of binder libraries using POIs makes the isolation of such binders increasingly straightforward.

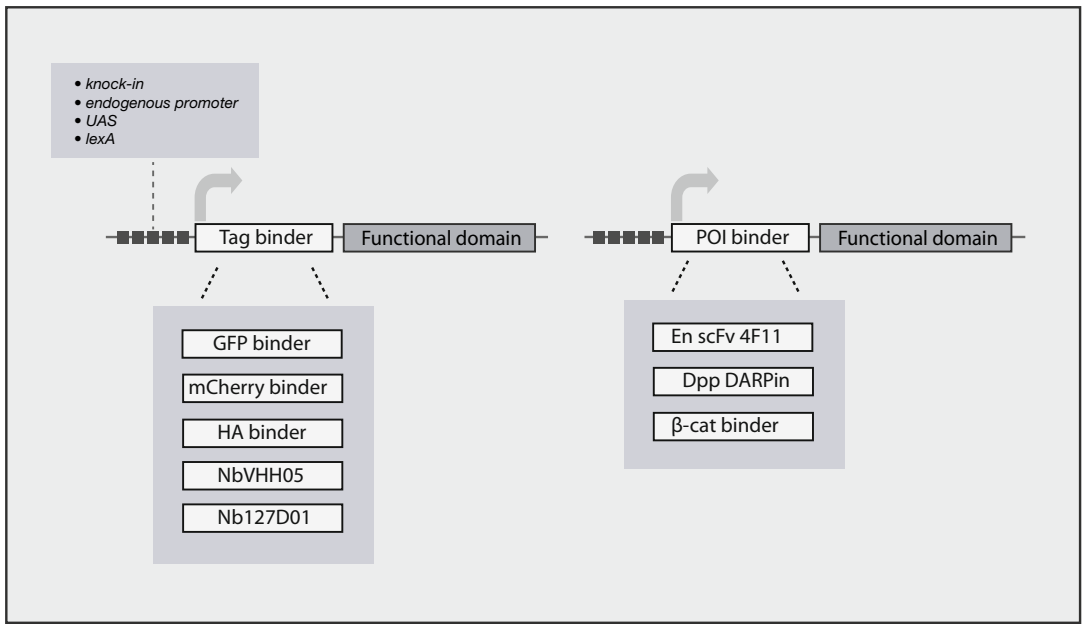


Fig. 2 Schematic representation of the target POI. Target POIs can be either under endogenous regulation or inducible sequences (UAS, LexA). For most applications, endogenous regulation is most recommendable, as changes in expression levels may cause functional defects. Target POIs can be fused in frame with a tag for which high-affinity protein binders are available. It is critical that the POI is functional after tagging. Tags used so far in *Drosophila* are listed. GFP variants Venus, EGFP, sfGFP, and EYFP are also recognized by several GFP antibodies (For more information *see* Ref. [36])

It does not come as a surprise that binders against fluorescent proteins, in particular, those recognizing GFP and related proteins, are most widely used in *Drosophila*. Fluorescent proteins have been used extensively as fusion partners to POIs in order to allow their visualization in vivo with fluorescence microscopy. Many large-scale initiatives have been launched to systematically tag endogenous proteins with GFP [13–16] or to generate GFP-tagged transgenes containing all regulatory elements [17, 18]. At present, more than thousand lines are available. These lines represent a huge resource for protein manipulations, since nanobodies against GFP allow to specifically recognize such GFP fusion proteins in vivo and eventually regulate their function. More recently, nanobodies and other binders against small peptide tags have been reported and tested in *Drosophila* cultured cells and in developing embryos and larvae [19, 20]

2.2 Availability or Generation of Targetable Proteins

Having a fully functional version of a tagged POI is essential if one wants to manipulate it in a meaningful way using protein binders. Since the in-frame fusion of a fluorescent protein to a POI allows its visualization in developing embryos using fluorescence microscopy, the most widely used tags are GFP, YFP, or other fluorescent proteins (Fig. 3).

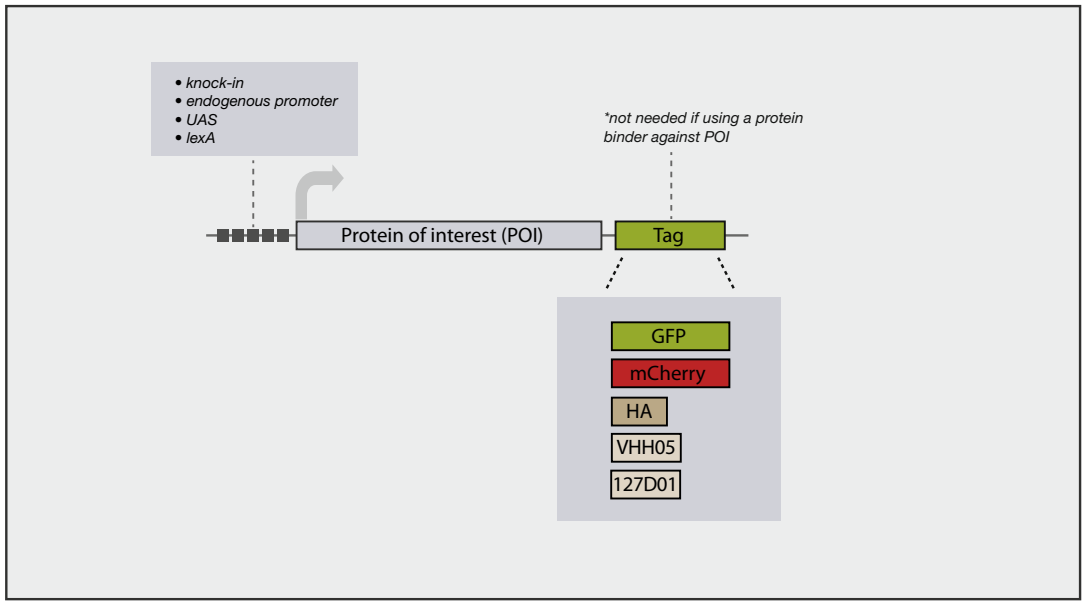


Fig. 3 Summary of protein binders available for protein manipulation in *Drosophila*. UAS or LexA are the most commonly used regulatory sequences to drive the expression of the functionalized protein binders. On the right side, scheme of a prototypical transgene, including UAS and/or LexA sequences followed by the coding region of the protein binder recognizing a protein tag and a functionalizing domain. As represented in the right side, some reagents are available to directly manipulate the POI without the need of a tag. For examples of functionalizing domains, *see* Fig. 4

The second most common fusion proteins are those bearing a short tag. Historically, these tags have been very popular for biochemical studies, chosen due to the availability of high-affinity antibodies that permit purification and visualization of fusion proteins via Western blot or immunohistochemistry. Recently, single-domain protein binders recognizing short peptides have been isolated and characterized, either by engineering some of these popular antibodies or by screening for new binders. Since such short tags are often less disruptive for the functionality of the tagged protein, we will likely see an increase in endogenous gene tagging with short peptides. Many novel methods have increased the efficiency with which genes can be endogenously tagged [21], allowing for a cheap and time-efficient generation of lines in which a POI is expressed in a version that can serve as a target for protein manipulation.

2.3 Driving the Expression of Functionalized Binders

2.3.1 GAL4 System

The use of the yeast GAL4 transcriptional system to direct gene expression in *Drosophila* was initially reported by Andrea Brand and Norbert Perrimon in 1993 [3]. In the meantime, hundreds of research projects have profited from this system for driving and manipulating gene expression in a multitude of cell- and tissue-specific patterns. More than 1000 GAL4 lines exist, and using these in conjunction with nanobody-based tools allows to manipulate proteins in distinct cells in living organisms [4]. To find suitable GAL4 lines, searching FlyBase might be the fastest way [22]. Additionally, the FliLight project provides an extensively annotated database of GAL4 lines: https://flweb.janelia.org/cgi-bin/flew.cgi [23]

2.3.2 Twists of the GAL4 System (GAL80, Split and Company)

The experimental design, or the cell type in which the nanobody construct should be expressed, may require tighter spatial or temporal control than any available GAL4 line can provide. However, the repertoire of tools that allow directed transcriptional control in *Drosophila* is vast. Currently, tools that permit the expression in the subset of cells in which two enhancers overlap are available (split-GAL4 [24, 25]). Equally important is the use of GAL80, a transcriptional repressor of the GAL4 system that permits refinement of the GAL4 patterns [26]. GAL80 has been further modified to allow strict temporal control of transcription, mediated by the expression of thermosensitive mutants ($GAL80^{ts}$ [27, 28]). Hence, simple temperature shifts are used to tightly control gene expression and, in the experimental setups discussed here, to regulate protein manipulation. Other transcriptional systems with analogous repressors and temporal control are available (QF/QUAS [29], LexA / LexO [30]) and can be used independently or in combination with GAL4. Indeed, some of the available protein binder tools in *Drosophila* have been generated with both LexO and UAS enhancer sequences [31, 32].

2.3.3 Specific Enhancers and mRNA Regulatory Elements

Another approach to limit the expression of a functionalized protein binder in time and space is to express it under the control of a specific promoter. Alternatively, distinct transcript localization can be achieved via specific sequences in the binder's UTRs. This has been done in several cases, in particular, with the aim to degrade proteins that are maternally provided (*βTub85D*-deGradFP [33]; mat-GAL4, matTub-GAL4-VP16 [34]; *hb*-deGradFP [35] or to degrade proteins uniquely in primordial germ cells (UASp-deGradFP-nos TCEpgc 3'UTR or nos-deGradFP [36]).

2.4 Different Functionalizations of Protein Binders

Although protein binders can disrupt protein function by direct interaction with the POI, the availability of a high-affinity binder for a POI or a tag fused to a POI by itself does not necessarily make this binder a great tool for basic research; it is the functionalization of the binder that endows this synthetic tool with a useful molecular or biochemical function. Fusion of a protein binder to a degron allows to degrade the target protein, while the fusion of the protein binder to a specific localization signal might allow to relocalize a POI to a novel spatial location. Below, we will refer to a number of different functionalizations that have been tested and published (Fig. 4).

2.4.1 Protein Degradation

deGradFP

As mentioned above, the degradation of a POI at a given time and place can be extremely helpful to elucidate its functional requirement during development and homeostasis. The fusion of an E3 ubiquitin ligase complex subunit (derived from the *Drosophila* protein Slmb) to a nanobody recognizing GFP (called VHH4) resulted in a functionalized GFP nanobody fusion which, when bound to a GFP fusion moiety of a POI, can induce the degradation of the POI [1]. The use of protein degradation induced via this fusion protein (called deGradFP) has been reported in many different contexts in *Drosophila* (for references, *see* [37]). Many *Drosophila* strains harboring UAS-deGradFP constructs on different chromosomes (as well as the corresponding control lines) are available from Bloomington. For experimental design and troubleshooting of experiments involving deGradFP, please refer to the detailed descriptions in [38, 39]. Similar degradation efficiencies have been obtained by substituting VHH4 by an anti-GFP DARPin [40]. The "deGrad" toolkit has recently been extended by the addition of a deGradHA, where the Slmb degron was coupled to an engineered anti-HA scFv [19].

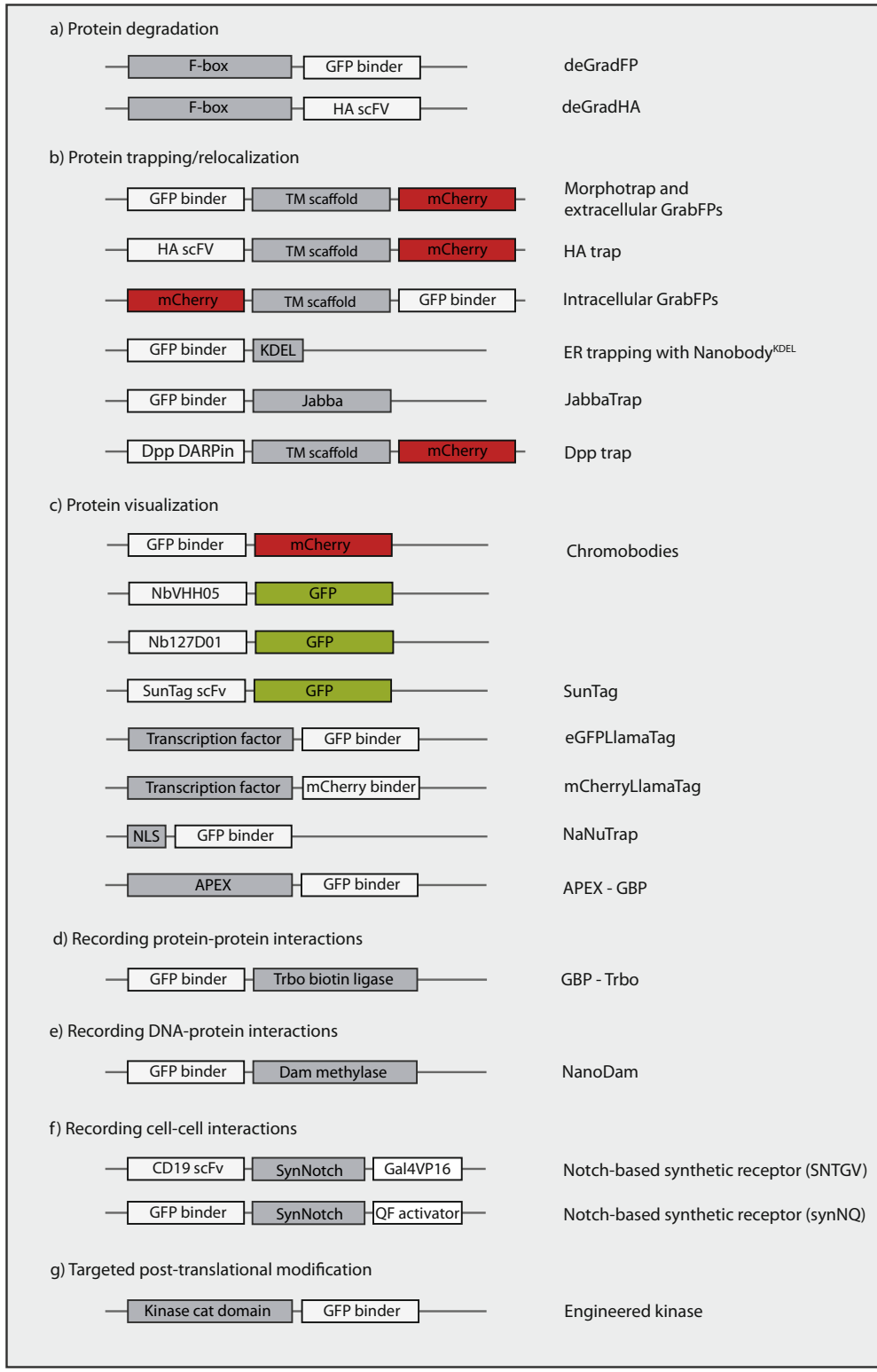


Fig. 4 Schematic representation of the different protein binder-based methods available in *Drosophila*. (**a**) Protein degradation: deGradFP and deGradHA. Upon binding, the GFP nanobody or HA-scFv fused to an F-box domain targets the GFP-tagged or HA-tagged POI for proteasomal degradation. (**b**) Protein trapping/

2.4.2 Protein Trapping/ Relocalization

Morphotrap and GrabFP

Protein relocalization and protein trapping have also been used in a number of different studies in *Drosophila*. Such studies include the trapping of secreted and dispersible ligands in order to better understand the requirement of dispersal for function. Furthermore, relocalization of proteins has been used to better understand their local requirement for particular functions. The basic design and initial testing of such strategies has been described in [31](Morphotrap) and in [32] (GrabFP). Many *Drosophila* strains harboring UAS-Morphotrap and UAS-GrabFP constructs on different chromosomes (as well as control lines) are available from Bloomington. For experimental design and troubleshooting of experiments involving Morphotrap and GrabFP, please refer to the detailed descriptions in Ref. [41]. As in the case of deGradFP, Morphotrap has also been recently altered, such as to be able to interfere with HA-tagged proteins, the new tool being baptized HA-trap [12].

ER Trapping with Nanobody ^{KDEL} Secreted proteins can not only be trapped in the extracellular space using Morphotrap or GrabFP tools (see above) but also be trapped in the ER using nanobodies fused to ER luminal localization signals (KDEL [42]). Such experiments will be helpful to study certain aspects of signaling pathways but might become even more useful in the future if trapped ligands could be released at a given time to study the kinetics of ligand secretion and dispersal. See Ref. [43].

Fig. 4 (continued) relocalization**:** protein binders fused to different scaffolds permit trapping/relocalization of POIs in different intra- and extracellular compartments. Morphotrap/HA-trap/Extracellular GrabFP: a GFP nanobody/HA-scFv fused to transmembrane scaffolds allows trapping of secreted POI on distinct regions of the cell surface. Intracellular GrabFP: a GFP nanobody fused to a transmembrane domain achieves the relocalization of intracellular POIs to novel locations at the membrane. Some constructs have been designed containing other fluorescent proteins (*see*: [32]). ER trapping with NanobodyKDEL: a GFP nanobody fused to ER luminal localization signal (KDEL) retains secreted POIs in the ER. JabbaTrap: a GFP nanobody fused to a lipid droplet-specific scaffold recruits soluble cytoplasmic proteins to the surface of lipid droplets. Dpp trap: a Dpp binder fused to CD8 with transmembrane domain permits simultaneous blocking of Dpp dispersal and signaling. (**c**) Protein visualization. Chromobodies: a nanobody recognizing a short tag fused to a fluorescent protein enables detection of POI carrying the short tag. LlamaTags: an anti-FP binder fused to a transcription factor permits immediate recruitment of a mature fluorescent protein (FP) upon synthesis. NaNuTrap: a GFP nanobody with nuclear localization sequence (NLS) sequesters cytoplasmic GFP to the nucleus. APEX-GFP nanobody fusion reveals the localization of the POI by electron microscopy. (**d**) Recording protein–protein interactions. GBP-Trbo: biotin ligase fused to anti-GFP nanobody permits interactome studies of GFP-tagged POIs in target tissues. (**e**) Recording protein–DNA interactions. NanoDam: a GFP nanobody brings a Dam methylase to the location of a tagged transcription factor. If the transcription factor is bound to DNA, the latter will be methylated in GATC sequences, permitting genome-wide binding profiling. (**f**) Recording cell–cell interactions: upon recognition of a certain membrane-tethered ligand in neighboring cells, the receptor is cleaved and activates transcription in the nucleus. (**g**) Engineered kinase: a GFP nanobody fused to a kinase catalytic domain induces phosphorylation of a GFP-tagged POI

JabbaTrap

In order to specifically delocalize and inactivate GFP-tagged nuclear proteins in the early *Drosophila* embryo, Seller and colleagues generated a fusion between an anti-GFP nanobody and the lipid droplet-binding protein Jabba and called the resulting fusion protein JabbaTrap [44]. Since lipid droplets are distributed throughout the cytoplasm in early *Drosophila* embryos, expression of JabbaTrap results in retention of GFP fusion proteins outside the nuclei, impairing transcriptional regulation. Delocalization to other cellular compartments might be a powerful approach to study protein function [44].

2.4.3 Protein Visualization

Chromobodies

In order to visualize the distribution of a POI and analyze its dynamics, genetically encoded protein binders recognizing the POI can be fused to a fluorescent protein [45–47]. Such fusions are called chromobodies [47] and permit POI visualization when expressed in cultured cells or during development. In *Drosophila*, transgenic lines expressing chromobodies have also been generated and used [20, 48].

LlamaTag

Certain biological processes happen in the minute range, meaning that they are faster than the maturation time of conventional fluorescent proteins; in this time range using fluorescent fusions as a readout is thus not advisable. To circumvent this problem, Bothma and colleagues [49] fused a transcription factor of interest not with GFP, but rather with the nanobody recognizing GFP. This allowed them to monitor the immediate relocalization of cytoplasmic GFP to the nucleus as the GFP-binding nanobody was being translated and moved into the nucleus [49–51] and to describe the dynamic behavior of the transcription factors during the early stages of embryogenesis.

NaNuTrap

A novel system for the visualization of nuclei during development sharing similarity in its experimental design to LlamaTag has recently been reported. In this case, a GFP nanobody was fused to a nuclear localization signal. Upon expression, this tool permitted the recruitment of cytoplasmic GFP to the nuclei, revealing their location. When GFP was provided maternally, nuclei could be accurately followed during early embryogenesis. The authors called this approach Nanobody Nuclear Trap, or NaNuTrap [52]. Interestingly, this approach might be useful in the future to delocalize cytoplasmic proteins to the nucleus.

SunTag/MoonTag In order to visualize single protein molecules, several studies in *Drosophila* made use of the SunTag system. SunTag is a two-component labeling system based on a GFP fusion of an scFv against SunTag, a 19aa peptide from the yeast GCN4 transcription factor, and a POI containing up to 24 copies of the GCN4 epitope. Due to the recruitment of over 20 GFP molecules to each POI, single proteins can be visualized in vivo [53]. The applications of SunTag in flies are manifold, including quantitative analyses of translational dynamics [54, 55]. Recently, a similar system called MoonTag was developed [56], and the combination of the two systems on the same transcript allows even more detailed molecular analyses.

GBP-APEX Most protein binder-based visualization methods have been designed to reveal protein localization under fluorescent light microscopes. Diffraction limit imposes a barrier for the resolution that can be achieved with these technologies. In spite the recent eruption of super-resolution methods, the electron microscope remains the go-to technique to reveal fine subcellular localization. Along these lines, Bernard and colleagues fused an anti-GFP nanobody to the engineered ascorbate peroxidase APEX [57]. Under specific conditions, APEX is able to generate electron-dense precipitates [58], visible under the electron microscope. When co-expressed with a GFP-tagged POI, the GBP-APEX fusion was able to detect POI localization in the *Drosophila* ovarioles.

2.4.4 Targeted Post-translational Modification

Nanobodies can also be used to target POIs for selective modification, such as, for example, phosphorylation. To achieve this, the GFP nanobody was fused to the constitutively active kinase domain of Rho kinase; upon binding to GFP-tagged Sqh, the proximity of the active kinase allowed to specifically phosphorylate and activate GFP-Sqh, a natural target of Rho kinase [59, 60]. Similar experiments were done with O-GlcNAc transferase in cultured cells [61], but to our best of knowledge, this later method has not been used in flies yet.

2.4.5 Recording Protein–Protein Interactions

GBP-Trbo

Analysis of protein interactions is classically done by co-immunoprecipitation of protein complexes with highly selective antibodies. Mass spectrometry has permitted the unbiased identification of interactors. While numerous crucial interactions have been found this way, it became obvious that many of the transient and/or low-affinity interactions were missed with such methods. To try to identify those, proximity-labeling techniques have been on the rise in the last ten years [62]. Proximity labeling relies on the promiscuous labeling of proteins by enzymes that catalyze the activation of Biotin. Normally, the enzyme is fused to a bait and expressed in the tissue of interest. Baker et al. have now generated *Drosophila* fly strains that permit targeted expression of a

promiscuous biotinylating enzyme (TurboID) to a GFP nanobody [63]. In theory, this tool permits tissue-specific interactome analyses of any GFP-tagged POI. Used in combination with endogenously tagged GFP proteins, this approach circumvents the possible localization perturbations caused by protein overexpression inherent to most of other tissue-specific approaches.

2.4.6 Recording Protein–DNA Interactions

NanoDam

Similar to the way GBP-Trbo can detect protein–protein interactions, the Brand lab has now proposed a nanobody-based strategy to uncover interactions between proteins and specific loci in the genome [64]. In this case, the GFP nanobody was fused to the *E. coli* Dam methylase. When the Dam methylase is brought in close proximity to DNA, it is able to methylate adenines within the GATC motif. While common in bacteria, adenine methylation is extremely rare in eukaryotes; hence, methylated adenines can be easily mapped in the genome [65]. In combination with the aforementioned endogenously GFP-tagged lines, this approach, named as NanoDam, provides a fast and robust alternative to chromatin immunoprecipitation for genome-wide binding profiling of POIs. Moreover, restricted NanoDam expression via the UAS/Gal4 system enables the dissection of protein–DNA interactions in specific cells.

2.4.7 Recording Cell–Cell Interactions

SynNotch

Protein binders are extensively used to enforce specific cell–cell interactions during cell-based treatments via engineered membrane receptors [66]. Usually, these systems introduce protein binders in the extracellular domain of known receptors, resulting in altered ligand specificity. The intracellular region can be further modified to program the outcome of ligand–receptor interaction. Receptor engineering has also broad applicability in model organisms. So far, in *Drosophila*, such receptors have been used mainly to record cell–cell interactions via synthetic Notch (synNotch) receptors. SynNotch exploits the core logic of the Notch receptor; upon binding of a membrane-tethered ligand displayed on the surface of a neighboring cell, the intracellular domain of the receptor is cleaved from the transmembrane core. This intracellular domain is in turn a transcriptional regulator, that, once soluble, modulates transcription in the nucleus. Two versions are available, containing either extracellular anti-CD19 scFv and intracellular GAL4 [67] or extracellular anti-GFP nanobody and intracellular QF (synNQ; [68]). In order to monitor cell contacts, a visible reporter was combined with these systems, so when ligand-expressing cells contact receptor-expressing cells, the reporter will be transcriptionally activated. These systems could potentially allow for any genetic manipulation in all cells in contact with a target cell population.

3 Future Directions

Using Unstable Versions of Binders

In certain experimental setups, it would be advantageous to only stabilize the functionalized nanobody when bound to the POI, such as to avoid or minimize secondary effects of the unbound fraction. Toward this aim, Tang et al. have isolated and characterized mutant versions of the anti-GFP nanobody which are degraded in the absence of the antigen (GFP in this case), thus diminishing the amount of unbound protein and avoid unwanted, non-specific effects [69]. The authors further demonstrated that the destabilizing mutations they identified can be introduced into the scaffold of other nanobodies, opening the door to the generalization of the approach. The application of such unstable protein binders might be particularly useful when dealing with chromobodies or directed protein modification, where the unbound fraction might impede visualization or cause off target effects, respectively (*see* [60] for an example of the successful use of unstable nanobodies). Interestingly, recent reports demonstrated that when expressed in the cytoplasm, nanobodies are intrinsically unstable in the absence of their antigen [70–72]. The authors termed this phenomenon antigen-mediated chromobody stabilization. Nanobody turnover could be further accelerated by modifying its N-terminal region, permitting the generation of chromobodies that faithfully monitor antigen levels.

Making Protein Manipulations Inducible and Reversible

Optogenetic methods enable light-dependent control of cellular processes with genetically encoded tools [73]. While optogenetics were first used in neuroscience to control neuronal activity in a light-dependent manner, this technology now includes a large number of tools that can be used to control a wide range of cellular functions. Optogenetic functionalities have been introduced into protein binder tools in many different ways; most of these novel methods, however, have not been applied in *Drosophila* as of today, and in most cases have been tested in cultured cells. Among the recent advances, tunable light- and drug-inducible depletion of target proteins was reported [74], light-activatable single-domain antibody fragments were developed [75, 76], split nanobody fragments were characterized that can be brought together and made functional by blue light-induced heterodimerization [77], and nanobodies were put under optogenetic control by using different light-switchable modules [78, 79]. In a similar fashion, protein binders have been fused to optogenetic modules to permit protein relocalization to different cellular compartments. It will be exciting to see these novel tools making their way into the fly system, profiting from the extensive toolbox allowing for restricted expression within developing organisms.

Use Protein Binders as Biosensors Another area where it is expected that protein binders will help unraveling protein networks is the field of biosensors. Using binders against different proteins, biosensors, such as Ca2+, or cAMP sensors, for example, can be brought to distinct cellular compartments [80, 81]. In addition, the activation state of distinct proteins (such as Rab GTPases [45], G-protein coupled receptors [82, 83], and others) can be monitored within their cellular environment in an unprecedented fashion. Again, such studies have not been done in *Drosophila* yet, but it is expected that this will occur in the near future.

Synthetic Approaches The expansion of the repertoire of available tools for protein manipulation in *Drosophila* will certainly help to better understand the immense complexity of the proteome of living organisms. In addition, the use of protein binders in synthetic biology and the incorporation of synthetic modules into existing circuits in vivo might help to disentangle the complexity of biological processes. This has recently been done in an elegant study by the laboratory of JP Vincent [84]. In this study, the potential of secreted GFP to function as a morphogen in the developing wing imaginal disc was investigated. GFP was used in place of Dpp, a well-studied morphogen involved in growth and patterning of the wing imaginal disc. The endogenous Dpp receptors were engineered, such as to be able to recognize GFP. In the presence of low-affinity receptors on the surface of wing imaginal disc cells, GFP could indeed replace Dpp to a large extent, suggesting a number of scenarios for GFP dispersal, capture, and signaling. Such studies allow for detailed analyses of quantitative parameters difficult to assess with classical approaches and allow to address long-standing issues that are important to better understand developmental processes.

References

1. Caussinus E, Kanca O, Affolter M (2011) Fluorescent fusion protein knockout mediated by anti-GFP nanobody. Nat Struct Mol Biol 19(1):117–121. https://doi.org/10.1038/nsmb.2180
2. Rothbauer U, Zolghadr K, Muyldermans S, Schepers A, Cardoso MC, Leonhardt H (2008) A versatile nanotrap for biochemical and functional studies with fluorescent fusion proteins. Mol Cell Proteom 7(2):282–289. https://doi.org/10.1074/mcp.M700342-MCP200
3. Brand AH, Perrimon N (1993) Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. Development 118(2):401–415
4. Caygill EE, Brand AH (2016) The GAL4 system: a versatile system for the manipulation and analysis of gene expression. Meth Mol Biol 1478:33–52. https://doi.org/10.1007/978-1-4939-6371-3_2
5. Toyama BH, Hetzer MW (2013) Protein homeostasis: live long, won't prosper. Nat Rev Mol Cell Biol 14(1):55–61. https://doi.org/10.1038/nrm3496
6. Pluckthun A (2015) Designed ankyrin repeat proteins (DARPins): binding proteins for research, diagnostics, and therapy. Annu Rev Pharmacol Toxicol 55:489–511. https://doi.org/10.1146/annurev-pharmtox-010611-134654

7. Helma J, Cardoso MC, Muyldermans S, Leonhardt H (2015) Nanobodies and recombinant binders in cell biology. J Cell Biol 209(5): 633–644. https://doi.org/10.1083/jcb.201409074
8. Beghein E, Gettemans J (2017) Nanobody technology: a versatile toolkit for microscopic imaging, protein-protein interaction analysis, and protein function exploration. Front Immunol 8:771. https://doi.org/10.3389/fimmu.2017.00771
9. Kaiser PD, Maier J, Traenkle B, Emele F, Rothbauer U (2014) Recent progress in generating intracellular functional antibody fragments to target and trace cellular components in living cells. Biochim Biophys Acta 1844(11): 1933–1942. https://doi.org/10.1016/j.bbapap.2014.04.019
10. Yeh JTH, Binari R, Gocha T, Dasgupta R, Perrimon N (2013) PAPTi: a peptide aptamer interference toolkit for perturbation of protein-protein interaction networks. Sci Rep 3(1):1156. https://doi.org/10.1038/srep01156
11. Layalle S, Volovitch M, Mugat B, Bonneaud N, Parmentier ML, Prochiantz A, Joliot A, Maschat F (2011) Engrailed homeoprotein acts as a signaling molecule in the developing fly. Development 138(11):2315–2323. https://doi.org/10.1242/dev.057059
12. Matsuda S, Schaefer JV, Mii Y, Hori Y, Bieli D, Taira M, Plückthun A, Affolter M (2021) Asymmetric requirement of Dpp/BMP morphogen dispersal in the Drosophila wing disc. bioRxiv 2020:2011.2023.394379. https://doi.org/10.1101/2020.11.23.394379
13. Morin X, Daneman R, Zavortink M, Chia W (2001) A protein trap strategy to detect GFP-tagged proteins expressed from their endogenous loci in Drosophila. Proc Nat Acad Sci 98(26):15050–15055. https://doi.org/10.1073/pnas.261408198
14. Buszczak M, Paterno S, Lighthouse D, Bachman J, Planck J, Owen S, Skora AD, Nystul TG, Ohlstein B, Allen A, Wilhelm JE, Murphy TD, Levis RW, Matunis E, Srivali N, Hoskins RA, Spradling AC (2007) The carnegie protein trap library: a versatile tool for Drosophila developmental studies. Genetics 175(3):1505–1531. https://doi.org/10.1534/genetics.106.065961
15. Nagarkar-Jaiswal S, Lee PT, Campbell ME, Chen K, Anguiano-Zarate S, Gutierrez MC, Busby T, Lin WW, He Y, Schulze KL, Booth BW, Evans-Holm M, Venken KJT, Levis RW, Spradling AC, Hoskins RA, Bellen HJ (2015) A library of MiMICs allows tagging of genes and reversible, spatial and temporal knockdown of proteins in Drosophila. eLife 4: e05338. https://doi.org/10.7554/eLife.05338
16. Nagarkar-Jaiswal S, DeLuca SZ, Lee P-T, Lin W-W, Pan H, Zuo Z, Lv J, Spradling AC, Bellen HJ (2015) A genetic toolkit for tagging intronic MiMIC containing genes. eLife 4: e08469. https://doi.org/10.7554/eLife.08469
17. Sarov M, Barz C, Jambor H, Hein MY, Schmied C, Suchold D, Stender B, Janosch S, Kj VV, Krishnan RT, Krishnamoorthy A, Ferreira IRS, Ejsmont RK, Finkl K, Hasse S, Kämpfer P, Plewka N, Vinis E, Schloissnig S, Knust E, Hartenstein V, Mann M, Ramaswami M, VijayRaghavan K, Tomancak P, Schnorrer F (2016) A genome-wide resource for the analysis of protein localisation in Drosophila. eLife 5:e12068. https://doi.org/10.7554/eLife.12068
18. Kudron MM, Victorsen A, Gevirtzman L, Hillier LW, Fisher WW, Vafeados D, Kirkey M, Hammonds AS, Gersch J, Ammouri H, Wall ML, Moran J, Steffen D, Szynkarek M, Seabrook-Sturgis S, Jameel N, Kadaba M, Patton J, Terrell R, Corson M, Durham TJ, Park S, Samanta S, Han M, Xu J, Yan K-K, Celniker SE, White KP, Ma L, Gerstein M, Reinke V, Waterston RH (2018) The ModERN Resource: genome-wide binding profiles for hundreds of Drosophila and Caenorhabditis elegans transcription factors. Genetics 208(3):937–949. https://doi.org/10.1534/genetics.117.300657
19. Vigano MA, Ell CM, Kustermann MMM, Aguilar G, Matsuda S, Zhao N, Stasevich TJ, Affolter M, Pyrowolakis G (2021) Protein manipulation using single copies of short peptide tags in cultured cells and in Drosophila melanogaster. Development 148(6). https://doi.org/10.1242/dev.191700
20. Xu J, Kim A-R, Cheloha RW, Fischer FA, Shun Li JS, Feng Y, Stoneburner E, Binari R, Mohr SE, Zirin J, Ploegh H, Perrimon N (2021) Protein visualization and manipulation in Drosophila through the use of epitope tags recognized by nanobodies. bioRxiv 2021: 2004.2016.440240. https://doi.org/10.1101/2021.04.16.440240
21. Kanca O, Bellen HJ, Schnorrer F (2017) Gene tagging strategies to assess protein expression, localization, and function in Drosophila. Genetics 207(2):389–412. https://doi.org/10.1534/genetics.117.199968

22. Marygold SJ, Crosby MA, Goodman JL, FlyBase C (2016) Using flybase, a database of drosophila genes and genomes. Meth Mol Biol 1478:1–31. https://doi.org/10.1007/978-1-4939-6371-3_1
23. Jenett A, Rubin Gerald M, Ngo T-TB, Shepherd D, Murphy C, Dionne H, Pfeiffer Barret D, Cavallaro A, Hall D, Jeter J, Iyer N, Fetter D, Hausenfluck Joanna H, Peng H, Trautman Eric T, Svirskas Robert R, Myers Eugene W, Iwinski Zbigniew R, Aso Y, DePasquale GM, Enos A, Hulamm P, Lam Shing Chun B, Li H-H, Laverty Todd R, Long F, Qu L, Murphy Sean D, Rokicki K, Safford T, Shaw K, Simpson Julie H, Sowell A, Tae S, Yu Y, Zugates Christopher T (2012) A GAL4-driver line resource for Drosophila neurobiology. Cell Rep 2(4):991–1001. https://doi.org/10.1016/j.celrep.2012.09.011
24. Luan H, Peabody NC, Vinson Charles R, White BH (2006) Refined spatial manipulation of neuronal function by combinatorial restriction of transgene expression. Neuron 52(3):425–436. https://doi.org/10.1016/j.neuron.2006.08.028
25. Pfeiffer BD, Ngo T-TB, Hibbard KL, Murphy C, Jenett A, Truman JW, Rubin GM (2010) Refinement of tools for targeted gene expression in Drosophila. Genetics 186(2):735–755. https://doi.org/10.1534/genetics.110.119917
26. Lee T, Luo L (1999) Mosaic analysis with a repressible cell marker for studies of gene function in neuronal morphogenesis. Neuron 22(3):451–461. https://doi.org/10.1016/S0896-6273(00)80701-1
27. McGuire SE, Le PT, Osborn AJ, Matsumoto K, Davis RL (2003) Spatiotemporal rescue of memory dysfunction in Drosophila. Science 302(5651):1765–1768. https://doi.org/10.1126/science.1089035
28. McGuire SE, Mao Z, Davis RL (2004) Spatiotemporal gene expression targeting with the TARGET and gene-switch systems in Drosophila. Science's STKE 2004(220):pl6. https://doi.org/10.1126/stke.2202004pl6
29. Potter CJ, Tasic B, Russler EV, Liang L, Luo L (2010) The Q system: a repressible binary system for transgene expression, lineage tracing, and mosaic analysis. Cell 141(3):536–548. https://doi.org/10.1016/j.cell.2010.02.025
30. Yagi R, Mayer F, Basler K (2010) Refined LexA transactivators and their use in combination with the Drosophila Gal4 system. Proc Nat Acad Sci U S A 107(37):16166–16171. https://doi.org/10.1073/pnas.1005957107
31. Harmansa S, Hamaratoglu F, Affolter M, Caussinus E (2015) Dpp spreading is required for medial but not for lateral wing disc growth. Nature 527(7578):317. https://doi.org/10.1038/nature15712
32. Harmansa S, Alborelli I, Bieli D, Caussinus E, Affolter M (2017) A nanobody-based toolset to investigate the role of protein localization and dispersal in Drosophila. Elife 6:e22549. https://doi.org/10.7554/eLife.22549
33. Sun MS, Weber J, Blattner AC, Chaurasia S, Lehner CF (2019) MNM and SNM maintain but do not establish achiasmate homolog conjunction during Drosophila male meiosis. PLoS Genet 15(5):e1008162. https://doi.org/10.1371/journal.pgen.1008162
34. Lin B, Luo J, Lehmann R (2020) Collectively stabilizing and orienting posterior migratory forces disperses cell clusters in vivo. Nat Commun 11(1):4477. https://doi.org/10.1038/s41467-020-18185-2
35. Vazquez-Pianzola P, Beuchle D, Saro G, Hernández G, Maldonado G, Brunβen D, Meister P, Suter B. Female meiosis II and pronuclear fusion require Bicaudal-D. bioRxiv. https://doi.org/10.1101/2020.12.22.423980
36. Gaskill MM, Gibson TJ, Larson ED, Harrison MM (2021) GAF is essential for zygotic genome activation and chromatin accessibility in the early Drosophila embryo. eLife 10:e66668. https://doi.org/10.7554/eLife.66668
37. Aguilar G, Vigano MA, Affolter M, Matsuda S (2019) Reflections on the use of protein binders to study protein function in developmental biology. Wiley Interdiscip Rev Dev Biol 8:e356. https://doi.org/10.1002/wdev.356
38. Caussinus E, Kanca O, Affolter M (2013) Protein knockouts in living eukaryotes using deGradFP and green fluorescent protein fusion targets. Curr Protoc Protein Sci 73:Unit 30 32. https://doi.org/10.1002/0471140864.ps3002s73
39. Caussinus E, Affolter M (2016) deGradFP: a system to knockdown GFP-tagged proteins. Meth Mol Biol 1478:177–187. https://doi.org/10.1007/978-1-4939-6371-3_9
40. Brauchle M, Hansen S, Caussinus E, Lenard A, Scholz AOEO, Sprecher SG, Plückthun A, Affolter M (2014) Protein interference applications in cellular and developmental biology using DARPins that recognize GFP and mCherry. Biol Open 3(12):1252–1261. https://doi.org/10.1242/bio.201410041

41. Matsuda S, Aguilar G, Vigano MA, Affolter M (2021) Nanobody-based GFP traps to study protein localization and function in developmental biology. Meth Mol Biol 2446:581
42. Yu JJS, Maugarny-Cales A, Pelletier S, Alexandre C, Bellaiche Y, Vincent JP, McGough IJ (2020) Frizzled-dependent planar cell polarity without secreted Wnt ligands. Dev Cell 54(5):583–592. e585. https://doi.org/10.1016/j.devcel.2020.08.004
43. Boncompain G, Divoux S, Gareil N, de Forges H, Lescure A, Latreche L, Mercanti V, Jollivet F, Raposo G, Perez F (2012) Synchronization of secretory protein traffic in populations of cells. Nat Meth 9(5):493–498. https://doi.org/10.1038/nmeth.1928
44. Seller CA, Cho CY, O'Farrell PH (2019) Rapid embryonic cell cycles defer the establishment of heterochromatin by Eggless/SetDB1 in Drosophila. Genes Dev 33(7-8):403–417. https://doi.org/10.1101/gad.321646.118
45. Nizak C, Martin-Lluesma S, Moutel S, Roux A, Kreis TE, Goud B, Perez F (2003) Recombinant antibodies against subcellular fractions used to track endogenous Golgi protein dynamics in vivo. Traffic 4(11):739–753. https://doi.org/10.1034/j.1600-0854.2003.00132.x
46. Nizak C, Monier S, del Nery E, Moutel S, Goud B, Perez F (2003) Recombinant antibodies to the small GTPase Rab6 as conformation sensors. Science 300(5621):984–987. https://doi.org/10.1126/science.1083911
47. Rothbauer U, Zolghadr K, Tillib S, Nowak D, Schermelleh L, Gahl A, Backmann N, Conrath K, Muyldermans S, Cardoso MC, Leonhardt H (2006) Targeting and tracing antigens in live cells with fluorescent nanobodies. Nat Meth 3(11):887–889. https://doi.org/10.1038/nmeth953
48. Kamiyama D, McGorty R, Kamiyama R, Kim MD, Chiba A, Huang B (2015) Specification of dendritogenesis site in Drosophila aCC motoneuron by membrane enrichment of Pak1 through Dscam1. Dev Cell 35(1):93–106. https://doi.org/10.1016/j.devcel.2015.09.007
49. Bothma JP, Norstad MR, Alamos S, Garcia HG (2018) LlamaTags: a versatile tool to image transcription factor dynamics in live embryos. Cell 173(7):1810–1822. e1816. https://doi.org/10.1016/j.cell.2018.03.069
50. Huang A, Rupprecht J-F, Saunders TE (2020) Embryonic geometry underlies phenotypic variation in decanalized conditions. eLife 9: e47380. https://doi.org/10.7554/eLife.47380
51. Soluri IV, Zumerling LM, Payan Parra OA, Clark EG, Blythe SA (2020) Zygotic pioneer factor activity of Odd-paired/Zic is necessary for late function of the Drosophila segmentation network. eLife 9:e53916. https://doi.org/10.7554/eLife.53916
52. Akos Z, Dunipace L, Stathopoulos A (2021) NaNuTrap: a technique for in vivo cell nucleus labelling using nanobodies. Development 148(18). https://doi.org/10.1242/dev.199822
53. Tanenbaum ME, Gilbert LA, Qi LS, Weissman JS, Vale RD (2014) A protein-tagging system for signal amplification in gene expression and fluorescence imaging. Cell 159(3):635–646. https://doi.org/10.1016/j.cell.2014.09.039
54. Dufourt J, Bellec M, Trullo A, Dejean M, De Rossi S, Favard C, Lagha M (2021) Imaging translation dynamics in live embryos reveals spatial heterogeneities. Science 372(6544):840–844. https://doi.org/10.1126/science.abc3483
55. Vinter DJ, Hoppe C, Minchington TG, Sutcliffe C, Ashe HL (2021) Dynamics of hunchback translation in real-time and at single-mRNA resolution in the Drosophila embryo. Development 148(18):dev196121. https://doi.org/10.1242/dev.196121
56. Boersma S, Khuperkar D, Verhagen BMP, Sonneveld S, Grimm JB, Lavis LD, Tanenbaum ME (2019) Multi-color single-molecule imaging uncovers extensive heterogeneity in mRNA decoding. Cell 178(2):458–472.e419. https://doi.org/10.1016/j.cell.2019.05.001
57. Bernard F, Jouette J, Durieu C, Le Borgne R, Guichet A, Claret S (2021) GFP-tagged protein detection by electron microscopy using a GBP-APEX tool in Drosophila. Front Cell Dev Biol 9(2092). https://doi.org/10.3389/fcell.2021.719582
58. Martell JD, Deerinck TJ, Sancak Y, Poulos TL, Mootha VK, Sosinsky GE, Ellisman MH, Ting AY (2012) Engineered ascorbate peroxidase as a genetically encoded reporter for electron microscopy. Nat Biotechnol 30(11):1143–1148. https://doi.org/10.1038/nbt.2375
59. Roubinet C, Tsankova A, Pham TT, Monnard A, Caussinus E, Affolter M, Cabernard C (2017) Spatio-temporally separated cortical flows and spindle geometry establish physical asymmetry in fly neural stem cells. Nat Commun 8(1):1383. https://doi.org/10.1038/s41467-017-01391-w

60. Lepeta K, Roubinet C, Kanca O, Ochoa-Espinosa A, Bieli D, Cabernard C, Affolter M, Caussinus E (2021) In vivo regulation of fluorescent fusion proteins by engineered kinases. bioRxiv 2021:2003.2026.433940. https://doi.org/10.1101/2021.03.26.433940
61. Ramirez DH, Aonbangkhen C, Wu HY, Naftaly JA, Tang S, O'Meara TR, Woo CM (2020) Engineering a proximity-directed O-GlcNAc transferase for selective protein O-GlcNAcylation in cells. ACS Chem Biol 15(4):1059–1066. https://doi.org/10.1021/acschembio.0c00074
62. Gingras A-C, Abe KT, Raught B (2019) Getting to know the neighborhood: using proximity-dependent biotinylation to characterize protein complexes and map organelles. Curr Opin Chem Biol 48:44–54. https://doi.org/10.1016/j.cbpa.2018.10.017
63. Baker FC, Neiswender H, Veeranan-Karmegam R, Gonsalvez GB (2021) In vivo proximity biotin ligation identifies the interactome of Egalitarian, a Dynein cargo adaptor. bioRxiv 2021:2007.2001.450794. https://doi.org/10.1101/2021.07.01.450794
64. Tang JLY, Hakes AE, Krautz R, Suzuki T, Contreras EG, Fox PM, Brand AH (2021) NanoDam identifies novel temporal transcription factors conserved between the Drosophila central brain and visual system. bioRxiv 2021:2006.2007.447332. https://doi.org/10.1101/2021.06.07.447332
65. Marshall OJ, Brand AH (2015) damidseq_pipeline: an automated pipeline for processing DamID sequencing datasets. Bioinformatics 31(20):3371–3373. https://doi.org/10.1093/bioinformatics/btv386
66. Rafiq S, Hackett CS, Brentjens RJ (2020) Engineering strategies to overcome the current roadblocks in CAR T cell therapy. Nat Rev Clin Oncol 17(3):147–167. https://doi.org/10.1038/s41571-019-0297-y
67. Huang T-H, Velho T, Lois C (2016) Monitoring cell-cell contacts in vivo in transgenic animals. Development 143(21):4073–4084. https://doi.org/10.1242/dev.142406
68. He L, Huang J, Perrimon N (2017) Development of an optimized synthetic Notch receptor as an in vivo cell-cell contact sensor. Proc Natl Acad Sci U S A 114(21):5467–5472. https://doi.org/10.1073/pnas.1703205114
69. Tang JC, Drokhlyansky E, Etemad B, Rudolph S, Guo B, Wang S, Ellis EG, Li JZ, Cepko CL (2016) Detection and manipulation of live antigen-expressing cells using conditionally stable nanobodies. Elife 5:e15312. https://doi.org/10.7554/eLife.15312
70. Keller BM, Maier J, Secker KA, Egetemaier SM, Parfyonova Y, Rothbauer U, Traenkle B (2018) Chromobodies to quantify changes of endogenous protein concentration in living cells. Mol Cell Proteomics 17(12):2518–2533. https://doi.org/10.1074/mcp.TIR118.000914
71. Keller BM, Maier J, Weldle M, Segan S, Traenkle B, Rothbauer U (2019) A strategy to optimize the generation of stable chromobody cell lines for visualization and quantification of endogenous proteins in living cells. Antibodies (Basel) 8(1). https://doi.org/10.3390/antib8010010
72. Traenkle B, Emele F, Anton R, Poetz O, Haeussler RS, Maier J, Kaiser PD, Scholz AM, Nueske S, Buchfellner A, Romer T, Rothbauer U (2015) Monitoring interactions and dynamics of endogenous beta-catenin with intracellular nanobodies in living cells*, [S]. Mol Cell Proteom 14(3):707–723. https://doi.org/10.1074/mcp.M114.044016
73. Manoilov KY, Verkhusha VV, Shcherbakova DM (2021) A guide to the optogenetic regulation of endogenous molecules. Nat Meth 18(9):1027–1037. https://doi.org/10.1038/s41592-021-01240-1
74. Deng W, Bates JA, Wei H, Bartoschek MD, Conradt B, Leonhardt H (2020) Tunable light and drug induced depletion of target proteins. Nat Commun 11(1):304. https://doi.org/10.1038/s41467-019-14160-8
75. Jedlitzke B, Yilmaz Z, Dorner W, Mootz HD (2020) Photobodies: light-activatable single-domain antibody fragments. Angew Chem Int Ed Engl 59(4):1506–1510. https://doi.org/10.1002/anie.201912286
76. Bridge T, Shaikh SA, Thomas P, Botta J, McCormick PJ, Sachdeva A (2019) Site-specific encoding of photoactivity in antibodies enables light-mediated antibody-antigen binding on live cells. Angew Chem Int Ed Engl 58(50):17986–17993. https://doi.org/10.1002/anie.201908655
77. Yu D, Lee H, Hong J, Jung H, Jo Y, Oh BH, Park BO, Heo WD (2019) Optogenetic activation of intracellular antibodies for direct modulation of endogenous proteins. Nat Meth 16(11):1095–1100. https://doi.org/10.1038/s41592-019-0592-7
78. Gil AA, Carrasco-Lopez C, Zhu L, Zhao EM, Ravindran PT, Wilson MZ, Goglia AG, Avalos JL, Toettcher JE (2020) Optogenetic control of protein binding using light-switchable nanobodies. Nat Commun 11(1):4044. https://doi.org/10.1038/s41467-020-17836-8
79. Joest EF, Winter C, Wesalo JS, Deiters A, Tampé R (2021) Light-guided intrabodies for on-demand in situ target recognition in human cells. Chem Sci 12(16):5787–5795. https://doi.org/10.1039/D1SC01331A

80. Prole DL, Taylor CW (2019) A genetically encoded toolkit of functionalized nanobodies against fluorescent proteins for visualizing and manipulating intracellular signalling. BMC Biol 17(1):41. https://doi.org/10.1186/s12915-019-0662-4
81. Hansen JN, Kaiser F, Klausen C, Stüven B, Chong R, Bönigk W, Mick DU, Möglich A, Jurisch-Yaksi N, Schmidt FI, Wachten D (2020) Nanobody-directed targeting of optogenetic tools to study signaling in the primary cilium. eLife 9:e57907. https://doi.org/10.7554/eLife.57907
82. Irannejad R, Tomshine JC, Tomshine JR, Chevalier M, Mahoney JP, Steyaert J, Rasmussen SGF, Sunahara RK, El-Samad H, Huang B, von Zastrow M (2013) Conformational biosensors reveal GPCR signalling from endosomes. Nature 495(7442):534–538. https://doi.org/10.1038/nature12000
83. Stoeber M, Jullié D, Lobingier BT, Laeremans T, Steyaert J, Schiller PW, Manglik A, von Zastrow M (2018) A genetically encoded biosensor reveals location bias of opioid drug action. Neuron 98(5):963–976. e965. https://doi.org/10.1016/j.neuron.2018.04.021
84. Stapornwongkul KS, de Gennes M, Cocconi L, Salbreux G, Vincent JP (2020) Patterning and growth control in vivo by an engineered GFP gradient. Science 370(6514):321–327. https://doi.org/10.1126/science.abb8205

Chapter 11

Anchor Away: A System for Fast Inhibition of Proteins in *Drosophila*

Pablo Sanchez Bosch

Abstract

Anchor away is a sequestering method designed to acutely and timely abrogate the function of a protein of interest by anchoring to a cell compartment different from its target. This method induces the binding of the target protein to the anchor by either the addition of rapamycin to *Drosophila* food or cell media. Rapamycin mediates the formation of a ternary complex between the anchor, which is tagged with the FK506-binding protein (FKBP12), and the target protein fused with the FKB12 rapamycin-binding (FRB) domain of mammalian target of rapamycin (mTOR). The rapamycin-bound target protein stays sequestered away from its compartment, where it cannot perform its biological function.

Key words Protein knockout, Loss of function, CRISPR, Rapamycin, TOR

1 Introduction

Decades of research with *Drosophila* have leveraged multiple methods to perform loss-of-function (LOF) experiments [1, 2]. Most of these methods, albeit efficient, act at the gene or mRNA levels, which introduces a delay from the time the LOF is induced, and the protein disappears from the tissue [3]. This is especially true for long-lived proteins. To overcome this problem, several approaches have been developed to target the protein directly, thus achieving a faster and more efficient LOF. Protein LOF can be achieved by targeted cleavage, targeted degradation, or sequestering [4–6].

Anchor away is a protein sequestering method developed in yeast to create LOF of nuclear proteins by binding them to ribosomes [5, 7]. It comprises three components (*see* Fig. 1): (A) an FKBP12-fused protein anchor, which has a specific intracellular localization (such as a ribosomal protein, a membrane receptor, or a mitochondrial protein) different from the localization of the target protein, (B) a target protein tagged with FRB, and (C) rapamycin to trigger the binding between the anchor and the

Christian Dahmann (ed.), *Drosophila: Methods and Protocols*, Methods in Molecular Biology, vol. 2540,
https://doi.org/10.1007/978-1-0716-2541-5_11,

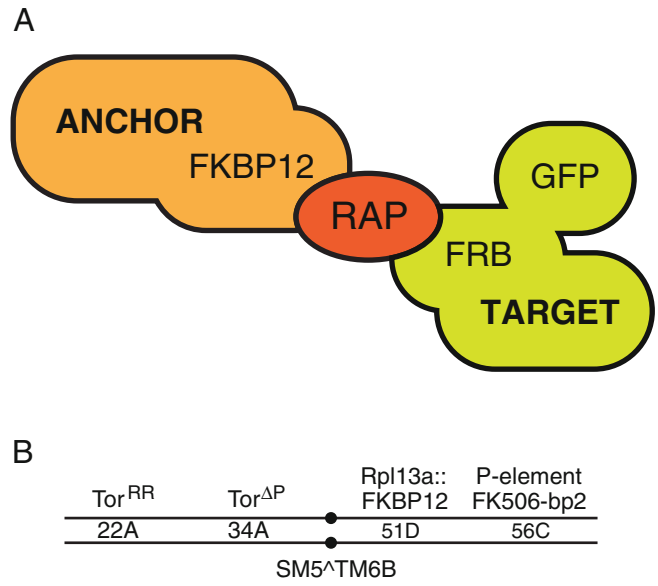


Fig. 1 Schematics of the anchor-away components. (**a**) The anchor protein is tagged with the FKBP12 peptide, while the target protein carries the FRB sequence, as well as GFP for detection. The addition of rapamycin creates a strong ternary complex comprised of the anchor, rapamycin, and target. (**b**) Location of the anchor-away transgenes and alleles in the second chromosome of *Drosophila*. All alleles are balanced over the compound balancer SM5^TM6b

target. It is an extremely efficient method in terms of speed and protein anchoring, giving it an edge over other sequestering or cleaving methods. All the components to elicit LOF are present in the cell, and the addition of rapamycin triggers a fast biochemical interaction that immediately captures the target protein and localizes it to another cellular compartment [8]. Moreover, because the anchor is present in extremely high numbers in the cell, it can efficiently trigger the anchoring and remain effective over the course of several hours.

In this chapter, a detailed protocol to achieve efficient protein anchoring in both live *Drosophila melanogaster* larvae and ex vivo wing disc cultures is described. The procedures to generate transgenic lines are briefly outlined in this protocol but are thoroughly presented in other chapters of this book (e.g., "CRISPR-/Cas9-Mediated Precise and Efficient Genome Editing in *Drosophila*" by Kevin Nyberg and Richard Carthew) [9].

2 Materials

2.1 Fly Lines, Fly Husbandry, and Dissection

1. *Drosophila* lines expressing the constructs of interest, i.e., an FRB-tagged target protein and the FKBP12-tagged anchor, recombine with a null allele of *FK506-bp2* to reduce

competition to rapamycin binding and a rapamycin-resistant Tor (Tor^{S1956T} or Tor^{RR}) to prevent rapamycin-induced lethality. A fly carrying all the components of the anchor away, except for the target (*see* **Note 1**), can be shared upon request from the Basler laboratory [8].

2. Fly vials.
3. Fly food.
4. Stereo microscope.
5. CO_2 fly station.
6. Forceps (Dumont #5).
7. Rapamycin stock solution: Dissolve rapamycin powder in dimethyl sulfoxide (DMSO) to a concentration of 2.5 mg mL^{-1} (2.74 mM). Alternatively, rapamycin can be purchased as 2.5 mg mL^{-1} ready-to-use solution. Make 20 µL aliquots from the stock solution and store at −80 °C.
8. Phosphate-buffered saline (PBS).
9. DMSO.

2.2 Fusing FRB::GFP with the Target Gene

This protocol follows standard cloning procedures such as restriction digest, polymerase chain reaction (PCR), DNA ligation, and bacterial transformation. Only the reagents specific to this protocol are listed in this section. Reagents common to any cloning procedure fall outside of the scope of this book chapter and should be readily available in any molecular biology laboratory.

1. pFRBGFP plasmid.
2. pCFD3 (*see* **Note 2**).
3. *Drosophila melanogaster* microinjection equipment. Alternatively, plasmids can be sent to a company to create the transgenic lines.
4. *nos-Cas9* transgenic flies (*see* **Note 2**).
5. Desalted primers specific to the target gene (**step 1** in Subheading 3.1).
6. ddH_2O.
7. AarI.
8. SapI.
9. Homology arm PCR primers. Design the primers in such a way that each arm will contain about 1 kb of homology **exactly** upstream and downstream of the target insertion region (*see* Fig. 2). When designing the PCR primers to amplify the fragments, ensure that the sequence upstream and downstream of the FRB::GFP sequence is kept in frame. Use the following primer template for the homology arms, where bold **N** marks

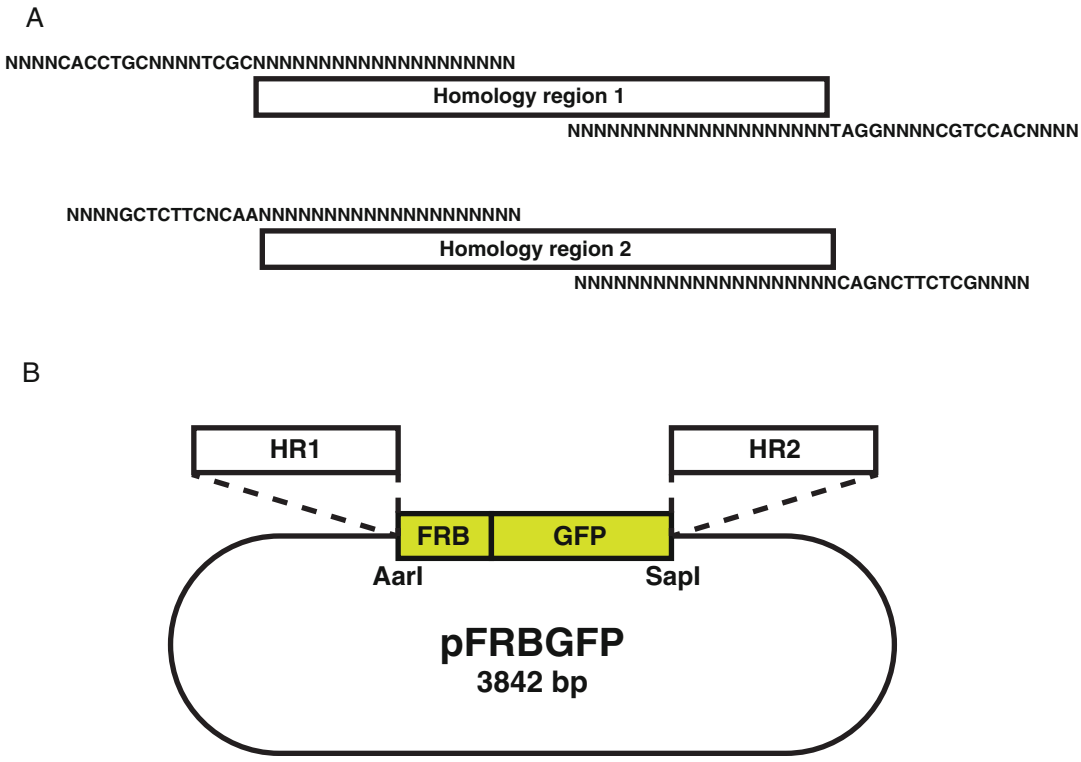


Fig. 2 Cloning of the homology arms into pFRBGFP. (**a**) The homology regions are amplified using the four primer sequences with the overhangs shown here. The overhangs are designed to ensure in-frame fusion. (**b**) The two homology arms created by PCR are cloned into pFRBGFP using the restriction enzymes AarI and SapI

the 20 bases from the primer that will bind to the target gene to amplify the homology region by PCR (*see* **Note 3**):

- AarI Fwd: NNNNCACCTGCNNNNTCGC**NNNNNNNNNNNNNNNN.**
- AarI Rev: NNNNCACCTGCNNNNGGAT**NNNNNNNNNNNNNNNN.**
- SapI Fwd: NNNNGCTCTTCNCAA**NNNNNNNNNNNNNNNNN.**
- SapI Rev: NNNNGCTCTTCNGAC**NNNNNNNNNNNNNNNNN.**

10. Sequencing primers:
    - FRBGFP Fwd: CCAGTAAAGGAGAAGAAC.
    - FRBGFP Rev.: TTTGTATAGTTCATCCATG.

2.3 Protein Anchoring in Live Drosophila

1. Gritman spatula or similar.
2. 4% Paraformaldehyde (PFA).
3. VECTASHIELD (Vector Labs).
4. Laser scanning confocal microscope.

2.4 Protein Anchoring in Cultured Wing Imaginal Discs

1. Wing medium 1 (WM1) [10]: Desired volume of Schneider's medium; 6.2 μg mL^{-1} insulin; 5% v/v fly extract. Prepare the fly extract as directed below. The fly extract can be kept frozen at −20 °C for several years.
2. Fly extract:
   (a). Collect a large amount of 2- to3-day-old *y, w* adult flies. Pour them into 50 ml tubes and store at −80 °C for at least 2 hours.
   (b). Weigh the flies and determine the total amount of M3 medium that will be used to make WM1. Prepare 6.18 mL of ice-cold M3 medium per gram of frozen flies.
   (c). Place the frozen flies in a sealable bag (zip bag or thermo-sealable plastic). Add just enough M3 medium to cover all the flies, never over the total amount calculated above.
   (d). Remove all air from the bag and seal it. Squash the flies with a large plastic pipette or a bottle by rolling it over the sealed bag. Repeat thoroughly until all the flies are homogenized.
   (e). Carefully open the bag and add the remaining ice-cold M3 medium. Pour the whole content of the bag into a 50 ml centrifuge tube.
   (f). Centrifuge at 1500× *g* at 4 °C for 15 min. In the meantime, prepare a water bath at 60 °C.
   (g). Collect all the supernatants, including surface fat, and discard the pellet.
   (h). Heat-inactivate the fly extract at 60 °C in a water bath for 5 min. Solution should turn cloudy.
   (i). Centrifuge at 1500× *g* at 4 °C for 90 min to pellet denatured proteins. Collect the supernatants and discard the pellet.
   (j). If the fly extract is still cloudy, filter through a 0.45 μm filter.
   (k). Filter the cleared fly extract through a 0.22 μm filter. Aliquot to the desired volume (recommended: 500 μl) and store for up to 2 years at −20 °C. It is best to avoid repeated freezing/thawing cycles, but aliquots of WM1 can be refrozen once or stored at 4 °C for 2 weeks.
3. Horizontal flow hood.
4. Sterilin 90 mm cell culture dish.
5. 6-Well plate.
6. 70% EtOH.
7. Ringer's solution.

8. Sterile 30 mm agar Petri dish.
9. Whatman filter paper.
10. Imaging Dish 1.0 (Zell-Kontakt) or glass bottom dish (MatTek).
11. Millicell standing inserts (MilliporeSigma).
12. Surgical scalpel.
13. Oil for embryo culture.
14. Spinning disc microscope.

3 Methods

The protocol described here is designed to anchor proteins to the ribosomes. This method works efficiently to abrogate nuclear proteins, for which it was designed, although it can potentially be applied to abrogate the function of other protein types, such as mitochondrial proteins and membrane receptors.

To deplete a protein of interest, it must first be tagged with FRB. To do so, it is recommended to fuse FRB with the endogenous gene. An outline of an efficient and straightforward clustered regularly interspaced short palindromic repeat (CRISPR)-mediated homologous recombination into the target gene locus is provided in **Note 2**, including the required fly stocks and plasmids. Subheading 3.1 describes the steps to generate the plasmids required to fuse FRB::GFP with the target gene.

The most efficient concentration of rapamycin for both rapamycin-containing food and the ex vivo culture of wing imaginal discs is 50 μM. Further diluting the rapamycin allows for mild anchoring of the target protein, but the results are variable and might affect reproducibility. Higher concentrations are not recommended, as they do not show an increased anchoring of the target. Rapamycin solution must be prepared fresh before use from a DMSO stock. To prepare 50 μM rapamycin, dilute the DMSO stock into an appropriate volume of PBS or WM1. Avoid multiple cycles of freezing/thawing.

3.1 Fusing FRB::GFP with the Target Gene

To integrate the FRB sequence into the target protein, use the donor plasmid pFRBGFP from the Basler laboratory ([8]; plasmid and DNA sequence are available upon request). pFRBGFP is based on the pDsRed-attP (Addgene #51019), but the DsRed-attP sequence was replaced by the FRB::GFP sequence. FRB::GFP is placed in such a way that it will be inserted in frame once the homology arms for the recombination are cloned into the plasmid.

1. To create the recombination plasmid, the homology arms are first amplified by PCR. Use the primers AarI Fwd and AarI Rev.

to create the homology arm 1 (5′ homology region) and the primers SapI Fwd and SapI Rev. to create the homology arm 2 (3′ homology region).

2. Once the PCR fragments are obtained, clone the homology region 1 digesting the PCR fragment and the plasmid with AarI, and then clone the homology region 2 by digesting both the PCR fragment and plasmid using SapI. If the first recombination region contains a SapI site, then the second arm must be cloned first.
3. Design gRNAs against the gene of interest and clone them in pCFD3 (*see* **Note 2**).
4. gRNA and pFRBGFP are co-injected into embryos from the line *nos-Cas9* [11], which expresses Cas9 in the germline.
5. Score survivor embryos and place them in a vial with fly food until adult flies are hatched.
6. Cross surviving adult flies in single fly crosses to a suitable balancer line.
   - Screen F1 flies by PCR to confirm the insertion of the FRB::GFP fragment into the gene of interest using the primers FRBGFP Fwd and FRBGFP Rev.
7. Although the CRISPR-mediated homologous recombination is highly efficient and rarely introduces single-point mutations, it is recommended to check the insertion to ensure that the FRB fragment has been inserted in frame. To do so, once a transgenic line has been established, use primers flanking the recombination region, around 200 bp upstream and downstream of the FRBGFP insertion site.

3.2 Protein Anchoring in Live Drosophila

1. Amplify a fly line carrying the constructs of interest (*see* Fig. 1). In the example given, the target protein is on the third chromosome. However, because the target protein should be expressed from the endogenous locus, it might be located on other chromosomes. Adapt the genetics accordingly to account for that.
2. Prepare rapamycin-containing food by embedding regular fly food with rapamycin 50 μM: create 1–2 cm deep cross-streaks in the fly food with a Gritman spatula or blunt forceps and add 300 μl of 50 μM rapamycin solution. Cover the vials with aluminum foil and store at 4 °C for at least 16 hours, until ready to use. Rapamycin-containing food can be stored up to 1 week at 4 °C.
3. Collect eggs from the transgenic flies in regular fly food, and let them develop until the developmental stage where the protein must be depleted. For this protocol, *Drosophila* can be maintained at any suitable standard temperature from 18 °C to 29 °C.

4. Place rapamycin-containing vials at room temperature (RT) for at least 1 hour before use to allow the food to warm to 20 °C. It is recommended to keep the fly food covered from light as much as possible to prevent degradation of rapamycin.
5. Transfer the embryos or larvae into the rapamycin-containing food, and return them to the fly incubator until end-point analysis. We recommend maintaining anchor-away *Drosophila* lines at 25 °C for optimal growth. Although possible if needed, it is not recommended to keep rapamycin-containing food at 29 °C or higher for extended periods of time, as higher temperatures increase rapamycin degradation.
6. Proceed to end-point analysis by dissecting the tissue of interest for microscopy or other downstream applications.
7. Fix the tissue with 4% PFA, and then wash with PBS and directly mount in VECTASHIELD, or perform immunostaining if desired.

3.3 Protein Anchoring in Cultured Wing Imaginal Discs

1. Prepare WM1 [10]. It is recommended to use fresh WM1, but if there is some unused WM1 after **step 17**, it can be stored at 4 °C for a week.
2. Amplify a fly line carrying the constructs of interest (*see* Fig. 1). In the example given, the target protein is on the third chromosome. However, given that the target protein should be expressed from the endogenous locus, it might be located on other chromosomes. Adapt the genetics accordingly to account for that.
3. Collect eggs from the transgenic fly line in regular fly food, and let them develop until the developmental stage where the protein should be depleted. For this protocol, *Drosophila* can be maintained at any suitable temperature between 18 °C and 29 °C.
4. Prepare rapamycin-containing WM1 by diluting stock rapamycin to 50 μM in WM1.
5. Prepare a 6-well plate with three wells containing Ringer's solution and a well with 70% EtOH.
6. Collect the larvae to dissect, and transfer them to the first well from the 6-well plate with Ringer's solution. Transfer larvae to the second well with Ringer's solution, and wash them for 1 min.
7. Sterilize the larval cuticle for 3 min by transferring the larvae to the well with 70% EtOH. Finally, wash away the EtOH by transferring the larvae to the final well with Ringer's solution. Place the sterile larvae on an agar Petri dish until ready to dissect.

8. For each larva, prepare the bottom of a Petri dish with three 30 μl drops of WM1. Place a single larva in the first WM1 drop.
9. Perform a quick section on the second third of the larval body, and remove the distal third of the larva from the WM1 drop. Invert the proximal side to expose the wing discs.
10. Remove the gut from the larva carefully, to avoid rupturing it.
11. Quickly cut the wing discs free by using fine forceps. Avoid stretching, scratching, or poking the imaginal discs.
12. Wash the discs by sequentially transferring them with a P20 micropipette to the second and third drops of WM1. Use 2–5 μl as the carrying volume. Alternatively, discs can be carefully transferred by holding them using closed forceps. This, however, increases the risk of damaging the discs. Discs can be stored on WM1 until enough specimens have been collected.
13. Prepare a wing disc imaging chamber by placing a piece of 10 x 0.5 cm^2 of Whatman paper on the wall of a live imaging dish. Humidify the filter paper with 300–500 μl of PBS. Prepare the Millicell filter by cutting off the bottom legs with a scalpel. Leave it aside for now.
14. Place a 20 μl drop of WM1 + rapamycin in the middle of the live imaging dish. Transfer the dissected discs into the drop with a 20 μl pipette in 2 μl of WM1. Set a timer or stopwatch to register the length of the rapamycin treatment. Protein trapping can be observed at around 30 min after the rapamycin treatment starts, and complete protein knockout is achieved between 1 and 2 hours.
15. Make sure that the discs are positioned with the peripodial membrane facing down. Use the forceps to align them correctly if needed.
16. Hold the Millicell filter with the forceps from the top, with the membrane side facing the discs, and place it gently but as rapidly as possible on the wing discs. Avoid lateral movements, as they might displace or damage the imaginal discs.
17. Add 200 μl of WM1 + rapamycin to the imaging chamber, until the Millicell filter chamber is fully covered with WM1.
18. Close the imaging chamber with the live imaging dish top. The wing discs are ready for imaging. Transport the chamber carefully to the microscope, as strong movements can displace the Millicell chamber and discs.
19. Avoid keeping temperatures below RT (21 °C) during imaging. If possible, use a microscope equipped with a temperature-controlled chamber.

4 Notes

1. Genetics of the Anchor away

   The anchor away requires the following transgenic constructs and alleles:

   - Mutant *tor* allele $Tor^{\Delta P}$ [12].
   - Transgenic rapamycin-resistant *tor* 22A-Tor^{S1956T} [12].
   - Mutant *fk506-bp2* allele, P{GSV6}GS10737, Kyoto stock #205244 [13].
   - Anchor *rpl13a*-tagged ribosomal protein, 56C-Rpl13a::FKBP12 [8].

2. A fly line carrying all the genetics components of the anchor away (Fig. 1) except for the target of interest. This line was generated in the Basler lab [8] and balanced over SM5^TM6B. This fly line is available upon request.

   To anchor a target protein, a single copy of the four alleles above achieves high efficiency, being the target protein the only one required in homozygosity.

3. About gRNA design and CRISPR-mediated homologous recombination

   Flies carrying FRB::GFP-tagged protein targets are generated using the CRISPR-mediated homologous recombination protocol from Port *et al* [11]. CRISPR gRNAs used in this protocol are cloned into the pCFD3 plasmid. The protocol from Port & Bullock from this same volume [9] thoroughly explains the design and cloning of the gRNAs against a target gene of interest. pCFD3 containing the gRNA specific to the gene of interest is co-transfected with the recombination plasmid designed following **steps 3.1.1** and **3.1.2** into embryos from flies containing *nos-Cas9* [11].

   Once the transgenic fly has been created, one must ensure that the transgene can become homozygous. Otherwise, the presence of a copy of the WT gene will greatly affect the efficiency of the anchoring. The WT copy will not be anchored at the ribosome upon rapamycin addition and the protocol will not be able to create loss of function alleles.

4. How to introduce the FRB::GFP in frame within the target gene

   To insert the FRB::GFP fragment in frame within the target, one must ensure that the sequence used to PCR the recombination arms 1 and 2 will align in frame with the FRB::GFP sequence from the plasmid. The FRB sequence starts with ATC. Therefore, the last codon from the homology arm 1 should be right before that sequence in the AarI reverse primer. For further detail, check the underlined sequence on the AarI Rev primer below:

N NNN CAC CTG CNN NNG GAT **NNN NNN NNN NNN NNN NNN NN.**

5. To maintain the rest of the target gene in frame, the second homology arm must start with the codon immediately after the last codon from homology arm 1. The GFP sequence ends with CAA. This sequence and the first codon of the second homology arm (which in the wild-type gene will be the one that immediately follows the last codon from the homology arm 1) are underlined in the SapI Fwd primer below:

   NNN NGC TCT TCN CAA **NNN NNN NNN NNN NNN NNN NN.**

6. It is recommended to place the FRB::GFP sequence at the target gene's N-terminal sequence. A C-terminal insertion is also feasible with this plasmid, but we advise to use a variant of the plasmid, pG17-FRBGFP, which includes a G17 linker right before the FRB sequence. pG17-FRBGFP is available at the Basler lab upon request [8]. When using the pG17-FRBGFP plasmid, use the following primer instead of AarI Rev (The last codon from the homology arm and the beginning of the G17 linker are underlined):

   AarI-G17 REV: N NNN CAC CTG CNN NNA GGC **NNN NNN NNN NNN NNN NNN NN.**

References

1. Cooley L, Kelley R, Spradling A (1988) Insertional mutagenesis of the Drosophila genome with single P elements. Science 239: 1121–1128
2. Adams MD, Sekelsky JJ (2002) From sequence to phenotype: reverse genetics in Drosophila melanogaster. Nat Rev Genet 3:189–198
3. Boutros M, Ahringer J (2008) The art and design of genetic screens: RNA interference. Nat Rev Genet 9:554–566
4. Harmansa S, Alborelli I, Bieli D, Caussinus E, Affolter M (2017) A nanobody-based toolset to investigate the role of protein localization and dispersal in Drosophila. Elife 6:99
5. Haruki H, Nishikawa J, Laemmli UK (2008) The anchor-away technique: rapid, conditional establishment of yeast mutant phenotypes. Mol Cell 31:925–932
6. Caussinus E, Kanca O, Affolter M (2011) Fluorescent fusion protein knockout mediated by anti-GFP nanobody. Nat Struct Mol Biol 19:117–121
7. Ding L, Laor D, Weisman R, Forsburg SL (2014) Rapid regulation of nuclear proteins by rapamycin-induced translocation in fission yeast. Yeast 31:253–264
8. Sanchez Bosch P, Pepperl J, Basler K (2020) Anchor away – a fast, reliable and reversible technique to inhibit proteins in Drosophila melanogaster. G3 (Bethesda) 10: 1745–1752
9. Port F, Bullock SL (2016) Creating heritable mutations in Drosophila with CRISPR-Cas9. Methods Mol Biol 1478:145–160
10. Restrepo S, Zartman JJ, Basler K (2016) Cultivation and live imaging of Drosophila imaginal discs. Methods Mol Biol 1478:203–213
11. Port F, Chen HM, Lee T, Bullock SL (2014) Optimized CRISPR/Cas tools for efficient germline and somatic genome engineering in Drosophila. Proc Natl Acad Sci U S A 111: E2967–76
12. Zhang H, Stallock JP, Ng JC, Reinhard C, Neufeld TP (2000) Regulation of cellular growth by the Drosophila target of rapamycin dTOR. Genes Dev 14:2712–2724
13. Toba G, Ohsako T, Miyata N, Ohtsuka T, Seong KH, Aigaki T (1999) The gene search system. A method for efficient detection and rapid molecular identification of genes in Drosophila melanogaster. Genetics 151:725–737

Chapter 12

Tagging *Drosophila* Proteins with Genetically Encoded Fluorophores

Jerome Avellaneda and Frank Schnorrer

Abstract

Proteins are typically not expressed homogeneously in all cells of a complex organism. Within cells, proteins can dynamically change locations, be transported to their destinations, or be degraded upon external signals. Thus, revealing the cellular and subcellular localizations as well as the temporal dynamics of a protein provides important insights into the possible function of the studied protein. Tagging a protein of interest with a genetically encoded fluorophore enables us to follow its expression dynamics in the living organism. Here, we summarize the genetic resources available for tagged *Drosophila* proteins that assist in studying protein expression and dynamics. We also review the various techniques used in the past and at present to tag a protein of interest with a genetically encoded fluorophore. Comparing the pros and cons of the various techniques guides the reader to judge the suitable applications possible with these tagged proteins in *Drosophila*.

Key words *Drosophila*, Tagging, Recombineering, Imaging, GFP, CRISPR-Cas9

1 Introduction

Drosophila has been a powerful genetic model for more than 100 years. Systematic forward genetic screens identified various adult and embryonic phenotypes, which, together with the subsequent identification of the affected genes, led to the functional annotation of a large part of the *Drosophila* genome [1–5]. The availability of genome-wide conditional genetic loss-of-function tools, such as genome-wide transgenic RNAi libraries and large-scale transposon collections, allows to functionally investigate most *Drosophila* genes in a cell type or at the developmental stage in a specific manner [6–11]. Thus, to date, we often have identified a functional contribution of a particular gene to a particular process; however, we are frequently missing the detailed mechanism of how the studied protein fulfils its function.

Christian Dahmann (ed.), *Drosophila: Methods and Protocols*, Methods in Molecular Biology, vol. 2540,
https://doi.org/10.1007/978-1-0716-2541-5_12,

Recent single-cell sequencing technologies that led to the fly cell atlas [12], together with large-scale mRNA localization studies, have revealed the tissue-specific expression of many *Drosophila* genes [13, 14], whose mRNAs are often localized to particular subcellular compartments [15, 16]. Thus, measuring gene expression and mRNA localization in *Drosophila* is rather simple and can be performed on a genome-wide basis, whereas studying the subcellular localization and dynamics of *Drosophila* proteins remains tricky to date. Only for about 5% of the *Drosophila* proteins, antibodies are readily available [10]. As antibody staining normally requires the fixation of the tissue of interest, it cannot reveal the dynamics of the studied proteins. Hence, the development of tools to determine protein localization and protein dynamics in *Drosophila* is of significant interest [17].

In this chapter, we review the currently available tools to monitor *Drosophila* protein expression and dynamics using well-characterized protein tags, which are in frame fusions to the protein of interest that can monitor localization in vivo. We particularly focus on fluorescent tags that can reveal protein dynamics in vivo. We first provide an overview of the currently available and ready-to-use tagged *Drosophila* collections and then detail the various methods available to newly tag a protein of interest with a genetically encoded fluorophore. In both cases, we discuss the important advantages and disadvantages of fly collections and tagging methods to guide a sensible application of the existing strains and to suggest the right tagging method when generating novel strains.

2 Available Fly Collections Expressing Fluorescently Tagged Proteins

Various strategies exist on how a fluorescent protein can be introduced into the *Drosophila* genome such that the protein of choice is tagged in frame and a fusion protein is produced. The tag can be introduced into the endogenous gene, and, thus, all copies of the protein are tagged, or it can be introduced as a transgene into the *Drosophila* genome, in addition to the untagged endogenous copy. Both strategies have been used to produce a large number of *Drosophila* strains that can be ordered from various stock centers (Table 1).

2.1 Endogenously Tagged Genes

Ideally, tagging a protein of choice at its endogenous gene locus should accurately recapitulate the endogenous expression pattern of the protein. However, one needs to be aware that all currently available genetically encoded fluorescent proteins suitable for in vivo protein tagging are relatively large (around 28 kDa and about 4 nm long) [18, 19] and may thus modify the dynamics or function of the tagged protein. Generally, tagging a protein of interest with a fluorescent tag has been tremendously useful for

Table 1
Publicly available (pre-)tagged *Drosophila* collections

Strategy	Method	Fly collection	No. of genes	Comment	Source	Studies
Endogenous gene	Pre-tag only	MiMIC CRIMIC	1862 1029	RMCE can insert fluorescent tags of choice	Bloomington	[10, 28, 29]
	Fluorescent tag	MiMIC	606	Internal tag, not all lines tag all protein isoforms		
		Fly trap	189		Bloomington and DGGR, Kyoto	[21–24]
		Kyoto fly trap	113		DGGR, Kyoto	Https://kyotofly.Kit.Jp/
		CPTI (Cambridge protein trap insertion)	374	Internal tag, not all lines tag all protein isoforms	DGGR, Kyoto	[25, 26]
		Rab-YFP	27	N-terminal tag, knock-in	Bloomington	[30]
Transgene	Endogenous control, genomic clone	fTRG (FlyFos)	847	C-terminal fusion, most tag all isoforms	VDRC, Vienna	[39]
	UAS control, cDNA clone	UAS-cDNA	309	Single isoform	Bloomington	Various labs
		FlyORF (UAS)	2858	Nonfluorescent, HA tag only, single isoform	FlyORF, Zurich	[32, 33]

various applications ranging from live in vivo microscopy to biochemical protein-complex purifications [17, 20].

Historically, the first endogenously tagged proteins were generated by the gene trap collections created by mobilization of various transposable elements, initially P-elements, which contained an exon coding for a green fluorescent protein (GFP) variant flanked by a splice acceptor and splice donor sites, together with other markers such as *white* + or *yellow* + that can be easily recognized. If such a construct jumped into a "coding" intron, a GFP fusion can be made when the reading frame is maintained (*see* **Note 1**) [21]. This approach was successfully expanded to about 300 different genes using piggyBac vectors, which are available from Bloomington or Kyoto stock centers (Table 1) [22–24] (*see* **Note 2**). A similar hybrid piggyBac–P-element vector (pigP) has been used more recently to create the Cambridge Protein Trap Insertion (CPTI) collection covering about 400 genes that are also available from Kyoto (Table 1) [25, 26]. These protein trap

lines share the advantage that the endogenous gene is tagged; however, they have the potential disadvantage that large parts of the transposon, including *white* or *yellow* markers, remain within the intron of the trapped gene. These may possibly interfere with the correct expression pattern of the tagged proteins.

All transposons have certain insertion biases in the genome. Hence, it is challenging to tag all proteins with reasonable effort. The advantage of the transposable Minos element is its ability to frequently insert into gene bodies; hence, it was chosen for a large-scale tagging project to create a large collection of fly lines, the MiMIC (Minos-mediated integration cassette) collection, including almost 2000 insertions located in "coding" introns (Table 1). The MiMIC cassette carries a splice acceptor, followed by a STOP cassette and a fluorescent red eye marker, which are flanked by two attP sites recognized by the recombinase phiC31 (*see* Fig. 2a) [10, 27]. The initial MiMIC line is thus a "pre-tagged" line that is not yet expressing GFP but rather results in a transcriptional stop and a loss-of-function allele (*see* **Note 3**). In the second step, the MiMIC marker cassette can be exchanged using recombinase-mediated cassette exchange (RMCE) with a tagging cassette of choice in the correct reading frame (*see* Subheading 4.2 for details). For a large set of genes, this cassette exchange has already been carried out, and fluorescently tagged lines for about 600 different genes are available from Bloomington (Table 1).

The Minos transposon technology has successfully tagged coding introns in many genes. However, the random nature of the transposon insertion makes further expansion of this tagged gene list extremely labor-intensive. Hence, with the availability of clustered regularly interspaced short palindromic repeat (CRISPR)-Cas9, systematic tagging efforts have made use of CRISPR-Cas9 to insert an attP-flanked STOP cassette containing a marker into the introns of choice. This allowed to systematically cover genes that had thus far escaped the Minos transposon method (*see* Subheading 4.2 for details). This technology, named CRIMIC (for CRISPR-mediated integration cassette), was used to generate a library of fly lines for about 1000 genes. These can also be considered "pre-tagged," as they do contain a STOP cassette followed by a T2A-GAL4 cassette, which allows to label the cell population expressing the gene of interest (*see* **Note 4**) [28, 29]. These pre-tagged lines can be used to generate fluorescently tagged protein lines using RMCE with the same cassettes as used for MiMIC (*see* Subheading 4.2 for details).

One possible concern of all the above-mentioned protein trap technologies is the insertion of an additional exon into the middle of the gene, resulting in the insertion of a fluorescent protein into the middle of the protein isoforms, instead of at its N- or C-terminus. However, it was shown that, indeed, many of these internally tagged proteins are functional [10].

Inserting the fluorescent tag at the N- or C-terminus of the protein might still be preferable for some proteins. To tag the endogenous protein precisely at the N- or C-terminus usually requires precise genome engineering within a particular exon. Even in the current CRISPR age, this has not yet been carried out to produce a large-scale resource of fluorescently tagged fly collections. One important N-terminal yellow fluorescent protein (YFP)-tagged fly collection is available for all the *Drosophila* Rab proteins, most of which are indeed functional (Table 1) [30]. This collection was produced in a heroic effort using the classical homologous recombination method, which, today, because of its intense workload, is not the method of choice anymore [30, 31].

2.2 Tagged Transgenes

An alternative to tagging the endogenous locus is to introduce a tagged transgene into the fly genome. Fluorescently tagged proteins expressed from a transgenic DNA construct can either be expressed under their endogenous control if upstream and downstream genomic sequences are included or under exogenous control using the upstream activation sequence (UAS)/GAL4 system. Various laboratories have deposited fly lines covering about 300 different genes in Bloomington, each expressing a different single protein isoform fused with a fluorescent protein under UAS control (Table 1). Importantly, the FlyORF consortium generated transgenic lines for almost 3000 UAS-cDNA constructs that can be ordered from FlyORF (Zurich), containing almost all *Drosophila* transcription-related factors [32, 33]. However, these proteins are not fluorescently tagged but are fused with only a small hemagglutinin (HA) tag (Table 1).

UAS constructs have the advantage that they are generally small and thus cloning and fly transgenesis are easy. However, they cannot recapitulate the endogenous protein dynamics as their transcription is controlled heterogeneously. With the development of recombineering techniques in bacteria and the availability of genome-wide libraries containing large pieces of *Drosophila* genomic DNA (Table 2), it has become possible to systematically tag one gene of choice within its native genomic context in bacteria [34–38]. One such tagging effort used the FlyFos library, which has an average genomic insert size of 36 kb, to tag each gene at its C-terminus with a superfolder-GFP (sGFP), in addition to small affinity tags, to generate the sGFP TransgeneOme (sGFP-TRG) FlyFos library (Table 2) (*see* **Note 5**) [39]. This sGFP-TRG library was used to generate transgenic lines for more than 800 GFP-tagged genes, the fly TransgeneOme (fTRG) collection, which are available from the Vienna Drosophila Resource Center (VDRC) in Vienna (Table 1). Functional tests from a representative set of genes suggested that at least two-thirds of these tagged proteins are functional as they rescue null mutants (*see* **Note 6**) [39].

Table 2
Publicly available *Drosophila* genomic (pre-)tagged clone collections

Library	Vector	No. of clones	Insert size	Gene coverage	Tagged?	Applications	Marker
CHORI-321	BAC (P[acman])	52,081	21 kb	88.9%	Untagged	Tagging of small genes	*White+*
CHORI-322	BAC (P [acman])	23,899	83 kb	99.3%	Untagged	Tagging of large genes; genome duplications	*White+*
RPCI-98	BAC	17,204	163 kb	>99%	Untagged	DNA source for CRISPR/ Cas9 tagging	Not suited for transgenesis
FlyFos	Fosmid (FlyFos)	15,204	36 kb	89.3%	Untagged	High-throughput tagging	3xP3-dsRed
Pre-tagged- TRG	Fosmid (FlyFos)	11,257	36 kb	84.0%	Spacer inserted for tagging	High-throughput tagging	3xP3-dsRed
sGFP-TRG	Fosmid (FlyFos)	9580	36 kb	71.5%	sGFP-tagged (C-terminal)	Rescue, protein localization, and dynamics purification	3xP3-dsRed
FlyFos *D. pseudoobscura*	Fosmid (FlyFos)	2592	36 kb	37.0%	Untagged	Trans-species rescue (for RNAi rescue)	3xP3-dsRed

2.3 Advantages and Limitations of the Existing Fluorescently Tagged Fly Collections

The strongest advantage of the UAS-based tagged collections is that the proteins of interest can be expressed with any GAL4 line of choice and hence be studied in a cell type or at a developmental stage of interest [40]. The induced overexpression of the tagged protein can make its visualization simpler. On the other hand, this frequently results in overexpression of artifacts; hence, results need to be interpreted with care, especially if conclusions on protein function or dynamics are to be drawn [41].

For many applications, the expression, localization, or dynamics of the fluorescently tagged protein should ideally recapitulate the behavior of the untagged endogenous protein as closely as possible. Hence, its expression should be under endogenous

control. This is optimally achieved by the libraries that insert the tag into the endogenous gene, as all transcriptional and splicing regulations are preserved. However, the fTRG FlyFos collection also preserves most of the transcriptional and all of the splicing regulations, as the entire gene body and at least 10 kb upstream and 5 kb downstream of the gene body are included in the clone. Nevertheless, for genes with complex cis-regulatory elements often present in transcription factors, the included regions might not be sufficient to fully recapitulate the native expression pattern in all cell types and, hence, the tagged construct will not fully rescue gene function [39]. The fTRG FlyFos collection used a particular fast-folding GFP variant, superfolder GFP, which should recapitulate the protein dynamics highly accurately, including in the rapidly developing early fly embryo (*see* **Note 7**) [39].

The MiMIC insertions are created in a random manner. Hence, not all insertions label all isoforms. From the 1862 pre-tagged genes listed in Tables 1, 1399 insertions would allow to label all isoforms because they are located in an intron shared by all isoforms and are considered "gold" lines [10, 17]. Even the current CRIMIC strategy cannot tag all proteins, as about 3000 *Drosophila* proteins are encoded in a single exon and, hence, these genes do not contain a coding intron to insert the generic tagging cassette [39]. Furthermore, the generic MiMIC cassette does rely on a strong splice acceptor from the *Mhc* gene, which might not be optimal to preserve splicing regulation in all cell types or at all developmental stages (*see* **Note 3**). Depending on the cell type, splicing can change dynamically during development [42, 43]. A fantastic advantage of the MiMIC/CRIMIC design is that the tagging cassette can be freely chosen using the same pre-tagged starter line (*see* also Subheading 4.2 for methodology).

Tagging the protein by N- or C-terminal fusion generally labels most isoforms, as alternative splicing is rarer at these locations. It may also have the advantage that the fluorescent tag is not located inside of the protein, although this likely depends on the individual protein [10, 39]. Tagging with generic cassettes that are inserted into introns is simpler to scale up, and only attR sites remain in the introns (*see* Subheading 4.2 for details on methodology).

Tagging the endogenous gene locus has additional advantages. The function of the tagged protein is revealed immediately by simply analyzing the phenotype of the homozygous line. Furthermore, imaging of the tagged protein reveals the dynamics of the entire protein pool, as the entire pool contains the fluorescent tag (with the isoform limitations discussed above). Hence, it is also easier to combine these fluorescently tagged lines with tools like the deGradeFP system or RNAi against GFP (in vivo GFP interference, iGFPi) that allows to remove the tagged protein at the protein or RNA levels, respectively, at any stage of interest. This allows to dynamically investigate protein function [9, 10, 44]. Anti-GFP

nanobodies containing ectopic localization signals can also be used to mislocalize the tagged GFPs and investigate the functional consequences [45–47]. These applications are also possible for the transgenic constructs expressing the tagged protein under endogenous control, such as the fTRG lines; however, their application is more complicated by the fact that they need to be crossed into a null mutant background (*see* **Note 6**).

3 Available Clone Collections

To tag genes under endogenous regulation using transgenes, one needs large genomic clones. Most *Drosophila* genes are included in the currently available genomic clone collections (Table 2). The larger the clone is, the harder it is to work with it. In particular, genomic integration of extremely large clones into the *Drosophila* genome using phiC31 recombinase is more tricky; however, it is possible [48]. The Roswell Park Cancer Institute (RPCI)-98 BAC library has extremely large inserts and covers more than 99% of all genes, but it neither contains a marker for transgenesis nor an attB site. Hence, it is not suitable for transgenesis. The P[acman] libraries CHORI-321 and CHORI-322 were used to duplicate most parts of the X chromosome and resulted in functional rescue for most genes (>90%) included in these clones [48]. However, these libraries are currently not available with an already inserted protein tag (Table 2).

The untagged FlyFos library contains clones for about 90% of the *Drosophila* genes (10 kb upstream and 5 kb downstream of the gene body included) [36]. For most of these genes, a "pre-tagged" version is available that can be used to insert a fluorescent tag of choice using one generic primer pair and a recombineering step in bacteria (*see* Subheading 4 for detailed methods) [39]. Furthermore, a superfolder-GFP tag was inserted at the C-terminus for more than 9500 genes present in the sGFP-TRG FlyFos library (Table 2). These clones can be directly used to generate new fluorescently tagged fly lines, using the same methodology as for the 800 existing ones [39]. Additionally, a related untagged fosmid library from *Drosophila pseudoobscura* exists that was used in trans-species rescue experiments to demonstrate RNAi specificity [49].

4 Methods to Fluorescently Tag a Protein of Choice in *Drosophila*

In the previous section, we discussed the existing fluorescently tagged fly lines and the available tagged clone collections. Since tagged fly lines or tagged clones are not yet existing for all genes in various colors, we briefly review the principal methods of how to tag a gene in vitro by recombineering or in vivo using CRISPR-

Cas9. We also review the principle of recombinase-mediated cassette exchange, which is extremely handy to insert various tags at the same location of a protein.

4.1 Tagging In Vitro by Recombineering

The available large genomic clones (Table 2) are suitable to recapitulate the endogenous expression patterns of the included genes. Their manipulation applies "recombineering" techniques, namely, homologous recombination in bacteria to insert any sequence of choice [50]. Detailed protocols on how to insert fluorescent tags into these large genomic clones have been published previously [35, 38, 51, 52]. Here, we briefly summarize the major steps for generating a fluorescently tagged clone using the FlyFos library as an example (*see* Fig. 1). To generate the clone library, sheared *Drosophila* genomic DNA is ligated into the linearized FlyFos vector, which contains an attB site for phiC31-mediated transgenesis and a 3xP3-dsRed cassette for identification in flies (red fluorescent eyes). After fosmid packing and bacterial infection, the untagged FlyFos library with the individual clones is prepared (36 kb average insert size for the current FlyFos library) (*see* Fig. 1a, Table 2).

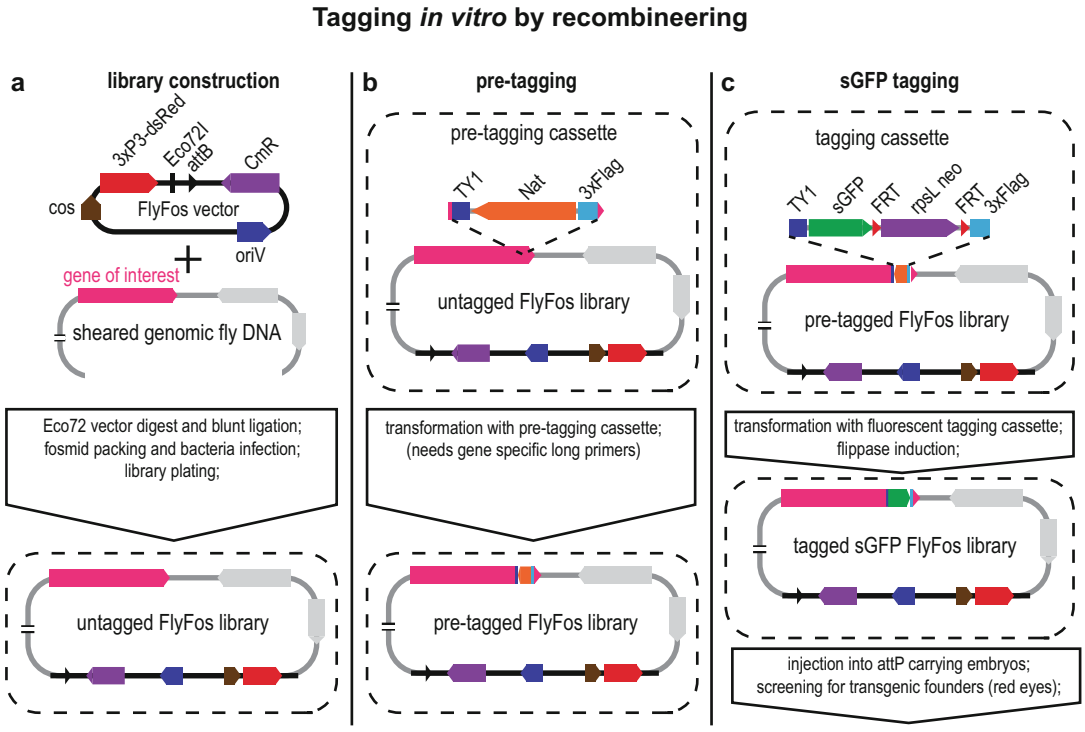


Fig. 1 Tagging in vitro by recombineering. Workflow illustration for generation of the sGFP FlyFos clone library. (**a**) Library construction. (**b**) Pre-tagging. (**c**) sGFP tagging

To tag a gene of choice at the C-terminus (or anywhere else), an appropriate pre-tagging cassette is amplified with gene-specific primer overhangs and inserted at the desired place by homologous recombination. Positive clones are easily identified by antibiotic resistance (the *Nat* gene results in resistance to nourseothricin). Thus, the pre-tagged FlyFos library is generated (*see* Fig. 1b). At the final step, the pre-tag is replaced by a fluorescent protein tag of choice, here a superfolder GFP, by amplifying the tagging cassette. This cassette can be amplified in bacteria and cut by restriction enzymes, as the 5-prime and 3-prime homology arms are the same for all clones (TY1 and 3xFlag tags present in the pre-tag, respectively). Hence, any polymerase chain reaction (PCR) errors are avoided. Transformation with the tagging cassette and selection for the rpsL neo cassette with kanamycin results in the replacement of the pre-tagging with the tagging cassette. At the final step, the expression of flippase is induced to remove the rpsL-neo cassette located between two flippase recognition target (FRT) sites, resulting in the tagged sGFP FlyFos library (*see* Fig. 1c, Table 2) [39]. These clones can be injected into attP carrying fly embryos expressing phiC31 recombinase, and transgenic founders are identified by red fluorescent eye color [39] (*see* **Note 5**).

Tagging by recombineering is highly efficient and can be performed on a genome-wide basis. Since both the pre-tagged and the sGFP-tagged FlyFos (sGFP-TRG) libraries are available from https://transgeneome.mpi-cbg.de/transgeneomics/, it is easy to generate new fTRG fly lines from the sGFP library by embryo injections or to insert different tags into the pre-tagged clones. Injection of the FlyFos clones is still labor-intensive since the integration of the large clones is challenging. A detailed and efficient protocol has been published [39]. New fly lines can be sent to the VDRC for distribution to the community.

4.2 Large-Scale In Vivo Tagging with Generic Cassettes Using MiMIC or CRIMIC

The large collection of MiMIC lines was generated by the in vivo mobilization of the Minos transposable element that contains a STOP cassette and a *yellow* body color marker flanked by two attP sites (*see* Fig. 2a). This Minos cassette can insert anywhere in the genome. If it inserts into a coding intron, a pre-tagged fly line is generated with one gene pre-tagged at its endogenous locus (*see* Fig. 2a). Generally, this line is a loss-of-function allele of the gene because the splice acceptor and transcriptional terminator result in transcriptional termination and no full-length RNA is produced (*see* **Note 8**). The lines are recognized by *yellow* + expression, and their insertion positions can be found in the GBrowse mode in Flybase [10, 27]. All the existing MiMIC pre-tagged lines can be easily ordered from the Bloomington stock center (Table 1).

As the insertion of the Minos element is random, it is not easy to generate a tagged insertion into a gene of choice for which no MiMIC insertion exists. Furthermore, the precise location of the insertion cannot be chosen. With the availability of CRISPR-Cas9

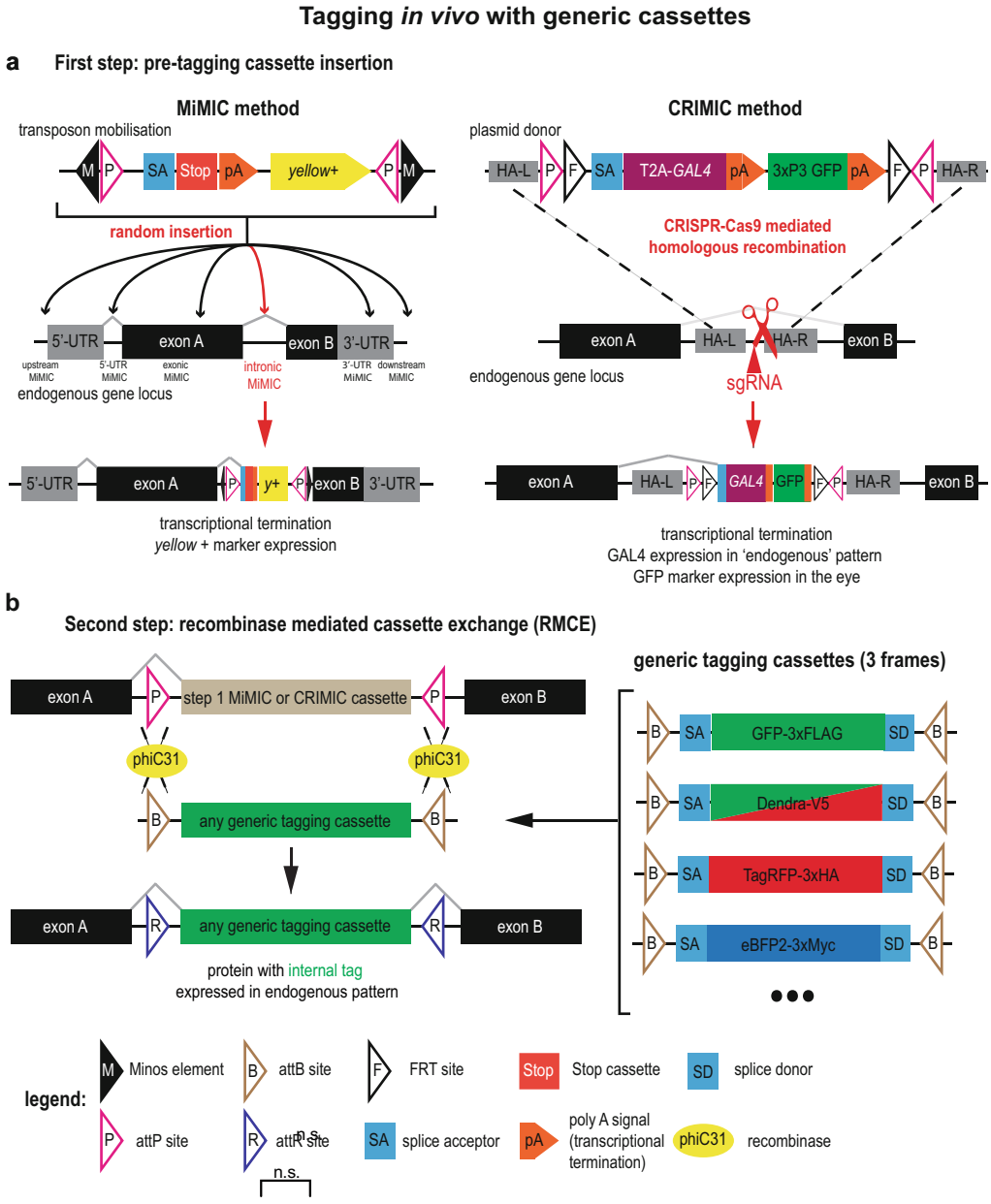


Fig. 2 Tagging in vivo with generic cassettes. Illustration of the MiMIC and CRIMIC two-step method to tag an endogenous gene in a random (MiMIC) or selected (CRIMIC) intron. Note the many available generic tagging cassettes for the RMCE step. (**a**) First step: pre-tagging cassette insertion. (**b**) Second step: recombinase mediated cassette exchange (RMCE)

and its adaptation to *Drosophila* [53–55], it has become, in principle, possible to insert cassettes of choice at any location of choice into the *Drosophila* genome. However, the integration of the fairly large cassettes needed for fluorescent protein tagging is still rather inefficient. Consequently, identification of the tagged founder flies can be labor-intensive if PCR screening is needed [55, 56].

Here, we describe an efficient method that copied the MiMIC design for protein tagging to insert either a dsRed or a GFP eye marker cassette (under control of the 3xP3 eyeless promotor) to allow simple identification of the founders using CRISPR-Cas9, recently dubbed CRIMIC (CRISPR-mediated integration cassette) (*see* Fig. 2a) [17, 28, 29, 56]. The CRIMIC cassette will be inserted into a suitable coding intron (*see* **Note 9**). After choosing a suitable target intron, the left and right homology arms (HA-L and HA-R, both between 1 and 2 kb) are amplified and cloned into the generic tagging cassette, which is then injected together with a suitable sgRNA into Cas9-expressing embryos (either *Act5C*-Cas9 or *nanos*-Cas9). The mosaic G_o flies are crossed to the balancer, and the founders are identified by their fluorescent eye color [28, 56]. A detailed crossing scheme as well as a timeline for cloning and fly work has been published [56]. Recently, a CRIMIC cassette plasmid has been designed that can be cleaved in vivo by co-injected sgRNAs to generate linear ds-DNA donors in vivo that integrate efficiently despite having only 100 bp homology arms. This makes the cloning steps more cost-effective [29]. In all cases, the pre-tagging cassette contains two attP sites that allow tag exchange by recombinase-mediated cassette exchange at step two (*see* Fig. 2a) [28, 56].

Many of the available CRIMIC step one pre-tagged fly lines contain a GAL4 after a ribosomal cleavage site (T2A) (*see* Fig. 2a). Thus, these pre-tagged lines can be used directly to label the cells expressing the pre-tagged gene, when crossed to a UAS reporter [28, 29]. Similar to the MiMIC step one flies, the CRIMIC step one flies also produce a truncated protein, which is likely not fully functional. Hence, in addition to assessing GAL4 expression, the loss-of-function phenotype of the homozygous line can be a good indication that the novel *Mhc* splice acceptor works effectively as designed (*see* Fig. 2a). Many of the pre-tagged CRIMIC lines can be ordered from Bloomington (Table 1).

Step one pre-tagged MiMIC or CRIMIC fly lines can be easily converted to fluorescently tagged lines by recombinase-mediated cassette exchange (RMCE) (*see* Fig. 2b). Several generic cassettes that allow the expression of various fluorescent proteins, including photo-activatable Dendra, are available in all three reading frames. These cassettes also contain small affinity tags useful for protein purification or antibody staining [27]. The plasmids are either injected together with a phiC31 plasmid into step one pre-tagged MiMIC or CRIMIC embryos [27, 56] or the pre-tagged flies are crossed to a transgenic RMCE donor source that is mobilized by flippase expression [57]. The latter is generally more effective and avoids time-consuming embryo injections. Mosaic G_o flies are crossed to the balancer, and successfully exchanged founders are identified by the absence of the fluorescent eye marker or by the *yellow* phenotype (excision of the *yellow* + cassette) (*see* Fig. 2b) (*see*

Note 10). A timeline and a crossing scheme for this second step have been published [56]. Converted step two lines expressing fluorescently tagged proteins are available for various different proteins from Bloomington (Table 1) [10].

4.3 In Vivo Tagging with CRISPR Using Gene-Specific Cassettes

Having access to large fly line collections expressing fluorescently tagged proteins at the endogenous locus is fantastic. However, for many applications, inserting the tag into a particular exon, instead of an intron, might be the preferred location to preserve protein function, or might be needed for particular goals, for example, to quantify mechanical tension across a protein of choice by a mechanosensitive fluorescence resonance energy transfer (FRET) sensor [58, 59]. There are various adaptations of CRISPR-Cas9 technology to insert fluorescent protein tags into *Drosophila* genes at the endogenous locus [55, 56, 60, 61]. Here, we focus our attention on a two-step method that is extremely similar to CRIMIC and combines the free choice of tag location with the advantages of flexible insertion of various different tagging cassettes [56].

To allow for a fully flexible choice of tag location in the protein, the two-step in vivo tagging method first selects a target exon that should contain the future location of the tag (*see* **Note 11**). This can either be the first or last coding exon to tag the protein at the N- or C-terminus, respectively, or it can be an internal exon to generate internal fluorescent protein fusions. At the first step, a donor plasmid with a pre-tagging cassette is injected into germline-expressing Cas9 embryos (*Act5*C-Cas9 or *nanos*-Cas9), this time together with two suitable sgRNAs that generate double-strand breaks on both sides of the target exon (*see* Fig. 3a). The donor plasmid contains right and left homology arms of 1–2 kb length that result in an effective Cas9-mediated homologous recombination. The transgenic founders are identified by red fluorescent eyes produced by 3xP3-dsRed in the pre-tagging cassette. Similar to the MiMIC system, the pre-tagging cassette contains a *Mhc* splice acceptor site and a transcriptional terminator, resulting in the production of a truncated protein (*see* **Note 12**). Importantly, the pre-tagging cassette is flanked by two attP sites that enable future cassette exchange. In summary, the step one pre-tagging cassette insertion generated a replacement of the target exon with a marker cassette (*see* Fig. 3a). A detailed protocol and timeline of the targeting step have been published [56].

Step two applies recombinase-mediated cassette exchange (RMCE) in a similar manner as explained above for MiMIC or CRIMIC; however, here, the entire target exon is included in the gene-specific tagging cassette. This allows to use the endogenous splice acceptor and donor and thus preserves the endogenous splicing pattern. It also enables to choose the exact position of the fluorescent tag in the target protein. However, this method is more demanding on cloning skills and possibly less effective when large

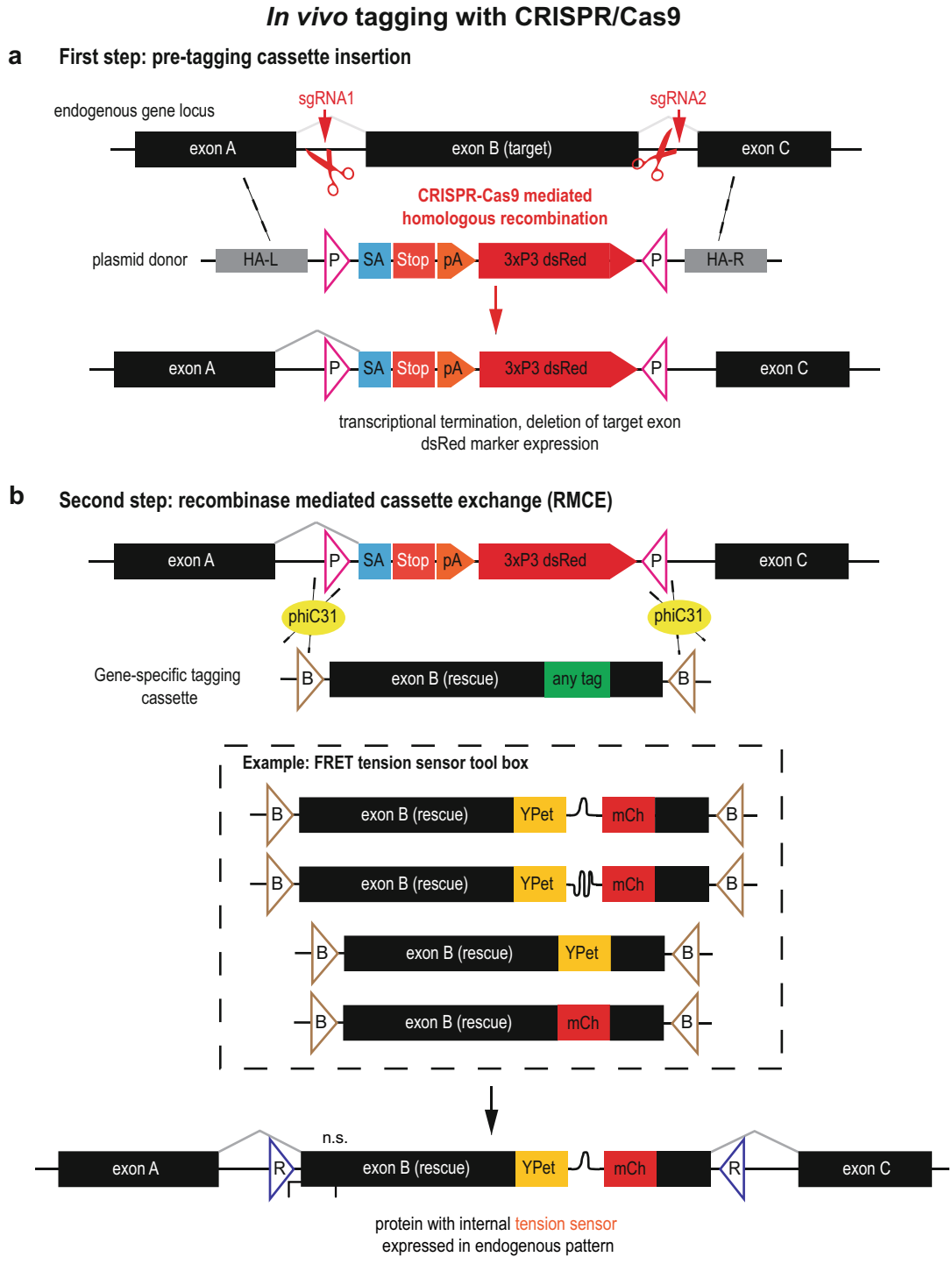


Fig. 3 Tagging in vivo with CRISPR-Cas9. Illustration of a flexible two-step tagging method that inserts a fluorescent tag of choice into a coding exon of choice. (**a**) First step: pre-tagging cassette insertion. (**b**) Second step: recombinase mediated cassette exchange (RMCE)

exons are replaced (we have effectively replaced exons of up to 8 kb with this methodology, F.S. unpublished data). This method is particularly powerful when several different tags should be inserted at the same location, which was successfully applied to generate a series of FRET-based molecular tension sensors in the Talin protein (*see* Fig. 3b). These sensors are all functional proteins and allow to quantify molecular tension across Talin in all cell types and at stages that express Talin [58]. This shows the particular strength of the two-step in vivo CRISPR method.

5 Outlook

Large-scale tagging projects have created an enormously rich resource for the fly community. In vivo CRISPR-Cas9-based tagging methodology is continuously improving, but we are still far from tagging a fly gene in vivo in two or three afternoons as is the case for yeast since many years. Continued funding of large resource projects and method developments will be instrumental for further improvements, making CRISPR-based tagging a routine for each *Drosophila* laboratory. Finally, sharing by sending all the generated fly lines to the established stock centers will allow the community to make the most use of these extremely valuable fly reagents.

6 Notes

1. A "coding" intron is an intron that is flanked by two protein-coding exons. Thus, the insertion of a new exon into this "coding" intron results in a fusion protein if the correct reading frame is used.
2. Not all gene traps listed under "Fly trap" are GFP fusions. Some only express GFP under control of an enhancer of a nearby gene. Only if inserted into the correct reading frame in a "coding" intron can a GFP-containing tagged protein be produced.
3. The MiMIC STOP cassette uses a strong splice acceptor site from the *Mhc* gene. However, this does not guarantee that the new artificial splice site is always used at its integrated location in each cell type that expresses the affected gene. One indication for efficient splicing is a strong loss-of-function phenotype of the MiMIC allele, e.g., lethality that is rescued when the pre-tagging cassette is exchanged with a tagging cassette. If the expected loss-of-function phenotype for the gene is not present, then the *Mhc* splice acceptor is likely not used effectively and tagging with a step two cassette might also not be highly efficient at this location.

4. T2A is a self-cleaving peptide that results in cleavage during ribosomal translation. In effect, this leads to free GAL4 protein production that is not fused with the remaining part of the tagged protein anymore.
5. The sGFP-TRG FlyFos library tagged all protein isoforms for most genes, since the fluorescent tag was inserted into the last coding exon before the stop codon and alternative C-termini resulting from alternative splicing are relatively rare.
6. Most available fTRG fly lines contain clones inserted into the third chromosome in attP VK00033. For genes located on the third chromosome, some have been inserted into the second chromosome to make rescue experiments simpler.
7. The fosmid clones contain a strong dsRed marker under control of the 3xP3 eyeless promotor. During folding of dsRed, a transient green fluorescence is present, which can be a disadvantage when imaging a tagged GFP in the developing brain. The dsRed fluorescence can be easily eliminated by expressing a sgRNA against dsRed together with Cas9 (F.S., reagent sent upon request).
8. The MiMIC step one cassette can be inserted in the sense or antisense direction compared to the gene. In both cases, it can be used to tag the gene in step two.
9. In the CRIMIC design, the target intron can be freely chosen. One should pay attention to the protein domain structure to select a suitable location of the internal fluorescent tag. The tag should not be located within a native protein domain and should ideally label all protein isoforms. Furthermore, placing the cassette into extremely small introns (smaller than 100 bp) increases the risk that the attR sites remaining at step two may affect the splicing and hence create a loss-of-function allele.
10. The exchange direction of RMCE is random. Thus, half the founders have the fluorescent cassette in the correct orientation. The orientation is easily verified with PCR.
11. If one of the flanking introns is smaller than 100 bp, it is better to replace more than one exon at step one to minimize the risk of splicing problems caused by the remaining attR sites in the introns.
12. As in the MiMIC system, the splice acceptor cassette produces a loss-of-function allele at step one.

References

1. Morgan TH (1932) The rise of genetics. Science 76:261–267
2. Lewis EB (1978) A gene complex controlling segmentation in Drosophila. Nature 276: 565–570
3. Nüsslein-Volhard C, Wieschaus E (1980) Mutations affecting segment number and polarity in Drosophila. Nature 287:795–801

4. St Johnston D (2002) The art and design of genetic screens: Drosophila melanogaster. Nat Rev Genet 3:176–188
5. Wieschaus E, Nüsslein-Volhard C (2016) The Heidelberg screen for pattern mutants of Drosophila: a personal account. Annu Rev Cell Dev Biol 32:1–46
6. Dietzl G et al (2007) A genome-wide transgenic RNAi library for conditional gene inactivation in Drosophila. Nature 448:151–156
7. Ni J-Q et al (2009) A Drosophila resource of transgenic RNAi lines for neurogenetics. Genetics 182:1089–1100
8. Schnorrer F et al (2010) Systematic genetic analysis of muscle morphogenesis and function in Drosophila. Nature 464:287–291
9. Neumüller RA et al (2012) Stringent analysis of gene function and protein-protein interactions using fluorescently tagged genes. Genetics 190:931–940
10. Nagarkar-Jaiswal S et al (2015) A library of MiMICs allows tagging of genes and reversible, spatial and temporal knockdown of proteins in Drosophila. elife 4:2743
11. Nagarkar-Jaiswal S, Manivannan SN, Zuo Z, Bellen HJ (2017) A cell cycle-independent, conditional gene inactivation strategy for differentially tagging wild-type and mutant cells. elife 6:1385
12. Li H et al (2021) Fly Cell Atlas: a single-cell transcriptomic atlas of the adult fruit fly. bioRxiv 375:eabk2432. https://doi.org/10.1126/science.abk2432
13. Tomancak P et al (2007) Global analysis of patterns of gene expression during Drosophila embryogenesis. Genome Biol 8:R145
14. Kvon EZ et al (2014) Genome-scale functional characterization of Drosophila developmental enhancers in vivo. Nature 512:91–95
15. Lécuyer E et al (2007) Global analysis of mRNA localization reveals a prominent role in organizing cellular architecture and function. Cell 131:174–187
16. Jambor H et al (2015) Systematic imaging reveals features and changing localization of mRNAs in Drosophila development. elife 4:e05003
17. Kanca O, Bellen HJ, Schnorrer F (2017) Gene tagging strategies to assess protein expression, localization, and function in Drosophila. Genetics 207:389–412
18. Hink MA et al (2000) Structural dynamics of green fluorescent protein alone and fused with a single chain Fv protein. J Biol Chem 275:17556–17560
19. Rodriguez EA et al (2017) The growing and glowing toolbox of fluorescent and photoactive proteins. Trends Biochem Sci 42:111–129
20. Dunst S, Tomancak P (2019) Imaging flies by fluorescence microscopy: principles, technologies, and applications. Genetics 211:15–34
21. Morin X, Daneman R, Zavortink M, Chia W (2001) A protein trap strategy to detect GFP-tagged proteins expressed from their endogenous loci in Drosophila. Proc Natl Acad Sci U S A 98:15050–15055
22. Kelso RJ et al (2004) Flytrap, a database documenting a GFP protein-trap insertion screen in Drosophila melanogaster. Nucleic Acids Res 32:D418–D420
23. Buszczak M et al (2007) The carnegie protein trap library: a versatile tool for Drosophila developmental studies. Genetics 175:1505–1531
24. Quiñones-Coello AT et al (2007) Exploring strategies for protein trapping in Drosophila. Genetics 175:1089–1104
25. Lowe N et al (2014) Analysis of the expression patterns, subcellular localisations and interaction partners of Drosophila proteins using a pigP protein trap library. Development 141:3994–4005
26. Lye CM, Naylor HW, Sanson B (2014) Subcellular localisations of the CPTI collection of YFP-tagged proteins in Drosophila embryos. Development 141:4006–4017
27. Venken KJT et al (2011) MiMIC: a highly versatile transposon insertion resource for engineering Drosophila melanogaster genes. Nat Methods 8:737–743
28. Lee P-T et al (2018) A gene-specific T2A-GAL4 library for Drosophila. elife 7:e35574
29. Kanca O et al (2019) An efficient CRISPR-based strategy to insert small and large fragments of DNA using short homology arms. elife 8:e51539
30. Dunst S et al (2015) Endogenously tagged rab proteins: a resource to study membrane trafficking in Drosophila. Dev Cell 33:351–365
31. Rong YS, Golic KG (2000) Gene targeting by homologous recombination in Drosophila. Science 288:2013–2018
32. Bischof J et al (2013) A versatile platform for creating a comprehensive UAS-ORFeome library in Drosophila. Development 140:2434–2442
33. Schertel C et al (2015) A large-scale, in vivo transcription factor screen defines bivalent chromatin as a key property of regulatory factors mediating Drosophila wing development. Genome Res 25:514–523

34. Venken KJT, He Y, Hoskins RA, Bellen HJ (2006) P[acman]: a BAC transgenic platform for targeted insertion of large DNA fragments in D. melanogaster. Science 314:1747–1751
35. Venken KJT et al (2008) Recombineering-mediated tagging of Drosophila genomic constructs for in vivo localization and acute protein inactivation. Nucleic Acids Res 36:e114
36. Ejsmont RK, Sarov M, Winkler S, Lipinski KA, Tomancak P (2009) A toolkit for high-throughput, cross-species gene engineering in Drosophila. Nat Methods 6:435–437
37. Venken KJT et al (2009) Versatile P[acman] BAC libraries for transgenesis studies in Drosophila melanogaster. Nat Methods 6:431–434
38. Ejsmont RK, Bogdanzaliewa M, Lipinski KA, Tomancak P (2011) Production of fosmid genomic libraries optimized for liquid culture recombineering and cross-species transgenesis. Methods Mol Biol 772:423–443
39. Sarov M et al (2016) A genome-wide resource for the analysis of protein localisation in Drosophila. elife 5:e12068
40. Elliott DA, Brand AH (2008) The GAL4 system: a versatile system for the expression of genes. Methods Mol Biol 420:79–95
41. Prelich G (2012) Gene overexpression: uses, mechanisms, and interpretation. Genetics 190:841–854
42. Spletter ML et al (2015) The RNA-binding protein Arrest (Bruno) regulates alternative splicing to enable myofibril maturation in Drosophila flight muscle. EMBO Rep 16:178–191
43. Spletter ML et al (2018) A transcriptomics resource reveals a transcriptional transition during ordered sarcomere morphogenesis in flight muscle. elife 7:1361
44. Caussinus E, Kanca O, Affolter M (2011) Fluorescent fusion protein knockout mediated by anti-GFP nanobody. Nat Struct Mol Biol 19:117–121
45. Harmansa S, Hamaratoglu F, Affolter M, Caussinus E (2015) Dpp spreading is required for medial but not for lateral wing disc growth. Nature 527:317–322
46. Harmansa S, Alborelli I, Bieli D, Caussinus E, Affolter M (2017) A nanobody-based toolset to investigate the role of protein localization and dispersal in Drosophila. elife 6:e22549
47. Harmansa S, Affolter M (2018) Protein binders and their applications in developmental biology. Development 145:dev148874
48. Venken KJT et al (2010) A molecularly defined duplication set for the X chromosome of Drosophila melanogaster. Genetics 186: 1111–1125
49. Langer CCH, Ejsmont RK, Schönbauer C, Schnorrer F, Tomancak P (2010) In vivo RNAi rescue in Drosophila melanogaster with genomic transgenes from Drosophila pseudoobscura. PLoS One 5:e8928
50. Murphy KC (1998) Use of bacteriophage lambda recombination functions to promote gene replacement in Escherichia coli. J Bacteriol 180:2063–2071
51. Ejsmont RK et al (2011) Recombination-mediated genetic engineering of large genomic DNA transgenes. Methods Mol Biol 772: 445–458
52. Hofemeister H et al (2011) Recombineering, transfection, Western, IP and ChIP methods for protein tagging via gene targeting or BAC transgenesis. Methods 53:437–452
53. Port F, Chen H-M, Lee T, Bullock SL (2014) Optimized CRISPR/Cas tools for efficient germline and somatic genome engineering in Drosophila. Proc Natl Acad Sci 111: E2967–E2976
54. Gratz SJ et al (2013) Genome engineering of Drosophila with the CRISPR RNA-guided Cas9 nuclease. Genetics 194:1029–1035
55. Baena-Lopez LA, Alexandre C, Mitchell A, Pasakarnis L, Vincent J-P (2013) Accelerated homologous recombination and subsequent genome modification in Drosophila. Development 140:4818–4825
56. Zhang X, Koolhaas WH, Schnorrer F (2014) A versatile two-step CRISPR- and RMCE-based strategy for efficient genome engineering in Drosophila. G3 (Bethesda) 4:2409–2418
57. Nagarkar-Jaiswal S et al (2015) A genetic toolkit for tagging intronic MiMIC containing genes. elife 4:166
58. Lemke SB, Weidemann T, Cost A-L, Grashoff C, Schnorrer F (2019) A small proportion of Talin molecules transmit forces at developing muscle attachments in vivo. PLoS Biol 17:e3000057
59. Fischer LS, Rangarajan S, Sadhanasatish T, Grashoff C (2021) Molecular force measurement with tension sensors. Annu Rev Biophys. https://doi.org/10.1146/annurev-biophys-101920-064756
60. Fletcher GC et al (2018) Mechanical strain regulates the hippo pathway in Drosophila. Development 145:dev159467–10
61. Pinheiro D et al (2017) Transmission of cytokinesis forces via E-cadherin dilution and actomyosin flows. Nature 545:103–107

Chapter 13

Optogenetic Methods to Control Tissue Mechanics in *Drosophila*

Daniel Krueger and Stefano De Renzis

Abstract

Optogenetics is a powerful technique that allows the control of protein function with high spatiotemporal precision using light. Here, we describe the application of this method to control tissue mechanics during *Drosophila* embryonic development. We detail optogenetic protocols to either increase or decrease cell contractility and analyze the interplay between cell–cell interaction, tissue geometry, and force transmission during gastrulation.

Key words Optogenetics, Tissue mechanics, Tissue morphogenesis, Actomyosin, Multiphoton, Cry2/CIB1, *Drosophila*, Gastrulation

1 Introduction

Tissue mechanics is of fundamental importance during animal development. It determines the three-dimensional (3D) organization of the body plan and it impacts on critical cell functions including proliferation, migration, and differentiation [1]. Tissue mechanics depends on forces generated by the contraction of cortical actomyosin filaments (i.e., filaments that are attached to the inner leaflet of the plasma membrane), which transmit across the tissue through adherens junctions [2–4]. The molecular pathways controlling cortical actomyosin assembly and contractility have been characterized and involve a complex interaction between plasma membrane phosphoinositides and Rho signaling, which in concert regulate actin polymerization, branching, cross-linking, and myosin II activation [5, 6]. A major challenge in studying the role of tissue mechanics in animal morphogenesis is the lack of methods to precisely modulate cell contractility with relevant spatiotemporal precision. Downregulation or overexpression of key pathway components perturbs protein function over longer timescales, often leading to secondary effects and cascade

Christian Dahmann (ed.), *Drosophila: Methods and Protocols*, Methods in Molecular Biology, vol. 2540,
https://doi.org/10.1007/978-1-0716-2541-5_13,

failure that are difficult to interpret [7]. Furthermore, genetic perturbations, only in rare cases, offer the possibility to perturb protein function in individual cells at will. Laser cutting of actin filaments has emerged as a powerful approach to acutely inhibit contractility [8–10]; however, this is an invasive technique, which can lead to undesired side effects and allow the experimentalist to inhibit contractility but not stimulate it. To circumvent these limitations, we developed optogenetic tools that enable us to specifically induce or inhibit cell contractility in embryonic tissues with subcellular and minute-scale precision. On the one hand, by controlling the recruitment of the 5-phosphatase OCRL [11] to the plasma membrane with light, we could induce cortical actin depolymerization and inhibition of contractility [12, 13] (*see* Figs. 1 and 2). On the other hand, by controlling the plasma

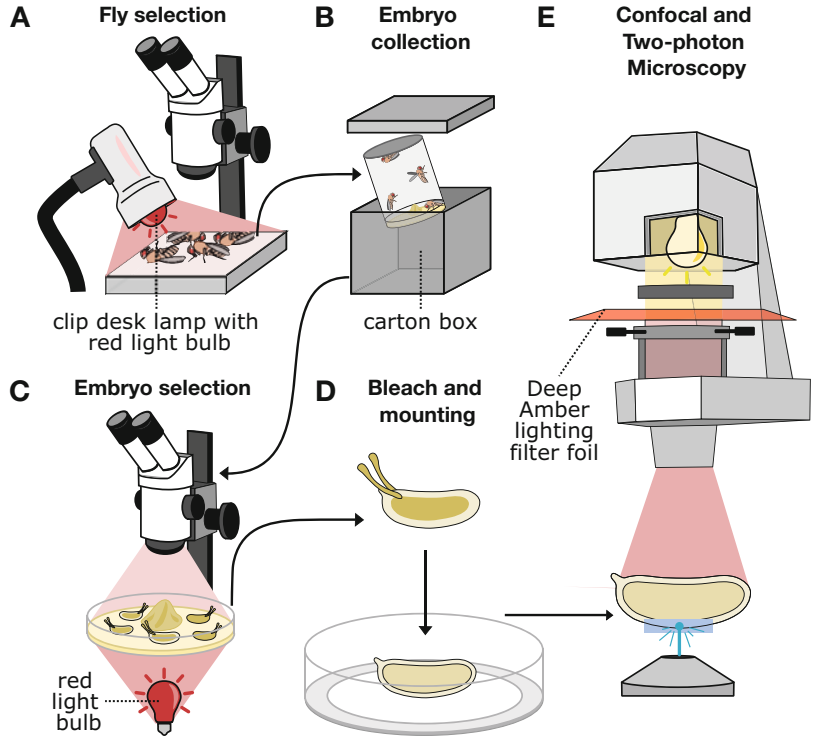


Fig. 1 Schematic of sample preparation for optogenetic experiments in fly embryos. Samples are prepared in the dark to avoid photoactivation. As only blue light excites the Cry2/CIB1 optogenetic system, red light sources can be used to handle the samples. (**a**) Selection of parental flies is done under a stereo microscope using a commercial clip desk lamp with a red light bulb. (**b**) Embryo collection cages are kept in optically opaque carton boxes, and the agar plates are changed regularly. (**c**) Embryo collection and mounting should be performed strictly in the dark. It is possible to replace white bulbs with yellow or red ones to operate a regular dissection stereo microscope. (**d**) Selected embryos are dechorionated by chemical bleach and mounted on a glass bottom dish. (**e**) Place the microscopy dish on the stage of an inverted confocal microscope. Inserting a red foil such as an amber lighting filter paper into the light path makes it possible to use filtered transmission light to position the sample

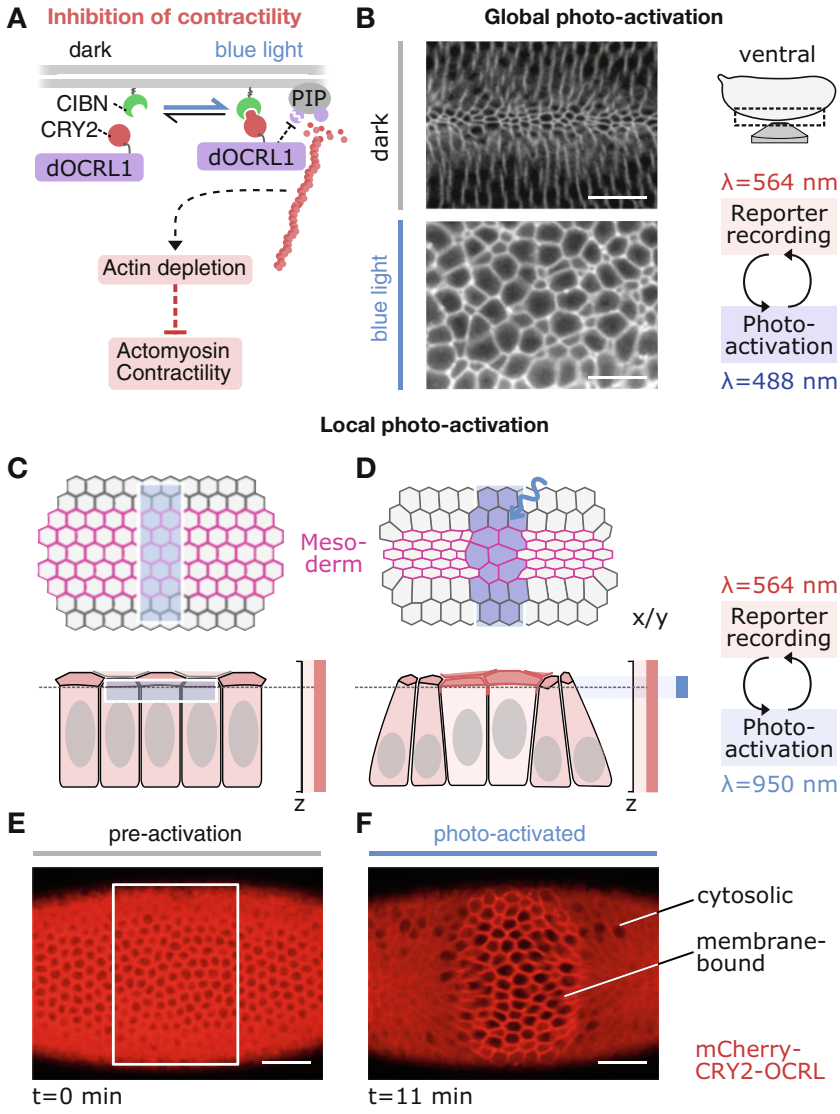


Fig. 2 Optogenetic activation of the inositol polyphosphate 5-phosphatase OCRL1 to inhibit cell contractility. (**a**) Illustration demonstrating the mechanisms of action of Cry2::OCRL. While Cry2::OCRL is cytoplasmic in the dark, blue light illumination induces the interaction of Cry2 with CIBN, which is anchored in the plasma membrane, causing membrane translocation of Cry2::OCRL. In the plasma membrane, OCRL converts PI(4,5) P_2 into phosphatidylinositol-4-phosphate (PI(4)P) that leads to actin depolymerization at the cell cortex, compromising cell contractility. (**b**) Still frames showing a control *Drosophila* embryo mounted with the ventral side facing the objective (schematics on the right) expressing only the plasma membrane anchor CIBN::GFP without Cry2::OCRL. (Images reproduced from Fig. 2 in Ref. [12] with permission from Cell Press). Ventral furrow invaginates normally (top). Global illumination with 488 nm light of an embryo co-expressing Cry2::OCRL causes inhibition of apical constriction and ventral furrow invagination (bottom). The embryo is shown 25 min after the onset of gastrulation. Schematic in the lower right shows the illumination scheme alternating between reporter recording at 564 m and photoactivation at 488 nm. (**c**, **d**) Schematic showing the setup of local photoactivation of ventral cells (magenta). Cells are shown from top

membrane recruitment of the Rho GTPase activating factor RhoGEF2 with light, we could induce myosin II activation and cell contraction (*see* Fig. 3) [14–17]. In the following paragraphs, we will detail how these two systems can be used to modulate cell contractility and gastrulation movements (ventral furrow formation). Both systems are based on the cryptochrome 2 (Cry2)/calcium- and integrin-binding protein 1 (CIB1) protein heterodimerization optogenetic module [18] and are composed of the plasma membrane anchor CIBN::pmGFP and the PHR domain of Cry2 tagged with mCherry fused with either the catalytic domain of the 5-phosphatase OCRL (mCh::Cry2::5-ptaseOCRL) or with the DHPH domain of RhoGEF2 (mCh::Cry2::RhoGEF2). Embryos co-expressing the two components can be photoactivated by blue light illumination, which causes a conformational change in Cry2 that allows its interaction with CIB1. Fluorescent tags are not required for optogenetic activation but are useful for visualizing the plasma membrane and for validating light-induced recruitment.

At ~5 hours post-fertilization (at 22 °C), the *Drosophila* embryo is composed of ~6000 cells that form a single-layered epithelium. Gastrulation begins with the upregulation of actomyosin activity and apical constriction in a stripe of mesodermal cells along the anterior–posterior axis on the ventral side of the embryo [19]. As cells undergo apical constriction and transform from a columnar into a conic shape, the tissue folds and invaginates within a period of ~15 min [20] (*see* Fig. 2). In this protocol, we describe optogenetic tools to either perturb endogenous apical constriction on the ventral side of the embryo or to induce ectopic tissue folds on the dorsal side with high spatiotemporal control (*see* Figs. 2 and 3).

Cry2/CIB1 interaction is triggered by blue light illumination [18]. Therefore, one-photon illumination at 488 nm efficiently induces photoactivation. However, due to the high sensitivity and kinetics of the Cry2/CIB1 interaction, light scattering does not allow localized activation patterns with cellular precision. To overcome this limitation, we developed a two-photon-based microcopy protocol to achieve high spatial control at the subcellular level

Fig. 2 (continued) view (top) and in cross section (bottom) before (**c**) and after photoactivation (**d**). The region of activation is centered at the apical surface (blue region), whereas the reporter signal can be recorded for the entire cell volume using a 564 nm laser. Within the tissue, a region of interest (blue box) is defined and illuminated with two-photon light at 950 nm. Photoactivation is alternated with acquisition of the reporter signal (illumination scheme on the right). Photoactivated cells (highlighted in purple) do not constrict the apical surface and fail to invaginate. (**e**, **f**) Still frames showing mCherry::CRY2::OCRL in a *Drosophila* embryo co-expressing CIBN::GFP. A region of photoactivation (white box) was defined in which cells were photoactivated using 950 nm two-photon light (**e**). Within the activation region, mCherry::CRY2::OCRL translocated to the plasma membrane (**f**). While the nonactivated cells constricted and internalized, the activated cells failed to constrict and did not invaginate. (**b**, **e**, **f**) Scale bars: 25 μm

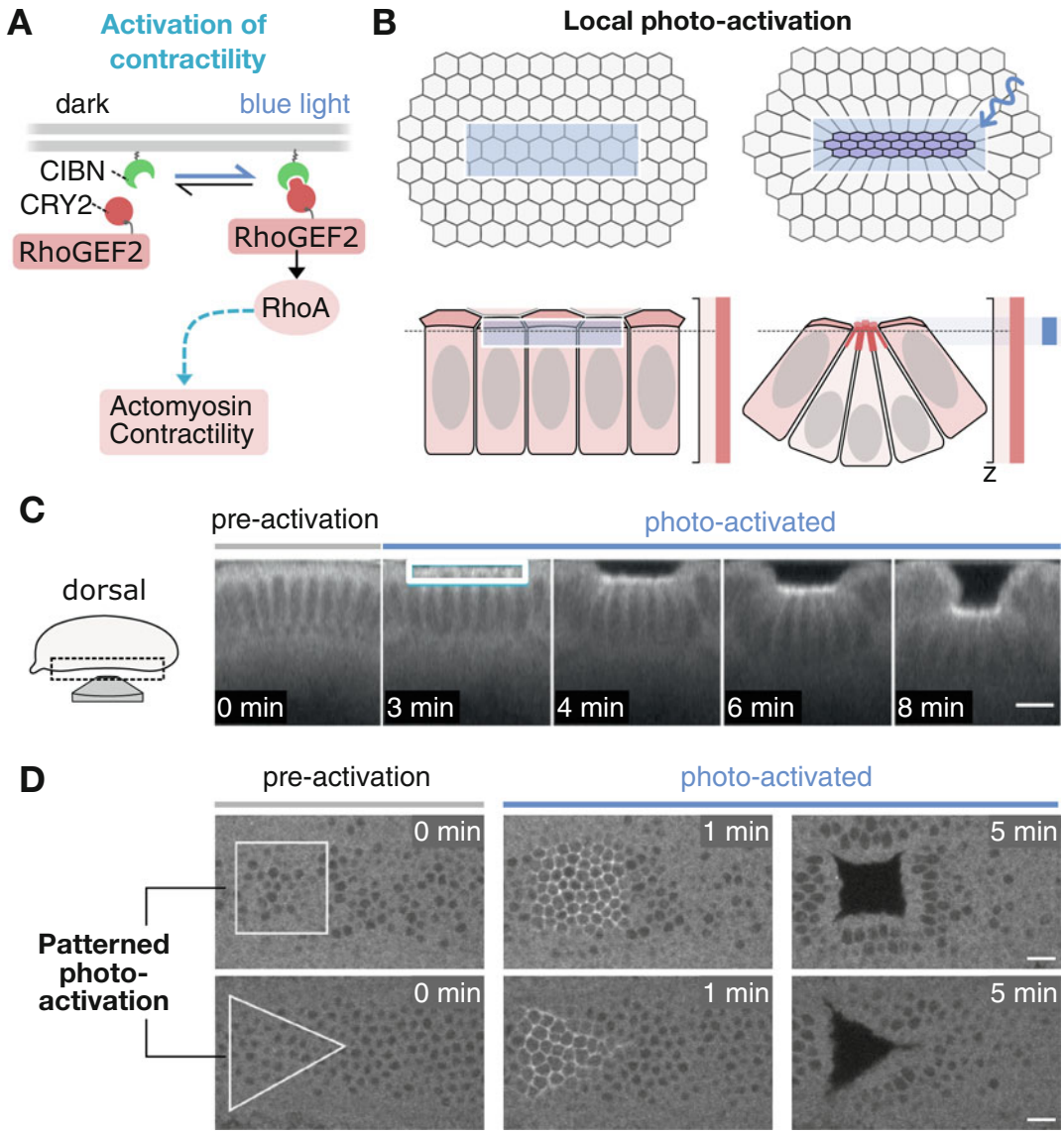


Fig. 3 Optogenetic induction of cell contractility through Rho signaling stimulation. (**a**) Illustration demonstrating the work mode of Cry2::RhoGEF2. In the dark, Cry2::RhoGEF2 is present in the cytoplasm. Blue light illumination causes its translocation to the plasma membrane through light-induced interaction between Cry2 and membrane-anchored CIBN. In the plasma membrane, RhoGEF2 activates Rho1 that starts a signaling cascade, resulting in the activation of myosin II that induces cell contractility. (**b**) Schematic showing the setup of local photoactivation of dorsal cells. Cells are shown from top view (top) and in cross section (bottom). The region of activation is centered at the apical surface (blue box) and is illuminated with 950 nm light (blue region), whereas the reporter signal can be recorded for the entire cell volume using a 564 nm laser. Photoactivated cells (highlighted in purple) constrict their apical surface and change their shape, inducing the formation of an ectopic invagination. (**c**, **d**) Still frames showing mCherry::Cry2::RhoGEF2 of an embryo co-expressing CIBN::GFP. The embryo was mounted with the dorsal surface facing the objective. (Images reproduced from Fig. 1 and Fig. 6 in Ref. [14] with permission from Nature Research). (**c**) Cells are shown in

[12]. Two-photon activation of Cry2 requires the simultaneous absorption of two 950 nm photons (lower-energy photons), which is extremely unlikely outside of the focused laser beam and thus restricts photoactivation to a small focal volume. Due to the slower kinetics of photoactivation using two-photon illumination, it is necessary to optimize photoactivation parameters to achieve levels of optogenetic activation comparable to those obtained with one-photon illumination (*see* Subheading 3.3).

2 Materials

2.1 Sample Preparation

1. Apple juice plates: Mix 22.5 g agar in 750 ml distilled water. In a 2-L Erlenmeyer flask, mix 25 g sucrose in 250 ml apple juice and stir until the sucrose is dissolved. Autoclave both solutions for 20 min. Gently pour the agar solution into the apple juice while stirring. Dissolve 1.5 g Nipagin (methyl-4-hydroxybenzoate) in the solution. Pour the solution into $35 \times 10\text{mm}^2$ dishes and allow it to cool down. Store the plates at 4 °C until use.
2. Yeast paste: Suspend 10 g dry yeast in distilled water. Store at 4 °C for up to 3 days.
3. Custom-made plastic embryo collection cage with a metal grid top.
4. Dissection microscope-equipped commercial clip desk lamp with a red light bulb as the light source.
5. Halocarbon oil 27.
6. 100% sodium hypochlorite solution.
7. Inox tweezers.
8. Filter paper with 9 cm diameter.
9. Double distilled water (ddH_2O).
10. Glass bottom dish of 35 mm (MatTek Corporation) with 0.16–0.19 mm coverslip thickness.
11. Phosphate-buffered saline (PBS) 1×, pH: 7.4.

Fig. 3 (continued) cross section before photoactivation and at different times after photoactivation of the apical cell surface (white box). Within the photoactivation region, mCherry::Cry2::RhoGEF2 translocates to the plasma membrane, causing apical constriction, cell shape changes, and an ectopic tissue invagination occurs. (**d**) Cells are shown from an apical top view before photoactivation, at 1 and 5 min after photoactivation of the apical cell surface. By varying the geometry of the region of photoactivation (white lined shapes), cells can be activated in defined patterns. Only within the region of photoactivation, mCherry::Cry2::RhoGEF2 translocates to the plasma membrane. After 5 min, photoactivated cells invaginate. (**c**, **d**) Scale bars: 10 μm

2.2 Microscopy

1. Confocal laser scanning microscope equipped with a multiline argon laser (458/488/514 nm), 561 nm He–Ne laser, and a femtosecond (140 fs) pulsed multiphoton laser with a 80 MHz repetition rate.
2. 40× C-Apochromat (NA 1.20) water immersion objective.
3. Double distilled water (immersion).
4. Deep amber lighting filter 104.
5. ZEN Black software for image acquisition.
6. Pipeline Constructor macro for sequential imaging protocols [21].

2.3 Fly Stocks

Fly lines expressing optogenetic transgenes were generated by microinjection of purified P-element-containing plasmids into eggs of w118 flies (harboring mutation in the *white* gene (w[*])). The following fly lines are needed for the described experiments and are available upon request to the authors:

1. w[*]; If/CyO; P[w+, UASp>5-ptase-OCRL::CRY2::mCherry]/TM3, Ser. Upstream activation sequence (UAS)-Gal4-driven 5-phosphatase domain of *Drosophila* OCRL1 fused with the photosensitive PHR domain of *Arabidopsis thaliana* cryptochrome 2 (CRY2) and the red fluorescent protein mCherry.
2. w[*]; If/CyO; P[w+, UASp> mCherry::Cry2::RhoGEF2]/TM3, Ser. UAS-Gal4-driven active DHPH domain of *Drosophila* RhoGEF2 fused with the photosensitive PHR domain of *Arabidopsis thaliana* cryptochrome 2 (CRY2) and the red fluorescent protein mCherry.
3. w[*]; P[w+, UASp>CIBN::pmGFP]/Cyo; Sb/TM3, Ser. UAS-Gal4-driven N-terminal domain of *Arabidopsis thaliana* CIB1 (CIBN) fused with the enhanced green fluorescent protein (EGFP) and the prenylation CAAX box signal peptide that serves as a plasma membrane tag.
4. w[*]; If/CyO; P[w+, Oskp>Gal4::VP16]/TM3, Ser. Gal4 transcription factor expressed under the maternally deposited *Oskar* promoter (Bloomington stock number 44242).
5. w[*]; P[w+, mat.tubulin>Gal4::VP16]; P[w+, mat.tubulin>Gal4::VP16]. Gal4 transcription factor expressed under the maternally deposited maternal tubulin promoter (Bloomington stock number 7062–7063).

3 Methods

3.1 Fly Crosses

1. Obtain flies that carry the plasma membrane anchor CIBN and the photosensitive protein CRY2 coupled to either (A) the inositol polyphosphate *5-phosphatase OCRL1 (*5-ptase-OCRL*)* or (B) the Rho signaling activator RhoGEF2. To this end, cross virgin females carrying the CIBN::pmGFP transgene on the second chromosome to males carrying either the (A) mCherry::Cry2::5-ptaseOCRL or (B) mCherry::Cry2:: RhoGEF transgene on the third chromosome (*see* **Note 1**):

   (A)
   $$\text{♀}\frac{\text{w}}{\text{w}};\ \frac{CIBN::pmGFP}{CyO};\ \frac{Sb}{TM3,Ser}\ (\text{X})\ \text{♂}\frac{\text{w}}{<};\ \frac{If}{CyO};\ \frac{5-ptase-OCRL::CRY2::mCherry}{TM3,Ser}$$

   (B)
   $$\text{♀}\frac{\text{w}}{\text{w}};\ \frac{CIBN::pmGFP}{CyO};\ \frac{Sb}{TM3,Ser}\ (\text{X})\text{♂}\frac{\text{w}}{<};\ \frac{If}{CyO};\ \frac{mCherry::Cry2::RhoGEF2}{TM6,Tb,Hu}$$

2. Keep the vials containing the crosses in the dark by placing them in an air-permeable paper box at 22 °C.
3. Males carrying both transgenes are crossed to virgin females expressing the maternally deposited Gal4 transcription factor (A) under the maternal tubulin promoter or (B) under the *Oskar* promoter (*see* **Note 2**):

   (C)
   $$\text{♀}\frac{\text{w}}{\text{w}};\ \frac{mat\ tub>Gal4}{mat\ tub>Gal4}\ (X)\text{♂}\frac{\text{w}}{<};\ \frac{CIBN::pmGFP}{CyO};\ \frac{5-ptase-OCRL::CRY2::mCherry}{TM3,Ser}$$
   $$\rightarrow \text{♀♂}\frac{\text{w}}{\text{w}/<};\ \frac{CIBN::pmGFP}{mat\ tub>Gal4};\ \frac{5-ptase-OCRL::CRY2::mCherry}{mat\ tub>Gal4}$$

   (D)
   $$\text{♀}\frac{\text{w}}{\text{w}};\ \frac{If}{Cyo};\ \frac{Osk>Gal4}{TM3,Ser}\ (\text{X})\text{♂}\frac{\text{w}}{<};\ \frac{CIBN::pmGFP}{CyO};\ \frac{mCherry::Cry2::RhoGEF2}{TM3,Ser}$$
   $$\rightarrow \text{♀♂}\frac{\text{w}}{\text{w}/<};\ \frac{CIBN::pmGFP}{Osk>Gal4};\ \frac{mCh::Cry2::RhoGEF2}{Gal4}$$

4. Keep the vials containing the crosses in the dark by placing them in an air-permeable paper box at 22 °C (*see* **Note 3**).
5. Select flies with the correct set of phenotypic markers, and transfer them to an embryo collection cage (*see* **Note 3**). To avoid premature photoactivation of the optogenetic system, it is recommended to select flies in the dark using a stereo microscope that is either equipped with a red light-emitting light

source or by providing the light externally using a commercial clip desk lamp with a red light bulb (*see* Fig. 1a).

6. Spread the center of an apple juice plate with a ~ 3 mm blob of yeast paste, and add it to the collection cage.
7. Keep the collection cages in the dark by placing them in an air-permeable carton box, and incubate them (A) at 18 °C or (B) at 22 °C (*see* Fig. 1b, **Note 3**). Change the apple juice plate twice a day. Flies can be used for experiments after an adaptation phase of 3 days and for about 10 days.

3.2 Embryo Collection and Mounting

Perform all the following procedures in a darkened room that can be illuminated by a red light-emitting lamp and at 20 °C (e.g., a microscope room).

1. Put a fresh apple juice plate supplemented with yeast paste onto the embryo collection cage, transfer it back to the carton box, and incubate it (A) at 18 °C or (B) at 22 °C for 5 h (*see* **Note 4**).
2. Collect the apple juice plate, and remove the yeast paste with a cotton swab.
3. Cover the plate with a thin layer of halocarbon oil 27.
4. Using an upright dissection microscope, screen the embryos on the plate and select stage five embryos at the end of cellularization (*see* Fig. 1c, **Note 5**). In the halocarbon oil, embryos at this stage display a dark core of yolk that is surrounded by an optically clear cortical layer of cytoplasm, a feature that makes them easily identifiable. To avoid photoactivation, replace the microscope's light source by an external lamp with a red light bulb (*see* **Note 6**).
5. Collect the embryos at the right stages gently using metal tweezers.
6. Transfer the embryos carefully onto a piece of Whatman filter paper to soak the bulk of halocarbon oil.
7. Transfer the embryos carefully onto a second, small (~2 cm^2) piece of Whatman filter paper and move the filter paper to soak it in a drop (250 μl) of 100% sodium hypochlorite. Incubate for 1 min (*see* Fig. 1d and **Note 7**).
8. Wash the embryos on top of the filter paper by dipping the filter paper into a drop of ddH_2O. Repeat with clean water drops 2 times, and dab the filter paper on a kitchen roll to remove excess of water.
9. Transfer the embryos to a 35 mm glass bottom microscopy dish using a brush or a gel-loading pipette tip.
10. Cover the microscopy dish with 2 mL of PBS.
11. Gently position the embryos in such a way that their (A) ventral side or (B) dorsal side is facing the glass bottom (*see* Fig. 1d).

3.3 Live Imaging

All imaging procedures are based on protocols developed for a Zeiss LSM780 NLO confocal laser scanning microscope equipped with a 561 nm He–Ne laser diode and a femtosecond (140 fs, at a repetition rate of 80 MHz) pulsed multiphoton laser (Chameleon, Coherent Inc.).

Live reporter acquisition was performed using a 561 nm laser, which excites mCherry and does not cause photoactivation. Actual photoactivation can be stimulated with either one-photon (458–488 nm) or two-photon (900–950 nm) illumination. In the following experimental settings, we describe three procedures: global, whole embryo activation of the *5-phosphatase OCRL1* using single-photon activation (Subheading 3.4), local activation of the *5-phosphatase OCRL1* (Subheading 3.5) and of RhoGEF2 (Subheading 3.6) using two-photon activation.

1. Place the microscopy dish on the stage of the inverted confocal microscope.
2. Position the sample in the center of view. Transmission light can be used without photoactivating the sample by inserting a sheet of amber lighting filter foil above the condenser centering that efficiently filters the transmission light source (*see* Fig. 1e, **Note 8**).
3. Use a C-Apochromat 40×/NA 1.2 water immersion objective.
4. Use the 561 nm laser exciting the mCherry reporter to orientate the sample.

3.4 Global Photoactivation of the 5-Phosphatase OCRL1

Photoactivation with one-photon illumination can be achieved with classical laser lines using light at $\lambda = 458$, 488, and 514 nm. In the early *Drosophila* embryo, 488 nm has been proved to be the most effective wavelength for triggering the translocation of Cry2 fusion proteins to the plasma membrane where CIBN::pmGFP is anchored [12, 13] (*see* Fig. 2a, b).

1. Set up the optogenetic experiment like a conventional dual-color imaging experiment with two separated tracks. The first track uses the 561 nm laser for mCherry reporter acquisition, and the second track uses the 488 nm laser for photoactivation. Switch tracks between z-stacks (*see* **Note 9**).
2. Optimize the imaging parameters for the mCherry reporter signal.
3. Set a z-stack spanning 5 μm from the apical surface with a 1-μm interval.
4. Set the parameters for one-photon activation using a continuous wave argon laser at $\lambda = 488$ nm, laser power: 6.9 μW (measured 1 cm from the objective) at a scanning speed of 1×10^{-6} s/pixel and a line averaging of 4 in a bidirectional scanning mode, which corresponds to a scan time of about 2.1 ms per cell and 1 s for the entire embryo (*see* **Note 10**).

5. Follow the development of the embryo using transmitted light or the mCherry reporter signal. Start the experiment ~5 min before ventral cells would normally begin to constrict in order to inhibit apical constriction.
6. Run a time-lapse imaging experiment for up to 30 min with a time interval of 30 s. The recording of the mCherry-tagged 5-ptase-OCRL at 561 nm will be followed by photoactivation using 488 nm light (*see* Fig. 2b, right). Due to scattering at the defined laser power, the entire embryo will be activated.

3.5 Local Photoactivation of the 5-Phosphatase OCRL1

Local photoactivation with cellular precision can be achieved with two-photon illumination using light between $\lambda = 900$ and 950 nm. In the early *Drosophila* embryo, 950 nm has proven to be the most effective wavelength for triggering the plasma membrane translocation of Cry2 fusion proteins [12] (*see* Figs. 2c–f and **Note 11**).

1. As this experiment requires inhomogeneous acquisition settings between the reporter acquisition and photoactivation mode, we use the Pipeline Constructor macro to alternate between different predefined acquisition modes. Newer versions of the Zeiss-owned ZEN software contain the ZEN Module Experiment Designer that can be used accordingly (*see* **Note 12**).
2. Optimize the imaging parameters, field of view, and resolution of the mCherry acquisition at 561 nm and define a z-stack (*see* Fig. 2c, d). Save the settings as reporter acquisition protocol (job) in the Pipeline Constructor macro.
3. Set a z-stack of 5 μm from the apical surface with 1 μm interval with the following parameters for two-photon activation at 950 nm: laser power: 3 mW (measured 1 cm from the objective), scanning speed of 1×10^{-6} s/pixel, line averaging of 2 in a bidirectional mode, which corresponds to a scan time of 500 ms per z-stack. Repeat the photoactivation for 8 consecutive repetitions, resulting in a total activation time of ~2.5 s (*see* **Note 10**). Save the settings as photoactivation protocol (job) in the Pipeline Constructor macro. Be aware that to prevent premature photoactivation, the 950 nm laser cannot be used while defining the settings (*see* **Note 9**).
4. Define the region that contains the cells that will be photoactivated, and find the apical cell surface using the 561 nm laser. Save the settings as photoactivation protocol.
5. Follow the development of the embryo using transmitted light or the mCherry reporter signal. Start the experiment ~5 min before ventral cells would normally begin to constrict in order to inhibit apical constriction (*see* Fig. 2e).

6. Run in alternating cycles the reporter acquisition and photoactivation protocol with a time interval of 30 s (*see* Fig. 2d). Acquire an initial mCherry stack before photoactivation.

3.6 Local Photoactivation of RhoGEF2

Local photoactivation of mCherry::Cry2::RhoGEF2 follows the same general workflow as described above in Subheading 3.5, and, due to its high photosensitivity, it only differs in the photoactivation parameters (*see* Figs. 3a–d and **Note 11**).

1. Load the Pipeline Constructor macro in addition to the ZEN software.
2. Optimize the imaging parameters, field of view, and resolution of the mCherry acquisition at 561 nm and define a z-stack. Save the settings as reporter acquisition protocol (job) in the Pipeline Constructor macro.
3. Set a z-stack with 3 slices, centered at 4 μm from the apical surface (*see* **Note 13**) with 1 μm interval with the following parameters for two-photon activation at 950 nm excitation: laser power: 10 mW (measured 1 cm from the objective), a pixel size of 300 nm, and a pixel dwell time of 1.27 μs in a bidirectional mode. Repeat the photoactivation for 8 consecutive repetitions, resulting in a total activation time of ~2 s (*see* **Note 10**). Save the settings as photoactivation protocol (job) in the Pipeline Constructor macro (*see* **Note 9**).
4. Define a rectangular region that will be photoactivated using the "Region" tool, and find the position of the apical cell surface using the 561 nm laser (*see* Fig. 3b). The activation region can be arbitrarily defined to probe the relationship between tissue geometry, apical constriction, and morphogenetic output (*see* Fig. 3d). Store the region of interest in the photoactivation protocol.
5. Run in alternating cycles the reporter acquisition and photoactivation protocol (*see* Fig. 2d, right). Acquire an initial mCherry stack before photoactivation.
6. As the photoactivated cells constrict their apical surface and invaginate (*see* Fig. 2c), manually shift the center z-position of the activation stack downward by entering a z-offset in the macro (*see* **Note 14**).

4 Notes

1. Maintain individual optogenetic fly stocks separately, and do not keep stocks expressing transgenes permanently. This can result in reduced expression and efficacy over time.

2. Expression levels of the different components must be optimized. Weak expression will result in weak photoactivation, and too high expressions will result in dark-state activity (effector protein activity also in the dark). The maternal *tubulin-Gal4* driver is stronger than the *Oskar-Gal4* driver. Although it is appropriate for expression of the 5-phosphatase OCRL, it is not appropriate for expression of the RhoGEF2 optogenetic module.
3. Keep vials containing the fly crosses in the dark by placing them in a paper box, and incubate at 22 °C to avoid unwanted activation (*see* Fig. 1b).
4. When using the Gal4 driver, the expression of transgenes is temperature-sensitive. With the maternal *tubulin-Gal4*, the expression is higher at 18 °C. If the expression is too high the cages can be switched to room temperature, i.e., ~22 °C.
5. Select embryos with a normal morphology and avoid embryos with an abnormal shape. Premature activation of the optogenetic system during oogenesis or embryos that stochastically express extremely high levels of the optogenetic transgenes can cause morphological abnormalities (e.g., short embryos).
6. Embryo collection and mounting should be performed strictly in the dark (*see* Fig. 1c). Even dim ambient light can result in unwanted photoactivation. In order to have a safe source of light, it is possible to replace white bulbs with yellow or red ones or cover light sources with amber lighting filters.
7. Avoid keeping the embryos in bleach for more than 1 min or letting them dry out for too long.
8. Blue light-filtered transmission light of a confocal microscope can be used to position the embryo by inserting a red foil such as the amber lighting filter paper (that can be bought at a low cost) into the light path, e.g., on top of the condenser centering (*see* Fig. 1e). Many filter foils from do-it-yourself stores are efficient enough to block blue light.
9. It is highly recommended to tune down the laser power of the 488 nm/950 nm laser to 0%, while setting up the experiment to prevent accidental photoactivation.
10. Efficient photoactivation has to be optimized, and the exact settings can vary between different microscopes. The most important parameters that affect photoactivation are laser power, number of repetitions, line averaging, resolution and pixel size, and imaging speed, i.e., dwell time, the size of the activated volume, its z-interval and number of slices, and the time interval between consecutive photoactivations. As a starting point, we advise using a low pixel size resolution (512 × 512) and rather increase the line averaging or number of repetitions than decreasing the scanning speed (low dwell

time). The most important parameter is laser power. When optimizing laser power, start with a fairly high power and gradually decrease the power to a minimal value that produces reproducible effects. Photoactivation at deeper z-positions requires higher laser power.

11. Localized photoactivation (e.g., one half of the embryo) can also be achieved by single-photon activation at 488 nm [22]. However, due to light scattering, it is not possible to achieve precise cellular resolution (in the *x/y* dimension); instead, a gradient of activation around the defined photoactivation region will be induced. The applied laser power is 1–5% of the power applied for the described global activation (~0.35 μW, measured 1 cm from the objective). The point scanning laser is a focused cone of light that penetrates all optical planes between the focal point and the objective. Therefore, this activation mode can only be used to photoactivate the apical-most surface of the embryo; otherwise, photoactivation of the entire cell volume will occur.
12. Localized photoactivation can also be achieved by fluorescence recovery after photobleaching (FRAP) or laser ablation microscopy settings using 488 or 950 nm light, respectively. Optimize laser power and iteration numbers that are suitable for photoactivation and do not result in photobleaching.
13. Selective photoactivation of a volume within the depth of the tissue can also be achieved using multiphoton light at 950 nm. To do so, a z-stack at the desired z-position has to be defined (the width of the stack and the number of slices have to be optimized), and the center position has to be stored (using the "Position" tool). Save the settings as photoactivation protocol.
14. By alternating cycles of photoactivation and darkness (~1 min intervals), it is possible to trigger non-ratcheted pulsatile apical constrictions. However, under this condition, tissue invagination is not induced.

Acknowledgments

We thank all members of the De Renzis laboratory for helpful discussions throughout the years. This work was funded by the European Molecular Biology Laboratory.

References

1. Heisenberg CP, Bellaiche Y (2013) Forces in tissue morphogenesis and patterning. Cell 153(5):948–962
2. Lecuit T, Lenne PF, Munro E (2011) Force generation, transmission, and integration during cell and tissue morphogenesis. Annu Rev Cell Dev Biol 27:157–184
3. Martin AC et al (2010) Integration of contractile forces during tissue invagination. J Cell Biol 188(5):735–749
4. Koenderink GH, Paluch EK (2018) Architecture shapes contractility in actomyosin networks. Curr Opin Cell Biol 50:79–85
5. Ennomani H et al (2016) Architecture and connectivity govern actin network contractility. Curr Biol 26(5):616–626
6. Agarwal P, Zaidel-Bar R (2018) Principles of actomyosin regulation in vivo. Trends Cell Biol 29:150
7. Hartmann J, Krueger D, De Renzis S (2020) Using optogenetics to tackle systems-level questions of multicellular morphogenesis. Curr Opin Cell Biol 66:19–27
8. Mayer M et al (2010) Anisotropies in cortical tension reveal the physical basis of polarizing cortical flows. Nature 467(7315):617–621
9. Rauzi M, Lenne PF, Lecuit T (2010) Planar polarized actomyosin contractile flows control epithelial junction remodelling. Nature 468(7327):1110–1114
10. de Medeiros G et al (2020) Cell and tissue manipulation with ultrashort infrared laser pulses in light-sheet microscopy. Sci Rep 10(1):1942
11. Idevall-Hagren O et al (2012) Optogenetic control of phosphoinositide metabolism. Proc Natl Acad Sci U S A 109(35):E2316–E2323
12. Guglielmi G et al (2015) An optogenetic method to modulate cell contractility during tissue morphogenesis. Dev Cell 35(5): 646–660
13. Guglielmi G, De Renzis S (2017) Optogenetic inhibition of apical constriction during Drosophila embryonic development. Methods Cell Biol 139:167–186
14. Izquierdo E, Quinkler T, De Renzis S (2018) Guided morphogenesis through optogenetic activation of Rho signalling during early Drosophila embryogenesis. Nat Commun 9(1): 2366
15. Krueger D et al (2018) Downregulation of basal myosin-II is required for cell shape changes and tissue invagination. EMBO J 37: e100170
16. Krueger D et al (2019) Cross-linker-mediated regulation of actin network organization controls tissue morphogenesis. J Cell Biol 218(8): 2743–2761
17. Krueger D et al (2020) βH-spectrin is required for ratcheting apical pulsatile constrictions during tissue invagination. EMBO Rep 21(8): e49858
18. Kennedy MJ et al (2010) Rapid blue-light-mediated induction of protein interactions in living cells. Nat Methods 7(12):973–975
19. Sweeton D et al (1991) Gastrulation in Drosophila: the formation of the ventral furrow and posterior midgut invaginations. Development 112(3):775–789
20. Martin AC, Kaschube M, Wieschaus EF (2009) Pulsed contractions of an actin-myosin network drive apical constriction. Nature 457(7228):495–499
21. Politi AZ et al (2018) Quantitative mapping of fluorescently tagged cellular proteins using FCS-calibrated four-dimensional imaging. Nat Protoc 13(6):1445–1464
22. Deneke VE et al (2019) Self-organized nuclear positioning synchronizes the cell cycle in drosophila embryos. Cell 177(4):925–941 e17

Chapter 14

Optochemical Control of Cell Contractility in *Drosophila* Embryos

Deqing Kong and Jörg Großhans

Abstract

Cell shape changes based on actomyosin contractility provide a driving force in tissue morphogenesis. The temporally and spatially coordinated constrictions of many cells result in changes in tissue morphology. Given the networks of complex and mutual cellular interactions, the mechanisms underlying the emergence in tissue behavior are challenging to pinpoint. Important in the analysis of such interactions are novel methods for noninvasive interference with single-cell resolution and sub-minute timescale temporal control. Here we characterize an optochemical approach of Ca^{2+} uncaging to control cell contractility in *Drosophila* embryos. We describe in detail the method of sample preparation, microinjection, Ca^{2+} uncaging, and data analysis.

Key words Optochemical, Ca^{2+}, Cell contractility, Epithelia, Morphogenesis, Live imaging

1 Introduction

The dynamic behaviors of cells, including cell shape changes, cell migration, cell proliferation, and cell rearrangement, drive tissue morphogenesis [1–4]. The mechanical forces generated by the contractile actomyosin cytoskeleton play a pivotal role in these types of cellular behaviors [5, 6]. Much less understood is how this set of cell behaviors leads to emergent tissue morphology. In addition to global interference affecting all cells of the tissue, such as in mutants, spatiotemporally controlled noninvasive interference is needed to dissect the mutual interactions between cells. For example, to understand how apical constriction of cells leads to a groove in the tissue, ideally, apical constriction is induced or inhibited in single cells, groups of cells, or cells selected by a chosen pattern. The kinetics of interference should be within the timescale of the process, at least.

The recent development of optogenetic and optochemical technologies has provided these long-sought-after possibilities to

Christian Dahmann (ed.), *Drosophila: Methods and Protocols*,
Methods in Molecular Biology, vol. 2540, https://doi.org/10.1007/978-1-0716-2541-5_14,

control cellular processes with high spatiotemporal resolution [7, 8]. Combined with live imaging and cell tracking, those novel noninvasive approaches offer avenues to dissect mutual cellular interactions in a highly spatial and temporal manner, e.g., how does a cell react to an adjacent contracting cell to finally understand tissue morphogenesis?

Intracellular calcium ions (Ca^{2+}) have been implicated in controlling cell contractility for a long time, not only in muscle cells [9] but also in mesenchymal and epithelial cells. In striated muscle cells, Ca^{2+} triggers contraction by binding the Ca^{2+}-binding protein troponin and releasing its inhibition of myosin. In smooth muscle [10] and nonmuscle myosin II [9], the myosin regulatory light chain is phosphorylated by a calcium–calmodulin-dependent myosin light-chain kinase. More recent reports have pointed to the role of Ca^{2+} in cell contraction in a variety of epithelial cells and tissues. Ca^{2+} has been implicated in contractility in cultured epithelial cells [11], in *Drosophila* amnioserosa cells during dorsal closure [12], during neural tube closure in chick embryos [13], in the folding morphogenesis of the neural plate in rat and chick embryos [14], and in the process of wound healing [15]. In *Drosophila* oogenesis, myosin II is activated by an induced intracellular increase of Ca^{2+} [16]. A transient increase in Ca^{2+} concentration is also associated with apical constriction in *Xenopus* neural tube cells [17]. The underlying mechanism for myosin activation has remained elusive.

An increase in intracellular Ca^{2+} can be induced noninvasively by so-called uncaging from a light-sensitive ethylene glycol tetraacetic acid (EGTA) derivative (*see* Fig. 1a). NP-EGTA is available in a form with linked fatty acid esters o-Nitrophenyl EGTA, AM (NP-EGTA, AM), which mediate membrane permeability. Inside the cell, the cage molecules are trapped because intracellular esterase cleaves the ester bonds of the fatty acid moieties. What is important is that NP-EGTA, AM is membrane permeable and does not need to be injected directly into cells. In neuroscience, optochemical approaches for the release of intracellular Ca^{2+} have been well established and widely employed for two decades [18, 19]. Despite its success, Ca^{2+} uncaging should not be applied to epithelial tissues wildly. Ca^{2+} has previously been implicated in several processes of actomyosin organization and activity in epithelial cells. Ca^{2+} uncaging offers a simple and spatiotemporally precise way to induce contractile activity. We have recently established a Ca^{2+} uncaging method employing a 355-nm pulsed UV laser to control epithelial cell contractility during tissue morphogenesis in *Drosophila* embryos with single-cell resolution [20]. Here we show that a continuous-wave (cw) laser at 375 nm can be used with similar efficiency and practicability (*see* Fig. 1). From a technical point of view, the 375-nm laser comes with several advantages.

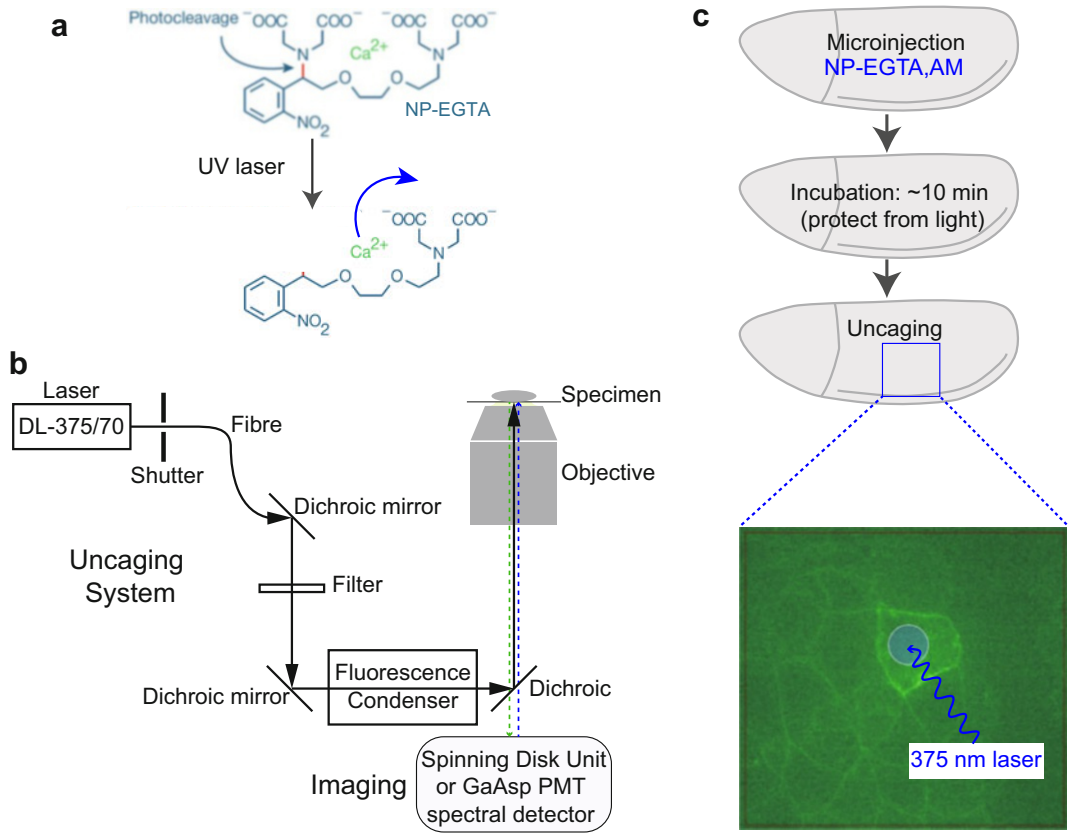


Fig. 1 Principle of optochemical control of cell contractility. (**a**) Structure of the cage molecule NP-EGTA. UV illumination cleaves the bond in red and releases Ca^{2+}. (**b**) Schematic drawing of the light path. DL-375/70 is a continuous-wave diode laser with 375 nm wavelength and 70 mW power DL-375/70 from Rapp OptoElectronic. **c.** Experimental scheme of the Ca^{2+} uncaging in *Drosophila* embryos. NP-EGTA, AM is microinjected into the embryos. In the cytoplasm, the cage is loaded with Ca^{2+}. A 375-nm UV laser-induced cleavage of NP-EGTA releases Ca^{2+} ("uncaging"), which can be visualized with a membrane-bound GCaMP6s fluorescent sensor, for example

Light transmission is possible by fiber optics compared to a 355-nm laser, which requires open transmission.

Furthermore, continuous-wave laser is less likely to induce ablation and tissue damage than pulsed laser due to its lower energy density. However, we have sporadically observed that Ca^{2+} is released after applying high doses of cw 375-nm laser even in the absence of NP-EGTA. It is conceivable that cw laser can cause damage to the Endoplasmic reticulum (ER), leading to the liberation of Ca^{2+} from an internal store. In any case, the intensity and exposure time of the UV laser applied to the target cell(s) has to be carefully titrated. The optimal dose will probably, in general, depend on tissue and cell type.

The origin of the released cytosolic Ca^{2+} as measured by a GCaMP sensor, for example, is unclear and needs to be studied

further. At least partially free Ca^{2+} originates from the opened EGTA cage. In addition, uncaged Ca^{2+} may trigger the opening of Ca^{2+}-gated Ca^{2+} channels, such as channels related to the ryanodine receptor or inositol 1,4,5-trisphosphate (IP3) receptors [21], which would lead to an influx from the extracellular space or intracellular stores such as the ER. The time traces of the GCaMP reporter show a peak only at about 3–6 s after uncaging. An immediate signal with a peak at 0 s would be expected for Ca^{2+} from the opened cage. The observed time lag may be due to the diffusion of the Ca^{2+} ions from the cytoplasm to the GCaMP sensor in the plasma membrane. In addition, the opening of Ca^{2+} channels in the ER or plasma membrane with a corresponding Ca^{2+} influx may contribute to the overall signal. The *Drosophila* genome encodes one RyR, which is also widely expressed in nonmuscle cells throughout development [22]. The RyR channel is a potential candidate for a mechanism enhancing the response to uncaged Ca^{2+}.

Some of our assays observed a gradual, monotonic dose-response relationship between the laser dose and the GCaMP signal. In some cases, our results suggest the presence of a threshold. At a low laser dose (20% intensity), the GCaMP signal remained unchanged after laser illumination and the cell did not constrict. The molecular events for uncaging are expected to be linearly dependent on the laser dose. Instead of 40%, 20% should uncage half of the molecules. A threshold effect may thus hint at the possibility that additional processes may be triggered, such as positive feedback loops involving the RyR Ca^{2+}-dependent Ca^{2+} channel, for example.

The noninvasive optochemical method of Ca^{2+} complements the recently developed toolbox for optogenetic interference to control selected single or multiple cell behavior with high spatial and temporal resolution. Future experiments will reveal to which cell types and species Ca^{2+} uncaging can be applied as a noninvasive optochemical method.

The protocols described here are an optochemical method to control cell contractility in *Drosophila* embryos with examples in the columnar epithelium (lateral epidermis) and squamous epithelium (amnioserosa) via Ca^{2+} uncaging in single cells and tissues (*see* Fig. 1c). Ca^{2+} cages are loaded into the embryos via microinjection. Ca^{2+} uncaging is performed by the continuous-wave diode laser with 375-nm wavelength mounted on the epiport of an inverted spinning disk confocal microscope or an inverted laser scanning confocal microscope. All the experiments are carried out at room temperature (≈22 °C). Here, we describe microinjection needle preparation, preparation of flies and embryos, microinjection, Ca^{2+} uncaging, and data analysis.

2 Materials

2.1 Sample Preparation

1. Stereomicroscope (for the alignment of embryos).
2. Fly cages.
3. Apple juice agar plates.
4. Yeast paste.
5. Glass slides.
6. Coverslips (24×24 mm^2).
7. Mounted needle for embryo alignment.
8. Brush.
9. Paper towel.
10. Basket 200 mesh, stainless steel.
11. 50% hypochlorite bleach: dilute the hypochlorite bleach with deionized water to 50%.
12. Deionized water in a squirt bottle.
13. Heptane glue: Soak a strip of sticky tape (1 m × 30 cm) in 50 mL heptane solution in a shaker overnight. Transfer the heptane to a falcon tube, and centrifuge at 10,000 rpm for 1 hour. Transfer the supernatant to a new tube and centrifuge at 10,000 rpm for 1 hour. The supernatant is ready to use.
14. Voltalef oil 10S.
15. Ophthalmic forceps.
16. Double-sided adhesive tape.
17. Coverslip holder.
18. Desiccation chamber.
19. Petri dish.

2.2 Drosophila Strains

1. E-cadherin-GFP (E-CadGFP) [23].
2. UASt-GCaMP6-myr [24].
3. Mat-Gal4–67;15 (St. Johnston Laboratory/Cambridge).
4. Amnioserosa-Gal4 (w[1118]; P{w[+mW.hs] = GawB}c381, Bloomington).

2.3 Microinjection

1. Inverted microscope including objectives (×10 and ×20).
2. Microinjector.
3. Manual micromanipulator.
4. Capillary (length 100 mm, outside dimeter 1.0 mm, inside dimeter 0.75 mm, with filament).
5. Micropipette puller.
6. Micro loader.

7. Capillary grinder.
8. Injection solution: 180 mM NaCl, 10 mM HEPES [pH 7.2], 5 mM KCl, 1 mM $MgCl_2$.
9. NP-EGTA, AM solution: Dissolve 50 mg NP-EGTA, AM (Invitrogen) in 30 µl (60 µl) injection solution for 2 mM (1 mM) stock solutions. Prepare 5 µl aliquots and store at −20 °C (*see* **Note 2**).

2.4 Ca^{2+} Uncaging

1. A continuous-wave diode laser with 375-nm wavelength and 70-mW power DL-375/70 (Rapp OptoElectronic). This laser is controlled with an independent scan module (Rapp UGA-42 Caliburn) and linked to the microscope via the epiport (*see* Fig. 1b). Uncaging is performed under the "sequence" or "click-and-fire" mode on the "REO-SysCon-Zen" platform (Rapp OptoElectronic).

2.5 Imaging

1. Imaging is conducted with the normal optical path of the inverted spinning disk confocal microscope or the inverted laser scanning confocal microscope with objectives (×40, ×63, and ×100) (*see* Fig. 1b) in parallel to the UV laser system.

2.6 Software

1. ZEN (blue edition, Zeiss).
2. REO-SysCon-Zen (Rapp OptoElectronic).
3. ImageJ/Fiji [25].
4. Tissue analyzer [26].

3 Methods

3.1 Microinjection Needle Preparation

1. The program with HEAT 515, PULL 150 (*see* **Note 1**), VEL 080, and TIME 250 in needle puller P97 (Sutter Instrument) works well, for example.
2. Load a capillary and pull the needle by pressing the "PULL."
3. Open the needle by a capillary grinder or manually with ophthalmic forceps under a stereomicroscope. The microinjection needle is ready for loading.

3.2 Preparation of Flies

GCaMP6-myr is a membrane-bound variant of the genetically encoded Ca^{2+} sensor GCaMP6s tagged with a myristoylation site that presents the intracellular Ca^{2+} concentration. The fly embryos expressing GCaMP6-myr are used to visualize the increasing Ca^{2+} concentration.

The fly embryos expressing E-CadGFP, which serves to label cell outlines, are used to visualize the cell size change.

1. Experiments in columnar epithelial cells (lateral epidermis) (stage 7–8): Cross Mat-Gal4–67;15 with UASt-GCaMP6-myr. Collect the F1 progenies into a fly cage. In parallel, transfer the E-CadGFP flies into a second fly cage.
2. Experiments in squamous epithelial cells (amnioserosa) (stage 14): Collect female virgins of AS-Gal4 and male flies of UASt-GCaMP6-myr and transfer both into a fly cage. In parallel, transfer the E-CadGFP flies into a second fly cage.

3.3 Preparation of Embryos

1. Set up cages a few days ahead. Feed the flies well with yeast. Keep the fly cages in a dark and quiet place at room temperature. This will help synchronize the egg-laying.
2. Allow the flies to lay eggs on the agar plate with little yeast for the required length of time depending on the embryonic stage of interest (*see* **Note 3**). Let the flies lay eggs for 1.5 h, and age the embryos for 1.5 h for stage 7 embryos. Let the flies lay eggs for 2 h, and age the embryos for 12 h (16 h at 18 °C) for stage 14 embryos.
3. To dechorionate the embryos, add deionized water onto the apple juice plate and resuspend the embryos with a brush. Transfer the embryos to a net and wash them thoroughly with deionized water. Dip the net with embryos into 50% hypochlorite bleach and incubate for 90 s. Then, wash the embryos thoroughly with water. Dry with a paper towel. Transfer the embryos with a fine brush to a piece of apple juice agar on a slide.
4. Select the embryos at the stage of interest and line them up under a stereomicroscope using a fine needle. Align multiple embryos in a row, head toward tail, and orient the region of interest (ROI) upward (will be toward the coverslip and objective). Remove the remaining embryos from the agar.
5. Spread a drop of heptane glue on the coverslip and wait until heptane has completely evaporated. Place the coverslip in the middle of the agar piece and gently tilt the coverslip on top of the embryos. Touch with a brush on top and gently apply some pressure to ensure that embryos stick to the coverslip. Gently flip back the coverslip. The embryos should stick to it.
6. Fix the coverslip with a double-sided adhesive tape to a holder fitting with the microscope stage. Consider the orientation and make sure that the microinjection can be down at the middle of the embryos.
7. Place the embryos in a desiccation chamber for about 10 min. The exact time depends on the material and humidity (*see* **Note 4**).
8. Cover the embryos with Voltalef oil 10S (use as little as possible). The embryos are ready for injection.

3.4 Microinjection

1. Thaw an aliquot of NP-EGTA, AM solution: 2 mM for columnar (lateral ectoderm) epithelial cells (stage 7–8) and 1 mM for squamous (amnioserosa) epithelial cells (stage 14). Load 5 μl NP-EGTA, AM solution to an injection needle.
2. Place a clear slide with a drop of Voltalef oil 10S on an inverted microscope. Place the injection needle on the manual micromanipulator and dip the needle tip into the Voltalef oil 10S gently. Set the pressure of the micromanipulator: pre-pressure: very little material may exit the needle – injection pressure: slightly higher (*see* **Note 5**). Find the needle tip under the inverted microscope with ×10 objective and press the "clear" button to clear the air from the needle tip completely; you can see a drop of liquid coming from the needle.
3. Replace the empty slide with the slide with embryos on the inverted microscope..
4. Inject the NP-EGTA, AM solution into the embryo/yolk from the dorsal side at the middle of the stage 7 embryos and from the lateral side at the middle of the stage 14 embryos. To ensure a uniform distribution of NP-EGTA, AM, we recommend deposition of the reagent to the middle of the embryos. Alternatively, the reagent can be injected into the extracellular perivitelline space, which is more effective but requires more skills. In this case, you may inject from either side of each embryo.
5. Place the injected embryos in a Petri dish with a damp paper towel and incubate them in the dark for 10 min. Then the embryos are ready for Ca^{2+} uncaging.

3.5 Uncaging Induces a Rapid Ca^{2+} Burst

Intracellular Ca^{2+} levels can be visualized with a Ca^{2+} sensor. (We employed a membrane-bound variant of the genetically encoded Ca^{2+} sensor GCaMP6s tagged with a myristoylation site. An increase of the GCaMP signal indicated a transient increase of intracellular Ca^{2+} levels in the target cells and revealed whether uncaging was only restricted to the target cells (*see* Fig. 2).)

In general, the uncaging experiments can be performed on a spinning disk microscope with a 375 laser for releasing Ca^{2+}. We used a Zeiss spinning disk confocal microscope and a Rapp system to perform all the experiments. The principle could be transferred to other microscope and laser systems.

1. Place the injected embryos on the inverted laser scanning microscope with the continuous-wave diode laser with 375-nm wavelength mounted on the epiport of the microscope (we used an inverted spinning disk confocal microscope [Zeiss] with 100x/oil objective). Use the 488-nm laser for GCaMP6s signal illumination and set up the time-lapse parameters

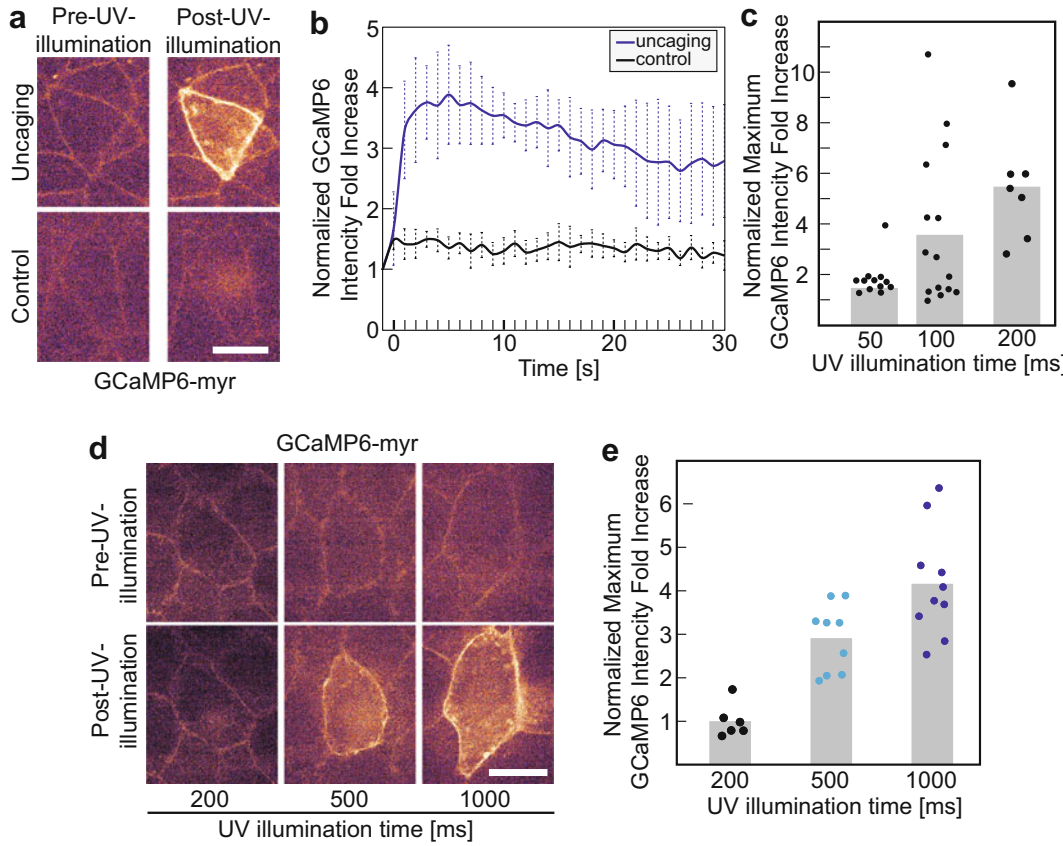


Fig. 2 Uncaging induces a swift Ca^{2+} burst. (**a–c**) Laser illumination for the indicated period in the columnar epithelium (stage 7 embryos, lateral epidermis). (**a**) Images from a time-lapse recording of an embryo in stage 7 (lateral epidermis) expressing membrane-bound GCaMP6s before and after UV laser illumination. Control embryos were injected with buffer without NP-EGTA, AM. (**b**) Time course of normalized GCaMP6 fluorescence in target and control cells. (**c**) Distribution of normalized maximal GCaMP6s signals. Bar indicates the mean. UV illumination time: 50 ms, $n = 12$ cells, 100 ms, $n = 15$ cells, 200 ms, $n = 7$ cells. (**d**, **e**) Laser illumination for the indicated period in the squamous epithelium (stage 14 embryos, amnioserosa). **(d)** Images from time-lapse recordings of embryos in stage 14 (amnioserosa) expressing membrane-bound GCaMP6s before and after UV illumination for the indicated period. (**e**) Distribution of normalized maximal GCaMP6s signals. Bar indicates the mean. UV illumination time: 200 ms, $n = 6$ cells, 500 ms, $n = 9$ cells, 1000 ms, $n = 10$ cells. Scale bars in (**a**) and (**d**) are 10 μm

according to the detector, e.g., laser intensity and exposure time (we used ZEN 3.0 [blue edition, Zeiss] for live imaging).

2. Switch on the continuous-wave diode laser with 375-nm wavelength for uncaging. The uncaging system is independently controlled by REO-SysCon-Zen (Rapp OptoElectronic), synchronized with ZEN 3.0 via UGA-42 Caliburn (Rapp). The uncaging can be performed under the "sequence" or "click-and-fire" mode. Under the "click-and-fire" mode, the relative UV laser illumination size can be controlled by "spot radius"

(we used 3–6 depending on the size of the target cell). UV laser illumination on the target cell is performed by clicking the mouse during imaging under ZEN. An ROI (region of interest) can be set up under the "sequence" mode. Set up the laser intensity and illumination time (*see* **Note 6**). (We used the "click-and-fire" mode for columnar epithelial cells and the "sequence" mode for squamous [amnioserosa] epithelial cells with 20% and 40% laser intensity, respectively, and different illumination times [*see* Fig. 2]).

3. Start the time-lapse imaging for GCaMP6s with ZEN and click "image transfer" (a ZEN macro). Then, the image is synchronized in ZEN and REO-SysCon-Zen. Perform the Ca^{2+} uncaging by 375-nm UV laser illumination on the target cell (s) during time-lapse imaging. The parameters for time-lapse imaging are:
   - Length of image acquisition (5 min).
   - Frequency of capture (1 image/s).
   - Laser and laser intensity (488-nm laser line at 50%).
   - Exposure time (300 ms).
   - Detector (AxiocamMR).
4. Analyze the data. The fluorescence intensity of GCaMP6-myr (*see* Fig. 2) is measured manually with Fiji. The mean density I_t is measured along the cell membrane. The mean density of the background I_b is measured from the cytoplasm. The normalized GCaMP6-myr fluorescence intensity increase is calculated as follows:

$$F/F_0 = (I_t - I_b)/(I_{t-1} - I_{bt-1})$$

where I_t is the mean intensity at time t, I_b is the mean intensity of the background at time t, I_{t-1} is the mean intensity at 1 second (one frame) just before UV illumination, and I_{bt-1} is the mean intensity of the background at 1 second (one frame) just before UV illumination.

3.6 Optochemical Control of Cell Contractility in the Columnar and Squamous Epithelia

In this protocol, E-CadGFP is used to visualize the cell outline. Optochemical control of cell contractility at single-cell resolution in the lateral epidermis (columnar epithelium) from stage 7 embryos and amnioserosa cells (squamous epithelium) from stage 14 embryos is performed (*see* Fig. 3).

1. Place the slide with the injected embryos (at the stage of interest) on the inverted laser scanning microscope. Set up the parameters of time-lapse and uncaging (*see* Subheading 3.6 for detailed information).
2. Start the time-lapse imaging for E-CadGFP with ZEN and click "image transfer." Then, the image is synchronized in

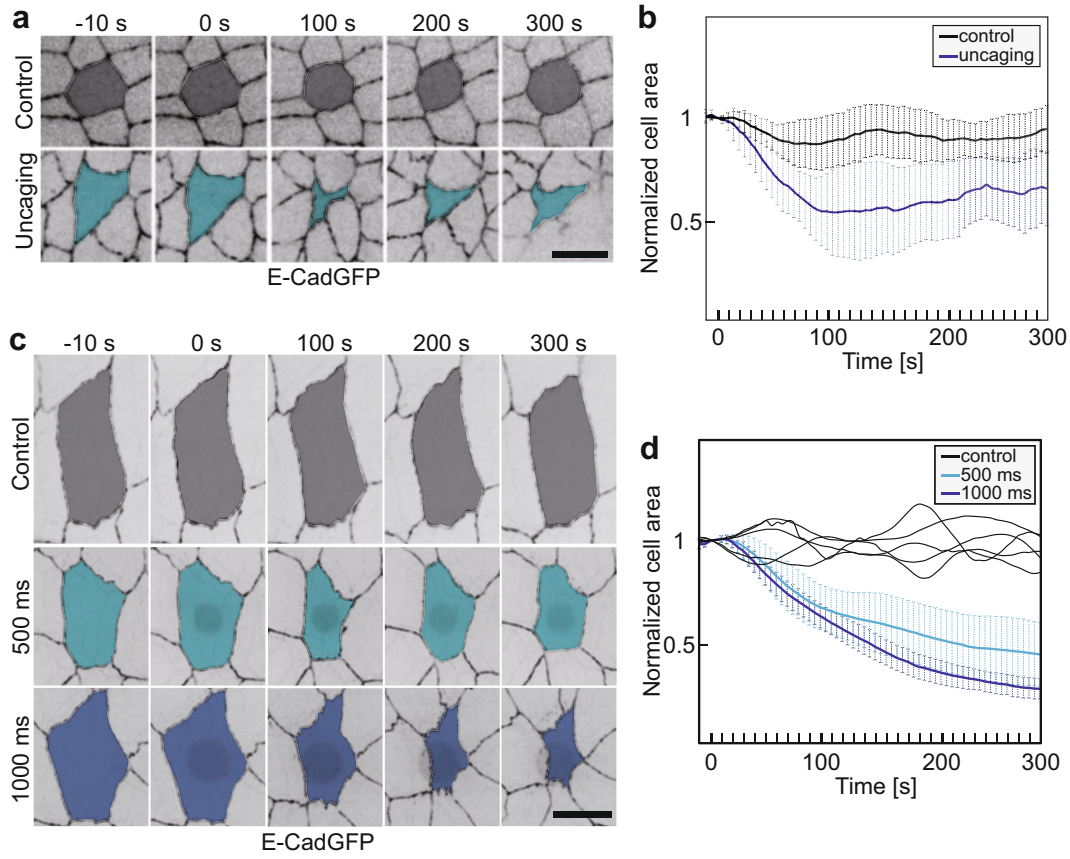


Fig. 3 Optochemical control of cell contractility in the epithelium. (**a**, **b**) Optochemical control of cell contractility in the columnar epithelium. (**a**) Images from a time-lapse recording of embryos in stage 7 (lateral epidermis) expressing the junction marker E-cadherin-GFP. Target cells are labeled by color. Control embryos are injected with buffer without NP-EGTA. (**b**) Time courses of the cross-sectional area of control ($n = 6$ cells) and target ($n = 5$ cells) cells. (**c**, **d**) Optochemical control of cell contractility squamous epithelium. (**c**) Images from a time-lapse recording of embryos (stage 14, amnioserosa) expressing the junction marker E-cadherin-GFP. Control and target cells are labeled by color. The control cells are from the same recording (next–next neighbor of target cells). UV illumination time as indicated. (**d**) Time courses of the cross-sectional area of the control ($n = 5$ cells) and target cells (each illumination, $n = 6$ cells). (**b**, **d**) Curves represent the mean with standard deviation. Scale bars in (**a**) and (**c**) are 10 μm

ZEN and REO-SysCon-Zen. Perform the Ca^{2+} uncaging by 375-nm UV laser illumination on the target cell(s) during time-lapse imaging. The parameters for time-lapse imaging are:

- Length of image acquisition (10 min).
- Frequency of capture (1 image/5 s).
- Laser and laser intensity (488-nm laser line at 50%).
- Exposure time (400 ms).
- Z-stacks (3–4 stacks with a 0.5 μm z-interval).
- Detector (AxiocamMR).

3. Analyze the data. The image stacks are processed with Fiji (Z-projected). Use the Fiji Plugin "Tissue Analyzer" to segment and track the time-lapse images and perform cross-sectional area measurement. (The value of the cell area from A_0 (the first frame after UV illumination) is set to 1, and all the cell areas are normalized by A_0 in our analysis (*see* Fig. 3).)

3.7 Optochemical Control of Tissue Deformation

In this protocol, E-CadGFP is used to visualize the cell outline. Optochemical control of contractility in multiple amnioserosa cells (squamous epithelium) is performed (*see* Fig. 4).

1. Place the slide with the injected embryos (stage 14) on the inverted laser scanning microscope. Set up the parameters of time-lapse and uncaging (*see* Subheading 3.6 for detailed information). A large ROI covering multiple cells for 375-nm UV

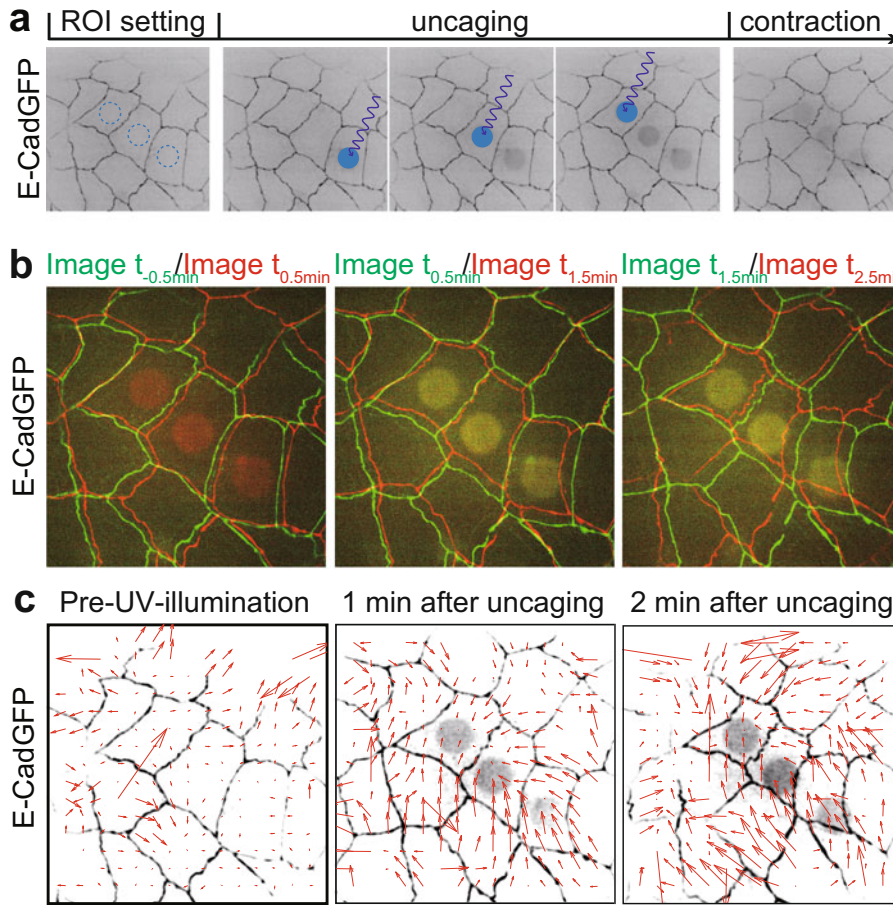


Fig. 4 Optochemical control of tissue deformation. (**a**) Experimental scheme. (**b**) Images from a time-lapse recording of an embryo in stage 14 (amnioserosa) expressing E-cadherin-GFP before and after uncaging at $t = 0$ in three adjacent target cells. Illumination period = 500 ms per cell. Overlays of two consecutive images (earlier green, later red) separated by 1 min. (**c**) PIV maps before and after uncaging as indicated

illuminations can optochemically induce multiple cell contractions simultaneously. Since 375-nm UV illuminations might introduce photobleaching of the cell outline maker, multiple small ROIs for 375-nm UV illuminations are suggested for every individual cell if you want to track the cell contraction.

2. Start time-lapse imaging for E-CadGFP with ZEN and click "image transfer." Then, the image is synchronized in ZEN and REO-SysCon-Zen. Perform the Ca^{2+} uncaging by 375-nm UV laser illumination on the target cells during time-lapse imaging. (We induced Ca^{2+} uncaging in an array of three adjacent cells in the amnioserosa.) The parameters for time-lapse imaging are:
   - Length of image acquisition (10 min).
   - Frequency of capture (1 image/10 s).
   - Laser and laser intensity (488 nm laser line at 50%).
   - Exposure time (400 ms).
   - Z-stacks (3–4 stacks with a 0.5-μm z-interval).
   - Detector (AxiocamMR).
3. Analyze the data. The image stacks are processed with Fiji (Z-projected). "PIVlab" in MATLAB is used for particle imaging velocimetry (PIV) analysis of E-CadGFP images. (We quantified the displacement by particle imaging velocimetry (PIV) using square interrogation windows of 64 × 64 pixels with a 10 s interval by the "PIVlab" in MATLAB (*see* Fig. 4). Prior to uncaging, a distribution of velocities is detected up to 4 μm min^{-1} with the majority below 1 μm min^{-1}. Following uncaging, the distribution of PIV velocities is shifted to larger values. The majority of vectors has a velocity higher than 1 μm min^{-1} with maximal values higher than 4 μm min^{-1}. This analysis shows that induced cell contraction has a striking and quantifiable impact on the neighboring tissue.).

4 Notes

1. In general, the higher the PULL, the smaller will be the needle's tip diameter and the longer the taper. The PULL should be set up according to the capillaries. The taper should not be too long because the needle might be too soft to lodge into the embryo.
2. NP-EGTA, AM is delivered as a powder with 50-mg aliquots. It is stable as a powder; however, once dissolved into the injection solution, repeated freezing and thawing can affect the effectiveness of the reagents. We strongly recommend preparing the NP-EGTA, AM solution freshly and storing the solution with aliquots and protecting it from light as much as possible.

3. Alternatively, allow the flies to lay eggs on the agar plate with yeast overnight. Collect the overnight agar plate and cover the embryos with Voltalef oil 3S. Pick up embryos at the stage of interest according to the morphology under a stereomicroscope using a needle. Transfer the wanted embryos to a net and place the net on a paper towel to remove the oil as much as possible. Then, dechorionate the embryos as mentioned above.
4. If the embryo is too soft during the injection process, this indicates that it is being over-dried. Conversely, the drying time must be extended if the fluid leaks out of the embryo quickly after injection.
5. Pre-pressure and injection pressure vary with different individual microinjection needles. In general, the smaller the needle's tip diameters, the higher the pre-pressure and injection pressure are.
6. Out of the single-mode fiber, the laser gets about 50% of the power (about 35 mW) before the objective of about 20 mW. Depending on the used objective with its transmission specifications, you usually get 10–15 mW after the objective. Besides setting up the percentage of power in REO-SysCon-Zen software, multiple filters are available for replacement, which allow 0.1%, 1%, 10%, and 100% laser to go to the objective.

Acknowledgments

This research was funded in part by Deutsche Forschungsgemeinschaft (project grants GR1945/10-2 and GR1945/15-1, and equipment grant INST160/718-1, No. 446967193) and VolkswagenStiftung A129197.

References

1. Bertet C, Sulak L, Lecuit T (2004) Myosin-dependent junction remodelling controls planar cell intercalation and axis elongation. Nature 429(6992):667–671. https://doi.org/10.1038/nature02590
2. Zallen JA, Blankenship JT (2008) Multicellular dynamics during epithelial elongation. Semin Cell Dev Biol 19(3):263–270. https://doi.org/10.1016/j.semcdb.2008.01.005
3. Kong D, Wolf F, Grosshans J (2017) Forces directing germ-band extension in drosophila embryos. Mech Dev 144(Pt A):11–22. https://doi.org/10.1016/j.mod.2016.12.001
4. Kong D, Großhans J (2020) Planar cell polarity and E-cadherin in tissue-scale shape changes in drosophila embryos. Front Cell Dev Biol 8:619958. https://doi.org/10.3389/fcell.2020.619958
5. Heisenberg CP, Bellaïche Y (2013) Forces in tissue morphogenesis and patterning. Cell 153(5):948–962. https://doi.org/10.1016/j.cell.2013.05.008
6. Martin AC, Goldstein B (2014) Apical constriction: themes and variations on a cellular mechanism driving morphogenesis. Development 141(10):1987–1998. https://doi.org/10.1242/dev.102228
7. Ankenbruck N, Courtney T, Naro Y, Deiters A (2018) Optochemical control of biological processes in cells and animals. Angew Chem

Int Ed Engl 57(11):2768–2798. https://doi.org/10.1002/anie.201700171

8. Hartmann J, Krueger D, De Renzis S (2020) Using optogenetics to tackle systems-level questions of multicellular morphogenesis. Curr Opin Cell Biol 66:19–27. https://doi.org/10.1016/j.ceb.2020.04.004
9. Somlyo AP, Somlyo AV (2003) Ca2+ sensitivity of smooth muscle and nonmuscle Myosin II: modulated by G proteins, kinases, and myosin phosphatase. Physiol Rev 83(4):1325–1358. https://doi.org/10.1152/physrev.00023.2003
10. Uehata M, Ishizaki T, Satoh H, Ono T, Kawahara T, Morishita T, Tamakawa H, Yamagami K, Inui J, Maekawa M, Narumiya S (1997) Calcium sensitisation of smooth muscle mediated by a Rho-associated protein kinase in hypertension. Nature 389(6654):990–994. https://doi.org/10.1038/40187
11. Lee HC, Auersperg N (1980) Calcium in epithelial cell contraction. J Cell Biol 85(2):325–336. https://doi.org/10.1083/jcb.85.2.325
12. Hunter GL, Crawford JM, Genkins JZ, Kiehart DP (2014) Ion channels contribute to the regulation of cell sheet forces during drosophila dorsal closure. Development 141(2):325–334. https://doi.org/10.1242/dev.097097
13. Lee H, Nagele RG (1986) Toxic and teratologic effects of verapamil on early chick embryos: evidence for the involvement of calcium in neural tube closure. Teratology 33(2):203–211. https://doi.org/10.1002/tera.1420330207
14. Ferreira MC, Hilfer SR (1993) Calcium regulation of neural fold formation: visualisation of the actin cytoskeleton in living chick embryos. Dev Biol 159(2):427–440. https://doi.org/10.1006/dbio.1993.1253
15. Hunter MV, Lee DM, Harris TJ, Fernandez-Gonzalez R (2015) Polarised E-cadherin endocytosis directs actomyosin remodeling during embryonic wound repair. J Cell Biol 210(5):801–816. https://doi.org/10.1083/jcb.201501076
16. He L, Wang X, Tang HL, Montell DJ (2010) Tissue elongation requires oscillating contractions of a basal actomyosin network. Nat Cell Biol 12(12):1133–1142. https://doi.org/10.1038/ncb2124
17. Suzuki M, Sato M, Koyama H, Hara Y, Hayashi K, Yasue N, Imamura H, Fujimori T, Nagai T, Campbell RE, Ueno N (2017) Distinct intracellular Ca(2+) dynamics regulate apical constriction and differentially contribute to neural tube closure. Development 144(7):1307–1316. https://doi.org/10.1242/dev.141952
18. Neher E, Zucker RS (1993) Multiple calcium-dependent processes related to secretion in bovine chromaffin cells. Neuron 10(1):21–30. https://doi.org/10.1016/0896-6273(93)90238-m
19. Delaney KR, Shahrezaei V (2013) Uncaging calcium in neurons. Cold Spring Harb Protoc 12:1115–1124. https://doi.org/10.1101/pdb.top079491
20. Kong D, Lv Z, Haring M, Lin B, Wolf F, Grosshans J (2019) In vivo optochemical control of cell contractility at single-cell resolution. EMBO Rep 20(12):e47755. https://doi.org/10.15252/embr.201947755
21. Eberhard DA, Holz RW (1988) Intracellular Ca2+ activates phospholipase C. Trends Neurosci 11(12):517–520. https://doi.org/10.1016/0166-2236(88)90174-9
22. Vázquez-Martínez O, Cañedo-Merino R, Díaz-Muñoz M, Riesgo-Escovar JR (2003) Biochemical characterization, distribution and phylogenetic analysis of Drosophila melanogaster ryanodine and IP3 receptors, and thapsigargin-sensitive Ca2+ ATPase. J Cell Sci 116(Pt 12):2483–2494. https://doi.org/10.1242/jcs.00455
23. Huang J, Zhou W, Dong W, Watson AM, Hong Y (2009) From the cover: directed, efficient, and versatile modifications of the drosophila genome by genomic engineering. Proc Natl Acad Sci U S A 106(20):8284–8289. https://doi.org/10.1073/pnas.0900641106
24. Chen TW, Wardill TJ, Sun Y, Pulver SR, Renninger SL, Baohan A, Schreiter ER, Kerr RA, Orger MB, Jayaraman V, Looger LL, Svoboda K, Kim DS (2013) Ultrasensitive fluorescent proteins for imaging neuronal activity. Nature 499(7458):295–300. https://doi.org/10.1038/nature12354
25. Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, Preibisch S, Rueden C, Saalfeld S, Schmid B, Tinevez JY, White DJ, Hartenstein V, Eliceiri K, Tomancak P, Cardona A (2012) Fiji: an open-source platform for biological-image analysis. Nat Methods 9(7):676–682. https://doi.org/10.1038/nmeth.2019
26. Aigouy B, Umetsu D, Eaton S (2016) Segmentation and quantitative analysis of epithelial tissues. Methods Mol Biol 1478:227–239. https://doi.org/10.1007/978-1-4939-6371-3_13

Chapter 15

Stiffness Measurement of *Drosophila* Egg Chambers by Atomic Force Microscopy

Uwe Töpfer, Karla Yanín Guerra Santillán, and Elisabeth Fischer-Friedrich

Abstract

Drosophila egg chamber development requires cellular and molecular mechanisms controlling morphogenesis. Previous research has shown that the mechanical properties of the basement membrane contribute to tissue elongation of the egg chamber. Here, we discuss how indentation with the microindenter of an atomic force microscope can be used to determine an effective stiffness value of a *Drosophila* egg chamber. We provide information on the preparation of egg chambers prior to the measurement, dish coating, the actual atomic force microscope measurement process, and data analysis. Furthermore, we discuss how to interpret acquired data and which mechanical components are expected to influence measured stiffness values.

Key words *Drosophila* egg chamber, Morphogenesis, Basement membrane, Atomic force microscope, AFM indentation, Extracellular matrix (ECM)

1 Introduction

The *Drosophila* egg chamber has emerged as a powerful model to study how organs and tissues are formed during development. *Drosophila* egg chambers are composed of a monolayered follicle cell epithelium, which covers the inner germline cells (the nurse cells and the oocytes). Follicle cells are covered by a basement membrane, which constitutes the outer envelope of the maturing egg chamber. This basement membrane is a stiff, solid-like layer of self-assembled extracellular matrix with a thickness of approximately 50–200 nm [1, 2].

The initially round egg chamber undergoes 14 developmental stages during maturation and elongates 2–2.5-fold [3]. Egg elongation depends on the planar cell polarity of follicle cells and requires the interaction of the cellular actin network with the surrounding basement membrane [4–9]. During egg chamber development, follicle cells undergo a collective cell migration. As

Christian Dahmann (ed.), *Drosophila: Methods and Protocols*, Methods in Molecular Biology, vol. 2540,
https://doi.org/10.1007/978-1-0716-2541-5_15,

a consequence, basal actin filaments and fiber-like structures within the basement membrane show a circumferential polarization in stage 8 [10]. This resulting circumferential scaffold of structural molecules is called molecular corset. The molecular corset prohibits radial outgrowth, causing the egg chamber to grow into an elongated ellipsoidal shape over time. Mutant egg chambers for the main basement membrane component collagen IV and the cell basement membrane receptor integrin fail to elongate [10]. Moreover, genetic manipulation of the basement membrane composition or secretion can result in aberrant egg elongation [11–13]. In addition, collagenase-treated egg chambers are less resistant to tissue expansion, as determined in an osmotic burst assay [11]. In conclusion, these observations point to a major functional role of the basement membrane in the mechanical stabilization and the shape regulation of the egg chamber.

Indentation experiments with an atomic force microscope (AFM) are a method to monitor the mechanical changes of the basement membrane, reflecting the changes of the basement membrane's architecture and molecular composition during egg chamber development. Previous AFM indentation measurements on *Drosophila* egg chambers reported apparent Young's moduli of egg chambers to be between 1 and 800 kPa from measurements of different indentation depths and indenter shapes [1, 2, 11, 14]. In those studies, the authors demonstrated that measured stiffness values were dominated by the basement membrane by showing that molecular perturbations of basement membrane components resulted in significant changes of measurement results (e.g., collagenase treatment led to a significant decrease of measured moduli) [1, 2, 11, 14]. Furthermore, during development, egg chambers were shown to stiffen as reflected by increasing Young's moduli for increasing stages [1, 2, 11]; measurements at the central, terminal, and pole regions of distinct egg chamber stages showed an initially uniform stiffness in round stages (3 and 4). At later stages (7 and 8), stiffness inhomogeneities started to form along the long axis of the egg chamber with a stiffness peak in the central region. This non-uniform increase leads to the generation of stiffness variations along the egg chamber with a stiffness peak in the center [11, 14]. Mechanical changes between stages 5 and 8 coincide with the establishment of the abovementioned molecular corset of the egg chamber.

Theoretical considerations suggest that apparent Young's moduli, as reported from shallow AFM indentation of egg chambers (described in this protocol), are influenced by several material parameters [15]: Young's modulus and the bending modulus of the basement membrane material, the thickness of the basement

membrane, the principal radii of curvature of the egg, and the pressure excess in the egg chamber.

As part of this protocol, we describe (1) how to prepare the dishes for mounting of the *Drosophila* egg chambers, (2) how to dissect the egg chambers from the flies, (3) how to set up an atomic force microscope for the measurement, (4) how to carry out the AFM indentation measurements, (5) how to perform data analysis, and (6) how to interpret the obtained measurement results.

2 Materials

2.1 Dish Preparation

1. Poly-D-lysine solution (0.1 mg ml^{-1}).
2. Cell culture dishes (35 mm diameter, e.g., FluoroDish FD35–100, World Precision Instruments, *see* **Note 1**).
3. Phosphate-buffered saline (PBS): 137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4.
4. 2 mL reaction tubes.
5. Compressed air or nitrogen spray gun.

2.2 Egg Chamber Dissection

1. Fetal bovine serum, sterile-filtered.
2. Penicillin/streptomycin (10,000 U mL^{-1} of penicillin and 10,000 μg mL^{-1} of streptomycin).
3. Insulin powder, from bovine pancreas.
4. Egg chamber culture medium [16]: 15% fetal bovine serum and 600 U ml^{-1} penicillin/ 600 μg ml^{-1} streptomycin in Schneider's medium (*see* **Note 2**). Adjust the pH to 7.0 (*see* **Note 3**). Store at 4 °C. Dissolve insulin in acidified water (37% HCl in water, 1:1000) and store at −20 °C (*see* **Note 4**). Immediately before usage, add insulin to a final concentration of 0.2 mg ml^{-1}.
5. Vial with fly food.
6. Baker's yeast paste: active dried baker's yeast dissolved in a small amount of water.
7. Stereomicroscope.
8. Light source.
9. Forceps (Dumont #55).
10. Block bowl or depression glass slide.
11. CO_2 source with a pad and fly gun.
12. 200 μl pipette.

2.3 AFM, Light Microscopy Setup

1. Standard atomic force microscope setup mounted on an inverted light microscope (*see* **Note 5**). The atomic force microscope setup should ideally offer a ~ 15 μm piezo-controlled z-range and a 100 μm × 100 μm x–y scan range.
2. Low-magnification air objective (10× or 20×).
3. Microscope camera for transmitted light pictures, ideally with a large chip and a corresponding large field of view (e.g., DMK 23 U445 from the Imaging Source) (*see* **Note 6**).

2.4 AFM Measurements on Egg Chambers

1. AFM cantilever holder for measurement in water (e.g., Bruker/JPK Supercut cantilever holder).
2. Bruker MLCT-C AFM cantilevers (at least a package of 10) (*see* **Note 7**).
3. A silicone/plastic foil ring (placed around the cantilever holder during the measurement to protect the AFM internal parts from condensation water).
4. Petri dish (35 mm diameter) for cantilever calibration filled with ~1 ml of Schneider's medium (ideally use the same dish type as for measuring egg chambers, e.g., FluoroDish FD35–100, World Precision Instruments).
5. AFM analysis software.

3 Methods

3.1 Dish Preparation

1. Aliquot poly-D-lysine solution in 2 mL reaction tubes and store them at −20 °C.
2. One day before the experiment, thaw a poly-D-lysine aliquot at room temperature. Once melted, dilute the poly-D-lysine in 18 mL of sterile PBS to prepare a 1:10 working solution.
3. Estimate the number of cell culture dishes required for the measurement of egg chambers (typically one per condition to be measured). Coat the surface of each cell culture dish with 2.5 mL of the above working solution and leave them overnight at 4 °C.
4. The next day, remove the solution from the dishes and let them dry for 1 hour at room temperature. Place them upside down or with the lid on to avoid dust or other kinds of particle contamination.
5. After 1 hour, remove the remaining solution by applying the compressed air or nitrogen spray gun, as excess of poly-D-lysine solution can be toxic to the biological samples.
6. Dry dishes can be used immediately or stored at 4 ° C and used within 2 weeks of coating.
7. If doing the measurement for the first time, prepare an uncoated dish for AFM cantilever calibration. To this end, mark the dish with a permanent marker pen from the inside in the dish center. This mark serves as a focusing aid. After the

measurement, this dish can be dried and reused for calibration during subsequent measurement days.

3.2 Egg Chamber Dissection

1. Collect 1–2-day-old female flies of the desired genotype in a standard food vial and add baker's yeast paste (*see* **Note 8**).
2. Incubate the flies (*see* **Note 9**) for 1–2 days to allow ovary growth.
3. Anesthetize one to three flies with CO_2 on a fly pad and transfer them to the egg chamber culture medium (in a block bowl or depression glass slide).
4. To dissect the ovary under a stereomicroscope with a light source, hold the abdomen at the dorsal region with one pair of forceps. Rip it into two pieces by pushing with the second forceps at the last but one posterior segment (*see* Fig. 1a, **Note 10**). Hold the residual abdomen with one forceps and pull with the other at the oviduct to separate the ovary (*see* Fig. 1b, c).

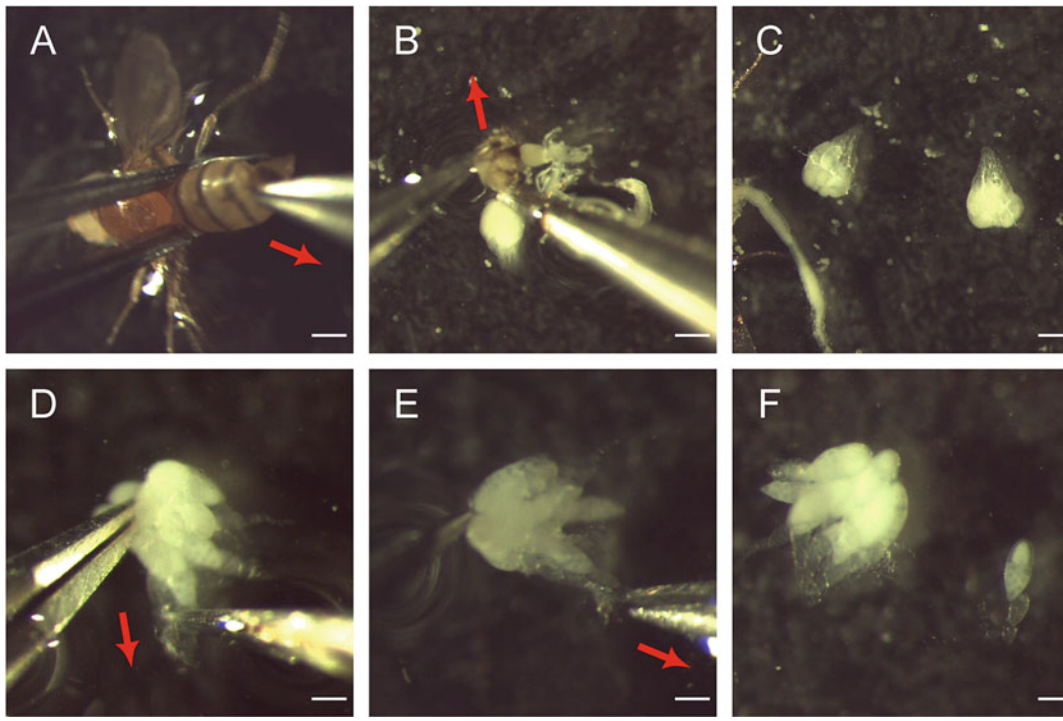


Fig. 1 Egg chamber dissection. (**a**) Grab the fly (dorsal to the top) with one forceps at the anterior central region of the abdomen and use the second forceps to grab and pull (red arrow for direction) at the last but one segment of the abdomen to move the ovary out of the fly. (**b**) Hold the ovary with one forceps and pull (red arrow for direction) at the residual cuticle with the other to remove it. In addition, remove other attached nonovarian tissues. (**c**) Only intact ovaries should be left. (**d**) Hold the ovary with one forceps and use one side of the second forceps to move between the ovarioles (red arrow for direction) to isolate them. (**e**) Still hold the ovary with one forceps. Grab the tip of one ovariole with another forceps and pull it out to isolate the egg chambers from the muscle sheet (red arrow for direction). (**f**) Isolated egg chambers on the right

Remove all dispensable tissues like the gut or fat body. Hold the ovary with one forceps and isolate the ovarioles by moving the second forceps between them (*see* Fig. 1d). Still hold the ovary and grasp the ovariole at its small tip. Gently, pull the egg chambers out of the muscle sheath (*see* Fig. 1e, **Note 11**). Repeat this step to collect more egg chambers.

5. Suck the egg chamber of the desired stage together with 20 µl egg chamber culture medium into the tip of a 200 µl pipette and release it in the center of a poly-D-lysine-coated glass-bottom dish. Once placed in the desired spot, quickly remove and add back the medium with the same pipette tip. The momentary lack of medium will adhere the egg chambers to the glass-bottom dish.
6. Transport the dish with the egg chambers to the AFM. Less medium while moving the dish will prevent the egg chambers from floating away. Before beginning AFM measurements, add 4 ml more of the medium.

3.3 AFM Preparation

1. Mount the cantilever on the cantilever holder. Mount the cantilever holder with the cantilever onto the AFM head and add a water protection ring if necessary.
2. Switch on the AFM/light microscope ensemble. Open the AFM software. Adjust the measurement mode to be *force spectroscopy* in *contact mode*. Tick the closed loop box in the software to use the piezo gauges.
3. Mount the calibration dish filled with ~1 ml egg chamber culture medium on the stage. Focus the microscope on the pen mark inside the calibration dish to find the upper side of the dish bottom glass. Now, keep the focus position of the objective.
4. Unmount the calibration dish with the dish holder. Mount the AFM head on the microscope. Lower the cantilever in 100 µm steps until it comes into focus. Double-check visually that the cantilever keeps sufficient distance from the objective top surface.
5. Place the cantilever tip in the center of the field of view. Position the AFM laser on the cantilever tip with the aid of the camera. Lift the cantilever by 2 mm with the stepper motors. Unmount the AFM head.
6. Mount the calibration dish again, mount the AFM head carefully avoiding physical contact between the cantilever and the dish. Then, lower the cantilever in steps of 100 µm with the stepper motors until the cantilever is ~100 µm above the dish surface (the cantilever tip is not quite in focus yet but is already visible).

7. Now, adjust the mirror in the AFM head to obtain maximum signal intensity on the AFM photodiode. Adjust the voltage signal to ~0 V vertical and horizontal deflection.
8. Do an approach with the cantilever on the glass surface of the calibration dish (Adjust the approach height appropriately. For a system with a ~15 μm piezo, an approach height of 7.5 μm is appropriate for calibration and measurements.)
9. Open the calibration manager of the AFM software to perform thermal calibration of the cantilever. Measure cantilever sensitivity: run a force spectroscopy measurement by lowering the cantilever against the glass bottom until half the maximum of the vertical deflection is reached at an extension speed of 2 μm s^{-1} (for a diode with a maximum deflection of 10 V, deflect until a 5 V signal is reached). Fit the slope of the voltage versus cantilever height curve to obtain the sensitivity estimate for the cantilever. Lift the cantilever by 200 μm with the stepper motors. Record the thermal resonance curve of cantilever fluctuations. Determine the cantilever spring constant by fitting this curve around a chosen resonance peak. For egg chamber indentation, we use the second peak of the thermal resonance curve and a corresponding correction factor of 0.25.
10. Lift the AFM cantilever by 2 mm and carefully unmount the AFM head.

3.4 AFM Measurements on Egg Chambers

1. Before measurements start, the AFM measurement process needs to be programmed in the AFM control software. For each egg chamber measurement, we put forward to probe the stiffness of the egg by indenting on several points on the top central area of the egg chamber. We recommend that the cantilever performs 64 indentations by scanning an 8 × 8 grid on a square *x*–*y* area of 10 μm × 10 μm. In case stiffness measurements on other parts of the egg chamber are of interest, the extension and localization of the force scan area can be adapted accordingly. By usage of a specialized PDMS holder, the egg chamber may be placed in an upright position, which allows stiffness probing of the egg chamber at the poles. Details of this procedure have been described in ref. [11].
2. The following indentation settings are recommended: an extension speed in the order of 0.5 μm s^{-1} (speed of cantilever lowering, *see* **Note 12**), indentation up to a peak force of 0.4 nN (*see* **Note 13**), retraction speed of 5 μm s^{-1} (speed of cantilever lifting), and z-length of 2 μm (this is the *z*-range, which will be available for data analysis in the fitting of force-indentation curves later). Once the force scanning has been programmed, save the scanning routine under a specified name.

3. Activate the autosave of the force-indentation curves with the AFM software and set a name for force spectroscopy files that are going to be saved during the measurement. We recommend to save force spectroscopy curves measured for each egg chamber in a separate folder.
4. Now, mount the Petri dish with dissected and adhered egg chambers on the microscope stage. Check that the egg chambers are properly adhered to the dish (*see* **Note 14**).
5. Move the microscope stage in the *x*–*y* direction until you find an egg chamber in the right developmental stage (*see* **Note 15**). Place the egg chamber in the center of the field of view and orient it such that the longitudinal axis is along the *y*-direction, i.e., perpendicular to the AFM cantilever (*see* Fig. 2). Mount

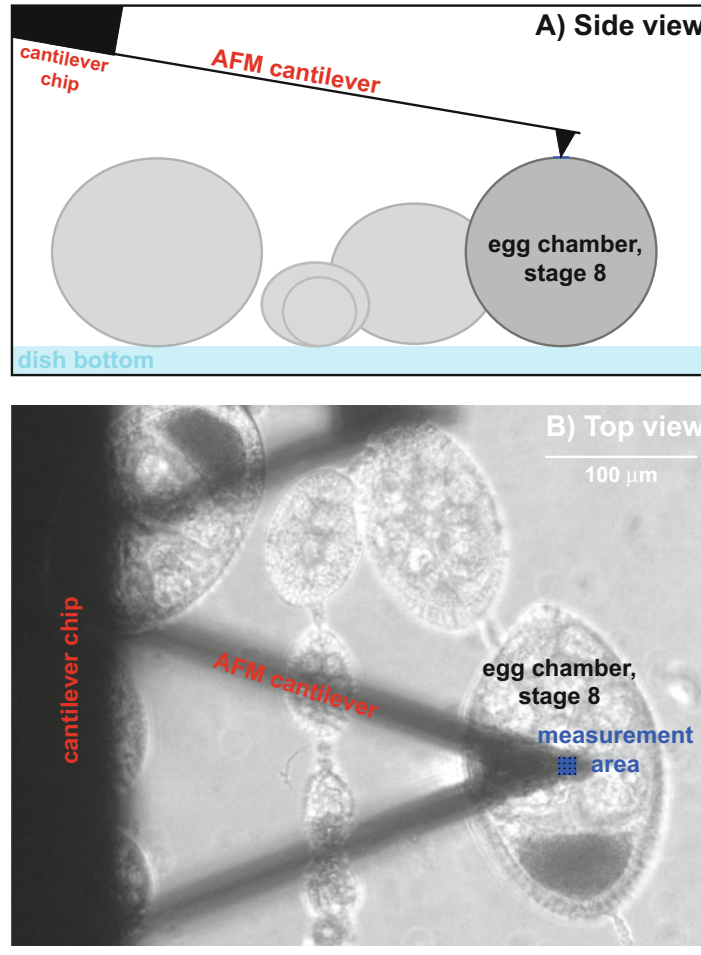


Fig. 2 AFM indentation measurement on a *Drosophila* egg chamber. (**a**) Schematic side view of the measurement setup (egg chambers are seen from the poles). (**b**) Transmitted light image of a top view of the measurement setup with a triangular MLCT-C cantilever over a stage 8 egg chamber

the AFM head and lower the cantilever in steps of 100 μm until approximately 100 μm above the dish surface.

6. Approach the cantilever in the center on the top surface of the chosen egg chamber. Set the center of the grid position. Start the force scan such that the 64 force-indentation curves are recorded on the different grid positions one after another.
7. Take a camera snapshot of the measured egg chamber if required and name the image file in correspondence to the force scan file.
8. Now, the next egg chamber can be measured. To that end, lift the AFM cantilever by 100 μm. If an appropriate egg chamber is available in the same field of view (or nearby) and in the correct orientation, then move the stage to center the new egg chamber in the field of view. Move the cantilever in the x–y direction above the center of the new egg chamber. Then, continue from **step 6**. Otherwise lift the AFM cantilever by ~1.5 mm and unmount the AFM head. Then move to point 5 (*see* **Note 16**).
9. We recommend to measure 5–15 egg chambers for each condition under consideration in order to obtain an appropriate statistic (*see* **Note 17**). In case of shallow stiffness changes, larger egg chamber numbers might be required in order to obtain statistical significance.
10. To change the dish and move to a second condition, lift the AFM cantilever ~1.5 mm above the dish bottom and unmount the AFM head.

3.5 Data Analysis of AFM Measurements

1. Load the force spectroscopy curves obtained from indentation of one egg chamber into the AFM analysis software to perform AFM curve fitting. Possible software choices include AFM-dedicated MATLAB routines [17], open-source software such as AtomicJ [18], or software from commercial AFM companies.
2. Open the batch of spectroscopy curves for one egg chamber in the curve-fitting software. For each curve,
   (a) Correct the cantilever tip height for cantilever bending.
   (b) Find the force baseline.
   (c) Fit the contact point of the force-indentation curve.
   (d) Fit the force-indentation curve: for a pyramidal indenter as given for the MLCT-C cantilevers, Hertzian contact mechanics predicts a force-indentation function $F_{\mathrm{AFM}} \approx \frac{E}{1-\nu^2} \frac{3 \tan\alpha}{4} \delta^2$ [19–21], (see **Note 18**). Here, α is the angle between the face of the pyramid and its central axis, E is an effective Young's modulus of the indented material, and ν is the Poisson ratio of the indented material, which is

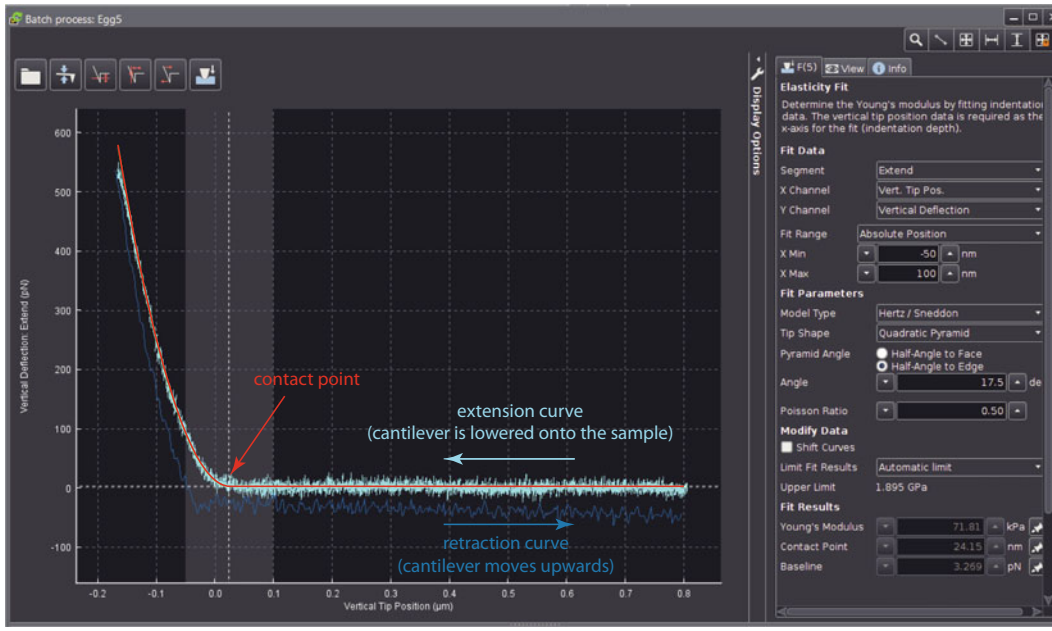


Fig. 3 Analysis of force-indentation curves by Hertz model fitting. The measured force is given by the vertical deflection of the cantilever (*y*-axis). The indentation depth is read out from the vertical tip position below the contact point (here at 0 μm, *x*-axis). The red curve represents fitting of the force-extension curve (bright blue) with the theoretical force-indentation relationship $F_{\mathrm{AFM}}(\delta) = \frac{E}{1-\nu^2}\,\frac{3\,\tan\alpha}{4}\,\delta^2$ for a pyramidal indenter [19]. The force-retraction curve (darker blue) is not fitted due to stronger influence of adhesion effects on this force curve

commonly set to 0.5 for biological samples. To probe the stiffness of the egg chamber basement membrane, force-indentation curves should be fitted only up to an indentation depth of 50 nm. An example fit is presented in Fig. 3.

(e) Output fit parameters into a text file, i.e., apparent Young's modulus E and the contact point.

3. Repeat **step 2** for batches of force-indentation curves for all egg chambers.

3.6 Interpretation of Measured Apparent Young's Moduli

AFM indentation experiments commonly use Hertzian contact mechanics to associate a specific material stiffness (as quantified by an apparent Young's modulus E) with measured force-indentation curves. In the analysis scheme described here (also used in ref. [1, 11]), a mechanical model is applied, which assumes indentation into an elastic half space with an indenter of a specific shape (e.g., pyramidal, parabolic, or spherical). In essence, this analysis assumption means that the force curves are interpreted as if indentation had been performed on the surface of a large, thick, isotropic, uniform, and nonadhesive workpiece (with a significant distance to the boundaries). Furthermore, incompressibility of the workpiece material is typically anticipated, i.e., the Poisson ratio ν is set

to 0.5. Clearly, the egg chamber does not fulfill all criteria of this assumption (for instance, the egg chamber is not made of a uniform material but has several mechanical components such as the outer basement membrane as well as the cytoskeleton and the cytoplasm of internal cells). Nevertheless, this simplifying interpretation scheme is applied and therefore provides a so-called apparent Young's modulus of the measured egg chamber.

More complex theoretical studies of indentation into pressurized elastic shells suggest that the force response of shallow shell indentations can be influenced by several material parameters of the shell [15, 22]: Young's modulus and the bending modulus of the shell, the thickness of the shell, the principal radii of curvature of the shell, and the hydrostatic pressure excess in the shell (if there is any). By applying this insight to the shallow indentation of *Drosophila* egg chambers as described here, we can conclude that the measured apparent Young's moduli are likely influenced by the thickness of the basement membrane, the radii of curvature of the egg chamber, the bending and Young's modulus of the basement membrane material, and the hydrostatic pressure excess inside the egg chamber.

4 Notes

1. The rim of the used dishes should not be too high since, otherwise, the AFM cantilever cannot reach the sample.
2. To avoid contaminations, we prepare the culture medium under a laminar flow hood.
3. For stage 9 egg chambers, an exact pH of the culture medium is critical [16]. If it exceeds a pH of 7.0, prepare a fresh one.
4. Single-use aliquots can be stored at −20 °C as a stock solution. Freeze-thaw cycles should be avoided.
5. Make sure that your AFM/light microscope setup is properly isolated against vibrations present in the room/building. Measurements described here are in the lower force range of AFM measurements and are vulnerable to disturbances by acoustic waves or small mechanical vibrations triggered by external sources such as nearby traffic. Prepare for the measurements to be carried out in silence.
6. Large microscope fields of view are beneficial for conducting AFM measurements (e.g., by choice of a low-magnification objective or a camera with a large chip size). In this manner, it is easier to locate the cantilever in the field of view. If the experimentalist is mistaken about the x–y position of the cantilever, then accidents with cantilever breakage during the measurement are much more likely.

7. It is possible to use other AFM cantilevers than the ones described here. However, much care is needed for the right choice. The cantilever stiffness needs to be adapted to the force range acquired during the measurement. This force range depends on the indenter shape – larger indenters will tend to give larger forces at a specific indentation depth. As a rule of thumb, a typical peak force during indentations should be no lower than ~5% but no larger than ~50% of the maximal force range for the cantilever under consideration.

   Furthermore, the indenter shape influences the functional shape of the curve fitted to the measured force-indentation curve. Standard analysis schemes are available for pyramidal, spherical, and conical indenters.

8. Age and nutrition are critical to ensure proper egg chamber development. Too old or malnourished flies stall in egg production and could show morphological defects [23, 24].

9. Incubation at 25 °C should be preferred for suitable egg production; however, an increase to 29 °C like for Gal4/upstream activation sequence (UAS) temperature-sensitive experiments will be tolerated. Too high- or low-temperature conditions may affect fly fitness and fecundity.

10. Correct grasp positions of the forceps are required to successfully isolate the intact ovaries. Positions more anterior will end in a ripping between the thorax and the abdomen and positions more posterior may damage the ovary.

11. Do not touch the dissected egg chambers of interest. Every damage could affect the integrity of the follicle cells and the measurements. If the egg chambers leak yolk, do not use them. Be sure that the egg chamber is no more covered by the muscle sheath. An overlying tissue will disturb the measurements.

12. Variation of AFM indentation speed may influence measured Young's moduli. It may seem attractive to increase indentation speed in order to perform measurements more quickly. However, in this case, AFM forces will be influenced to a larger extent by the excitation of viscous flows during AFM indentation. These may be (1) viscous flows inside the medium surrounding the moving cantilever or (2) flows inside the viscous cytoplasm. In particular, the former does not provide any meaningful force contribution to the measurement, and respective force contributions should be avoided. The current choice of AFM indentation speed is a good compromise between the speed of measurement and avoidance of viscous contributions during AFM measurements. Furthermore, it is a standard speed value in the field.

13. Indenting up to a higher peak force is an option and increases the accessible force-indentation range for analysis. For analysis

up to 50 nm indentation, this increase is, however, not required. A peak force increase might call for an accompanying increase of the z-length of the AFM indentation depending on the stiffness of the sample.

14. Proper adhesion of the egg chambers to the dish bottom is crucial for the measurements. If the egg chambers are only loosely attached or not attached at all, then this will influence measured Young's moduli. In particular, failure in adhesion will yield modulus values that are smaller than values from measurements on the adhered egg chambers. To avoid corresponding artifacts in the measurement results, test whether the egg chambers are properly adhered prior to the measurement. To this end, gently (!!) agitate the microscope mechanically when the dish is mounted (ideally, without the AFM head mounted). The liquid in the dish will start to move gently in response. Egg chambers that are not attached properly will be seen to move with the liquid when monitored by the ocular or by the camera.
15. Detailed protocols of how to identify the distinct stages of egg chamber development with or without fluorescent markers have been previously published [25, 26].
16. After finishing the AFM measurement on one egg chamber, you may want to change the field of view to measure another egg chamber. If this second egg chamber is further apart than the extension of your field of view, or if the egg chamber does not have the proper orientation with respect to the AFM cantilever (the longitudinal axis of the egg chamber is perpendicular to the longitudinal axis of the AFM cantilever), then lift your AFM cantilever ~1.5 mm away from the dish bottom and unmount the AFM head. Only after this, change the field of view of the dish and/or turn the dish to achieve proper egg orientation. Then, the AFM head can be mounted again, and the cantilever can be lowered and approached on the sample. Why is this necessary? Moving the stage excessively with the AFM head mounted may lead to driving the cantilever holder into the rim of the dish. This may (1) displace your cantilever and thereby ruin your cantilever calibration or (2) immediately break the cantilever. Furthermore, dish rotation is not possible without AFM unmounting.
17. We suggest to compare stiffnesses of the egg chambers under different conditions on one and the same day of measurement using the same cantilever and the same calibration of this cantilever. What are the reasons for this? AFM cantilevers of a specific kind are subject to a certain degree of variability even within one batch of fabrication. This concerns the tip curvature and tip height as well as cantilever stiffness. Furthermore,

cantilever stiffness determined by thermal calibration is accurate in a range of approximately 10% [27] and may vary from one calibration to the other. These effects may introduce systematic changes of measured Young's moduli from one measurement day to the other even in the absence of actual stiffness changes. If not enough data for a satisfactory statistic can be produced on one day, then the measurement needs to be repeated on forthcoming days, and data from different days can then be pooled later. If this pooling option is chosen, the egg chambers from each condition should be measured at approximately the same number on each measurement day to not introduce biases. Moreover, the order of conditions during the measurement should be permuted at different measurement days to avoid possible effects of sample waiting times.

18. For some cantilevers only the angle β between the outer edge of the pyramid and the central axis is given by the manufacturer. In this case, the alternative formula $F_{\mathrm{AFM}} \approx \frac{E}{1-\nu^2} \frac{3 \tan\beta}{4\sqrt{2}} \delta^2$ for the force-indentation curve can be used.

Acknowledgments

E.F.-F. was supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC-2068 – 390729961– Cluster of Excellence Physics of Life of TU Dresden.

References

1. Chlasta J, Milani P, Runel G et al (2017) Variations in basement membrane mechanics are linked to epithelial morphogenesis. Development 144:4350–4362
2. Díaz de la Loza MC, Díaz-Torres A, Zurita F et al (2017) Laminin levels regulate tissue migration and anterior-posterior polarity during egg morphogenesis in Drosophila. Cell Rep 20:211–223
3. King RC (1970) Ovarian development in Drosophila melanogaster. Academic Press, New York
4. Gates J (2012) Drosophila egg chamber elongation. Fly (Austin) 6:213–227
5. Horne-Badovinac S, Bilder D (2005) Mass transit: Epithelial morphogenesis in theDrosophila egg chamber. Dev Dyn 232:559–574
6. Horne-Badovinac S (2014) The Drosophila egg chamber–a new spin on how tissues elongate. Integr Comp Biol 54:667–676
7. Viktorinová I, König T, Schlichting K et al (2009) The cadherin Fat2 is required for planar cell polarity in the Drosophila ovary. Development 136:4123 LP–4124132
8. Barlan K, Cetera M, Horne-Badovinac S (2017) Fat2 and Lar define a basally localized planar signaling system controlling collective cell migration. Dev Cell 40:467–477.e5
9. Frydman HM, Spradling AC (2001) The receptor-like tyrosine phosphatase lar is required for epithelial planar polarity and for axis determination within drosophila ovarian follicles. Development 128:3209–3220
10. Haigo SL, Bilder D (2011) Global tissue revolutions in a morphogenetic movement controlling elongation. Science 331: 1071–1074
11. Crest J, Diz-Muñoz A, Chen D-Y et al (2017) Organ sculpting by patterned extracellular matrix stiffness. Elife 6:e24958
12. Isabella AJ, Horne-Badovinac S (2015) Dynamic regulation of basement membrane

protein levels promotes egg chamber elongation in Drosophila. Dev Biol 406:212–221

13. Isabella AJ, Horne-Badovinac S (2016) Rab10-mediated secretion synergizes with tissue movement to build a polarized basement membrane architecture for organ morphogenesis. Dev Cell 38:47–60
14. Töpfer U, Guerra Santillán KY, Fischer-Friedrich E, Dahmann C (2022) Development 149(10):dev200456. https://doi.org/10.1242/dev.200456
15. Vella D, Ajdari A, Vaziri A et al (2011) The indentation of pressurized elastic shells: from polymeric capsules to yeast cells. J R Soc Interface 9:448
16. Prasad M, Jang AC-C, Starz-Gaiano M et al (2007) A protocol for culturing Drosophila melanogaster stage 9 egg chambers for live imaging. Nat Protoc 2:2467–2473
17. Huth S, Sindt S, Selhuber-Unkel C (2019) Automated analysis of soft hydrogel microindentation: Impact of various indentation parameters on the measurement of Young's modulus. PLoS One 14:e0220281
18. Hermanowicz P, Sarna M, Burda K et al (2014) AtomicJ: an open source software for analysis of force curves. Rev. Sci Instrum 85:63703
19. Bilodeau GG (1992) Regular pyramid punch problem. J Appl Mech 59:519–523
20. Sneddon IN (1965) The relation between load and penetration in the axisymmetric boussinesq problem for a punch of arbitrary profile. Int J Eng Sci 3:47–57
21. Rosenbluth MJ, Lam WA, Fletcher DA (2006) Force microscopy of nonadherent cells: a comparison of leukemia cell deformability. Biophys J 90:2994–3003
22. Vella D, Ajdari A, Vaziri A et al (2012) Indentation of ellipsoidal and cylindrical elastic shells. Phys Rev Lett 109:144302
23. Drummond-Barbosa D, Spradling AC (2001) Stem cells and their progeny respond to nutritional changes during drosophila oogenesis. Dev Biol 231:265–278
24. Zhao R, Xuan Y, Li X et al (2008) Age-related changes of germline stem cell activity, niche signaling activity and egg production in Drosophila. Aging Cell 7:344–354
25. Spradling AC (1993) Developmental genetics of oogenesis. In: The development of Drosophila melanogaster, pp. 1–70 Cold Spring Harbor Laboratory.
26. Jia D, Xu Q, Xie Q et al (2016) Automatic stage identification of Drosophila egg chamber based on DAPI images. Sci Rep 6:1–12
27. Matei GA, Thoreson EJ, Pratt JR et al (2006) Precision and accuracy of thermal calibration of atomic force microscopy cantilevers. Rev Sci Instrum 77:83703

Chapter 16

Cultivation and Live Imaging of *Drosophila* Imaginal Discs

Natalie A. Dye

Abstract

In this chapter, I present a method for the ex vivo cultivation and live imaging of *Drosophila* imaginal disc explants using low concentrations of the steroid hormone 20-hydroxyecdysone (20E). This method has been optimized for analyzing cellular dynamics during wing disc growth and leverages recent insights from in vivo experiments demonstrating that 20E is required for growth and patterning of the imaginal tissues. Using this protocol, we directly observe wing disc proliferation at a rapid rate for at least 13 h during live imaging. The orientation of tissue growth is also consistent with that inferred from indirect in vivo techniques. Thus, this method provides an improved way of studying dynamic cellular processes and tissue movements during imaginal disc development. I first describe the preparation of the growth medium and the dissection, and then I include a protocol for mounting and live imaging of the explants.

Key words *Drosophila*, Explant, Culture, Ecdysone, Growth, Eversion, Live imaging, Dynamics, Wing disc

1 Introduction

Drosophila imaginal discs have proven to be excellent model systems for studying many aspects of animal tissue development and cell biology [1]. These tissues are flat epithelial sacs that grow during larval stages of a fly's life cycle before undergoing complex morphogenesis during pupal stages to eventually generate most of the adult tissues of the animal. By the third instar, the imaginal discs are large enough to visualize with a stereomicroscope and can be dissected away from the rest of the animal by carefully severing only one or two thin attachments to their surroundings. Most often, the tissues are then fixed and analyzed with immunofluorescence, as staining patterns are easy to visualize in this relatively simple, flat geometry and many genetic perturbations are available. Keeping the dissected tissues alive in ex vivo culture, however, is required for studying dynamic processes at high spatial and temporal resolution as well as for making perturbations with physical and pharmacological methods. In this case, care must be taken to ensure that the

Christian Dahmann (ed.), *Drosophila: Methods and Protocols*,
Methods in Molecular Biology, vol. 2540, https://doi.org/10.1007/978-1-0716-2541-5_16,

culture conditions replicate the in vivo environment as much as possible, including important signals that imaginal discs get from elsewhere in the animal.

One circulating signal that is critically important for many aspects of fly development is the steroid hormone ecdysone, which is produced in the prothoracic gland and then converted into its active form, 20-hydroxyecdysone (20E), by distal tissues. A large peak in the circulating levels of ecdysone triggers pupariation at the end of the third instar. Early work has shown that the addition of 20E to a culture medium could support ex vivo eversion morphogenesis and cuticular differentiation of the wing and leg imaginal discs [2–4]. More recently, it has become clear that 20E signaling is also required in vivo throughout the larval stages to support the morphogen signaling systems that pattern gene expression in the wing disc [5, 6] and allow morphogenetic furrow progression in the eye disc [7, 8]. Furthermore, it was shown that the addition of relatively low concentrations of 20E to a culture medium can support wing disc proliferation for up to 24 h ex vivo [5]. In summary, 20E acts directly on imaginal discs, both in vivo and ex vivo, to promote their growth, development, and morphogenesis.

The method presented here was developed to enable quantification of the cellular dynamics underlying the oriented tissue growth of the wing disc using live imaging at fairly high spatial and temporal resolution [5, 9]. For this purpose, it is important to maintain proliferation for as long as possible, while preserving the patterning systems that are known to be required for the proper amount and orientation of growth. Thus, it is important to include low levels of 20E in the culture medium. Insulin, a well-conserved anabolic growth hormone that is often added to a culture medium [10–13], is not absolutely necessary for ex vivo proliferation when low concentrations of 20E are supplied [5]. In fact, 20E can support more proliferation in ex vivo culture than insulin [5]. Although the imaginal discs are exposed in vivo to circulating *Drosophila* insulin-like peptides under fed conditions, they also continue proliferating at least partially during periods of starvation, a phenomenon called "sparing" [14, 15]. Whereas loss of 20E signaling in vivo causes patterning defects [5, 6], loss of insulin signaling causes reduced but generally well-proportioned growth [16], suggesting a limited requirement for insulin for wing growth orientation. For reasons that are not yet clear, extended proliferation in culture does not occur when adding both insulin and 20E together [5, 13], and therefore I recommend using 20E alone. The conditions described here may more accurately reflect a "starved" disc but otherwise recapitulate the conditions required for proper growth orientation. Thus, the method is well suited to studies of morphogenesis and less useful for studying processes that may be heavily insulin-dependent.

For long-term live imaging, I describe a gentle immobilization technique to reduce drift of the sample during the experiment. This technique involves laying a porous filter over the discs, using a double-sided tape as a convenient spacer to hold the filter in place. The spacer is slightly thicker than the discs to avoid compression, and the filter reduces medium flows enough to keep the discs in place during transport, microscope setup, and imaging. This setup has the added benefit of allowing medium exchange. In addition, all steps of a standard immunofluorescence protocol with paraformaldehyde fixation can be performed with the discs mounted under the filter, and subsequent imaging can then be done in the same dish (as done in ref. 9). Thus, this method has some added benefits compared to others that have been presented in the literature (*see* **Note 1**). It is of course possible to culture the discs off the microscope if live imaging is not required.

With this method, it is possible to directly observe proliferation occurring at a rapid rate in mid-third instar wing discs for up to ~13 h of live imaging [5]. Growth orientation measured directly as the sum of cellular behaviors observed during live imaging is entirely consistent with indirect methods for tracking growth and cell division in vivo [17–22]. Thus, ex vivo culture in 20E can be used to visualize and quantify cellular dynamics – such as divisions, rearrangements, shape changes, and deaths – which are otherwise impossible to fully capture in live animals, even with newer methods [22].

After the original report describing the cell dynamics underlying tissue growth orientation in wild-type wing discs [5], the method was successfully used to study the dynamics of fold formation in very early third instar wing discs [23], the effect of particular genetic mutations on wing growth [24, 25], and the spatial patterns of ATP concentration upon pharmacological inhibition in late third instar wing discs [26]. In addition, by increasing the concentration of 20E, I have had success using the same setup to study imaginal wing and leg eversion, a process whereby the tissues unfold and change shape, removing the peripodial membrane and exposing the apical surface of the disc proper epithelium to the outside. I am optimistic that it can be used to study other imaginal discs, as I have unpublished data indicating that proliferation, morphogenetic furrow progression, and ommatidial rotation occur in cultured mid-third instar eye discs. Thus, the method described here is a useful new tool for studying dynamic cell biological and morphogenetic processes, further enhancing the power of these classic genetically tractable model tissues.

Nonetheless, ex vivo methods always have limits. Proliferation in the described growth medium does not continue indefinitely, even when starting with very young discs, and the reasons why they stop dividing are not yet understood. Using mid-third instar larvae, we never observed a cell divide more than once. This result is not

unexpected, given estimates of the cell cycle duration from indirect methods in well-fed animals [17, 22, 27, 28]. However, it does prevent very long-term analysis of lineage, for example. It must also be noted that this method has not been used to study regenerative growth, although I have occasionally observed some minor wound healing during imaging. I expect that the ex vivo requirements for regenerative growth are actually quite different, as injured discs in vivo secrete a signal, Dilp8, which inhibits ecdysone production [29, 30]. Thus, there is still room for future improvement of ex vivo culture methods. Nevertheless, ex vivo culture should be viewed as a powerful tool, enabling certain types of experiments that are impossible to perform in vivo, either due to physical inaccessibility or complex interactions that are difficult to precisely manipulate genetically.

Here, I present this improved method for imaginal disc cultivation and imaging, which incorporates the knowledge that 20E is required in vivo for proper growth and development. I first describe the preparation of the medium and the sample. For completeness, I include a description of the dissection protocol for the wing disc, as this disc has been used most often with this method. However, other methods of dissection can be used if preferred, particularly if this method is used to study different discs such as the eye. Lastly, I describe the method of sample mounting and provide guidelines for live imaging.

2 Materials

Prepare and store all reagents/materials at room temperature, unless otherwise indicated. Use gloves to protect the sample.

2.1 Culture Medium

1. 0.5 M Bis-Tris stock solution, pH 6.8.
2. Bottle vacuum filters with a hole size of 0.2 μm.
3. Modified Grace's insect medium from Sigma (Table 1 for composition; *see* **Notes 2** and **3**). Add the premixed powder to ~800 mL of deionized water in a large beaker by stirring continuously and, then, rinse the bottle with more water to remove all the powder and add this to the beaker. Adjust the pH to 6.6–6.7 with sodium hydroxide, first with 10 N and then with 1 N as you near the target pH. Be very careful as you approach the target pH, as an extremely high pH will cause precipitants to form and the solution to become cloudy. Add NaOH dropwise while stirring the solution. Add 10 mL of 0.5 M Bis-Tris stock solution and, then, bring the total volume to 1 L with deionized water. Notice that we do not add sodium bicarbonate. Sterilize the filter and store at 4 °C for up to 1 month (*see* **Note 2**).

Table 1
Composition of Grace's Insect medium from Sigma

Component	g L^{-1}
Inorganic salts	
$CaCl_2$ (anhydrous)	1
$MgCl_2$ (anhydrous)	1.068189
$MgSO_4$ (anhydrous)	1.357858
KCl	2.24
NaH_2PO_4	0.876923
Amino acids	
β-Alanine	0.2
L-Alanine	0.225
L-Arginine–HCl	0.7
L-Asparagine	0.35
L-Aspartic acid	0.35
L-Cystine–HCl	0.025
L-Glutamic acid	0.6
L-Glutamine	0.6
Glycine	0.65
L-Histidine	2.5
L-Isoleucine	0.05
L-Leucine	0.075
L-Lysine–HCl	0.625
L-Methionine	0.05
L-Phenylalanine	0.15
L-Proline	0.35
L-Serine	0.55
L-Threonine	0.175
L-Tryptophan	0.1
L-Tyrosine–2Na	0.07202
L-Valine	0.1
Vitamins and others	
D-Biotin	0.00001
Choline chloride	0.0002

(continued)

Table 1
(continued)

Component	g L^{-1}
Folic acid	0.00002
Myo-inositol	0.00002
Niacin	0.00002
D-Pantothenic acid–1/2 Ca	0.00002
p-Aminobenzoic acid	0.00002
Pyridoxine–HCl	0.00002
Riboflavin	0.00002
Thiamine	0.00002
Dextrose anhydrous	0.7
D-(-)-Fructose	0.4
Fumaric acid	0.055
α-Ketoglutaric acid	0.37
L-(-)-Malic acid	0.67
Succinic acid, free acid	0.06
Sucrose	26.68

4. 20-Hydroxyecdysone (20E) stock solution: Dissolve 20E powder in ethanol to make a 2 mM stock solution. Make a 1000× working stock solution of 20uM by diluting the 2 mM stock 1:100 in ethanol. Do not aliquot, as ethanol can easily evaporate at low volumes. Store both in tightly sealed containers at −20 °C.
5. GFBE (*G*race's medium, *FB*S, 20*E*): Modified Grace's insect medium, 5% fetal bovine serum (FBS, *see* **Note 4**), 20 nM 20-hydroxyecdysone (*see* **Note 5**), and 1× penicillin–streptomycin solution for cell culture (100 U penicillin and 0.1 mg mL^{-1} streptomycin at final concentration). Prepare just before use and bring to 25 °C.

2.2 Imaginal Disc Dissection

1. *Drosophila* larvae with the genotype of interest. For live imaging, a suitable fluorescent marker is required. For marking the apical junctional network, E-cadherin–green fluorescent protein (GFP) recombined into its native locus is excellent [5, 31, 32].
2. 70% ethanol (v/v).

3. Glass staining blocks for dissecting and culturing discs off the microscope. We use staining blocks made of clear glass molded with a semicircular depression of 32 mm in diameter. Rinse first with 70% ethanol, followed by sterile water.
4. Two fine dissection forceps: Dumont #55 or similar, rinsed before use with 70% ethanol, water, and medium.
5. Small paintbrush (for handling the larvae).
6. Razor blade.

2.3 Mounting for Live Imaging

1. Uncoated glass-bottom 35 mm culture dishes. We typically use dishes with a 20 mm opening so that there is plenty of room to add the double-sided tape and filter for imaging several discs, but smaller openings are also fine, depending on the size of your hole puncher.
2. Porous filter for immobilization: Whatman cyclopore polycarbonate membranes with a hole diameter of 5–8 μm.
3. Good-quality hole puncher. The device should cleanly cut the edges without too many ripples, which can cause variations in the sandwich thickness or allow bubbles to form. The diameter of the hole can be anywhere from 2 to 6 mm, depending on how many discs you need to image in the same chamber.
4. Photo-quality (acid-free) double-sided tape: Tesa 5338 Doppelband Fotostrip (~100 μm thick).
5. Scissors.

3 Methods

Carry out all procedures at room temperature, unless otherwise indicated.

3.1 Preparation of the Larvae for Dissection

1. Remove *Drosophila* third instar larvae (*see* **Note 6**) from their food vial with a paintbrush and transfer them to a glass staining block containing ~1 mL of water. Transfer up to ~15–20 larvae per dissecting dish (*see* **Note 7**).
2. Stir the larvae with the paintbrush, trying to remove as much food as possible from the body wall.
3. Rinse the larvae by removing the water and food from the glass block and by adding another 1–2 mL of water. This step can be done with a transfer pipette or a p1000. Repeat as needed until all the food is visibly gone.
4. (Optional, *see* **Note 8**): Surface-sterilize the larvae by rinsing with 70% ethanol briefly, followed again by sterile water.
5. Remove the water and replace with the GFBE medium (*see* **Note 9**).

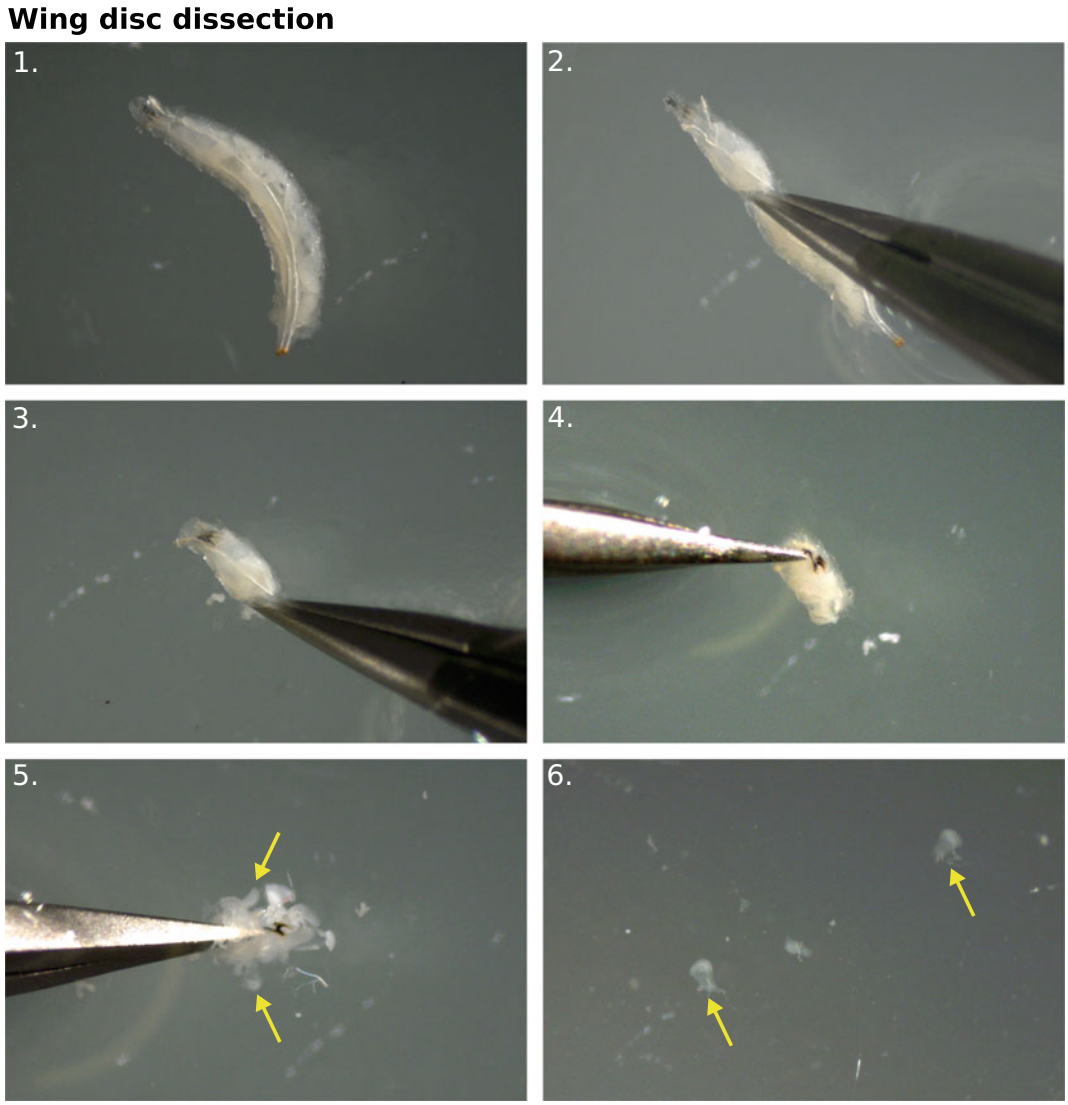


Fig. 1 Dissection of wing imaginal discs. Shown here are the steps of the wing disc dissection described in the text. In this series, the dissection was done by a right-handed person. The right hand grabs the larva (1–2), while the left hand bisects it (3). The right hand then gently inverts the larva over the tip of the forceps in the left hand (4–5). Finally, the wing discs are severed from the larva carefully with the right hand (6). Wing discs are pointed out in 5–6 by the yellow arrows

3.2 Dissection

1. Using forceps with hand 1 (*see* **Note 10**), grab a larva at its midsection (roughly a third of the way from its anterior end) and fully close the forceps firmly (*see* Fig. 1, **steps 1** and **2**).
2. Using hand 2 with another set of forceps, slice off the posterior end of the larva while keeping hand 1 firmly holding the larva (*see* Fig. 1, **step 3**).

3. Remove the posterior part (using hand 2) from the dissecting well.
4. With hand 2, use the forceps to firmly grab the larva's mouth and do not let go (*see* Fig. 1, **step 4**).
5. Release hand 1, causing some of the larva's interior to expel. You can gently remove the pieces of gut and fat body if you wish but be careful not to stretch and pull too much to avoid damaging the imaginal discs.
6. Invert the remaining anterior end of the larva by gently rolling back the larval epidermis onto the forceps in hand 2, which are still firmly holding the mouth (keep forceps in hand 2 closed throughout) (*see* Fig. 1, **step 5**).
7. Locate the discs and gently move the forceps from around the perimeter of the discs to loosen it from the rest of the animal.
8. Separate the discs from the rest of the body wall by clipping them as far up the stalk region as possible, aiming for the least amount of physical disturbance and touching as possible (*see* Fig. 1, **step 6**). Take special care not to pull, pierce, or tear the discs with the forceps. It is not necessary to remove all traces of the trachea.
9. Prepare a pipette tip for disc transfer: cut off the end of a p100 or p200 pipette tip with the razor blade and block the surface by pipetting up and down a few times with FBS so that the discs do not stick to the walls of the pipette tip when you transfer them.
10. Using this blocked pipette tip, transfer the imaginal discs in a volume of ~20 μl to a clean glass block containing ~1 mL of the GFBE medium to rinse the discs.
11. Continue dissecting all the larvae and collecting the discs in the clean glass block (*see* **Note 11**).
12. Wash the discs by rinsing with another 1 mL of the medium.

3.3 Disc Culture Without Imaging

1. If the discs are meant to be cultured off the microscope, replace the medium with 0.5 mL fresh GFBE medium.
2. Place the glass block in a humidified chamber. The humidified chamber can be a simple covered Petri dish with wet paper towels lining the edges to prevent medium evaporation.
3. Place the humidified chamber in a 25 °C incubator for culture (*see* **Note 12**).
4. Remove the discs at the time point of interest and process as needed, depending on the experiment (e.g., immunofluorescence or RNA/protein analysis).

3.4 Preparation of the Live Imaging Chamber

1. (*See* **Notes 13** and **14**) Cut a strip of the double-sided tape, leaving the protective cover on one side (*see* Fig. 2a, **step 1**).
2. Punch a hole in the tape with the hole puncher (*see* Fig. 2a, **step 1**).
3. Trim the tape around the punched hole to generate a piece that can fit within the inner diameter of the exposed cover glass in the glass-bottom dish (*see* Fig. 2a, **steps 1** and **2**).
4. Adhere the hole-punched tape to the cover glass using the exposed sticky side, leaving the protective coating on the other side (*see* Fig. 2a, **step 3**).
5. Make sure that the exposed glass under the hole in the center of the tape is completely free of the tape glue. If not, it is best to start over.
6. With a blunt tool, such as the blunt end of the razor blade or forceps, press on the back of the tape to firmly adhere it to the glass, focusing especially on the edges of the cut hole so that it is nicely flush against the surface of the glass.
7. Remove the protective covering on the tape to expose the second sticky side (*see* Fig. 2a, **step 4**).
8. Trim a piece of the porous filter to roughly match the size of the tape spacer in the dish and set it aside until you are ready for mounting the discs (*see* Fig. 2a, **step 1**).

3.5 Mounting of the Discs in the Live Imaging Chamber

1. Transfer the discs to the center of the hole in the tape using the blocked pipette tip in a volume of ~20 μL (*see* **Note 15**) (*see* Fig. 2b, **steps 1** and **2**).
2. With the pipette tip, gently move the medium around to cover the entire surface enclosed by the hole with the medium. Make sure that you do not touch the tape or make the tape wet, or else it will not stick well to the filter. The medium will make a rounded top (*see* Fig. 2b, **step 3**).
3. Remove ~15 μL of the medium so that the discs are closer to the surface of the glass.
4. Examine the discs under the dissecting microscope and arrange them so that the surface of interest (usually the apical side) is facing down, closest to the glass and the objective.
5. Spread out the discs so that they are not too close together to prevent them from overlapping one another after the filter is added.
6. Acting quickly, carefully remove most of the remaining liquid surrounding the discs (all but a couple of microliters). Then, gently cover them with the filter by first separating the filter from its protective blue cover and then dropping it gently onto the tape, centered on the discs (shiny side down, toward the tape) (*see* **Note 16**) (*see* Fig. 2a, **step 5**; Fig. 2b, **step 4**).

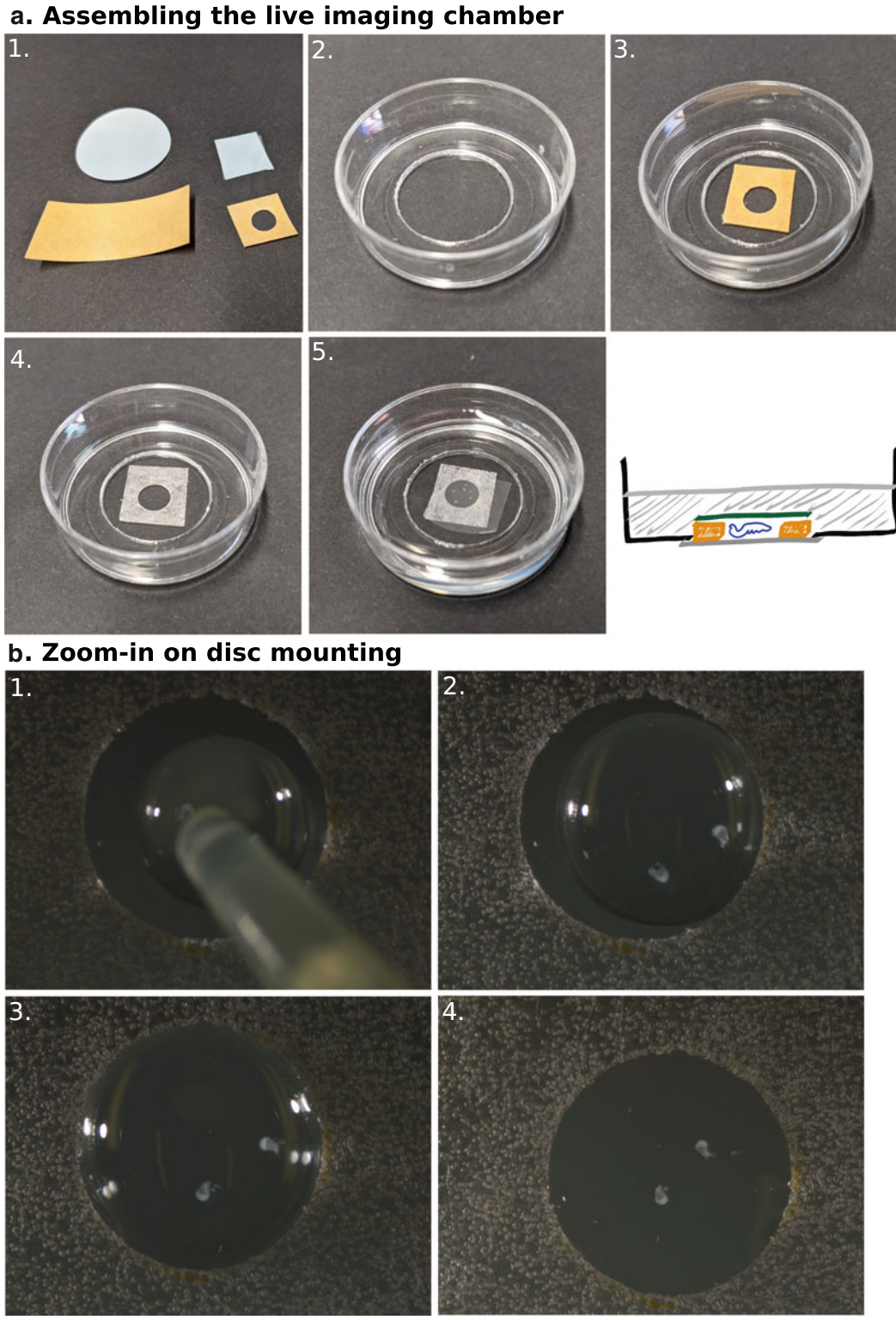


Fig. 2 Assembly of the live imaging chamber and mounting of the discs. Shown here is the assembly and mounting of the discs for live imaging (**a**), with a close-up view of the addition of the sample to the dish in (**b**). (**a**) The materials needed include a glass-bottom dish, a double-sided Tesa tape, and a Whatman cyclopore

7. As soon as you add the filter, press hard on the filter with the blunt end of the razor blade or forceps where it overlaps with the tape to seal the filter firmly to the sticky tape spacer (*see* **Note 17**).
8. Fill the dish with the GFBE medium to sufficiently cover the discs and filter (1–2 mL).

3.6 Live Imaging

1. Preset the microscope stage to 25 °C.
2. Put the live chamber containing the discs on the microscope and bring the discs into focus.
3. Inspect the discs under high magnification, looking for evidence of damage, adherence, or compression (*see* **Note 18**). If found, do not image those discs.
4. Perform time-lapse live imaging as required for the experiment. Most applications will require a spinning disc microscope to image fast enough and expose less light intensity to the sample overall. Take care to minimize light exposure, ideally with a laser power of <0.1 mW (*see* **Notes 19** and **20**).
5. (Optionally): Fix the tissues with 4% paraformaldehyde and perform immunofluorescence according to the standard protocols, keeping the discs immobilized under the filter until ready for imaging.

4 Notes

1. This filter method is similar to that presented in Zartman et al. [10, 33]. Aldaz et al. use methylcellulose to create a very viscous medium that impedes disc movement [4]. I found this to be suboptimal, as it can be difficult to optimally position the wing discs. In order to perform high-resolution imaging, the discs need to be flat, with their apical side down, and very close to the glass coverslip. Dragging them through the viscous methylcellulose medium to position them in this manner is time consuming and often mechanically disruptive for the discs. Thus, I did not achieve good results with this method.

Fig. 2 (continued) filter (1–2). The tape and filter are trimmed to fit inside the dish on the cover glass, and a hole is cut in the tape (1). The tape is attached to the dish on the cover glass (3), and the yellow protective coating is removed (4). The samples are added, followed by the filter (without the blue protective coating). To the right of (5) is a cartoon depiction of the setup, with the spacer in orange, the wing disc (apical side down) in blue, the filter in green, the cover glass in dark gray, and the medium in light gray. (**b**) The discs are added with the blocked pipette (1–2), and the medium is moved around to cover the hole in the tape spacer (3). The excess liquid is removed, and the samples are covered with the filter (4). The whole dish is then filled with the medium

Furthermore, this method does not allow for immunofluorescence post imaging or for medium exchange and rapid pharmacological perturbations. Poly-L-lysine (PLL)-coated glass works to immobilize the discs for short times. It is again sometimes difficult to position the discs optimally with PLL, however, as they tend to stick as soon as they touch the surface and repositioning them is tricky. Tsao et al. imbedded the discs in 0.75% agarose [12]. Such a method may work well for the eye disc, which was their primary goal, but it is less optimal for the wing disc, as it is likely to impose a mechanical barrier to growth.

2. Be aware that the "Grace's Insect medium" is sold by different companies with slightly different compositions. The medium composition that we have tested comes from Sigma and is listed in Table 1, according to their product information sheet. When correctly prepared, the solution will be clear and colorless or very light yellow. Discard if the solution becomes visibly yellow or cloudy. Do not freeze. Although it may be tempting to subdivide the prepared powder mix to make <1 L at a time, this practice is not recommended as it can result in extremely variable and unpredictable medium compositions, with highly detrimental effects on disc growth and viability.
3. The wing disc shrinks and curls slightly in every culture medium that we have tried (including Schneider's, Shields and Sang, and Grace's), independent of hormonal composition. Grace's is the least dramatic, and that is why it was selected. Nonetheless, even in the GFBE medium, there is a consistent shrinkage that is measurable in the apical cell area during the first ~2 h of culture [5]. We call this the "adaption phase." Thereafter, size stabilizes and then grows in GFBE. It is always recommended to dissect in a full medium, rather than in phosphate-buffered saline (PBS), even if culture times are kept short, in order to minimize any disruptions due to changes in the medium's environment.
4. Other groups have reported adding a fly extract to a growth medium (an aqueous extract of homogenized adult female flies) rather than FBS [10–12]. We have found the fly extract to be far less optimal for growth and much more variable.
5. Our original report [5], as well as others that have come later [9, 23, 25], uses 20 nM 20E for live imaging of proliferation in wing discs younger than ~100 h after egg laying (AEL). Zhou et al. [24], however, thought they could see premature eversion of older discs with 20 nM 20E (personal communication) and used instead 10 nM 20E to visualize cell divisions during live imaging. To induce eversion morphogenesis of late third instar wing and leg discs, I have had consistent results using 400 nM 20E (added from the 2 mM stock solution).

6. Younger larvae that are still in the food can be easily isolated by floating them up in 30% sucrose or 30% glycerol. Add this solution to the food vial (breaking up the food with a transfer pipette) or to the glass block and wait a minute or two for the food to sink (the larvae will float). Wash the larvae thereafter in water.
7. You can dissect one animal at a time while all the larvae share the same well. Alternatively, you can move a single larva to a separate well for dissection so that it is alone in the dissecting well and you are not disturbed by other larvae in the dish. It depends purely on your preference and how proficient you are at dissecting live animals. The larvae should not spend more than ~15 min in the dish.
8. Long-term disc culture is prone to microbial contamination, even with antibiotics added to the medium. A brief surface sterilization with 70% ethanol helps alleviates this problem, particularly for yeast contamination likely originating from the food. This step was used in our original report without visible disruption to wing disc proliferation [5]. However, subsequently, our group noticed that this rinse causes mitochondrial fragmentation, even when extremely brief. Thus, for short-term cultures or experiments investigating metabolism in the discs, this step should be omitted, and extra care should be taken to dissect and handle the discs with sterile techniques.
9. Many people chill the larvae by placing them on ice for some time before dissection. This immobilizes the larvae and makes them easier to dissect. Here, however, we want to perturb the imaginal discs as little as possible to prevent any growth delay or defect. Thus, we dissect them live, at room temperature.
10. I describe the dissection protocol with "hand 1" and "hand 2," as people can have different dominant hands. If you are right-handed, hand 1 is your right and hand 2 is your left. In this way, you will be using your dominant hand for the delicate work of removing the discs from the body wall.
11. Keep in mind that the more larvae you dissect, the longer the discs will be waiting in the culture medium. Keep track of how long it takes to dissect all the larvae so that the variation around the "time 0 h" start of culture is known.
12. In our original report [5], we replaced the culture medium every 2–8 h when growing off the microscope and circulated fresh medium during the course of the live imaging. We did this to ensure that the medium was not being depleted of any nutrients or that there was no toxic component accumulating. Nonetheless, we have subsequently performed live imaging experiments without circulating the medium, and we did not see dramatic differences [9]. Thus, medium circulation may

not be necessary. Interestingly, Strassburger and Lorbeer et al. report that wing discs can become hypoxic within 3 h of culture and arrest growth when the medium is not mixed [13]. Their base medium was different from ours, however, and they did not include a serum supplement. Our transcriptomic data on discs cultured in GFBE do not suggest that they become hypoxic [5].

13. Preparation of the live chambers can and should be done ahead of time to reduce the time that the discs are sitting in culture waiting to be mounted. Use gloves whenever handling any of the parts to prevent contamination.
14. Typically, one layer of tape is sufficient, but if you are imaging larger discs (for example, for eversion) and want less chance of the discs being compressed, you can make a double layer. Cut two pieces of the tape and adhere them together with their exposed sticky sides. Then, punch a hole, trim to fit in the dish, remove one side of the protective covering, and adhere it to the dish, just as described for the single layer in **steps 2**–**7**.
15. The volumes needed will depend on the size of the punched hole in the tape spacer as well as the thickness of the tape spacer (whether you have a single layer of tape or double). Here, we report volumes for a 6 mm hole with a single layer of a sticky tape spacer. Likewise, the number of discs you can add depends on the size of the hole. For a 6 mm hole, do not add more than 10–12 discs; otherwise, they may be too close to one another or may overlap one another once the filter is added.
16. This step is a little tricky and requires some practice. Aim for leaving a visible thin layer of the medium. The discs should not move if you move the surface of the liquid using forceps or a pipette, but there should still be some liquid. If you remove too much of the medium before adding the filter, you risk the discs drying out or being compressed too much by the filter. If you remove too little of the medium, however, the filter may become wet and not properly seal to the tape, causing the discs to drift during imaging. Removing too much or too little of the medium can also cause air bubbles to form when adding the filter. Note, however, that the filter does not need to be exactly aligned to the tape spacer when you place it – it just needs to cover the hole containing the discs. For example, notice that the filter is not exactly aligned to the tape in Fig. 2a, **step 5**. That is fine, as it still covers the entire hole.
17. If you accidentally fold the filter while adhering it to the tape, it is best to start over, since the ripples will cause the discs to drift during imaging. If you only have a few air bubbles at the interface between the filter and the tape, it is usually not a problem.

18. Damage can be a hole or a tear in the tissue from dissection. Compression of or adherence to the filter can sometimes be seen when the discs are observed with transmitted light, where the discs look stretched between the holes of the filter or the thickness of the disc is considerably less than it should be (normally ~30–50 μm, depending on the age of the disc). Excessive compression of the discs can cause cell death and cell division arrest in the wing disc and arrest of the morphogenetic furrow progression in the eye disc. Sometimes, however, these defects are only apparent after projection and analysis of the time-lapse movie.
19. Light exposure is an extremely important consideration [34]. Most often, when live imaging experiments do not produce good results, it is because the discs received too much light. Spinning disc microscopes are much preferred over traditional point scanning confocals, but even then, care must be taken. To determine whether light damage is producing suboptimal results, one can either compare drastically different frame rates or even prepare two samples, with one being imaged and the other kept off the microscope.
20. As a reference, our long-term live imaging of wing disc cell dynamics during growth was performed using Z-stacks of 65–85 frames captured with a 350 ms exposure time every 5 min for ~13 h, using a laser power of 0.04–0.05 mW [5]. We also tiled 2 × 2 across the wing disc to make sure that the disc was captured in the field of view even if there were any drifts or movements.

References

1. Beira JV, Paro R (2016) The legacy of Drosophila imaginal discs. Chromosoma 125: 573–592
2. Fristrom JW, Logan WR, Murphy C (1973) The synthetic and minimal culture requirements for evagination of imaginal discs of Drosophila melanogaster in vitro. Dev Biol 33: 441–456. https://doi.org/10.1016/0012-1606(73)90149-8
3. Milner MJ (1977) The eversion and differentiation of Drosophila melanogaster leg and wing imaginal discs cultured in vitro with an optimal concentration of β-ecdysone. Development 37(1):105–117
4. Aldaz S, Escudero LM, Freeman M (2010) Live imaging of Drosophila imaginal disc development. Proc Natl Acad Sci U S A 107: 14217–14222. https://doi.org/10.1073/pnas.1008623107
5. Dye NA, Popović M, Spannl S et al (2017) Cell dynamics underlying oriented growth of the Drosophila wing imaginal disc. Development 144:4406–4421. https://doi.org/10.1242/dev.155069
6. Parker J, Struhl G (2020) Control of Drosophila wing size by morphogen range and hormonal gating. Proc Natl Acad Sci 117(50): 31935–31944. https://doi.org/10.1073/pnas.2018196117
7. Brennan CA, Ashburner M, Moses K et al (1998) Ecdysone pathway is required for furrow progression in the developing Drosophila eye. Development 125:2653–2664. https://doi.org/10.1016/s0012-1606(74)80016-3
8. Brennan CA, Li TR, Bender M et al (2001) Broad-complex, but not ecdysone receptor, is required for progression of the morphogenetic furrow in the Drosophila eye. Development

128:1–11. https://doi.org/10.1016/s0092-8674(00)80094-x
9. Dye NA, Popovic M, Venkatesan Iyer K et al (2021) Self-organized patterning of cell morphology via mechanosensitive feedback. Elife 10:e57964. https://doi.org/10.7554/eLife.57964
10. Zartman J, Restrepo S, Basler K (2013) A high-throughput template for optimizing Drosophila organ culture with response-surface methods. Development 140(3):667–674
11. Handke B, Szabad J, Lidsky PV et al (2014) Towards long term cultivation of Drosophila wing imaginal discs in vitro. PLoS One 9: e107333. https://doi.org/10.1371/journal.pone.0107333
12. Tsao C-K, Ku H-Y, Lee Y-M et al (2016) Long term ex vivo culture and live imaging of drosophila larval imaginal discs. PLoS One 11: e0163744. https://doi.org/10.1371/journal.pone.0163744
13. Strassburger K, Lorbeer FK, Lutz M et al (2017) Oxygenation and adenosine deaminase support growth and proliferation of ex vivo cultured Drosophila wing imaginal discs. Development 144:2529–2538. https://doi.org/10.1242/dev.147538
14. Britton JS, Edgar BA (1998) Environmental control of the cell cycle in Drosophila: nutrition activates mitotic and endoreplicative cells by distinct mechanisms. Development 125(11):2149–2158
15. Cheng LY, Bailey AP, Leevers SJ et al (2011) Anaplastic lymphoma kinase spares organ growth during nutrient restriction in drosophila. Cell 146:435–447. https://doi.org/10.1016/j.cell.2011.06.040
16. Rulifson EJ, Kim SK, Nusse R (2002) Ablation of insulin-producing neurons in flies: growth and diabetic phenotypes. Science 296(5570): 1118–1120
17. González-Gaitán M, Capdevila MP, García-Bellido A (1994) Cell proliferation patterns in the wing imaginal disc of Drosophila. Mech Dev 46:183–200. https://doi.org/10.1016/0925-4773(94)90070-1
18. Resino J, Salama-Cohen P, García-Bellido A (2002) Determining the role of patterned cell proliferation in the shape and size of the Drosophila wing. Proc Natl Acad Sci U S A 99: 7502–7507. https://doi.org/10.1073/pnas.072208199
19. Baena-López LA, Baonza A, García-Bellido A (2005) The orientation of cell divisions determines the shape of Drosophila organs. Curr Biol 15:1640–1644. https://doi.org/10.1016/j.cub.2005.07.062
20. Mao Y, Tournier AL, Hoppe A et al (2013) Differential proliferation rates generate patterns of mechanical tension that orient tissue growth. EMBO J 32:2790–2803. https://doi.org/10.1038/emboj.2013.197
21. Worley MI, Setiawan L, Hariharan IK (2013) TIE-DYE: a combinatorial marking system to visualize and genetically manipulate clones during development in Drosophila melanogaster. Development 140(15):3275–3284
22. Heemskerk I, Lecuit T, LeGoff L (2014) Dynamic clonal analysis based on chronic in vivo imaging allows multiscale quantification of growth in the Drosophila wing disc. Development 141(11):2339–2348
23. Sui L, Alt S, Weigert M et al (2018) Differential lateral and basal tension drive folding of Drosophila wing discs through two distinct mechanisms. Nat Commun 9:4620. https://doi.org/10.1038/s41467-018-06497-3
24. Zhou Z, Alégot H, Irvine KD (2019) Oriented cell divisions are not required for Drosophila wing shape. Curr Biol 29:856–864, e3. https://doi.org/10.1016/j.cub.2019.01.044
25. Lavalou J, Mao Q, Harmansa S et al (2021) Formation of polarized contractile interfaces by self-organized Toll-8/Cirl GPCR asymmetry. Dev Cell 56:1574–1588, e7. https://doi.org/10.1016/j.devcel.2021.03.030
26. Spannl S, Buhl T, Nellas I et al (2020) Glycolysis regulates Hedgehog signalling via the plasma membrane potential. EMBO J 39: e101767. https://doi.org/10.15252/embj.2019101767
27. Garcia-Bellido A, Merriam JR (1971) Parameters of the wing imaginal disc development of Drosophila melanogaster. Dev Biol 24: 61–87. https://doi.org/10.1016/0012-1606(71)90047-9
28. Bryant PJ, Levinson P (1985) Intrinsic growth control in the imaginal primordia of Drosophila, and the autonomous action of a lethal mutation causing overgrowth. Dev Biol 107: 355–363. https://doi.org/10.1016/0012-1606(85)90317-3
29. Colombani J, Andersen DS, Léopold P (2012) Secreted peptide Dilp8 coordinates drosophila

tissue growth with developmental timing. Science 336(6081):582–585

30. Garelli A, Gontijo AM, Miguela V et al (2012) Imaginal discs secrete insulin-like peptide 8 to mediate plasticity of growth and maturation. Science 336:579–582. https://doi.org/10.1126/science.1216735
31. Huang J, Zhou W, Dong W et al (2009) Directed, efficient, and versatile modifications of the Drosophila genome by genomic engineering. Proc Natl Acad Sci 106:8284–8289. https://doi.org/10.1073/pnas.0900641106
32. Etournay R, Popović M, Merkel M et al (2015) Interplay of cell dynamics and epithelial tension during morphogenesis of the Drosophila pupal wing. elife 4:e07090. https://doi.org/10.7554/eLife.07090
33. Restrepo S, Zartman JJ, Basler K (2016) Cultivation and live imaging of drosophila imaginal discs. Methods Mol Biol 1478:203–213. https://doi.org/10.1007/978-1-4939-6371-3_11
34. Icha J, Weber M, Waters JC, Norden C (2017) Phototoxicity in live fluorescence microscopy, and how to avoid it. BioEssays 39(8):1700003

Chapter 17

Sample Preparation and Imaging of the Pupal *Drosophila* Abdominal Epidermis

Daiki Umetsu

Abstract

The epithelium is one of the best studied tissues for morphogenesis, pattern formation, cell polarity, cell division, cell competition, tumorigenesis, and metastatic behaviors. However, it has been challenging to analyze real-time cell interactions or cell dynamics within the epithelia under physiological conditions. The *Drosophila* pupal abdominal epidermis is a model system that allows to combine long-term real-time imaging under physiological conditions with the use of powerful *Drosophila* genetics tools. The abdominal epidermis displays a wide range of stereotypical characteristics of the epithelia and cellular behaviors including cell division, cell death, cell rearrangement, apical constriction, and apicobasal/planar polarity, making this tissue a first choice for the study of epithelial morphogenesis and relevant phenomena. In this chapter, I describe the staging and mounting of pupae and the live imaging of the abdominal epidermis. Moreover, methods to combine live imaging with mosaic analysis or drug injection will be presented. The long-term live imaging of the pupal abdominal epidermis is straightforward and opens up the possibility to analyze cell dynamics during epithelial morphogenesis at an unprecedented resolution.

Key words Live imaging, Epithelium, Tissue remodeling, Histoblasts, Proliferation, Apoptosis

1 Introduction

The epidermis of the *Drosophila* pupal abdomen is emerging as an excellent model to study the various aspects of epithelial tissue morphogenesis at both the cellular and the tissue-wide scales. The adult epidermis of the abdomen is derived from imaginal cells called histoblasts, which are born during embryogenesis, being mitotically quiescent until the onset of the pupal stage [1]. Histoblasts form discrete cell groups, called histoblast nests, which are embedded in the larval epidermal tissue during development [2, 3] (*see* Fig. 1). Meanwhile, larval epidermal cells (LECs) cover the

Supplementary Information The online version contains supplementary material available at [https://doi.org/10.1007/978-1-0716-2541-5_17].

Christian Dahmann (ed.), *Drosophila: Methods and Protocols*, Methods in Molecular Biology, vol. 2540,
https://doi.org/10.1007/978-1-0716-2541-5_17,

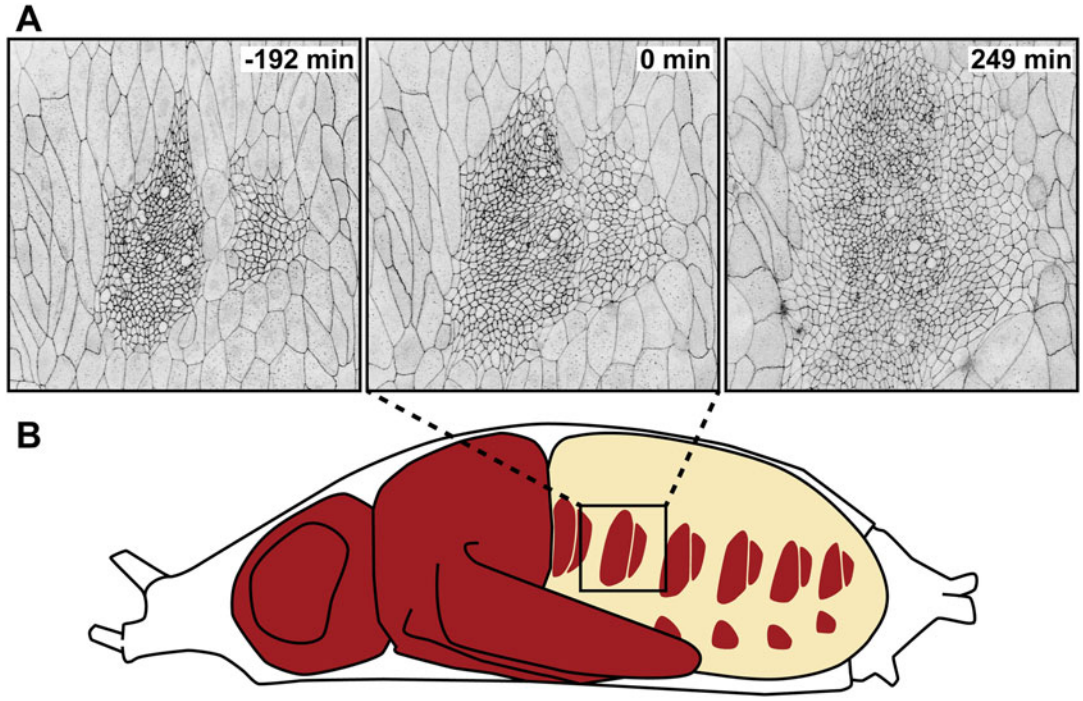


Fig. 1 Abdominal epidermis in the *Drosophila* pupa. (**a**) Proliferating histoblasts. Anterior and posterior dorsal histoblast nests are shown. Smaller cells in the center are histoblasts, and the surrounding larger cells are larval epidermal cells (LECs). Times indicated at the upper right corner represent the time to the fusion of the anterior and posterior dorsal histoblast nests. Cells are visualized with the E-cadherin::green fluorescent protein (GFP) knock-in [27]. (**b**) Illustration of the pupa around the time of the anterior and posterior dorsal histoblast nest fusion. Histoblast nests are diploid cells like the cells consisting of other body parts at this stage. They are highlighted in brown. LECs are polypoloid cells and shown in beige. Anterior is to left, dorsal is up in all panels

epidermis of the larval abdomen [2, 3]. As the histoblast nests expand, driven by both vigorous cell proliferation and migration, the surrounding LECs are removed from the epithelium by apoptosis and extrusion [2, 4]. This tissue remodeling is a rapid process, taking only 20 h from the onset of the histoblast proliferation to the complete removal of LECs [2, 5].

Since cellular events such as cell proliferation and cell death can be easily monitored in histoblast nests (*see* Fig. 1), live imaging makes this tissue a very powerful system to study basic cellular activities and tissue-scale morphogenesis. Unlike most of the other live imaging systems, this tissue does not need any cultivation buffers. The live imaging of this tissue can be performed for 24 h and potentially more. The pupa hatch after imaging without any problem in their morphology or health states. Studies of abdominal epidermis development have revealed an intriguing coordination between histoblasts and LECs. The proliferation of histoblasts and the removal of LECs are tightly coupled. When the apoptosis of

LECs is prevented by overexpressing p35, which is a baculoviral protein that inhibits caspases, the cell growth of histoblasts is nonautonomously reduced and a frequent histoblast delamination is induced [6, 7]. On the other hand, when the proliferation of histoblasts is blocked by genetic means or an optical method, the extrusion of LECs is nonautonomously suppressed [7, 8]. This tight coordination would guarantee the preservation of tissue integrity in accidental situations. Although the detailed mechanisms which coordinate histoblast proliferation and LEC elimination are yet to be revealed, it is believed to occur through both mechanical and chemical interactions.

Cell sorting is another interesting process that benefits from the study in histoblasts [9, 10]. The dorsal part of the abdominal segments is derived from anterior and posterior dorsal histoblast nests (*see* Fig. 1b). Cells from these nests are not intermingled and maintain a sharp boundary between them, establishing anterior and posterior compartments. The establishment of straight boundaries is commonly found in many other tissues and is important for tissue patterning since compartment boundaries act as a signaling center [11–13]. The apoptosis and extrusion process of LECs have also been studied. Image analysis is relatively easy for LECs since they are gigantic (up to 70 micrometers in diameter). For example, the extrusion process has been extensively studied by quantitatively measuring area reduction [5, 14]. LECs also display an intriguing migratory behavior, which is planar oriented [3, 15].

The combination of live imaging and *Drosophila* genetics renders the tissue an even more attractive system for wider applications. Powerful *Drosophila* genetics such as mosaic clone analysis has been used in the studies of this tissue [16–20]. The molecular mechanisms that regulate planar cell polarity [17, 19], patterning [18], coordination of cell cycle progression and cell growth [4], uniform cell alignments [21], and cell sorting [10, 16] have been identified using mosaic analysis of this tissue. Drug application, which enables an acute inhibition of protein functions, is also applicable [10]. Moreover, laser ablation methods, a new biophysics approach, have been introduced to reveal the mechanical properties of histoblasts and LECs [7, 9, 14, 21, 22]. Many more interesting biological processes observed in this tissue can be considered for a detailed study in the future, including bristle formation, fold formation, and pigmentation, which are often challenging to study live in other tissues. The availability of a wide range of techniques in the tissue raises the stakes for a better understanding of tissue morphogenesis at the cellular and molecular levels.

Not only live imaging but also classical imaging methods such as whole-mount immunohistochemistry have been performed, and such a protocol is available online (https://www.jove.com/t/3139/drosophila-pupal-abdomen-immunohistochemistry). Please

refer to the online protocol for the dissection and immunohistochemistry of fixed samples. This chapter features the live imaging technique for the study of histoblasts and LECs in combination with other methods. The major procedures described here will be live imaging and its combination with mosaic clone analysis and drug application. Mosaic clone analysis allows the tissue-specific knockdown or overexpression of any genes of interest, and drug application allows the acute inhibition of protein function at the time of imaging. The combinatory usage of these methods enables the sophisticated analysis of gene and protein function in regulating cell dynamics and tissue morphogenesis.

2 Materials

2.1 Live Imaging of Histoblasts and LECs

1. A fly line harboring fluorescently tagged protein(s) for visualization of the tissues of interest (Table 1).
2. Stereomicroscope.
3. Forceps.
4. Adhesion tape (15 mm width).
5. Double-sided adhesion tape (12 mm width).
6. Scissors.
7. Thin paintbrush.
8. Delicate task wipe (Kimwipes).
9. Coverslip (24 mm × 32 mm or wider).
10. Confocal microscope.
11. 35 mm petri dish.
12. Deionized water.
13. Water bath (for live imaging in combination with mosaic analysis).

2.2 Live Imaging in Combination with Drug Application

In addition to the materials listed in Subheading 2.1:

1. Glass capillary (e.g., Microcaps 30 μL, Drummond).
2. Rotating grinder (e.g., Microgrinder EG-400, NARISHIGE).
3. Microinjector (e.g., FemtoJet 4i, Eppendorf).
4. Micromanipulator (e.g., M-152, NARISHIGE).
5. Chromic acid mixture (H_2SO_4 58.0–59.6%, CrO_3(IV) 2.3–2.7 w/v%, toxic).

Table 1
A list of fly lines useful for the visualization of histoblasts or LECs

Fly line	Targeted tissue	Advantages	Weakness	References
esg-GAL4	Histoblasts	Highly specific to histoblasts	Expression level reduces at later stages	[4, 14, 22, 28, 29]
Ecad::GFP knock-in	Histoblasts, LECs	Visualizing cell outline	Expression level reduces in LECs at later stages	[5, 7, 14, 22]
Nrg::GFP protein trap	Histoblasts, LECs	Visualizing cell outline	A little bit fuzzy	[5, 14]
pnr-GAL4	LECs	Very strong	Expression is missing in a few rows at the segment boundaries. Initially specific to LECs but eventually expressed in histoblasts in later stages	[5, 15]
Eip71CD-GAL4	LECs	Specific to LECs; broader in each segment than *pnr*-GAL4	Slightly weaker than *pnr*-GAL4	[14, 30]
32B-GAL4	LECs	n.d.	n.d.	[7, 31]
tsh-GAL4	LECs	Very strong compared to *pnr*-GAL4; Broader broader expression in LECs than *pnr*-GAL4	Initially specific to LECs but eventually expressed in histoblasts at later stages (similar to *pnr*-GAL4); Expressed expressed in many other tissues	[14, 22]

3 Methods

3.1 Live Imaging

3.1.1 Cultivation of Properly Staged Pupae

1. Make and amplify a fly line, which visualizes histoblasts or LECs with the expression of fluorescent proteins (Table 1).
2. Mark outside of a fly vial where prepupae are present.
3. Pour deionized water in a 35 mm dish to fill about a half of its volume.
4. Wet a thin paintbrush with water in the dish.
5. Use the wet paintbrush to place a drop of water onto each prepupa and wait for a minute until the pupa can be easily taken away from the vial.
6. Pick the wet pupa from the fly vial with the paintbrush and put it into the dish filled with water (*see* Fig. 2a).

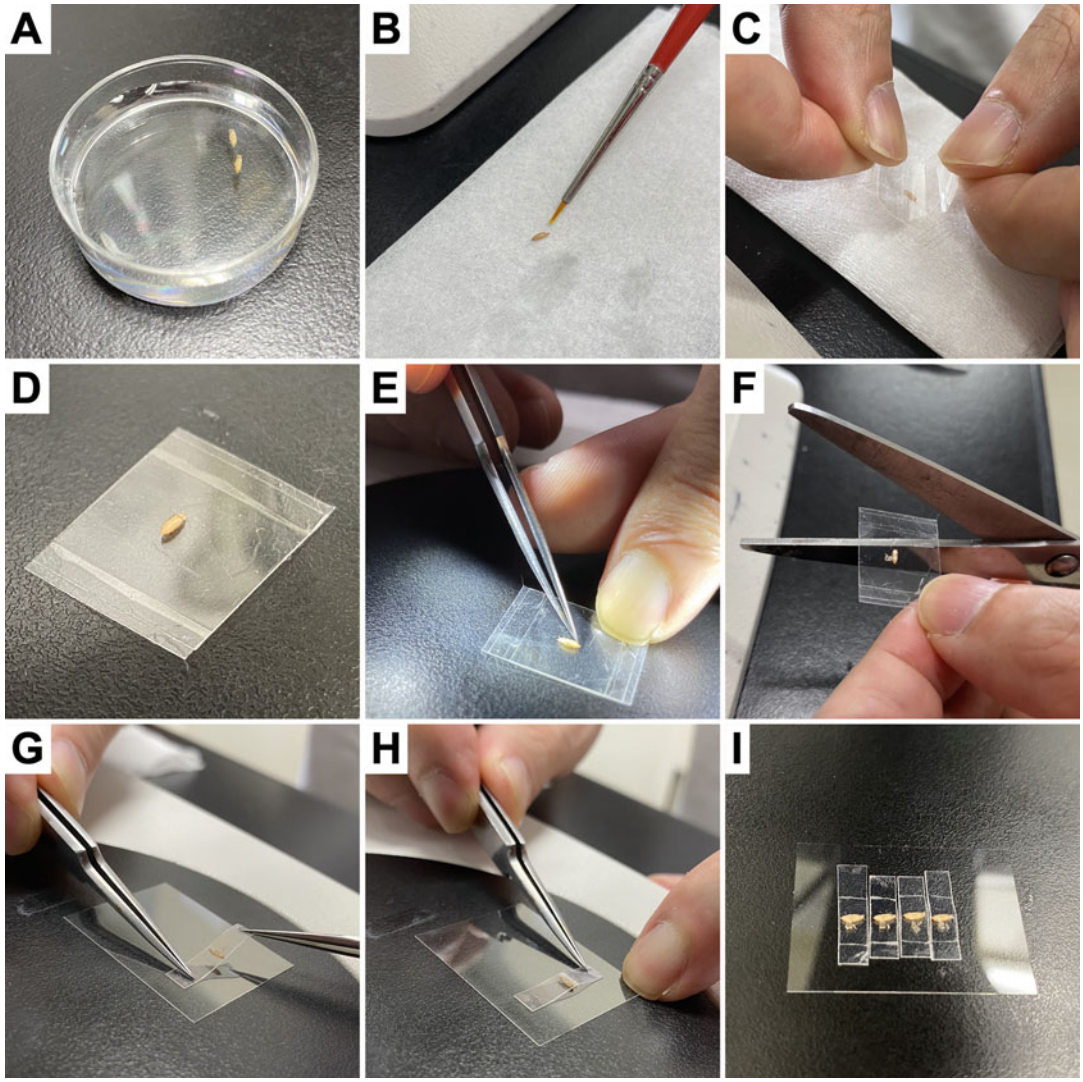


Fig. 2 Flow of sample preparation for live imaging of the abdominal epidermis. (**a**–**b**) Staged pupae are washed in water (**a**) and then dried (**b**) before dissection. (**c**) The pupa is placed onto a dissection tape. (**d**) The pupa is placed on the tape at a right angle to the dorsal midline facing up for the imaging of LECs. Ventral side is down. (**e**) Dissection with forceps on the dissection tape. (**f**) The tape is cut into a narrow strip. (**g**–**h**) Place the strip onto a coverslip. (**i**) Pupae are mounted for multiposition imaging, which enables automated live imaging of multiple samples on a slide

7. Let the pupa soak in water for a minute so that water permeates into the debris sticking on the pupal case.
8. Gently brush the pupa in water with the paintbrush to remove the debris.
9. Stick the pupa onto the wall of a new fly food vial (*see* **Note 1**).
10. Place the vial in an incubator at a proper temperature for the desired time period.

3.1.2 Dissecting Staged Pupae

1. Make a stack of three adhesion tapes. Cut the tape stack into a small piece of about 1 cm long. This is the dissection tape used for dissection of pupae (*see* **Note 2**). Use one piece of the dissection tape for each pupa.
2. Pour water in a 35 mm dish to fill about a half of its volume.
3. Wet a thin paintbrush with water in the dish.
4. Use the wet paintbrush to wet a pupa and wait for a minute until the pupa can be easily taken away from the vial.
5. Pick the pupa with the paintbrush and place it onto a piece of a delicate task wipe (*see* Fig. 2b).
6. Roll the pupa on the delicate task wipe with the paintbrush to remove water and the remaining debris (*see* **Note 3**).
7. Dry the paintbrush using the delicate task wipe.
8. Lay the pupa on its dorsal side on the delicate task wipe (*see* Fig. 3a, b, **Note 4**).
9. Take the dissection tape and gently touch the pupa with the adhesive side so that the pupa sticks on the tape on its ventral side (*see* Fig. 2d, **Note 5**).
10. Place the dissection tape with the pupa's dorsal side up under a stereomicroscope (*see* Fig. 2d).
11. Take the forceps with your right hand (if you are right-handed) and hold the dissection tape with your left index finger (use the right index finger if you are left-handed) and, then, navigate to the field of view under the microscope (*see* Figs. 2e and 3c).
12. Scratch or gently poke the pupal case near the joint between the thorax and the abdomen where there is a space between the pupal body and the pupal case (*see* Fig. 3d, Video 17.1) with the forceps to make a tiny hole (*see* **Note 6**).
13. Insert the tip of the forceps and tear the pupal case little by little toward the tail to about three-quarters of the abdomen (*see* Fig. 3e).
14. Strip off the torn pupal case by grabbing it with the forceps. Place the dissected pupa onto the tape on the other side of the body to expose the abdominal epidermis (*see* Fig. 3f, g).

3.1.3 Mounting and Live Imaging

1. Cut the dissection tape into narrow strips with the width of the pupal body length (*see* Fig. 2f).
2. Place the tape strips with pupae onto a wide coverslip (*see* Fig. 2g–i).
3. Refill the 35 mm dish with clean water.
4. Mount the dish onto an upright confocal microscope stage (*see* **Note 7**).
5. Place the coverslip with pupae on top of the dish in a hanging position.
6. Start imaging.

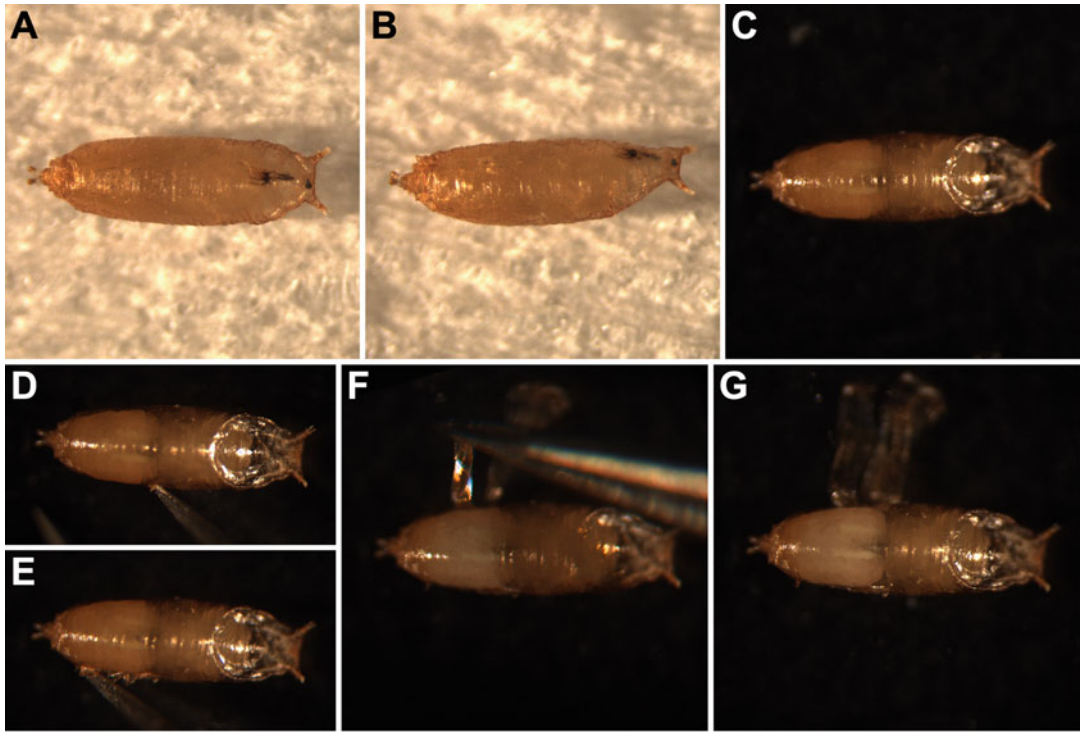


Fig. 3 Dissection of the pupa for live imaging of the abdominal epidermis. (**a**–**b**) Pupa placed on its dorsal side. For the imaging of the LECs, the pupa is placed on the position at a right angle to the dorsal midline facing up when placed on the adhesion tape (**a**). For the imaging of the entire anterior and posterior dorsal histoblast nests, the pupa is placed in a position slightly tilted (**b**). (**c**) Pupa pasted on the adhesion tape at a right angle to the dorsal midline facing up. (**d**–**e**) Removal of the dorsal part of the pupal case of the abdomen. Inserting the tip of the forceps (**d**) and tearing the pupal case little by little toward the tail (**e**). (**f**) Removal of the pupal case. Strip off the pupal case with the forceps and place it on the tape on the other side of the pupa. (**g**) The abdominal epidermis is exposed and is ready for imaging

3.2 Live Imaging in Combination with Mosaic Analysis

Application of mosaic clone analyses is a strength of this tissue as a model system of epithelial morphogenesis. Mosaic analyses broaden the application of live imaging. The conditions for clone induction in histoblasts and LECs are different due to the difference in ploidy and the timing of proliferation. Interchromosomal recombination-based methods such as twin spot analysis and MARCM (*m*osaic *a*nalysis with a *r*epressible *c*ell *m*arker) [23, 24] are useful for the analysis of gene function specific to histoblasts, which are highly proliferative in later development. Marked clones of cells can be also generated in LECs with this method. Flippase (FLP)-out techniques [25, 26] can be used in both histoblasts and LECs. The clones can be generated predominantly in LECs since LECs are more prone to generate clones due to their polyploidy (more copies of FLP-out elements in each cell).

3.2.1 Interchromosomal Recombination (Twin Spot Analysis and MARCM)

1. Set up an appropriate cross to perform clonal analysis with the flippase recognition target (FRT)/FLP system or similar techniques.
2. Let the flies lay eggs in a normal fly food vial and grow them to third instar larvae.
3. Heat-shock at 37 °C for 90 min (*see* **Note 8**).
4. Pick up prepupae for proper staging and dissect them at the timing of interest.
5. Follow the protocol for live imaging as detailed in Subheading 3.1.

3.2.2 Overexpression Analysis with FLP-Out Clones

1. Set up an appropriate cross to perform clonal analysis with an FLP-out system (e.g., Act5C > CD2 > GAL4).
2. Let the flies lay eggs in a normal fly food vial and grow them for 2 more days.
3. Heat-shock at 37 °C for 8 min for LECs and at 37 °C for 10 min for histoblasts (*see* **Note 9**).
4. Pick up prepupae and dissect them at the timing of interest.
5. Follow the protocol for live imaging as detailed in Subheading 3.1.

3.3 Live Imaging in Combination with Drug Application

Drug application makes it possible to acutely inhibit protein function in vivo. The effect can be analyzed live in combination with the live imaging, making it a powerful method for the analysis of protein function in the regulation of cell dynamics and morphogenesis.

3.3.1 Needle Preparation

1. Prepare a capillary needle using a micropipette puller.
2. Grind the tip of the capillary needle to make a hole and sharpen using a rotating grinder equipped with a stereomicroscope (*see* Fig. 4a).
3. Set the sharpened capillary needle to a micromanipulator under the microscope and wash the tip in a drop of chromic acid mixture by repeatedly pulling in and pushing out the solution and then rinsing in a drop of water.

3.3.2 Drug Injection

1. Follow the protocol detailed in Subheadings 3.1.1 and 3.1.2 for staging of pupae and dissection, respectively.
2. Fill the capillary needle with a chemical solution of interest using a microloader tip.
3. Set up the capillary needle on a microinjector device (*see* Fig. 4b, e.g., FemtoJet 4i, Eppendorf).
4. Paste a double-sided adhesion tape on a microscope slide (*see* Fig. 4c).

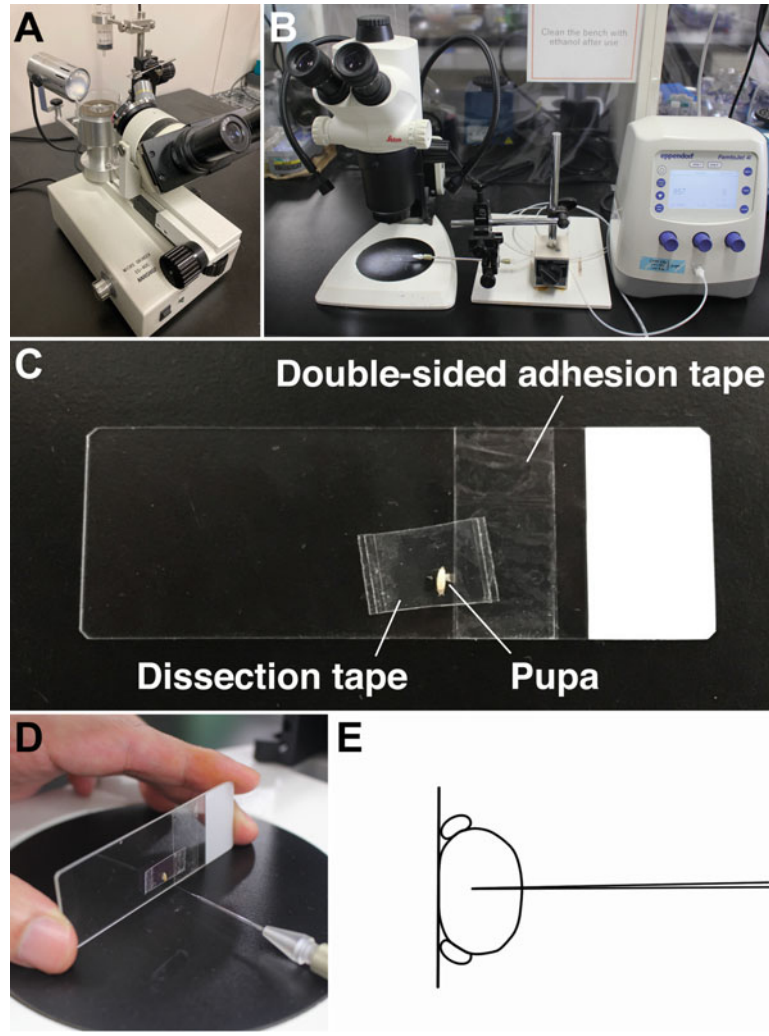


Fig. 4 Drug injection in the pupal abdomen. (**a**) Microgrinder used for sharpening capillary needles. (**b**) Injection setup. Stereomicroscope, micromanipulator, and microinjector from left to right. (**c**) Sample preparation for injection. (**d**) Injecting drug under the stereomicroscope. (**e**) Schematic illustration of a pupa stabbed with a capillary needle for drug injection viewed from top

5. Place the dissection tape with a pupa on the microscope slide (*see* Fig. 4c).
6. Hold the microscope slide with the pupa and push it toward the capillary needle. Stab the needle into the dorsal abdomen at an angle perpendicular to the body axis (*see* Fig. 4c, d).
7. Inject the drug as soon as the capillary needle is inserted into the pupa.
8. Pull out the capillary needle gently while applying pressure on the needle (*see* **Note 10**).

9. Start imaging by following the protocol detailed in Subheading 3.1.3.

4 Notes

1. For the incubation of pupae, use a normal fly food vial. The food provides pupae with proper humidity.
2. Prepare two pieces of the single-sided adhesion tape of about 15 cm long and a piece of the double-sided adhesion tape with the length slightly longer than that of the single-sided tape. Place a piece of the single-sided tape on a flat surface (adhesive side is up) and stack the double-sided tape on top of it. Stack the other piece of the single-sided tape on top of the stack of the tapes with the same orientation (adhesive side is up) so that the double-sided adhesion tape is sandwiched between two single-sided adhesion tapes.
3. Debris should be cleaned completely. Remaining debris prevents the pupa from sticking well to the dissection tape.
4. Make sure that the pupa is oriented right up without any angle for the imaging of LECs. For histoblast imaging, tilt the pupa by about 30 degrees to its right if you are right-handed (*see* Fig. 3b).
5. Do not stick the pupa tightly to the dissection tape so that you can adjust the angle of the pupa easily.
6. Make sure that the forceps are sharpened well. If not, use a piece of sandpaper (#1000, #2000) to sharpen. The sharpness of the forceps substantially affects the efficiency of dissection.
7. If you use an inverted confocal microscope, skip this step and directly set the coverslip on the microscope stage with an appropriate holder.
8. This heat-shock condition induces clones in histoblasts arrested at G2. Clones are specifically generated in histoblasts but not in LECs. For the clone induction in LECs, collect eggs on agar plates supplied with yeast for 2 h at 25 °C, then allow them to develop for 3 more hours at 25 °C and, finally, heat-shock for 1 h at 37 °C.
9. Play around the heat-shock condition at 35–37 °C for 4–10 min for the induction of clones in LECs. For the induction of clones in histoblasts, play around with a wider range of heat-shock conditions to have more or less clones depending on your purpose.
10. If the pressure is not applied while the needle is still inside the body, the body fluid backflows inside the needle. The body fluid often clogs the needle and makes it difficult to reuse the

needle for the following samples. In case the needle is clogged, immediate washing with chromic acid mixture will resolve the clog.

Acknowledgments

I thank Kevin Yuswan for useful comments on the manuscript, Fumiya Sato for providing pictures, and Izumi Nagai for providing images. D.U. was supported by JST FOREST J210000474, JSPS KAKENHI Grant Numbers 21K06144, 21H05105, Takeda Science Foundation, and Program for Creation of Interdisciplinary Research from FRIS at Tohoku University.

References

1. Roseland CR, Schneiderman HA (1979) Regulation and metamorphosis of the abdominal histoblasts ofDrosophila melanogaster. Wilhelm Roux's Arch Dev Biol 186:235–265. https://doi.org/10.1007/BF00848591
2. Madhavan MM, Madhavan K (1980) Morphogenesis of the epidermis of adult abdomen of Drosophila. J Embryol Exp Morphol 60:1–31
3. Bischoff M (2012) Lamellipodia-based migrations of larval epithelial cells are required for normal closure of the adult epidermis of Drosophila. Dev Biol 363:179–190. https://doi.org/10.1016/j.ydbio.2011.12.033
4. Ninov N, Chiarelli DA, Martin-Blanco E (2006) Extrinsic and intrinsic mechanisms directing epithelial cell sheet replacement during Drosophila metamorphosis. Development 2006:367–379. https://doi.org/10.1242/dev.02728
5. Hoshika S, Sun X, Kuranaga E, Umetsu D (2020) Reduction of endocytic activity accelerates cell elimination during tissue remodeling of the Drosophila epidermal epithelium. Development 147(7):dev179648. https://doi.org/10.1242/dev.179648
6. Ninov N, Menezes-Cabral S, Prat-Rojo C et al (2010) Dpp signaling directs cell motility and invasiveness during epithelial morphogenesis. Curr Biol 20:513–520. https://doi.org/10.1016/j.cub.2010.01.063
7. Prat-Rojo C, Pouille P, Buceta J, Martin-Blanco E (2020) Mechanical coordination is sufficient to promote tissue replacement during metamorphosis in Drosophila. EMBO J 39(3):e103594. https://doi.org/10.15252/embj.2019103594
8. Nakajima YI, Meyer EJ, Kroesen A et al (2013) Epithelial junctions maintain tissue architecture by directing planar spindle orientation. Nature 500(7462):359–362. https://doi.org/10.1038/nature12335
9. Umetsu D, Aigouy B, Aliee M et al (2014) Local increases in mechanical tension shape compartment boundaries by biasing cell intercalations. Curr Biol 24(15):1798–1805. https://doi.org/10.1016/j.cub.2014.06.052
10. Iijima N, Sato K, Kuranaga E, Umetsu D (2020) Differential cell adhesion implemented by Drosophila Toll corrects local distortions of the anterior-posterior compartment boundary. Nat Commun 11(1):1–2. https://doi.org/10.1038/s41467-020-20118-y
11. Dahmann C, Oates AC, Brand M (2011) Boundary formation and maintenance in tissue development. Nat Rev Genet 12(1):43–55
12. Umetsu D, Dahmann C (2015) Signals and mechanics shaping compartment boundaries in Drosophila. Wiley Interdiscip Rev Dev Biol 4:407–417. https://doi.org/10.1002/wdev.178
13. Wang J, Dahmann C (2020) Establishing compartment boundaries in Drosophila wing imaginal discs: an interplay between selector genes, signaling pathways and cell mechanics. In: Seminars in cell & developmental biology, vol 107. Academic Press, pp 161–169
14. Teng X, Qin L, Le Borgne R, Toyama Y (2017) Remodeling of adhesion and modulation of mechanical tensile forces during apoptosis in Drosophila epithelium. J Cell Sci 144(1):95–105. https://doi.org/10.1242/jcs.201343
15. Arata M, Sugimura K, Uemura T (2017) Difference in Dachsous levels between migrating cells coordinates the direction of collective cell

migration. Dev Cell 42(5):479–497. https://doi.org/10.1016/j.devcel.2017.08.001

16. Lawrence PA, Casal J, Struhl G (1999) The hedgehog morphogen and gradients of cell affinity in the abdomen of Drosophila. Development 126(11):2441–2449
17. Lawrence PA, Casal J, Struhl G (1999) Hedgehog and engrailed: pattern formation and polarity in the Drosophila abdomen. Development 126:2431–2439. https://doi.org/10.1242/dev.126.11.2431
18. Struhl G, Barbash DA, Lawrence PA (1997) Hedgehog acts by distinct gradient and signal relay mechanisms to organise cell type and cell polarity in the Drosophila abdomen. Development 124:2155–2165. https://doi.org/10.1242/dev.124.11.2155
19. Struhl G, Barbash DA, Lawrence PA (1997) Hedgehog organises the pattern and polarity of epidermal cells in the Drosophila abdomen. Development 124:2143–2154. https://doi.org/10.1242/dev.124.11.2143
20. Kopp A, Duncan I (2002) Anteroposterior patterning in adult abdominal segments of Drosophila. Dev Biol 242:15–30. https://doi.org/10.1006/dbio.2001.0529
21. Mangione F, Martín-Blanco E (2018) The Dachsous/fat/four-jointed pathway directs the uniform axial orientation of epithelial cells in the drosophila abdomen. Cell Rep 25(10):2836–2850. https://doi.org/10.1016/j.celrep.2018.11.036
22. Michel M, Dahmann C (2020) Tissue mechanical properties modulate cell extrusion in the drosophila abdominal epidermis. Development 147(5):dev179606. https://doi.org/10.1242/dev.179606
23. Lee T, Luo L (1999) Mosaic analysis with a repressible neurotechnique cell marker for studies of gene function in neuronal morphogenesis. Neuron 22:451–461. https://doi.org/10.1016/S0896-6273(00)80701-1
24. Luo L, Wu JS (2007) A protocol for mosaic analysis with a repressible cell marker (Marcm) in drosophila. Nat Protoc 1(6):2583–2589. https://doi.org/10.1038/nprot.2006.320
25. Golic KG, Lindquist S (1989) The FLP recombinase of yeast catalyzes site-specific recombination in the drosophila genome. Cell 59:499–509. https://doi.org/10.1016/0092-8674(89)90033-0
26. Germani F, Bergantinos C, Johnston LA (2018) Mosaic analysis in drosophila. Genetics 208:473–490. https://doi.org/10.1534/genetics.117.300256
27. Huang J, Zhou W, Dong W et al (2009) Directed, efficient, and versatile modifications of the Drosophila genome by genomic engineering. Proc Natl Acad Sci U S A 106:8284–8289. https://doi.org/10.1073/pnas.0900641106
28. Ninov N, Manjón C, Martín-Blanco E (2009) Dynamic control of cell cycle and growth coupling by ecdysone, egfr, and PI3K signaling in Drosophila histoblasts. PLoS Biol 7(4):e1000079. https://doi.org/10.1371/journal.pbio.1000079
29. Ninov N, Manjon C, Cabral S et al (2010) Dpp signaling directs cell motility and invasiveness during epithelial morphogenesis. Curr Biol 20(6):513–520. https://doi.org/10.1016/j.cub.2010.01.063
30. Sekyrova P, Bohmann D, Jindra M, Uhlirova M (2010) Interaction between Drosophila bZIP proteins Atf3 and Jun prevents replacement of epithelial cells during metamorphosis. Development 137(1):141–150. https://doi.org/10.1242/dev.037861
31. Kester RS, Nambu JR (2011) Targeted expression of p35 reveals a role for caspases in formation of the adult abdominal cuticle in Drosophila. Int J Dev Biol 55:109–119. https://doi.org/10.1387/ijdb.103109rk

Chapter 18

FlyClear: A Tissue-Clearing Technique for High-Resolution Microscopy of *Drosophila*

Marko Pende, Saiedeh Saghafi, Klaus Becker, Thomas Hummel, and Hans-Ulrich Dodt

Abstract

Fluorescently labeled transgenic lines of *Drosophila melanogaster* are a powerful routine tool in fly laboratories. The possibility to fluorescently visualize individual cell populations or entire tissues and the constantly improving microscopy technologies such as two-photon or light-sheet applications, with deep tissue imaging, hold great potential to address central biological questions at an organismic level. However, strong pigmentation and the opaque nature of the *D. melanogaster* cuticle hinder the penetration of visible light into internal tissues, thereby limiting the application of fluorescent microscopes to analyses of the outermost surfaces of intact samples. In addition, tissue-induced light scattering and optical aberrations quickly blur the view and, hence, require tissue sectioning for further investigation. We have developed a tissue-clearing and depigmentation approach (FlyClear), which preserves endogenous fluorescent signals and is applicable to various developmental stages ranging from larvae to adult fruit flies (Pende et al. Nature communications 9:4731, 2018). In this chapter, we provide a detailed protocol of the experimental steps involved.

Key words Tissue clearing, Depigmentation, Transgenic fluorescent signal preservation, Fruit fly

1 Introduction

Although the fruit fly *D. melanogaster* has been used as a research model for several decades, three-dimensional (3D) visualization of internal structures remains a challenge because – unlike in some other animals such as zebrafish larvae or *Caenorhabditis elegans* – all developmental stages of *Drosophila* are opaque. This opacity relies on light being absorbed by different types of pigments and scattered due to inhomogeneities in tissue composition. Both the scattering and absorption of light increase with imaging depth. This, in turn, reduces image sharpness and limits excitation of fluorophores [2–4].

Christian Dahmann (ed.), *Drosophila: Methods and Protocols*,
Methods in Molecular Biology, vol. 2540, https://doi.org/10.1007/978-1-0716-2541-5_18,

Thus, most analyses of *D. melanogaster* samples rely on two-dimensional (2D) information obtained from thin histological sections. The translation of these 2D data into a comprehensible, three-dimensional (3D) reconstruction is an error-prone and laborious process. As an alternative, *D. melanogaster* organs are dissected before mounting and imaging, which can damage the tissue and often lead to tissue deformation, making conclusive investigation of biological mechanisms difficult. This is particularly true when studying structures or processes that extend across different tissues and organs such as the long-range projections of peripheral nerves that span the entire body, linking different body parts to the central nervous system.

To overcome the deleterious light-absorbing and scattering properties of the biological specimens, samples can be rendered transparent using certain chemicals – a concept referred to as tissue clearing. The principle of optical clearing is to even out refractive index (RI) mismatches at the interface between different cellular components, creating a homogeneous medium for the transmission of light. This is achieved by replacing low RI components such as the water-based cytoplasm with a high RI solution and removing strongly scattering molecules such as lipids. In this way, tissue clearing can improve both imaging depth and contrast.

However, implementation of a working clearing method, which preserves the fluorescent signal, is difficult, as the chemicals used in this process are usually harsh to the tissue and their combination and incubation times are very much tissue dependent [5, 6]. Furthermore, tissue-clearing protocols were optimized for use in rodent organs such as the brain [7–20] or even in whole animals [21–26]. Few attempts of tissue clearing were published for fruit flies but with minimal [27] or no possibility of transgenic fluorescent signal preservation [28–30].

We developed a clearing technique tailored to *D. melanogaster* termed "FlyClear," which reliably clears intact samples from different developmental stages while preserving the fluorescent transgenic signal (*see* Fig. 1a, b). FlyClear has several advantages: (1) It is fast, simple, and requires immersion of flies in only two solutions after fixation (*see* Fig. 1c). (2) Animal and tissue morphology is preserved after the procedure (*see* Fig. 1b, c). (3) The method facilitates removal of most body and eye pigments without the use of any bleaching agents, which are known to damage the tissue and quench fluorescent signals (*see* Fig. 1d). (4) The protocol was established with endogenously expressed green fluorescent protein (GFP) and mCherry; however, it is most likely to also be compatible with other related fluorophores (*see* Fig. 1e). (5) After the clearing procedure, the endogenous fluorescence signal is preserved for at least 1 month without significant changes in intensity (*see* Fig. 1f). In this chapter we provide a detailed description of the steps involved in the FlyClear protocol.

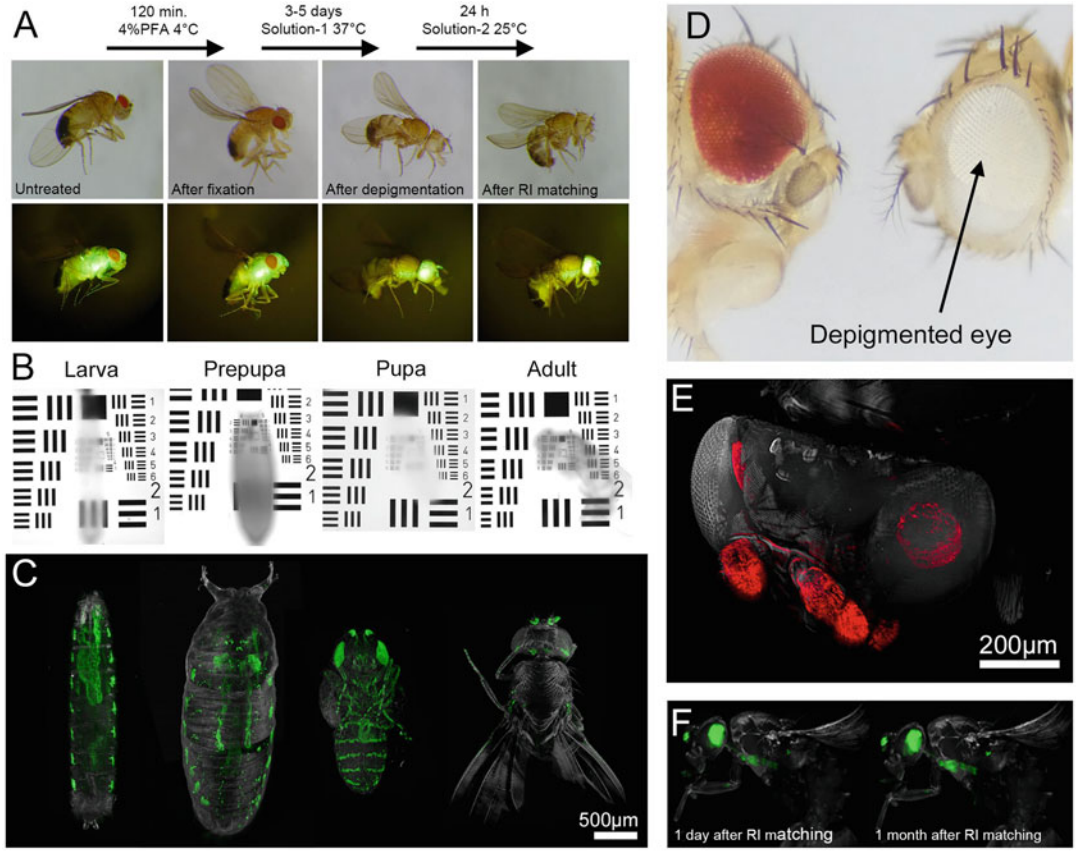


Fig. 1 Application and properties of the FlyClear procedure. (**a**) Major steps of the FlyClear protocol. (**b**) Wide-field image of different optically cleared fruit fly developmental stages placed on top of a USAF1951 chart in Solution-2 demonstrates the level of overall transparency. (**c**) Corresponding specimens from (**b**) after clearing and light-sheet imaging showing the GFP signal in *Peb-Gal4 UAS-mCD8::GFP* flies. (**d**) Adult fly head cut into two pieces. PBS (pH 8.5)-treated side on the left; Solution-1-treated side on the right (image was taken in PBS). (**e**) Light-sheet image of adult;;*R88E12-Gal4 UAS-mCherry* fly showing mCherry preservation. (**f**) Light-sheet image of *dscam-Gal4/CyO; UAS-mCD8::GFP* 1 day and 1 month after the FlyClear procedure. (Pictures are modified from Pende et al. (2018) [1])

2 Materials

2.1 Tools for Sample Handling

1. Dumont #5 forceps.
2. Pins with round heads.
3. 3-ml disposable plastic transfer pipettes.
4. Clear acrylic sticks.
5. Glass slides.
6. Coverslips.
7. Cuvettes.
8. Petri dish.

9. Ultraviolet (UV) glue.
10. UV lamp.
11. Blu Tak.
12. Glass snap cap vials, 50 × 30 mm, 25 ml.
13. Shaking water bath or shaking incubator.
14. Platform rocker.
15. Rotator.

2.2 Tools for Solution Measurements and Preparation

1. Benchtop pH meter.
2. Refractometer (e.g., Abbe).
3. Magnetic stirrer with a hot plate.
4. Stirring magnets.

2.3 Fixative

1. Phosphate-buffered saline (PBS), pH 8.5 (1 L): NaCl, 8 g, KCl, 0.2 g, Na_2HPO_4, 1.44 g, KH_2PO_4, 0.24 g. Adjust the pH to 8.5 with HCl.
2. 4% Formaldehyde solution (pH 8.5) (*see* Fig. 2): Prepare a fresh fixative under a fume hood by adding paraformaldehyde (PFA) pellets or powder to PBS, and stir on a magnetic stirrer with a hot plate. Set the temperature at 60–70 °C. Place a pH meter into the solution and adjust the pH by adding 4 M NaOH until reaching a pH between 9 and 11. The increased temperature and pH facilitate fast dissolution of the PFA. After the solution turns clear, adjust the pH of the fixative to 8.5 using concentrated HCl (decrease pH) or 4 M NaOH (increase pH) (*see* **Note 1**).

2.4 Permeabilization

1. Acetone.
2. Proteinase.

2.5 Depigmentation and Clearing Solution

1. Solution-1: 8% (v/v) *N*,*N*,*N*′,*N*′-tetrakis(2-hydroxyethyl) ethylenediamine (THEED), 5% (v/v) Triton® X-100, and 25% (w/v) urea in dH_2O. Prepare a Solution-1 stock in a 50 ml Falcon tube by mixing 4 ml THEED, 2.5 ml Triton® X-100, and 12.5 g urea filled up to 50 ml with dH_2O. THEED is very viscous, so it should be loaded into a plastic syringe and then added to the Falcon tube. The solution can be reused multiple times (*see* **Note 2**).

2.6 Refractive Index Matching Medium if Mounted on Slides

1. VECTASHIELD®.
2. Slides.
3. Coverslips.
4. Blu Tak (as a spacer).

A

Fixation time	Fixation quality	GFP emission
30min	+	+++
60min	++	+++
120min	+++	+++
240min	+++	+

B

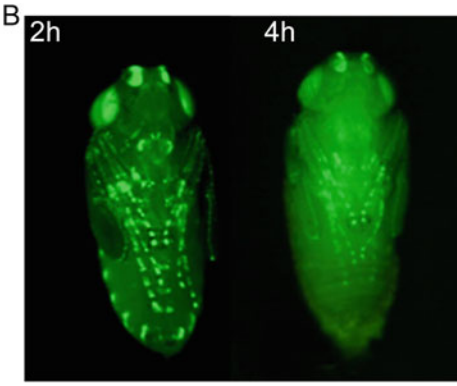


Fig. 2 Fixation of fruit fly samples. (**a**) Comparison of fixation quality and GFP emission after different incubation times in 4% formaldehyde at 4 °C (+, poor; +++, good). (**b**) *Peb-Gal4 UAS-mCD8::GFP* pupa after 2 and 4 hours of formaldehyde fixation

2.7 Refractive Index Matching Solution if Mounted in a Cuvette

1. Phosphate-buffered saline (PBS), pH 8.5.
2. Solution-2: 50 wt% meglumine diatrizoate in PBS (pH 8.5) adjusted to a refractive index (RI) of 1.45. Prepare a Solution-2 stock by adding 10 g of meglumine diatrizoate to 10 ml of PBS. Add further meglumine diatrizoate while monitoring the RI of the solution with a refractometer until an RI of 1.45 is reached. The initial pH of PBS defines the final pH of Solution-2. Hence, dissolving 10 g of meglumine diatrizoate in PBS (pH 8.5) will result in Solution-2 with a pH of 8.5 (*see* **Note 3**).
3. Diluted Solution-2 (optional): 25 wt% meglumine diatrizoate in PBS (pH 8.5). Prepare by adding 5 g of meglumine diatrizoate to 10 ml of PBS.

3 Methods

3.1 FlyClear Protocol (Third Instar Larva)

1. Place the samples in pre-warmed PBS (pH 7.5) with 0.03% proteinase (*see* **Note 4**) and incubate for 1 hour to make the larva surface permeable.
2. Fix the larva in 4% formaldehyde solution (pH 8.5) for 2 h at 4 °C under gentle shaking (*see* **Note 5**).
3. Wash the samples 3 × for 20 mins in PBS (pH 8.5) at 4 °C.
4. Place the larva in a 50 × 30 mm, 25 ml glass snap cap vial and incubate in 10 ml of Solution-1 at 37 °C under gentle shaking for 4–5 days using either a water bath or an incubator.
5. Wash the samples 3 × throughout 1 day in PBS (pH 8.5) at room temperature.

6. Incubate the samples in diluted Solution-2 (25 wt% meglumine diatrizoate) or diluted VECTASHIELD® (50% VECTASHIELD® in dH_2O) – depending on the later imaging approach – to avoid deformation (*see* **Note 6**).
7. Incubate the samples in Solution-2 or VECTASHIELD® for 24 hours at room temperature for the final RI matching (*see* **Notes 7** and **8**).

3.2 FlyClear Protocol (Prepupa)

1. Make a small rupture between T1 and T2 of each prepupal case using two Dumont #5 forceps to enable better penetration of chemicals and to avoid tissue deformation during later clearing steps.
2. Place the samples in pre-warmed PBS with 0.03% proteinase (*see* **Note 4**) and incubate for 1 hour to digest the prepupal case surface.
3. Fix the prepupa in 4% formaldehyde solution (pH 8.5) for 2 h at 4 °C under gentle shaking (*see* **Note 5**).
4. Wash the samples 3 × for 20 mins in PBS (pH 8.5) at 4 °C.
5. For permeabilization, place the samples in pre-chilled acetone and incubate for 2 hours at −20 °C (*see* **Note 9**).
6. Wash the samples once for 20 min in PBS (pH 8.5) at 4 °C.
7. Place the prepupa in a 50 × 30 mm, 25 ml glass snap cap vial and incubate in 10 ml of Solution-1 at 37 °C under gentle shaking for 4–5 days using either a water bath or an incubator.
8. Wash the samples 3 × throughout 1 day in PBS (pH 8.5) at room temperature.
9. Incubate the samples in diluted Solution-2 (25 wt% meglumine diatrizoate) or diluted VECTASHIELD® (50% VECTASHIELD® in dH_2O) – depending on the later imaging approach – to avoid deformation (*see* **Note 6**).
10. Incubate the samples in Solution-2 or VECTASHIELD® for 24 hours at room temperature for the final RI matching (*see* **Notes 7** and **8**).

3.3 FlyClear Protocol (Pupa)

1. Put pupa in PBS (pH 7.4) and remove the pupal case gently with Dumont #5 forceps.
2. Fix pupa in 4% formaldehyde solution (pH 8.5) for 2 h at 4 °C under gentle shaking (*see* **Note 5**).
3. Wash the samples 3 × for 20 min in PBS (pH 8.5) at 4 °C.
4. Put pupa in a 50 × 30 mm, 25 ml glass snap cap vial and incubate in 10 ml of Solution-1 at 37 °C for 3–5 days under gentle shaking using either a water bath or an incubator.
5. Wash the samples 3 × throughout 1 day in PBS (pH 8.5) at room temperature.

6. Incubate the samples in Solution-2 or VECTASHIELD® for 24 hours at room temperature for the final RI matching (*see* **Notes 7** and **8**).

3.4 FlyClear Protocol (Adult)

1. Anesthetize adult flies with CO_2.
2. Fix the fly samples in 4% formaldehyde solution (pH 8.5) for 2 h at 4 °C under gentle shaking (*see* **Note 5**).
3. Wash the samples 3 × for 20 min in PBS (pH 8.5) at 4 °C.
4. Place the flies in a 50 × 30 mm, 25 ml glass snap cap vial and incubate them in 10 ml of Solution-1 at 37 °C for 3–5 days under gentle shaking using either a water bath or an incubator (*see* **Note 10**).
5. Wash the samples 3 × throughout 1 day in PBS (pH 8.5) at room temperature.
6. Incubate the samples in Solution-2 or VECTASHIELD® for 24 hours at room temperature for the final RI matching (*see* **Notes 7** and **8**).

3.5 Mounting on Slides

1. Make four small Blu Tak spacers and put them in each corner of the slide.
2. Place the cleared samples on slides and add VECTASHIELD®.
3. Place a coverslip on the Blu Tak spacer and gently apply pressure to avoid squeezing of the sample.

3.6 Mounting in a Cuvette

1. Use a holder for a needle, e.g., a round-headed pin.
2. Keep the cleared samples in a Petri dish with Solution-2.
3. Place a drop of UV glue on a needle tip.
4. Immerse a small portion of the samples in the UV glue and illuminate the UV glue using a UV lamp for solidification.
5. Add Solution-2 in a cuvette using a pipette, e.g., a 3 ml disposable plastic transfer pipette, and remove eventual air bubbles, e.g., using a clear acrylic stick.
6. Place the samples in the cuvette.

4 Notes

1. To ensure proper fixation, the samples should be fully immersed in the fixative. However, sometimes, the samples float on the surface of the fixative. In this case, add a few drops of detergent such as Triton X-100 to lower the surface tension. Alternatively, 69 nl of 9% formaldehyde (pH 8.5) can be injected into the samples using a microinjector. The samples should be washed 3 × for 10 min in PBS (pH 8.5) at RT after 10 min of 9% formaldehyde incubation.

2. The pH of the solution is around 11 and decreases after every use, thereby reducing depigmentation efficiency. When the pH of the solution drops to about 10, a new solution should be prepared.
3. Meglumine diatrizoate is an aromatic iodine compound, which is present in a highly saturated state in Solution-2. Therefore, it tends to precipitate when in contact with air. Furthermore, the RI can change when water evaporates from the solution. Thus, Solution-2 should be used in sealed containers like a cuvette.
4. The activity of proteinase is very much pH, batch/lot, and vendor dependent. In our case, we use the proteinase from Sigma (Sigma, P8038-250MG), which worked best at pH 7.5.
5. Fixation is one of the most crucial steps when it comes to preservation of the fluorescent signal and tissue morphology. However, fixation also induces autofluorescence. Fixation times longer than 2 hours should be avoided as they lead to a substantial loss of signal and increased autofluorescence (*see* Fig. 2a, b).
6. Larvae and prepupae can be subject to tissue deformations induced by high differences in osmolarity. Thus, gradual adjustment of the RI is advised.
7. The final RI matching usually depends on the imaging modality the sample is intended for. If the samples are to be mounted on a microscope slide, it is advised to use VECTASHIELD® or a similar glycerol-based mounting medium and spacers such as Blu Tak to avoid squeezing of the specimen (*see* Fig. 3a). If the samples are imaged in sealed chambers (e.g., light-sheet imaging), then they can be mounted in Solution-2 using UV glue (*see* Fig. 3b, c).
8. The samples can be stored for more than a year at RT in the RI matching solution.
9. Acetone incubation is associated with fluorophore quenching; however, if pre-chilled acetone is used and the samples are kept at −20 °C, even overnight incubation steps have no visible impact on fluorophore stability (tested for transgenic GFP).
10. Incubation in Solution-1 already renders the samples transparent. However, transgenic signal and sample integrity are destroyed over time. Furthermore, Solution-1 is not able to remove certain types of pigments, such as the melanin in the male adult fly abdomen or the whole-body pigment in the ebony fly line.

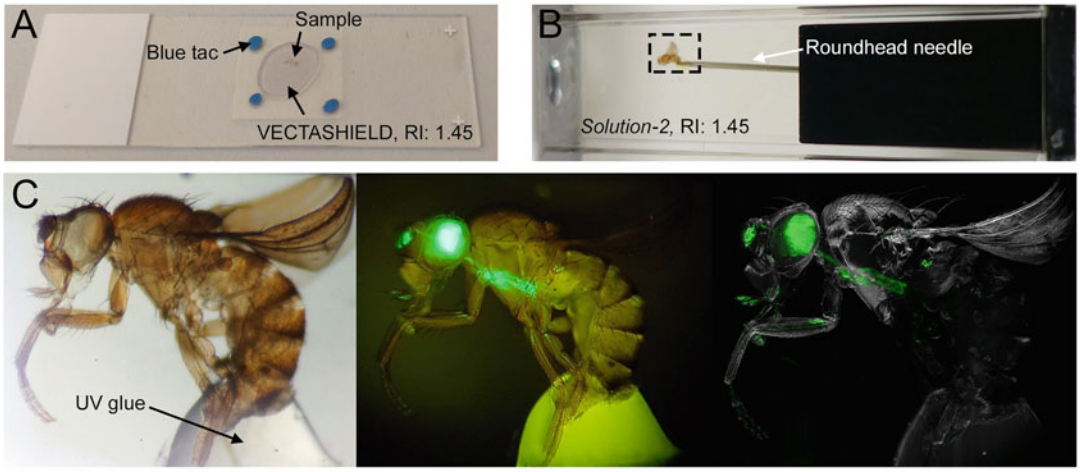


Fig. 3 Mounting of fruit fly samples. (**a**) Mounting of an adult fruit fly on a glass slide. A Blu Tak spacer is used to avoid sample squeezing. (**b**) Adult fly mounted with UV glue on a needle tip and placed in a cuvette with Solution-2. (**c**) Higher magnification of the rectangle area in (**b**) showing adult *dscam-Gal4/CyO; UAS-mCD8::GFP* fly imaged with white light (left) and blue laser (488 nm) (middle) on a stereoscope and with a light-sheet microscope (right)

Acknowledgments

We thank members of the Hans Ulrich Dodt and Thomas Hummel laboratories for their inputs to improve the protocol. The study was funded by the Austrian Science Fund (FWF) Project P 23102-N22 and Project P 25134.

References

1. Pende M, Becker K, Wanis M, Saghafi S, Kaur R, Hahn C, Pende N, Foroughipour M, Hummel T, Dodt HU (2018) High-resolution ultramicroscopy of the developing and adult nervous system in optically cleared Drosophila melanogaster. Nat Commun 9(1):4731. https://doi.org/10.1038/s41467-018-07192-z
2. Lichtman JW, Conchello JA (2005) Fluorescence microscopy. Nat Methods 2(12): 910–919. https://doi.org/10.1038/nmeth817
3. Lichtman JW, Denk W (2011) The big and the small: challenges of imaging the brain's circuits. Science (New York, NY) 334(6056): 618–623. https://doi.org/10.1126/science.1209168
4. Richardson DS, Lichtman JW (2015) Clarifying tissue clearing. Cell 162(2):246–257. https://doi.org/10.1016/j.cell.2015.06.067
5. Tainaka K, Kuno A, Kubota SI, Murakami T, Ueda HR (2016) Chemical principles in tissue clearing and staining protocols for whole-body cell profiling. Annu Rev Cell Dev Biol 32: 713–741. https://doi.org/10.1146/annurev-cellbio-111315-125001
6. Tainaka K, Murakami TC, Susaki EA, Shimizu C, Saito R, Takahashi K, Hayashi-Takagi A, Sekiya H, Arima Y, Nojima S, Ikemura M, Ushiku T, Shimizu Y, Murakami M, Tanaka KF, Iino M, Kasai H, Sasaoka T, Kobayashi K, Miyazono K, Morii E, Isa T, Fukayama M, Kakita A, Ueda HR (2018) Chemical landscape for tissue clearing based on hydrophilic reagents. Cell Rep 24(8):2196–2210.e2199. https://doi.org/10.1016/j.celrep.2018.07.056
7. Dodt HU, Leischner U, Schierloh A, Jahrling N, Mauch CP, Deininger K, Deussing JM, Eder M, Zieglgansberger W, Becker K (2007) Ultramicroscopy: three-dimensional

visualization of neuronal networks in the whole mouse brain. Nat Methods 4(4):331–336. https://doi.org/10.1038/nmeth1036

8. Becker K, Jahrling N, Saghafi S, Weiler R, Dodt HU (2012) Chemical clearing and dehydration of GFP expressing mouse brains. PLoS One 7(3):e33916. https://doi.org/10.1371/journal.pone.0033916
9. Erturk A, Becker K, Jahrling N, Mauch CP, Hojer CD, Egen JG, Hellal F, Bradke F, Sheng M, Dodt HU (2012) Three-dimensional imaging of solvent-cleared organs using 3DISCO. Nat Protoc 7(11):1983–1995. https://doi.org/10.1038/nprot.2012.119
10. Chung K, Wallace J, Kim SY, Kalyanasundaram S, Andalman AS, Davidson TJ, Mirzabekov JJ, Zalocusky KA, Mattis J, Denisin AK, Pak S, Bernstein H, Ramakrishnan C, Grosenick L, Gradinaru V, Deisseroth K (2013) Structural and molecular interrogation of intact biological systems. Nature 497(7449):332–337. https://doi.org/10.1038/nature12107
11. Hama H, Kurokawa H, Kawano H, Ando R, Shimogori T, Noda H, Fukami K, Sakaue-Sawano A, Miyawaki A (2011) Scale: a chemical approach for fluorescence imaging and reconstruction of transparent mouse brain. Nat Neurosci 14(11):1481–1488. https://doi.org/10.1038/nn.2928
12. Kuwajima T, Sitko AA, Bhansali P, Jurgens C, Guido W, Mason C (2013) ClearT: a detergent- and solvent-free clearing method for neuronal and non-neuronal tissue. Development 140(6):1364–1368. https://doi.org/10.1242/dev.091844
13. Susaki EA, Tainaka K, Perrin D, Kishino F, Tawara T, Watanabe TM, Yokoyama C, Onoe H, Eguchi M, Yamaguchi S, Abe T, Kiyonari H, Shimizu Y, Miyawaki A, Yokota H, Ueda HR (2014) Whole-brain imaging with single-cell resolution using chemical cocktails and computational analysis. Cell 157(3):726–739. https://doi.org/10.1016/j.cell.2014.03.042
14. Ke MT, Fujimoto S, Imai T (2013) SeeDB: a simple and morphology-preserving optical clearing agent for neuronal circuit reconstruction. Nat Neurosci 16(8):1154–1161. https://doi.org/10.1038/nn.3447
15. Murakami TC, Mano T, Saikawa S, Horiguchi SA, Shigeta D, Baba K, Sekiya H, Shimizu Y, Tanaka KF, Kiyonari H, Iino M, Mochizuki H, Tainaka K, Ueda HR (2018) A three-dimensional single-cell-resolution whole-brain atlas using CUBIC-X expansion microscopy and tissue clearing. Nat Neurosci 21(4): 625–637. https://doi.org/10.1038/s41593-018-0109-1
16. Ke MT, Nakai Y, Fujimoto S, Takayama R, Yoshida S, Kitajima TS, Sato M, Imai T (2016) Super-resolution mapping of neuronal circuitry with an index-optimized clearing agent. Cell Rep 14(11):2718–2732. https://doi.org/10.1016/j.celrep.2016.02.057
17. Yu T, Zhu J, Li Y, Ma Y, Wang J, Cheng X, Jin S, Sun Q, Li X, Gong H, Luo Q, Xu F, Zhao S, Zhu D (2018) RTF: a rapid and versatile tissue optical clearing method. Sci Rep 8(1):1964. https://doi.org/10.1038/s41598-018-20306-3
18. Hou B, Zhang D, Zhao S, Wei M, Yang Z, Wang S, Wang J, Zhang X, Liu B, Fan L, Li Y, Qiu Z, Zhang C, Jiang T (2015) Scalable and DiI-compatible optical clearance of the mammalian brain. Front Neuroanat 9:19. https://doi.org/10.3389/fnana.2015.00019
19. Hama H, Hioki H, Namiki K, Hoshida T, Kurokawa H, Ishidate F, Kaneko T, Akagi T, Saito T, Saido T, Miyawaki A (2015) ScaleS: an optical clearing palette for biological imaging. Nat Neurosci 18(10):1518–1529. https://doi.org/10.1038/nn.4107
20. Chen F, Tillberg PW, Boyden ES (2015) Optical imaging. Expansion microscopy. Science (New York, NY) 347(6221):543–548. https://doi.org/10.1126/science.1260088
21. Susaki EA, Tainaka K, Perrin D, Yukinaga H, Kuno A, Ueda HR (2015) Advanced CUBIC protocols for whole-brain and whole-body clearing and imaging. Nat Protoc 10(11): 1709–1727. https://doi.org/10.1038/nprot.2015.085
22. Yang B, Treweek JB, Kulkarni RP, Deverman BE, Chen CK, Lubeck E, Shah S, Cai L, Gradinaru V (2014) Single-cell phenotyping within transparent intact tissue through whole-body clearing. Cell 158(4):945–958. https://doi.org/10.1016/j.cell.2014.07.017
23. Tainaka K, Kubota SI, Suyama TQ, Susaki EA, Perrin D, Ukai-Tadenuma M, Ukai H, Ueda HR (2014) Whole-body imaging with single-cell resolution by tissue decolorization. Cell 159(4):911–924. https://doi.org/10.1016/j.cell.2014.10.034
24. Pan C, Cai R, Quacquarelli FP, Ghasemigharagoz A, Lourbopoulos A, Matryba P, Plesnila N, Dichgans M, Hellal F, Erturk A (2016) Shrinkage-mediated imaging of entire organs and organisms using uDISCO. Nat Methods 13(10):859–867. https://doi.org/10.1038/nmeth.3964
25. Treweek JB, Chan KY, Flytzanis NC, Yang B, Deverman BE, Greenbaum A, Lignell A,

Xiao C, Cai L, Ladinsky MS, Bjorkman PJ, Fowlkes CC, Gradinaru V (2015) Whole-body tissue stabilization and selective extractions via tissue-hydrogel hybrids for high-resolution intact circuit mapping and phenotyping. Nat Protoc 10(11):1860–1896. https://doi.org/10.1038/nprot.2015.122

26. Jing D, Zhang S, Luo W, Gao X, Men Y, Ma C, Liu X, Yi Y, Bugde A, Zhou BO, Zhao Z, Yuan Q, Feng JQ, Gao L, Ge WP, Zhao H (2018) Tissue clearing of both hard and soft tissue organs with the PEGASOS method. Cell Res 28(8):803–818. https://doi.org/10.1038/s41422-018-0049-z
27. Masselink W, Reumann D, Murawala P, Pasierbek P, Taniguchi Y, Bonnay F, Meixner K, Knoblich JA, Tanaka EM (2019) Broad applicability of a streamlined ethyl cinnamate-based clearing procedure. Development (Cambridge, England) 146(3): dev166884. https://doi.org/10.1242/dev.166884
28. McGurk L, Morrison H, Keegan LP, Sharpe J, O'Connell MA (2007) Three-dimensional imaging of Drosophila melanogaster. PLoS One 2(9):e834. https://doi.org/10.1371/journal.pone.0000834
29. Jahrling N, Becker K, Schonbauer C, Schnorrer F, Dodt HU (2010) Three-dimensional reconstruction and segmentation of intact Drosophila by ultramicroscopy. Front Syst Neurosci 4:1. https://doi.org/10.3389/neuro.06.001.2010
30. Saghafi S, Becker K, Hahn C, Dodt HU (2014) 3D-ultramicroscopy utilizing aspheric optics. J Biophotonics 7(1–2):117–125. https://doi.org/10.1002/jbio.201300048

Chapter 19

Preparation of *Drosophila* Tissues and Organs for Transmission Electron Microscopy

Olympia-Ekaterini Psathaki and Achim Paululat

Abstract

Transmission electron microscopy (TEM) is the method of choice to image the ultrastructure of cells or tissues. TEM allows the visualization of molecular complexes up to an atomic resolution. Thus, TEM data have led to important conclusions about cellular processes and supported findings obtained by functional analyses. In this chapter, we describe the preparation of *Drosophila* tissues for TEM and provide reliable step-by-step protocols for applying classical chemical fixation or high-pressure freezing–freeze substitution (HPF–FS) to preserve cellular structures.

Key words Chemical fixation, *Drosophila*, Transmission electron microscopy, High-pressure freezing, Freeze-substitution

1 Introduction

To study the ultrastructure of cells under a transmission electron microscope, some type of fixation is always necessary. Two main fixation methods are widely used in our days, chemical fixation at room temperature and cryogenic fixation. Classical chemical fixation with aldehydes such as glutaraldehyde, is based on slow penetration of the fixative into the tissue by diffusion, which causes the cross-linking of proteins. Proper chemical fixation of samples often works excellently and results in remarkable structural details that allow for a comprehensive ultrastructural analysis. Thousands of scientifc publications with chemically fixed tissue have been published, providing deep insights into morphological details. However, during chemical fixation, artefacts such as degradation, shrinking or other processes may occur and remain unnoticed because of the slow fixation process.

Aiming to eliminate artefacts, fixation based on ultrafast freezing of biological samples within milliseconds is now considered the method of choice. High-pressure freezing–freeze substitution

Christian Dahmann (ed.), *Drosophila: Methods and Protocols*,
Methods in Molecular Biology, vol. 2540, https://doi.org/10.1007/978-1-0716-2541-5_19,

(HPF–FS) results in the preservation of cellular structures, such as luminal compartments and membranes, and thus offers a higher resolution of ultrastructural details, such as cellular membrane bilayers. HPF–FS requires technical equipment and experience that may not be available at every institute.

The protocols described in this chapter have been used successfully in our laboratory for many years to prepare *Drosophila* embryos, larvae, pupae, and adult flies. They allow the reliable acquisition of detailed TEM images with a relatively high contrast. However, these protocols have plenty of room for optimization, such as by testing different fixation mixtures, buffers or different incubation times. In addition, numerous methods exist for sample embedding and trimming, including the type of knife to use. Here, we provide protocols that have been established in our laboratory and used to achieve favorable results when imaging *Drosophila* tissues. Our protocols are also suitable for novices to quickly learn sample preparation of *Drosophila* tissue for transmission electron microscopy. If researchers know which organelle in which cell type they would like to image at the beginning of their studies, it is always wise to search both the recent and older literature to learn about efficient techniques that previous authors have recommended for fixatives, buffer composition, fixation times, and other parameters. For example, it has been observed that *Drosophila* eyes and ovaries do not freeze well; thus, optimized chemical fixation may be the best choice for these tissues [1]. When analyzing *Drosophila* mutant phenotypes, proper control is always needed. The *white*^{1118} line is presumably the most abundant fly line in the *Drosophila* community and it is often used as a reference. However, it is important to note that *white*^{1118} is a mutant on its own and has been inbred for decades, although, in most cases, this is not an issue for ultrastructural studies.

2 Materials

It is crucial to use only freshly made buffers, fixatives, and contrasting solutions from trusted suppliers. Some chemicals can be purchased in a quality grade that is particularly suitable for electron microscopy (marked as EM grade). When using formaldehyde (FA) (*see* **Note 1**), it is best to use EM grade FA from freshly opened ampoules and not to use opened formaldehyde ampoules of unknown age and source.

2.1 Chemicals

1. Acetone, water-free.
2. Aqua destillata (a.d.).
3. Artificial hemolymph: 108 mmol NaCl, 5 mmol KCl, 2 mmol $CaCl_2$, 8 mmol $MgCl_2$, 1 mmol NaH_2PO_4, 4 mmol $NaHCO_3$,

5 mmol HEPES, pH 7.1; prior to use, supplement the buffer with sucrose (final concentration, 10 mmol) and trehalose (final concentration, 10 mmol).

4. Bovine serum albumin (BSA) buffer: 20% bovine serum albumin (BSA) in phosphate-buffered saline (PBS) buffer, pH 7.4.
5. Cacodylate buffer: pH 7.4, 0.2 M, EM grade (ready-to-use buffer) or use sodium cacodylate trihydrate to prepare 0.2 M cacodylate buffer, pH 7.4.
6. Clorox®, DanKlorix, Colgate-Palmolive, Hamburg, Germany.
7. Dental wax.
8. Ethanol (C_2H_6O).
9. Epon.
10. Extran®.
11. Dental wax.
12. Fixation buffer: 2% glutaraldehyde (made from 25% glutaraldehyde in ampoules, EM grade) in 0.05 M cacodylate buffer, pH 7.4.
13. Formvar® solution: 0.3% in chloroform (v/v).
14. FS (freeze-substitution) solution: 1% osmium tetroxide, 0.1% uranyl acetate, and 5% water in anhydrous acetone. Use ampoules with 1 g osmium tetroxide crystals, EM grade. Carefully break open an ampoule, and add 1 mL of anhydrous acetone. Use a pipette and a disposable tip, mix the solution thoroughly by pipetting up and down 2–3 times and then transfer the osmium tetroxide solution to a new 15 mL Falcon tube. Next, bring the solution to a total volume of 10 mL with anhydrous acetone to obtain a 1% osmium tetroxide solution, and pipette 950 µL of this solution into a 1.5 mL Sarstedt screw cap tube.
15. Glutaraldehyde-saturated heptane: Mix 1 mL glutaraldehyde stock solution (25% glutaraldehyde, EM grade) in 0.05 M cacodylate buffer, pH 7, with 4 mL *n*-heptane. Shake vigorously, allow phases to separate, and use the upper *n*-heptane phase for fixation.
16. Hexadecene.
17. Isotonic embryo wash buffer: 0.7% NaCl, 0.05% Triton X-100.
18. Lead citrate ($Pb(NO_3)_2$): 2.66 g lead citrate, 3.52 g sodium citrate, in 100 mL aqua destillata (Millipore), pH 12.
19. Nitric acid (HNO_3).
20. Nitrogen, liquid.
21. Osmium tetroxide, crystallized, EM grade.
22. Osmium tetroxide, 4%, 1 mL ampoules, EM grade.

23. Paraformaldehyde, 16%, ampoules, EM grade.
24. PBS buffer: 137 mM NaCl, 2.7 mM KCl, 10 mM NaH_2PO_4, 1.8 mM KH_2PO_4, pH 7.4.
25. Post-fixation buffer I: 2% glutaraldehyde and 1% osmium tetroxide in 0.05 M cacodylate buffer, pH 7.4.
26. Post-fixation buffer II: 1% osmium tetroxide in 0.05 M cacodylate buffer, pH 7.4.
27. Sylgard®, Dow Corning, Midland, MI, USA.
28. Toluidine blue staining solution: 1% toluidine blue, 1% sodium borate in aqua destillata (Millipore).
29. Uranyl acetate 2%: 4 g uranyl acetate in 200 mL aqua destillata (Millipore).
30. Yeast paste: Mix dry yeast with 10% methanol until a stiff yeast paste is prepared.

2.2 Equipment

1. Stereo microscope.
2. Contrasting machine e.g., Nanofilm Surface Analysis, leica EM AC20, Wetzlar, Germany.
3. Diamond knife, Histo, 4 mm, DiATOME, Nidau, Switzerland.
4. Diamond knife, Ultra 35°, 2 mm, DiATOME, Nidau, Switzerland.
5. Double-sided adhesive Scotch® tape.
6. Dumont tweezers, style 7, HP grade.
7. Egg basket: Prepare as follows: Cut off 0.5 cm of a 1.5 mL microcentrifuge tube, keeping the part with the cap. Place a 1 × 1 cm piece of 120 μm synthetic mesh on a hot plate, and melt the microcentrifuge ring to the mesh. Cut off any protruding material. Larger baskets can be prepared using Falcon tubes.
8. Egg collection bottle (for *Drosophila*).
9. Embedding mold.
10. Fine-haired paintbrush.
11. Freeze-substitution, e.g., Leica EM AFS2, Wetzlar, Germany.
12. Forceps.
13. Gelatine-coated slide.
14. Glass microscope slide.
15. Glass Petri dish.
16. Glass screw-top jar (5 ml).
17. Grid boxes.
18. Grids: copper slot grids, 1 × 2 mm, ready-to-use.
19. Heating plate, e.g. Medax, Neumünster, Germany.

20. High-pressure freezer, e.g., M. Wohlwend, Sennwald, Switzerland.
21. Loop, steel loop e.g. Perfect Loop, Electron Microscopy Sciences, Hatfield, PA, USA.
22. Minutien pins.
23. Micro-scissors.
24. Oven.
25. Parafilm®.
26. Petri dish.
27. Pipettes, disposable, 3 mL.
28. Planchettes for HPF (3 mm diameter, 0.15–0.2 mm depth of deepening) e.g. Model 241, 242, 353, M. Wohlwend, Sennwald, Switzerland.
29. Razor blades, carbon steel.
30. Razor blades.
31. Screw cap Sarstedt tubes (1.5 mL).
32. Tissue Tack®/Biobond® glue.
33. Transmission electron microscope e.g. Zeiss TEM 120kV Leo912, Zeiss TEM 120kV Libra120, Jeol TEM 200kV JEM2100Plus.
34. Tungsten needle.
35. Ultramicrotome e.g. Leica Ultracut UC7RT, UC7cryo, Leica EM UC6, Wetzlar, Germany.
36. Ultrasonic bath e.g. Bandelin, Berlin, Germany.

2.3 Safety Issues

Some of the chemicals used are toxic, so the local safety regulations applicable to biology laboratories must be strictly observed. This includes using gloves and working in a safety fume hood. Further information on the use of chemicals in histological techniques can be found in textbooks [2].

Aldehyde All aldehydes are toxic and can cause dermatitis on contact, among other harmful effects. Moreover, the inhalation of aldehyde vapors is hazardous. Therefore, aldehydes must be handled with care.

Cacodylate This buffer contains arsenic, which is a health hazard. If absorbed through the skin, it can cause dermatitis and liver and kidney inflammation. The hands should be protected by wearing disposable gloves made of impermeable material. A fume hood should be used while weighing the reagent and preparing the buffer solution. To prevent the release of arsenic gas, strong acids must not be used in conjunction with cacodylate buffers.

Osmium Tetroxide This is a strong oxidizing agent with a high vapor pressure; therefore, it volatilizes very easily at room temperature. Its vapors are harmful to the eyes, throat, and nose. Osmium tetroxide easily penetrates through the skin, so the hands and other parts of the body should not be exposed to this reagent. Osmium tetroxide must always be handled in a fume hood, and gas-tight goggles should be worn when handling it. For storage, use a refrigerator that fits in the fume hood, such as a small refrigerator from a camping supplier.

Uranyl Acetate Uranium and its compounds are radioactive and extremely toxic. Uranyl acetate is a double hazard, being both chemically and radiologically toxic. The inhalation of powdered uranyl compounds can cause severe damage to the respiratory tract, lungs, and liver. The hazards of uranium compounds have been discussed elsewhere [3]. Check the local safety regulations before handling uranyl acetate.

3 Methods

3.1 Chemical Fixation of Drosophila *Specimens*

3.1.1 Preparation and Chemical Fixation of Embryos

Numerous protocols have been successfully used for the fixation and sample preparation of *Drosophila* embryos for ultrastructural analysis. These protocols often differ in their details (e.g., composition and concentration of fixatives and incubation times) and manual handling methods (e.g., how the vitelline membrane is perforated and removed). Here, we describe the method routinely used in our laboratory, which has been assembled by combining numerous protocols [4–14].

1. Collecting eggs: Set up a *Drosophila* egg collection bottle fitted with a Petri dish lid containing apple juice agar. Flies will deposit eggs onto the agar. Replace the apple juice agar dish approximately every 2 h, and incubate the eggs and embryos in a 20 °C incubator for a defined period. This allows the embryos to reach the desired developmental stage. The work is greatly facilitated if the age of the embryos or larvae is determined as accurately as possible before starting the fixation and embedding process.
2. Pick up the embryos with a fine-haired paintbrush, and transfer them into an egg basket (*see* Subheading 2.2). Wash the embryos several times with copious amounts of fresh tap water to remove food remains from the apple juice plate.
3. Dechorionation: Place the egg basket in a larger glass Petri dish or crystallizing dish filled with 50% Clorox (Clorox:water, v/v). Rinse the embryos thoroughly with the bleach solution using a Pasteur pipette. The dechorionation process should be

performed under a dissecting microscope. It typically takes 2–3 min to dechorionate more than 75% of the embryos.

4. Immediately continue with extensive washing using the isotonic embryo wash buffer. Finally, wash all the embryos from the basket onto a single piece of mesh.
5. Transfer the embryos with a fine-haired paintbrush from the mesh into a 5 mL glass screw-top jar (*see* **Note 2**) filled halfway with glutaraldehyde-saturated heptane (*see* Subheading 2.2). Mix well and incubate the embryos in this solution at room temperature for 20 min with gentle agitation in an orbital shaker. Heptane permeabilizes the vitelline membrane, whereas glutaraldehyde penetrates the embryo and preserves the internal tissues.
6. Transfer the embryos onto a glass microscope slide with a Pasteur pipette. Heptane typically evaporates within 30–60 s, and the embryos remain on the glass slide. Adhere the embryos extremely carefully onto one side of the double-sided adhesive Scotch® tape, and apply this to a Petri dish (or a new microscope slide) with the embryo on top.
7. Drop the fixation buffer onto the embryos. To remove the vitelline membrane, check the embryos under a dissecting microscope while poking a small hole in the vitelline membrane at the anterior tip of the embryos using a self-made, extremely thin glass or tungsten needle. Push the embryos out of the vitelline membrane by applying gentle pressure from the posterior. The vitelline membrane will peel off and adhere to the tape. The membrane has a silvery shimmer, whereas the embryos appear white.
8. Pre-fixation: Using a glass pipette, transfer the embryos into an Eppendorf cup filled with the post-fixation buffer I (*see* **Note 3**) in 0.05 M cacodylate buffer, pH 7.4. Incubate the specimens for 2 h at 4 °C (use an ice bucket or a small refrigerator placed in the fume hood). Wash the embryos 3 times for 10 min in 0.05 M cacodylate buffer, pH 7.4, at room temperature.
9. Post-fixation: Remove the wash buffer, and add the post-fixation buffer II to the embryos. Incubate for 1 h at 4 °C. Afterward, wash the embryos 3 times for 10 min in 0.05 M cacodylate buffer, pH 7.4, at room temperature. After the second wash, transfer the embryos into a new Eppendorf cup and continue washing (*see* **Note 4**).
10. Dehydrate the embryos in a series of ethanol:cacodylate buffer solutions for 10 min per step: 30%:70% (4 °C), 50%:50% (4 °C), 70%:30% (room temperature), 2 × 95%:5% (room temperature), and 2 × 100% ethanol (room temperature) (*see* **Note 5**).

11. The embryos are now ready for infiltration with Epon resin and embedding (*see* Subheading 3.3).

3.1.2 Dissection and Chemical Fixation of Third Instar Wandering Larvae and Isolated Organs and Tissues

1. Prepare small Petri dishes (e.g., with a diameter of 3 cm) half-filled with Sylgard® resin (*see* **Note 6**). Sylgard dissection plates remain usable for several weeks when stored at 4 °C, and dehydration is avoided (*see* Fig. 1).
2. Pin down anesthetized third instar wandering larvae (*see* **Note 7**) on Sylgard plates with insect needles and Minutien pins. Prepare in artificial hemolymph or PBS buffer. Open each larva with micro-scissors, starting from the posterior pole and working toward the anterior without destroying the tissue of interest. This method effectively prepares structures such as the somatic muscles, heart, and other tissues connected to external structures for fixation. Tungsten needles and fine-tipped forceps are helpful for the preparation of the larval tissues. Ensure that the larva is consistently covered with hemolymph solution.
3. For studies of specific organs or tissues, such as the salivary glands, malpighian tubules, fat body, gut, brain, testis, and other easily accessible organs, isolate and remove them from the larvae and transfer them into an Eppendorf cup with the fixation buffer for further processing.
4. If proceeding with tissue prepared on Sylgard plates, replace the artificial hemolymph with the fixation buffer, and fix the specimen for 1.5 h at room temperature. The fixative must cover the specimen at all times. Gentle agitation is recommended. Extracted organs or tissues are fixed in the same manner but in an Eppendorf cup.
5. Wash the larvae or organs 3 times for 10 min with 0.05 M cacodylate buffer, pH 7.4.
6. Post-fix the larvae or organs for 1 h in the post-fixation buffer II in the same Sylgard plate/Eppendorf cup.
7. Wash the larvae or organs 3 times for 10 min with 0.05 M cacodylate buffer, pH 7.4.
8. Transfer the larvae or organs into a 5 mL glass screw-top jar filled with 0.05 M cacodylate buffer, pH 7.4, at 4 °C.
9. Dehydrate the samples in a series of ethanol:cacodylate buffer solutions for 10 min per step: 30%:70% (4 °C), 50%:50% (4 °C), 70%:30% (room temperature), 2 × 95%:5% (room temperature), and 2 × 100% ethanol (room temperature). Do not allow the tissue to dry out.
10. The samples are now ready for infiltration with epoxy resin and embedding (*see* Subheading 3.3).

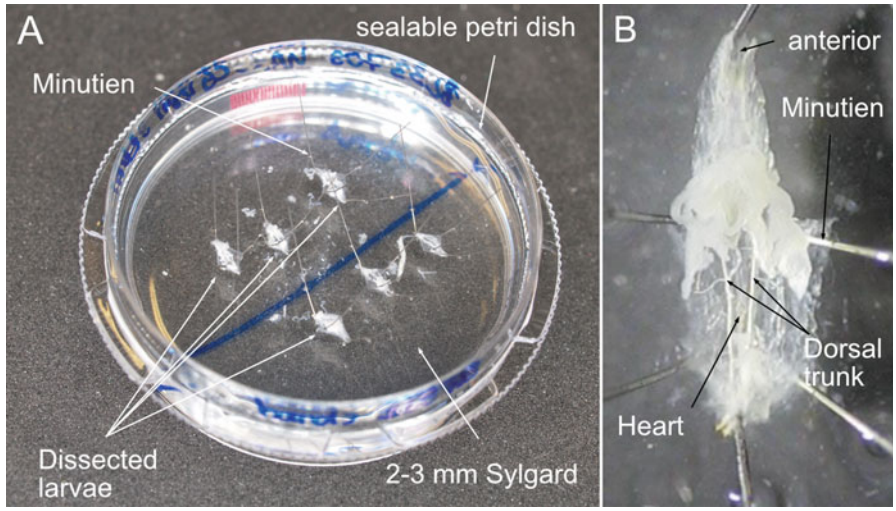


Fig. 1 Dissecting third instar wandering larvae. (**a**) Larvae were pinned down onto a 3 cm Sylgard plate with Minutien pins. (**b**) Heart preparation of a third instar larva

3.1.3 Dissection and Chemical Fixation of Pupae

As a holometabolic insect, *Drosophila* undergoes a complete metamorphosis in which most of the larval tissue is histolyzed and most of the organs of the adult differentiate from imaginal cells, which are mostly organized in imaginal discs. Tissue reorganization during metamorphosis makes it difficult to provide a simple preparation protocol for the generation of pupae TEM analysis. However, we have studied the differentiation of several cell types during metamorphosis, including the heart tissue, wing hearts, and nephrocytes; on the basis of our experience, we suggest a simple protocol for the sample preparation of pupal and adult tissues.

1. Collect pupae of the desired age (*see* Fig. 2). If needed, one can refer to the efficient method for collecting larvae and pupae at defined developmental stages described by Gerlitz and colleagues [15].
2. Transfer individual pupae to a small Petri dish, and glue the pupae ventral side down to the bottom of the dish using double-sided adhesive Scotch® tape. Alternatively, Tissue Tack®/Biobond® glue may be used.
3. Dissect older pupae with sharp micro-scissors and forceps as illustrated in Fig. 2. If preparing several specimens simultaneously, avoid dehydration of the specimens (e.g., by dissecting them in artificial hemolymph).
4. Once the dissection is finished, replace the artificial hemolymph with the fixation buffer and incubate the specimens overnight at 4 °C with gentle agitation.
5. Wash the specimens 3 times for 10 min with 0.05 M cacodylate buffer, pH 7.4. Continue with post-fixation.

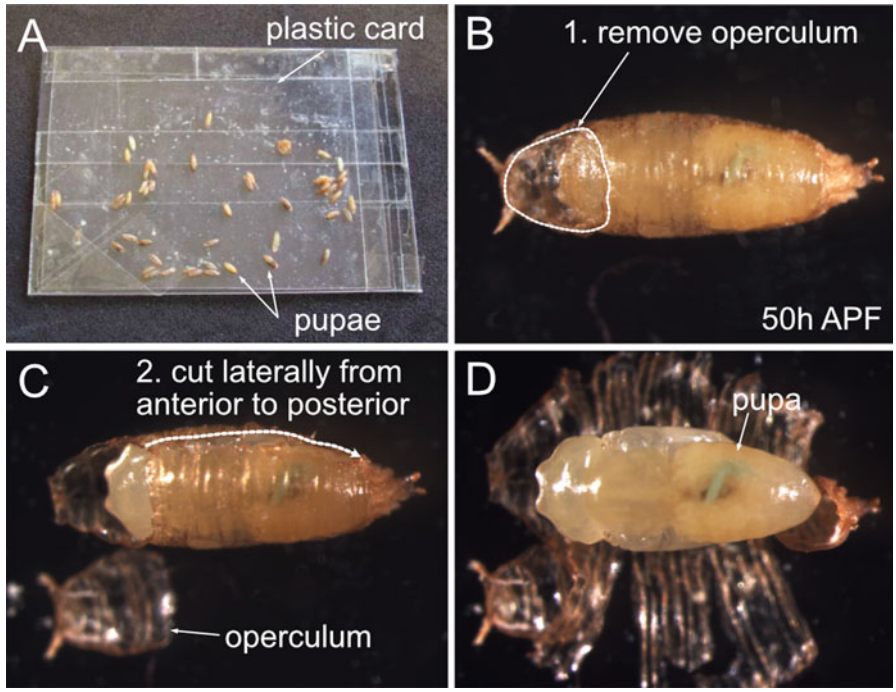


Fig. 2 Dissecting pupae. (**a**) A piece of plastic, placed on the inner surface of a *Drosophila* breeding vial, was used to collect pupae. (**b–d**) Removal of the operculum and the pupal case. APF after puparium formation

6. Post-fixation: Post-fix the specimens in the post-fixation buffer II for 4 h at 4 °C with gentle agitation. This step may be shortened to 2–3 h for small samples, such as isolated salivary glands or imaginal discs.
7. Wash the specimens 3 times for 10 min with 0.05 M cacodylate buffer, pH 7.4.
8. Next, dehydrate the samples in a graded ethanol series (*see* Subheading 3.1.2, **step 9**).
9. The samples are now ready for infiltration with Epon resin and embedding (*see* Subheading 3.3).

3.1.4 Dissection and Chemical Fixation of Adult Flies

Adult flies are anesthetized with carbon dioxide before dissection.

1. Use a Sylgard® plate for preparation, and fix the specimens with Minutien pins.
2. Dissect the adult flies in artificial hemolymph. For example, for studying the dorsal vessel, remove the ventral body side and the head with a razor blade. Specific protocols have been established for the preparation of some adult organs, such as the eyes or ovaries, for TEM [1, 16].

3. Replace the dissection solution with the fixation buffer, pH 7.4. The fixative should cover the specimens at all times. Fix the specimens for 1.5 h at room temperature with gentle agitation.
4. Transfer the flies from the Sylgard plates to an Eppendorf cup, and wash them 3 times for 10 min with 0.05 M cacodylate buffer, pH 7.4.
5. Next, post-fix the flies for 2 h in the post-fixation buffer II at 4 °C.
6. Wash the specimens in the same Eppendorf cup 3 times for 10 min with 0.05 M cacodylate buffer, pH 7.4.
7. Transfer the flies into an Eppendorf cup filled with 0.05 M cacodylate buffer, pH 7.4, at 4 °C.
8. Dehydrate the samples in a series of ethanol:cacodylate buffer solutions for 10 min per step: 30%:70% (4 °C), 50%:50% (4 °C), 70%:30% (room temperature), 2 × 95%:5% (room temperature), and 2 × 100% ethanol (room temperature). Do not allow the tissue to dry out.
9. The samples are now ready for infiltration with epoxy resin and final embedding (*see* Subheading 3.3).

3.2 High-Pressure Freezing and Freeze-Substitution of Drosophila *Specimens*

The idea of using high pressure (approximately 2000 atmospheres or 200 MPa) to freeze cells or tissues was conceived in the late 1960s [17]. The value of high pressure freezing (HPF) is that samples with relatively large volumes up to 200 μm thickness, including large cells and whole organisms, can be frozen without visible ice crystal damage. Several cryo methods can be used to freeze cells or tissues, but they all aim to remove heat at a rate fast enough so that the water molecules form amorphous, noncrystalline ice [18]. This can be achieved because at these high pressures, ice crystal nucleation and growth are slowed down [18]. HPF retains cellular structures in their original positions with fewer distortions; moreover, the molecules remain in a near-native state. In larger specimens of several microns, or in organisms that have natural diffusion barriers, such as *Drosophila* larvae or embryos, diffusion by chemical fixation throughout all the cells is relatively slow [18]. During that time, e.g. autolysis can occurs, and the cell structures change. Another problem is that cross-linking reactions are selective. For example, glutaraldehyde only reacts with certain amino acid residues on proteins and does not cross-link nucleic acids or carbohydrate molecules. As a consequence, these molecules are extracted from the cell during the rinse steps and especially during dehydration at room temperature [19]. High-pressure freezing and freeze-substitution reduce the artifacts caused by conventional EM specimen preparation.

Embryos, larvae, and organs are collected and prepared using the same protocol as that described for the classical chemical fixation of *Drosophila*.

1. Immediately after the dechorionation of the embryos and the preparation of the larvae in hemolymph, carefully lift off the specimen, immerse it briefly into the BSA buffer, and transfer it onto an aluminum planchette 3 mm in diameter with a 100 μm or 150 μm cavity. The organs of the larvae are processed in the same manner (*see* Fig. 3).
2. Especially for the embryos, the use of yeast paste instead of the BSA buffer as a filling material is recommended. Transfer one or several dechorionated embryos to the yeast paste using a paintbrush. Mix the embryos and the yeast paste carefully, until they form a homogeneous yeast ball with embryos. Paint the probe onto planchette. More details are described by McDonald [20].
3. After placing the specimen on the planchette, fill the cavity with the sample completely with hexadecene. Use another planchette of the same diameter without a cavity on the sample side and a cavity of 300 μm on the outside to cover the sample. Take care that no air bubbles are enclosed. The hexadecene should overflow when the cover planchette is closed and attached to the HPF sample holder (*see* Fig. 3).
4. Next, close the HPF sample holder and start high pressure freezing; this can be carried out, e.g., in a Compact 03 from Martin Wohlwend (Sennwald, Switzerland) (*see* Fig. 3a).
5. After HPF, immediately transfer the sample holder into a custom-made basin filled with liquid nitrogen placed nearby (*see* Fig. 3b).
6. Open the HPF sample holder in the liquid nitrogen basin, press the planchette sandwich out of the holder with pre-cooled tweezers, and then transfer the planchette sandwich to the cryo storage container (*see* Fig. 3b–d). Ensure that the sample is always in liquid nitrogen after HPF. To store the samples in cryo containers, we recommend using custom baskets made from 500 μL Eppendorf cups and mesh tissues, which fit perfectly in the cryo containers (*see* Fig. 3d). This ensures that the individual samples can be identified at any time and will not be lost in the liquid nitrogen basin.
7. To proceed to freeze-substitution, transfer the planchette into a small basin inside a larger basin filled with liquid nitrogen (*see* Fig. 3e–g) and open the planchette sandwiches in the liquid nitrogen with a scalpel or tweezers under a stereo microscope. It is critical for the liquid nitrogen level of the small basin to be

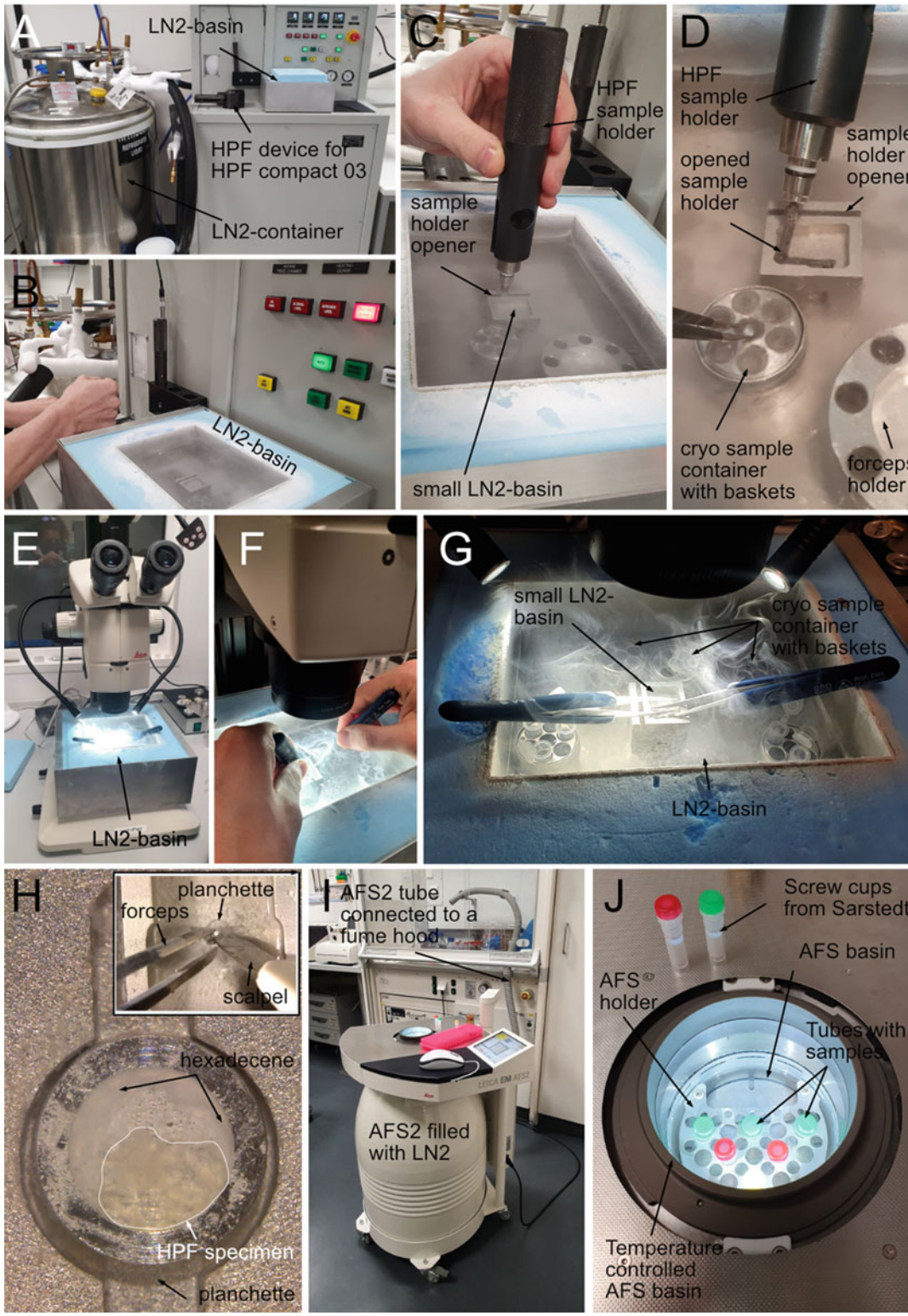


Fig. 3 High-pressure freezing. (**a**) HPF device, Wohlwend Compact 03 with a self-filling LN_2 container. (**b**) The sample holder with a clamped sample is frozen at high pressure and immediately transferred to a prepared

slightly above that of the larger basin; otherwise, turbulence will be present on the liquid nitrogen surface in the small basin, obscuring the sample from view.

8. As soon as the planchette sandwich is opened, hexadecene, which also escapes from the planchettes, hinders the visibility in liquid nitrogen. Therefore, it is necessary to keep adding liquid nitrogen to the small basin without sweeping away the sample. After opening the planchette, remove the lid and continue processing the planchette with the sample. To do this, first, carefully scrape away the frozen hexadecene with a scalpel under a stereo microscope without damaging the sample. Hexadecene appears white, and the sample should be light brown to yellow and transparent (*see* Fig. 3h). Caution must be exercised in this step; if the hexadecene is not adequately removed, the freeze-substitution solution will not be able to penetrate the sample because hexadecene shields the sample.
9. Once the hexadecene has been removed and the sample is clearly visible (*see* Fig. 3h), the sample can be transferred to the freeze-substitution (FS) system (in our case, the AFS2 from Leica (Germany); *see* Fig. 3i).
10. First, fill the dewar of the FS system with liquid nitrogen (approximately 40 L) and set the system to −90 °C, the starting temperature for the FS. When the set temperature is reached, pour approximately 300 mL of 100% absolute ethanol into the sample basin so that 1.5 mL screw cap Sarstedt tubes are immersed into the alcohol up to the 1 mL mark. Cool the alcohol to −90 °C, and then place the sample stand in the basin.
11. In the meantime, prepare the FS solution (*see* Subheading 2.1) in a fume hood. For *Drosophila* third instar larvae and embryos, we have obtained the best results with the FS solution. Add 50 μL of a previously prepared and filtered 2% aqueous uranyl acetate solution, and then mix and briefly cool the mixture in liquid nitrogen. Next, place the tubes in the FS sample stand of the AFS2 (*see* Fig. 3j).

Fig. 3 (continued) LN_2 bath. (**c**) In the LN_2 bath, the specimen holder is opened and unfolded. (**d**) The planchette doublet with the sample simply falls out of the holder into the small LN_2-filled basin or can be pushed out from behind with pre-cooled forceps. It can then be transferred with pre-cooled forceps into the sample baskets of the LN_2 storage container (**d**). Prior to freeze-substitution, the planchette doublet must be opened in LN_2. This is performed using binoculars in the small LN_2 basin within the larger LN_2 basin (**e–h**). The planchette doublet is opened using a pre-cooled scalpel and tweezers (**h**, inset). The white, opaque hexadecene is carefully scraped off the sample with the scalpel to reveal the sample (**h**), which ideally looks yellowish and glassy. The Sarstedt tubes with the substitution solution are pre-cooled in the basket holder of the Leica AFS2 freeze-substitution system to −90 °C (**i**, **j**). Freeze-substitution is then started. HPF high-pressure freezing; LN_2 liquid nitrogen

12. We have achieved good results with the following temperature curves: 12 h at −90 °C, 12 h at −60 °C, 12 h at −30 °C, and 10 h at 0 °C.
13. When the FS is finished, remove the samples from the FS, place them on ice, and wash them 4 times with ice-cold anhydrous acetone. Then, infiltrate the samples overnight in a 50% Epon–acetone mixture. The following day, infiltrate the samples for at least 4 h in a 70% Epon–acetone mixture and then overnight in 100% Epon.
14. The following day, transfer the samples into fresh 100% Epon for at least 4 h. Finally, transfer the samples into fresh 100% Epon in embedding molds and cure them in an oven at 60 °C for 48 h. The samples are now ready for ultrathin sectioning and post-staining, as described in Subheadings 3.4, 3.5, and 3.6.

3.3 Embedding of Chemically Fixed Embryos, Larvae, Pupae, or Adult Tissues

Because misalignment of the embedded specimens cannot be corrected by trimming and requires re-embedding, it is crucial to align the embryo, larva, or tissue in the desired position in the embedding mold as precisely as possible before the Epon is cured.

1. All chemically fixed samples were collected in 100% ethanol (*see* Subheading 3.1). Now, replace the 100% ethanol with 100% acetone and incubate for 15 min. Repeat this step. Note that acetone evaporates very quickly; ensure that the samples never become dry. Because acetone reacts with polystyrene, do not use classical cell culture dishes or Petri dishes. Eppendorf cups and Falcon tubes made of polypropylene, as well as glassware, can be used with acetone.
2. Replace acetone with a 1:1 mixture of Epon–acetone, and incubate the samples at room temperature overnight with gentle agitation.
3. Replace the 1:1 Epon–acetone mixture with 100% Epon, and incubate the samples at room temperature for 4 h (leave the Eppendorf cup open). Transfer the specimens manually into a new cup with fresh Epon, and incubate overnight with gentle agitation (without rotation).
4. Transfer the embryos, larvae, pupae, adult flies, or dissected organs individually into fresh Epon in embedding molds. Orient them with a fine needle or a sharpened wooden toothpick so that they are at the top of the mold in the desired position (use a stereo microscope; *see* Fig. 4).
5. Polymerize the resin in an oven for 48 h at 60 °C. The samples are now ready for trimming and cutting.

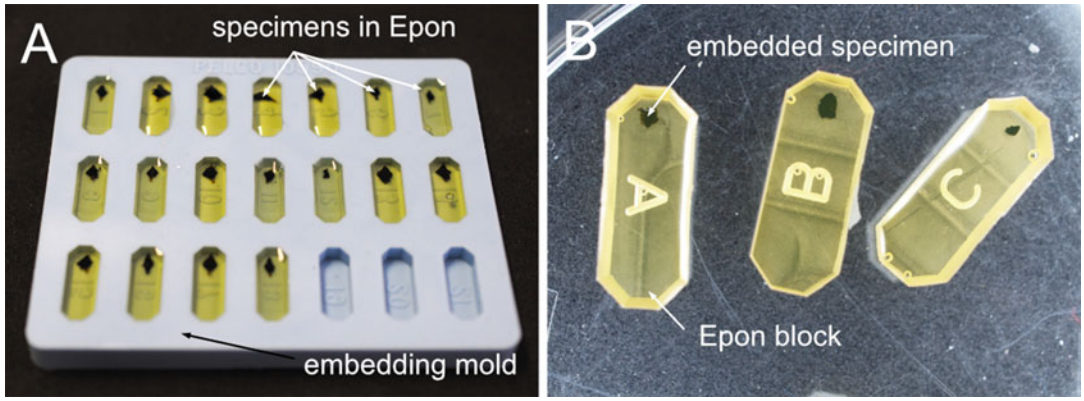


Fig. 4 Embedding molds. (**a**) Silicone rubber mold with (**b**) Epon-embedded thoraces

3.4 Trimming

Once the specimens are embedded, the Epon block must be trimmed (*see* Fig. 5). Trimming is essential to achieve a defined size and trapezoidal shape of the tip of the Epon block in which the desired part of the sample is positioned. Only if this is achieved, is it possible to use the diamond knife to generate a series of ultrathin sections arranged as ribbons on the water surface of the diamond basin in the ultramicrotome (*see* Fig. 5f). Therefore, excess Epon must be removed (or trimmed) away from the sides and the front of the Epon block with a sharp razor blade or a trimming diamond.

1. First, clamp the sample holder with the Epon-embedded specimen in a vice (*see* Fig. 5a, b). Fine trimming of the sides and surface should be performed extremely carefully under a dissecting microscope. We use a custom-made trimming setup for this purpose (*see* Fig. 5c). Use a single-edged razor blade to remove the resin from the sides. Care must be taken not to splinter the Epon block. Therefore, always cut thinly from top to bottom and up to the side of the specimen. The cuts must not be thicker than approximately 200–300 μm and not deeper than approximately 300–400 μm. A trapezoid must be generated before the sample is installed in the ultramicrotome for ultrathin sectioning. The trapezoid should be no wider or taller than 500–700 μm (*see* Fig. 5d, e). However, the height should be greater than the width and the edges should be relatively steep. It is easier to make serial sections from a smaller pyramid. The top and bottom edges must be parallel. An optimally trimmed trapezoid is crucial for well-executed, ultrathin sections (*see* Fig. 5f).
2. To approach the sample from above, use a new double-sided razor blade. Break the blade into two beforehand by bending it in the paper sleeve containing the razor blade. With one half of the blade, carefully cut off thin sections of Epon. The surface should be transparent after cutting. A rough surface indicates

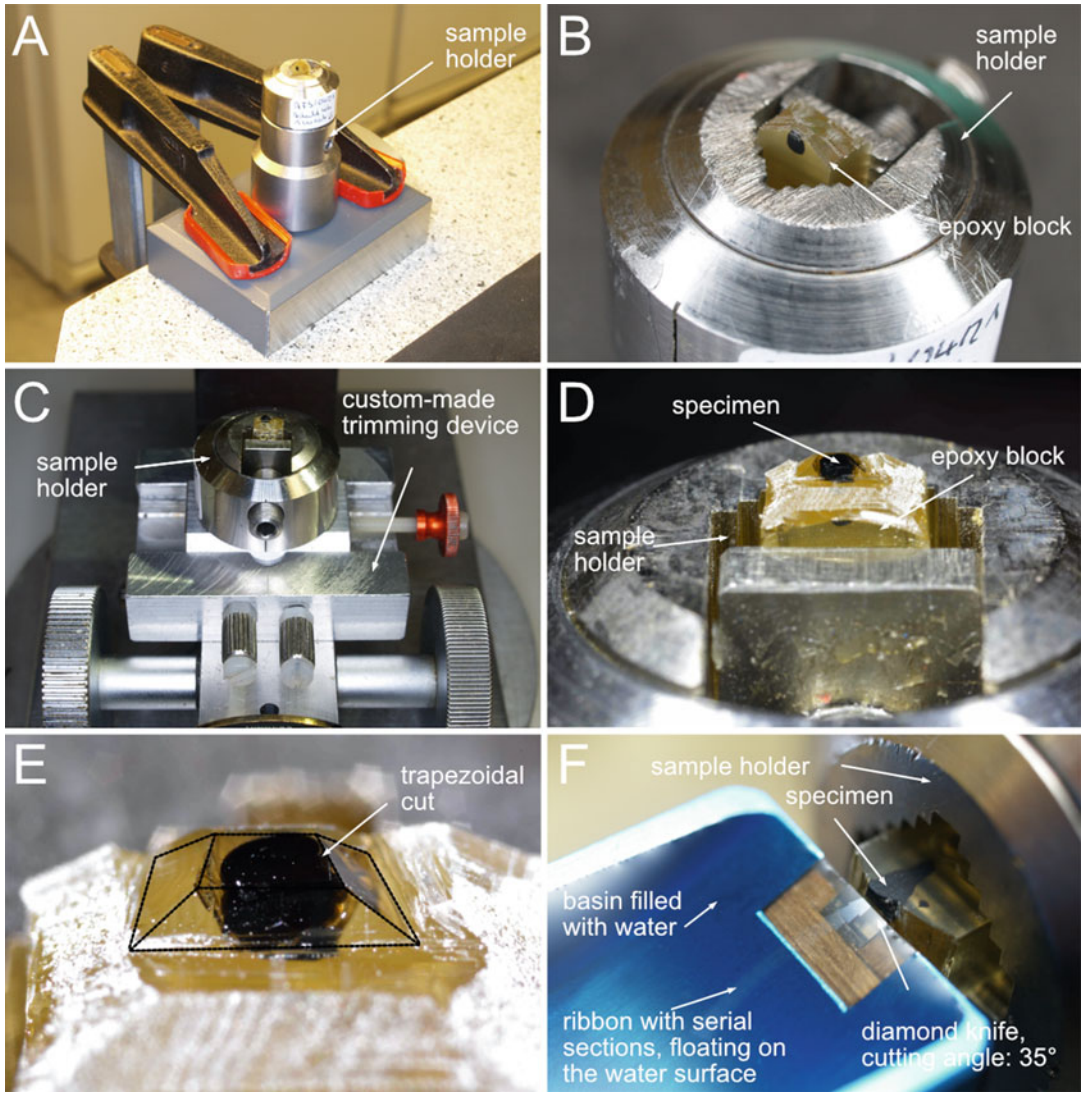


Fig. 5 Trimming and sectioning of an Epon embedded sample. (**a**) Pre-trimming. Excess Epon next to the specimen is removed with the single-sided steel blade. (**b**) A trapezoidal shape is created around the specimen. (**c**) The specimen is clamped in the custom-made trimming device for fine trimming. The excess Epon has already been removed, as can be seen in (**d**). The double-sided razor blade is now used to precisely trim the trapezoid around the specimen (**e**). The upper and lower edges must be parallel; in addition, the surface should be trimmed to create a largely smooth surface with the double-sided razor blade, and it should be transparent. (**f**) The sample is now ready to be sectioned in the ultramicrotome. Start with the DiATOME Histo knife and cut semithin sections (250 nm)

that the blade is no longer sharp. At a certain angle of illumination, the dark sample can be immediately recognized in the section or on the Epon surface. The last 100 µm in front of the sample or in front of the region of interest should be trimmed with a trimming diamond followed by the Histo diamond knife in the ultramicrotome.

3. Another option for trimming is to use a trimming diamond (20° or 45° diamond) to trim the surface shortly before the last 100 μm in front of the sample is reached and then to trim the sides. This is much more accurate and protects the ultra-diamond knife; splinters from the razor blades on the block can damage the diamonds.

3.5 Sections

Once the probe is clamped into the ultramicrotome probe holder, sectioning generates trapezoid-shaped single sections with an identifiable and defined orientation of the embedded embryo, larva, pupa, or adult fly.

1. Use the Histo diamond knife to produce semithin sections of 200–400 nm thickness.
2. It is necessary to start with semithin sections to determine whether the desired region of the specimen has been reached. For this purpose, stain individual sections with toluidine blue for light microscopic examination. Transfer a selected section using a steel loop (either a self-made loop or one available for purchase, such as the Perfect Loop®) onto a drop of water (deionized distilled water) on a gelatine-coated slide. To adhere the section, place the slide on a hot plate (80 °C) until the water droplet has evaporated. Next, add a drop of filtered (filter pore size 0.2 μm) toluidine blue staining solution, and leave the slide on the hot plate at 80 °C for 2–3 min (*see* **Note 8**).
3. Next, rinse the slide with distilled water to remove excess toluidine blue. The slide can now be examined under a light microscope. Add a coverslip for objectives with 20× and 40× magnifications (it is noted on each microscope objective whether a coverslip, usually 0.17 mm thick, is required). Continue producing semithin sections and staining with toluidine blue until the region of interest is reached.
4. Once the desired region is reached, prepare a sufficient number of ultrathin sections of either 60 or 70 nm with the Histo diamond knife. Alternatively, switch to an ultra-diamond knife to produce the highest quality ultrathin sections. The sections, arranged in ribbons, float on the surface of the water in the collection vessel of the knife and are now ready to be transferred onto grids.
5. Using fine tweezers, dip a Formvar® coated grid into the catching vessel of the knife under the water surface at a sufficient distance from the sections. Carefully move the coated grid under the sections. Some practice will allow the sections to be caught easily by lifting the grid slowly and then positioning them on the Formvar film of the grid. In other laboratories, a steel loop is used to approach the ultrathin sections from above with a hanging drop of water. If the droplet of water

touches the sections, then they remain attached to the water droplet and can be deposited on the coated grids.

6. All *Drosophila* probes analyzed thus far in our laboratory have been placed on 1 × 2 mm custom-made Formvar®-coated copper grids (*see* Subheading 3.8). However, we strongly recommend buying manufactured, coated, ready-to-use grids from a trusted TEM supplier. TEM grids are available with different materials, mesh sizes, and thicknesses from various suppliers, such as Science Services (Munich, Germany). It requires time to gain sufficient experimental experience to produce grids. It is only worthwhile to produce grids independently if an imaging facility with experience in electron microscopy and sample preparation is available on site. Always use special grid tweezers to handle grids, and use dedicated clean grid boxes for long-term storage.

3.6 Post-Staining (Contrasting with Heavy Metal Ions)

Staining the sections is necessary to enhance the contrast of the electron microscope image. During this process, atoms with a high atomic number, such as lead, uranium, or both, are introduced into the samples. They scatter the electron beam much more strongly than atoms with a low atomic number (which naturally dominate in the tissue) and thus increase the contrast of the resulting electron microscopic image.

1. For safety reasons, the post-staining of the sections (already placed on the grids) with contrasting agents should be carried out in a specially designed space in the EM laboratory or the fume hood. We use staining machines (the nanofilm surface analysis system and a Leica EM AC 20) for post-staining. However, many EM groups perform contrasting manually (*see* Fig. 6).
2. Stain sections first for 30 min in 2% uranyl acetate and then for 20 min in 3% lead citrate.
3. After post-staining the grids, store them in grid storage boxes.

3.7 From Post-Staining to Transmission Electron Microscopy Images

Once the *Drosophila* samples are embedded, sectioned, and post-stained, they can be analyzed in a TEM (see Fig. 7). For imaging, we use a Zeiss 120 kV TEM (Leo 912), a Zeiss 120 kV TEM (Libra 120), or a Jeol 200 kV TEM (JEM 2100 Plus).

3.8 Grids

We As mentioned earlier, we strongly recommend buying ready-to-use Formvar®-coated TEM grids from a trusted TEM supplier. However, a brief description of our own protocol for preparing Formvar®-coated slot grids is as follows:

1. Clean microscope slides (*see* **Note 9**) for 5 min in an ultrasonic bath in 2% Extran®.

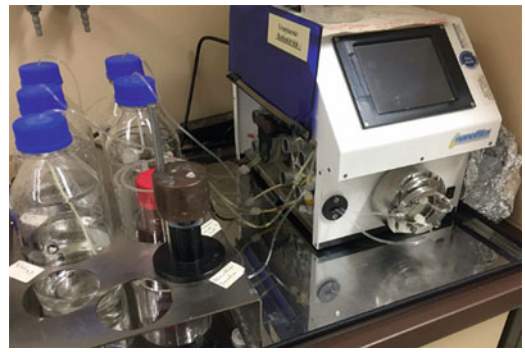


Fig. 6 Staining machine. Post-staining of the sections placed on grids in the staining device

2. Next, wash the slides in the ultrasonic bath with deionized water twice for 5 min each, transfer them into 70% ethanol for 5 min in the ultrasonic bath, and, then, wash them two more times in deionized water for 5 min each.
3. Dry the slides in a gas burner flame, and store them in a clean beaker with a piece of filter paper on the bottom. To keep dust out, close the beaker with the lid of a Petri dish.
4. Prepare an oval glass cuvette filled with Formvar solution. Then immerse the slides (held with tweezers with flattened tips) into the Formvar solution to within 1 cm of the edge of the slides and then slowly and evenly pull them out of the solution. The faster this step is performed, the thicker the Formvar® film will be. The optimum speed must be determined individually through experience.
5. Next, carefully remove the Formvar® foil from the slides. To achieve this, immerse the Formvar®-coated slides in water. This is best accomplished by filling an extremely clean large glass dish with Millipore water so that a "water mountain" forms. Wipe the water surface with a clean glass rod to completely remove any lint from the surface.
6. Next, take a coated slide from the beaker, grasping it only at the top uncoated area, and score the sides and bottom with a sharp single-edged razor blade (cutting, not scraping). Do not touch the surface with your fingers.
7. Then dip the slide once quickly into the water and immediately afterwards place the lower edge very slowly on the water surface and then dip the slide very slowly. Ideally, the film will detach from both sides of the slide. Often, the whole film gets stuck or comes off on only one side. Whether this occurs depends on many parameters, including humidity, slide batch, the Formvar® solution, and the slide cleaning procedure. If the

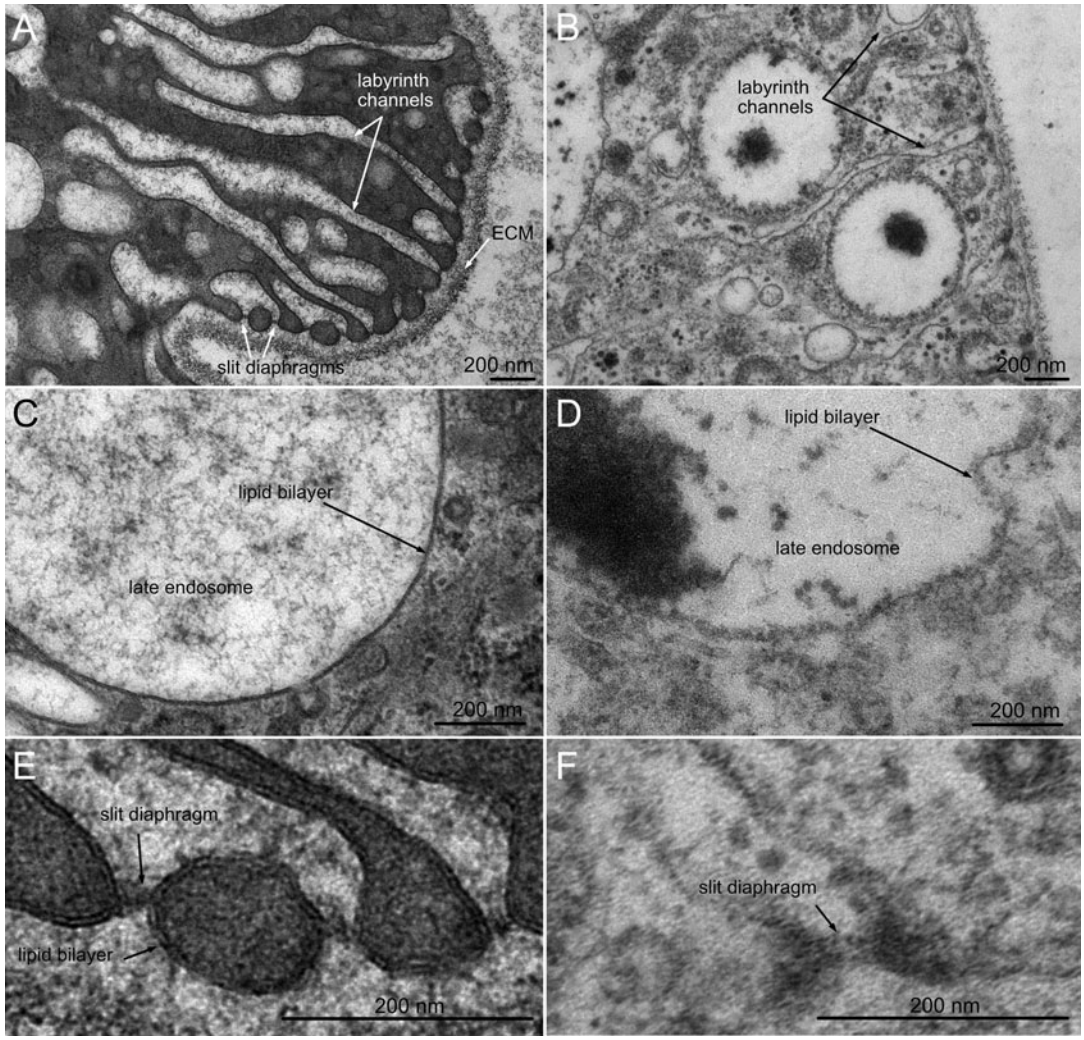


Fig. 7 Transmission electron microscopy images of nephrocytes from the third instar larvae. (**a, c, e**) High-pressure freezing–freeze substitution (HPF–FS), (**b, d, f**) chemically fixed samples. Note the preservation of the lipid bilayer of the membranes as double black lines in the HPF–FS samples (**c, d,** arrows) compared with the chemically fixed samples, in which the membranes are visible as one line (**d**). Particularly striking in the HPF–FS samples is the good preservation of the vesicular structures without shrinkage artifacts (**a**, **c,** endosome), which occur in vesicular organelles during chemical fixation (**d,** endosome). The so-called slit diaphragms (filtration apparatus of the nephrocytes) are also well preserved in the chemical- and HPF-prepared samples (**e, f**). ECM extracellular matrix

film sticks in one place, tweezers can be used in an attempt to detach the film; however, this often does not work, and the film is lost.

8. When the Formvar® film floats on the water surface, it can be covered with grids. Using tweezers, place the grids in short, regular intervals on the Formvar® film. A reasonable number

of grids is 30–50 on one film. Currently, our most commonly used grids are 1 × 2 mm. The film should have a silvery sheen. If it is golden, it is too thick; if it is transparent, it is too thin. Place the grids only on the smooth surfaces of the film, not on the waves.

9. The grids have a glossy and a matte side. We always place the matte side of the grids down so that the glossy side is facing us. The grids should always be placed evenly on the film, and not too close together.
10. Use the clean part of a piece of Parafilm® that was covered with paper to cover the plastic films holding the grids. The film, including the grids, should lift from the water surface and stick to the Parafilm®.
11. Store the pieces of Parafilm® with grids and plastic film in Petri dishes and let them dry.
12. To use the grids, cut them out with fine tweezers.

3.9 Knives

In general, two types of knives are available: glass and diamond. Ultrathin sections can be produced perfectly with a glass knife, which accepts all types of embedding media, including the commonly used resins Durcupan®, Lowicryl®, and Epon®. Beginners should always practice making thin sections on the ultramicrotome with a glass knife before starting to work with diamond knives. Glass knives can be purchased commercially. Alternatively, they can be produced independently with a knife maker. We use a 7800 Histo Knifemaker from LKB (Germany).

However, for ultrathin sectioning diamond knives are preferable for obtaining optimal results. Diamond knives are high-precision instruments provided by specialized companies such as DiATOME (Switzerland). They are available with various specifications, optimized for semithin, ultrathin, or cryo-sections. To approach the relevant region within a sample, it is best to use a Histo diamond knife first. In practice, the Histo diamond knife is often used for ultrathin sectioning as well. However, the Histo diamond knife is typically only changed to an ultra-diamond knife to create excellent sections, such as when generating images for publication.

Extreme caution must be exercised with diamond knives because even the slightest faulty touch will damage the cutting edge. Great care should also be taken when trimming the Epon block with a razor blade. As soon as pressure is applied with the razor blade, tiny metal particles can be deposited on the surface of the Epon block, which leads to the chipping of the diamond. Switching from glass knives to diamond knives for the same sample is not recommended, as residual glass debris can damage the diamond knife edge and lead to undesirable marks in the subsequent ultrathin section. Thus, it is better to use a trimming diamond first.

4 Notes

1. Aldehyde is a polymerization product of formaldehyde in powder form. It depolymerizes to formaldehyde in water when heated. Such solutions are used as fixatives for sample preparation in histology.
2. Embryos stick to plastic; use glassware instead (a scintillation vial is suitable).
3. Osmium tetroxide fixation is used to preserve membranes. Because of its extremely slow penetration efficiency, osmium tetroxide is almost always used as a secondary fixative after primary fixation in aldehyde.
4. After this step, the procedure can be paused, and the samples can be stored overnight at 4–6 °C in the refrigerator in the wash buffer.
5. In practice, an alcohol:water series is often used instead of an alcohol:buffer series. This works in many cases but is an additional burden on the samples, especially in the 70:30 and 50:50 mixtures. In addition, the incubation times depend on the sizes of the objects; 10 min is only sufficient for small samples.
6. Sylgard 184 (Dow Corning) is mixed according to the manufacturer's instructions (typically from an elastomer/hardener in a mixing ratio of 10:1, w/w) and poured into small glass or plastic Petri dishes. The dishes are filled approximately halfway with the elastomer. The pinned prepared animals must be able to be completely covered by liquid.
7. Larvae are easier to dissect if they are placed on ice for at least 10 min before dissection. However, we recommend anesthetizing as shortly as possible to avoid any kind of damage.
8. The ideal duration for staining a section to obtain a clear image depends on the temperature of the heating block, the amount and concentration of the toluidine blue stain, and the thickness of the section. This step must be optimized to meet individual needs and instruments. Keep in mind that an understained section can be restained.
9. We often observe that, for unknown reasons, some microscopic slide batches are not suitable for the production of defect-free Formvar® films. This is possibly due to interference from microscopic scratches on the glass surface. Therefore, we recommend checking Formvar® films made with a particular batch of slides under an electron microscope to ensure that they are free of defects. Only flawless films should be used.

Acknowledgments

We thank Kerstin Etzold, Birgit Hemmis, Mechthild Krabusch, Martina Biedermann, and Werner Mangerich for expert technical assistance. Furthermore, we thank Christian Meyer and Jonas Olbrich for sharing their TEM expertise with us. This work was supported by grants from the DFG (Deutsche Forschungsgemeinschaft) to A.P. (PA517/13-1, PA517/15-1, PA517/16-1, SFB 944-TP7, SFB 944 Z-Project). Our TEM protocols were initially compiled from Tepass and Hartenstein (1994); McDonald, Sharp, and Rickoll (2000); Lehmacher (2009, 2012), and many other authors.

References

1. Hurd TR, Sanchez CG, Teixeira FK, Petzold C, Dancel-Manning K, Wang JY, Lehmann R, Liang FX (2015) Ultrastructural analysis of *drosophila* ovaries by electron microscopy. Methods Mol Biol 1328: 151–162. https://doi.org/10.1007/978-1-4939-2851-4_11
2. Suvarna SK, Layton C, Bancroft JD (2019) Bancroft's theory and practice of histological techniques, 8th edn. Elsevier
3. Darley JJ, Ezoe H (1976) Potential hazards of uranium and its compounds in electron microscopy: a brief review. J Microsc 106:85–96
4. Tepass U, Hartenstein V (1994) Epithelium formation in the *Drosophila* midgut depends on the interaction of endoderm and mesoderm. Development 120:579–590
5. McDonald KL, Sharp DJ, Rickoll W (2000) Preparation of thin sections of *Drosophila* for examination by transmission electron microscopy. In: Sullivan W, Ashburner M, Hawley RS (eds) *Drosophila* protocols. Cold Spring Harbor Laboratory Press, New York, pp 245–271
6. Zhang S, Chen EH (2008) Ultrastructural analysis of myoblast fusion in *Drosophila*. Methods Mol Biol 475:275–297. https://doi.org/10.1007/978-1-59745-250-2_16
7. Lehmacher C, Abeln B, Paululat A (2012) The ultrastructure of *Drosophila* heart cells. Arthropod Struct Dev 41:459–474. https://doi.org/10.1016/j.asd.2012.02.002
8. McDonald KL, Sharp DJ, Rickoll W (2012) Transmission electron microscopy of thin sections of *Drosophila*: conventional chemical fixation of embryos using trialdehyde. Cold Spring Harb Protoc 7:516–520. https://doi.org/10.1101/pdb.prot068411
9. McDonald KL, Sharp DJ, Rickoll W (2012) Transmission electron microscopy of thin sections of *Drosophila*: high-pressure freezing and freeze-substitution. Cold Spring Harb Protoc 7:510–515. https://doi.org/10.1101/pdb.prot068403
10. McDonald KL, Sharp DJ, Rickoll W (2012) Transmission electron microscopy of thin sections of *Drosophila*: preparation of embryos using n-heptane and glutaraldehyde. Cold Spring Harb Protoc 7:1100–1103. https://doi.org/10.1101/pdb.prot068460
11. Wang S, Meyer H, Ochoa-Espinosa A, Buchwald U, Onel S, Altenhein B, Heinisch JJ, Affolter M, Paululat A (2012) GBF1 (Gartenzwerg)-dependent secretion is required for *Drosophila* tubulogenesis. J Cell Sci 125: 461–472. https://doi.org/10.1242/jcs.092551
12. Psathaki OE, Dehnen L, Hartley PS, Paululat A (2018) *Drosophila* pericardial nephrocyte ultrastructure changes during ageing. Mech Ageing Dev 173:9–20. https://doi.org/10.1016/j.mad.2018.04.006
13. Beyenbach KW, Schöne F, Breitsprecher LF, Tiburcy F, Furuse M, Izumi Y, Meyer H, Jonusaite S, Rodan AR, Paululat A (2020) The septate junction protein Tetraspanin 2A is critical to the structure and function of Malpighian tubules in *Drosophila melanogaster*. Am J Physiol Cell Physiol 318:C1107–C1122. https://doi.org/10.1152/ajpcell.00061.2020
14. Dehnen L, Janz M, Kumar Vermar J, Psathaki OE, Langemeyer L, Fröhlich F, Heinisch JJ, Meyer H, Ungermann C, Paululat A (2020) A trimeric metazoan Rab7 GEF complex is crucial for endocytosis and scavenger function. J Cell Sci 133:jcs247080. https://doi.org/10.1242/jcs.247080

15. Gerlitz O, Nellen D, Ottiger M, Basler K (2002) A screen for genes expressed in *Drosophila* imaginal discs. Int J Dev Biol 46: 173–176
16. Wolff T (2011) Preparation of *Drosophila* eye specimens for transmission electron microscopy. Cold Spring Harb Protoc 2011: 1386–1388. https://doi.org/10.1101/pdb.prot066514
17. Moor H, Riehle U (1968) Snap-freezing under high pressure: a new fixation technique for freeze-etching. Proc 4th Eur Reg Conf Electron Microsc 2:33–34
18. McDonald KL, Auer M (2006) High-pressure freezing, cellular tomography, and structural cell biology. BioTechniques 41:137–143. https://doi.org/10.2144/000112226
19. Small JV (1981) Organization of actin in the leading edge of cultured cells: influence of osmium tetroxide and dehydration on the ultrastructure of actin meshworks. J Cell Biol 91:695–705. https://doi.org/10.1083/jcb.91.3.695
20. McDonald KL (1994) Electron microscopy and EM immunocytochemistry. Method Cell Biol 44:411–444. https://doi.org/10.1016/s0091-679x(08)60926-7

Chapter 20

Segmentation and Quantitative Analysis of Epithelial Tissues

Benoit Aigouy and Benjamin Prud'homme

Abstract

Epithelial tissues regulate exchanges with the environment. They are highly dynamic and can acquire virtually any shape. At the cellular level, they are composed of cells tightly connected by junctions. Most often epithelia are amenable to live imaging; however, the vast number of cells composing an epithelium makes large-scale studies tedious. Here, we present Tissue Analyzer (TA), an open-source tool that can be used to segment epithelia and monitor cell and tissue dynamics.

Key words Epithelia, Quantitative biology, Tissue Analyzer, Cells, Segmentation, Tracking, 2D, 3D, Deep learning, Open source

1 Introduction

Epithelial tissues cover body structures, organs, and cavities. They act as barriers and regulate the passage of molecules between the internal milieu and the environment. *Drosophila* epithelia often consist of a single layer of cells tightly connected by junctions. These monolayers are highly dynamic and can acquire almost any shape. Their plasticity is driven by a series of cellular events, including cell death, proliferation, shape changes, and neighbor exchanges. Fully understanding epithelial development, therefore, requires the quantitative description of cellular events as well as collective behaviors. Such a quantitative analysis is facilitated, nowadays, by the advent of deep learning segmentation that largely automates cell segmentation. The latter can then, in combination with various computer science algorithms, be used to convert the pixel information of images into biological knowledge. Quantitative biology is not only accurate and unbiased but also allows for the development and in-depth validation of theoretical modeling of biological processes [1–6].

Christian Dahmann (ed.), *Drosophila: Methods and Protocols*,
Methods in Molecular Biology, vol. 2540, https://doi.org/10.1007/978-1-0716-2541-5_20,

Here, we introduce Tissue Analyzer (TA), an open-source Python software, capable of segmenting and analyzing single-layered epithelia. TA conveniently integrates, in a single package, all the tools necessary to study epithelial development, from cell segmentation using deep learning up to quantitative analysis of cell properties such as area and shape as well as neighbor relationships. In this chapter, we describe software installation (*see* Subheading 2.1) and image pre-processing (*see* Subheading 3.2) and introduce epithelial segmentation (*see* Subheading 3.3). Later, we discuss cell tracking (*see* Subheading 3.5). We finally provide a brief overview of the data collected by TA (*see* Subheading 3.7) and demonstrate some of the plotting capabilities of the software (*see* Subheading 3.8).

2 Software

2.1 Software Installation

1. The software can be installed on any recent computer capable of running Conda and Python 3.7 (or superior) (*see* **Notes 1** and **2**). Ideally, the computer should be equipped with a CUDA®-compatible graphic card (*see* **Note 4**).
2. Open the system command line (*see* **Note 3**).
3. Then, type "conda create -n TA python==3.7", without the quotes, to create and activate a Conda environment for Tissue Analyzer.
4. Type "conda deactivate" to leave the default Conda environment.
5. Enter the TA environment by typing "conda activate TA".
6. Finally, install Tissue Analyzer by running the following command: "pip install —upgrade epyseg".

2.2 Run Tissue Analyzer

1. Open a terminal (*see* **Note 3**).
2. Type "conda activate TA" in the terminal to activate the environment.
3. Type "python –m epyseg.ta" to open the TA graphical user interface.

2.3 Input Data

1. TA supports single and multichannel images.
2. TA supports pixel intensities up to 32 bits per channel.
3. TA reads TIF, CZI, LIF, LSM, JPEG, and PNG images (*see* **Notes 5** and **6**).
4. TA can create two-dimensional (2D) projections by extracting the surface of epithelial cells in three-dimensional (3D) stacks using deep learning [7].
5. TA can segment two-dimensional (2D) images of epithelial cells.

3 Applications

TA can be used to segment still images of immunostained epithelia or image sequences of a living sample expressing a fluorescently tagged protein.

3.1 Image Acquisition Guidelines

Ideally, the images to be segmented and analyzed should have a good signal-to-noise ratio. However, nowadays, good results can be obtained with images exhibiting a low signal-to-noise (*see* **Notes 7** and **8**).

1. TA is designed to detect cell outlines in single-layered epithelia – cells need to be labeled with a membrane marker (e.g., E-cadherin in *Drosophila* and zonula occludens-1 (ZO-1) in vertebrates).
2. If pixel intensity measurements are important for the study, make sure to avoid both bleaching and pixel saturation.
3. Post-acquisition, ensure the epithelial integrity of the live samples by verifying that the animal continues to develop normally.

3.2 Image Pre-Processing

Classically, epithelial images are acquired as 3D stacks; however, tissue segmentation is often faster, more reliable, and best suited for publication when performed on 2D images. Therefore, creating a 2D projection from a 3D epithelium is usually a good enough approximation of the epithelial tissue. This is especially true since most fly epithelia are monolayers lying in a plane, and this avoids the hassle of having to correct segmentation errors in 3D, which is long and difficult.

Such a 2D image can easily be created from a 3D stack using the maximum projection algorithm; however, this method accumulates the noise of the entire stack, generating a 2D image that is extremely noisy. To avoid noise accumulation and to select only the relevant planes carrying the epithelial signal within a wide and noisy 3D epithelial stack, several algorithms or deep learning methods have been developed [8–16, 7, 17]. Only the deep learning-based method for surface extraction [7] is currently implemented in TA (*see* **Note 9**).

However, there are rare cases, i.e., when the epithelial monolayer is not flat (e.g., the tissue is curved or bent), where the 2D projection can lead to severe underestimation of several cell metrics (e.g., cell area and perimeter) and needs be corrected. Importantly, fixing 2D cellular metrics to accommodate for nonflat tissues is easy to achieve; it requires a 2D segmentation mask along with a height map (depth localization) of the cells in the tissue [18, 12]. In parallel to extracting 2D cell cortices from 3D epithelial stacks, TA can also map cell heights. Combining 2D cell segmentation with a height map allows to obtain reliable 3D cell metrics

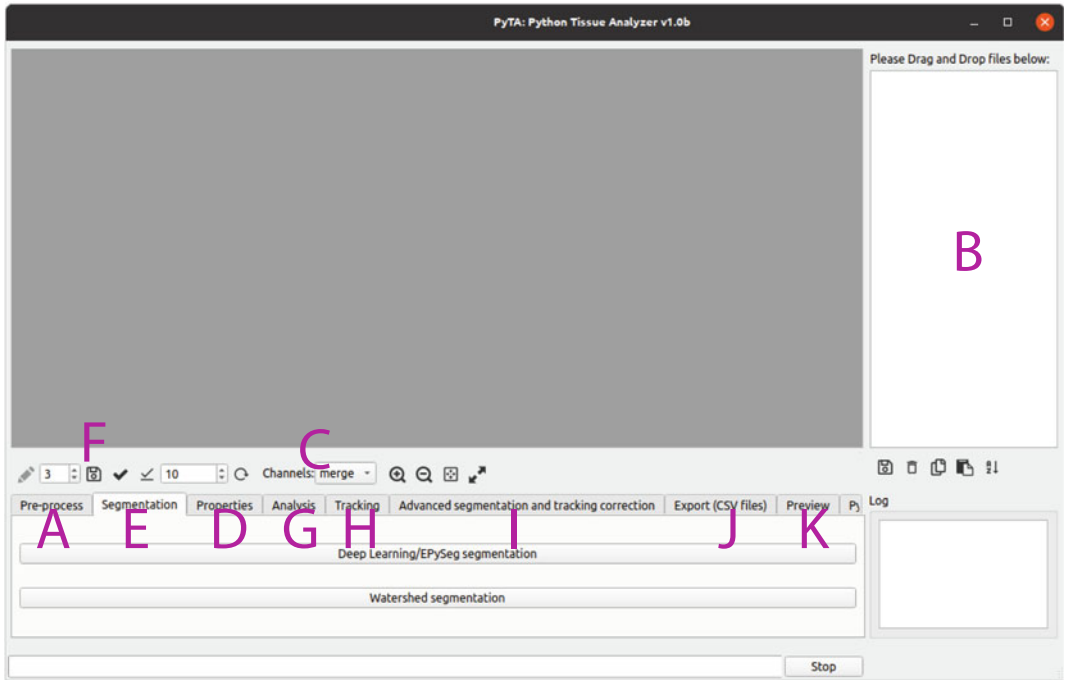


Fig. 1 The Tissue Analyzer user interface. (**a**) "Pre-processing" tab. (**b**) Image list where images can be dropped. (**c**) Channel selection tool. (**d**) "Properties" tab. (**e**) "Segmentation" tab. (**f**) "Save" button. (**g**) "Analysis" tab. (**h**) "Tracking" tab. (**i**) "Advanced segmentation and tracking correction" tab. (**j**) "Export (CSV files)" tab. (**k**) "Preview" tab

[12]. Below is a description of the procedure to perform surface extraction in TA using the content-aware image restoration (CARE) model to generate height maps from 3D epithelial images. Please note, however, that if the epithelium you study is perfectly flat or if you have already obtained a 2D projection using a third-party software (*see* **Note 9**), you may move directly to Subheading 3.3.

1. Select the "Pre-process" tab (*see* Fig. 1a).
2. Drag and drop single time points of epithelia 3D stacks over the TA interface (*see* Fig. 1b).
3. Select the channel containing the cell outlines (*see* Fig. 1c and **Note 10**).
4. Press "Surface projection" (*see* **Notes 11** and **12**).

3.3 Automated Segmentation

1. TA relies on EPySeg [19], a deep learning software, to segment epithelial cells. We will now guide you through the segmentation procedure.
2. Click on the "Segmentation" tab (*see* Fig. 1e).

3. If you skipped the pre-processing step (*see* Subheading 3.1), drag and drop 2D projections of epithelial cells over the TA interface, otherwise move to **step 4**.
4. Select the channel that needs to be segmented (*see* Fig. 1c).
5. Press the "Deep learning/EPySeg segmentation" button and wait until the segmentation mask is overlaid.

3.4 Correction of Segmentation Masks

Even though deep learning segmentation is globally excellent (compare Fig. 2a and b), few errors may remain. TA, therefore, allows editing segmentation masks (*see* Fig. 2).

1. Select an image from the list, and scroll over it to detect the missing or erroneous cell junctions (*see* **steps 2** and **3** and **Notes 13** and **14**).
2. Adding missing cell contacts: Left-click the centroid of all the cells surrounding the missing junctions (compare Fig. 2c and d) and press "Ctrl/Cmd+M" to run the watershed algorithm [20] locally and have the missing junctions appear (*see* Fig. 2e).
3. Removing false cell contacts: Right-click any pixel of the junction (compare Fig. 2i and j) and press "Enter" or "Shift +Enter." All the pixels contained between the two vertices of the targeted junction will disappear (*see* Fig. 2k).
4. Once editing the segmentation mask of an image is done, press the "Save" button (*see* Fig. 1f) to store the changes.
5. Finalize segmentation, and compute cell metrics.
   (a) Repeat **steps 2–5** for all the images in the list until there are no segmentation errors left.
   (b) Select the "Analysis" tab (*see* Fig. 1g), and press the "Measure cell properties" button.

3.5 Cell Tracking

Identifying the same cells in time and space is essential to understand morphogenesis. Cell tracking is achieved in TA by matching cells in consecutive frames and assigning them a unique identity (represented by a unique color) throughout the movie (*see* **Note 15**).

1. Please complete "Epithelial cell segmentation" before tracking cells.
2. Select the "Tracking" tab (*see* Fig. 1h).
3. Press "Track cells."

3.6 Correction of Cell Tracks

The tracking algorithm in TA is usually robust, especially when the segmentation is good. It can also accommodate for large time gaps between consecutive frames. However, tracking may, in rare cases, contain two types of errors: (i) cell swapping errors where two adjacent cells exchange their identities (compare Fig. 3d and e)

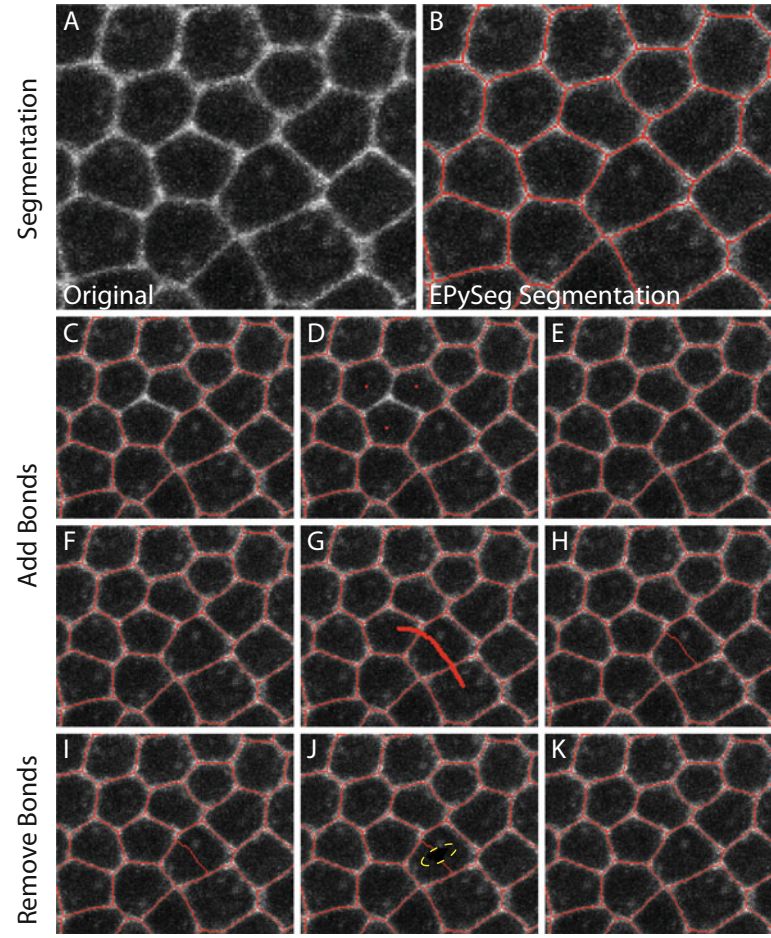


Fig. 2 Epithelial segmentation and mask editing. (**a**) Original image. (**b**) The red overlay is the result of deep learning (EPySeg) segmentation of the image shown in (**a**). (**c**) Image with three missing junctions. (**d**) The same as (**c**), where cell centroids have been drawn. (**e**) Locally running the watershed algorithm ("Ctrl/CMD+M") using seeds drawn in (**d**) restores the missing contacts. (**f**) Segmented image. (**g**) The same as in (**f**), except that a new junction bisecting a cell has been drawn manually (note that the new junction is intersecting two existing junctions). (**h**) The software removes extra mask pixels when "Enter" is pressed. (**i**) Segmented image. (**j**) The same image as in (**i**), except that a small piece of a cell contact mask has been removed using the mouse right-click (dashed outline). (**k**) The entire junction is removed after pressing "Enter"

and (ii) interrupted tracks where the identity of a cell is changed from one frame to the next (compare Fig. 3e and f). TA now ships with a tool to rapidly spot tracking and segmentation errors (*see* Fig. 3). This tool measures the similarities of the cellular environments in consecutive frames (*see* Fig. 3g, h). A low score (i.e., a dark color-coded cell) using this metric indicates that the cell environment is rapidly changing; this is often (*see* **Note 16**) due to a

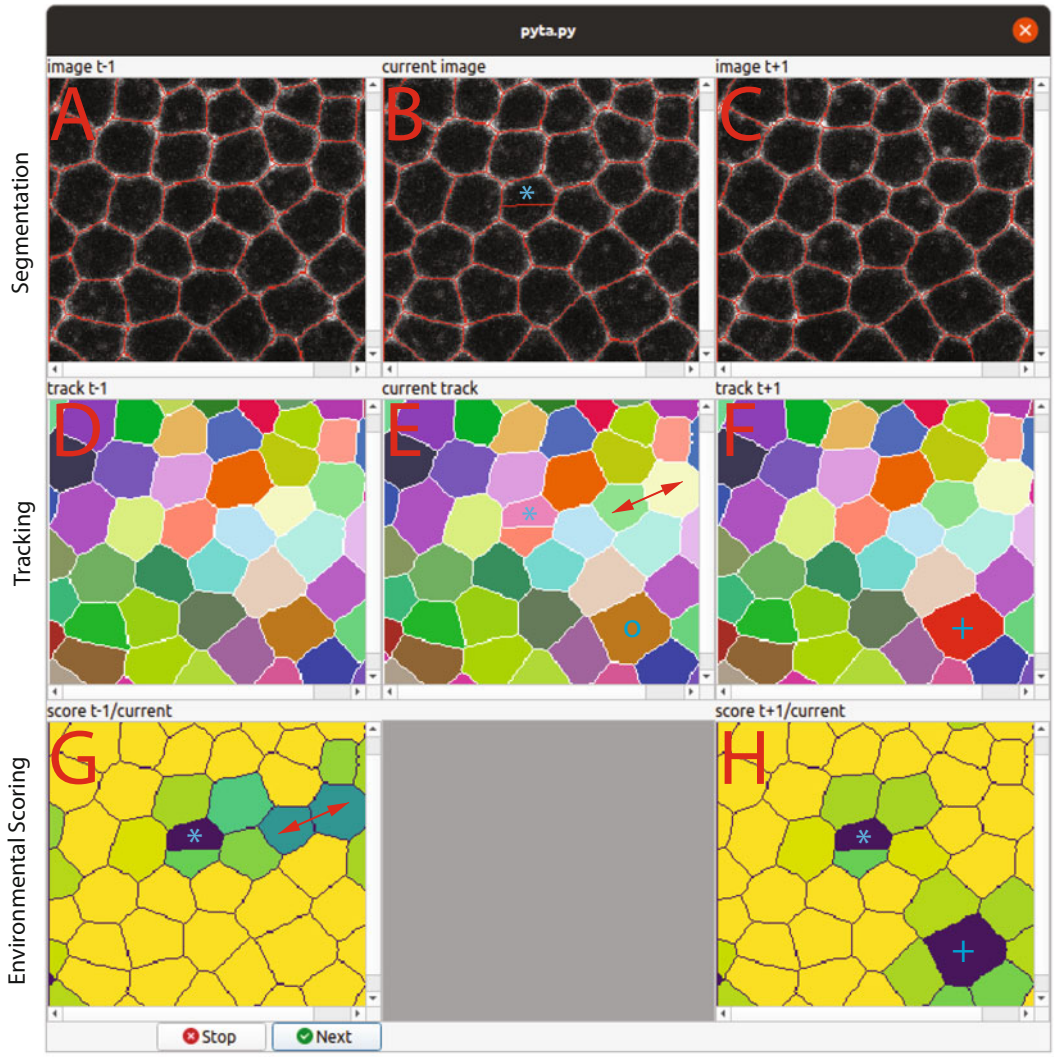


Fig. 3 Tool to rapidly detect tracking and segmentation errors. (**a**–**c**) Three consecutive time points of the same epithelial tissue overlaid with their segmentation masks in red. (**d**–**f**) Cell tracks corresponding to (**a**–**c**). (**g**, **h**) Color code of the environmental score obtained when comparing the current track with the previous (**g**) and next tracks (**h**). Any dark-colored cell in (**g**, **h**) has undergone important environmental changes between consecutive frames and needs be scrutinized for potential errors. In contrast, bright cells in (**g**, **h**) have kept a similar environment between consecutive frames and can be ignored safely. For example, the asterisk in (**g**) indicates an oversegmented cell. This corresponds to a newly appeared cell (asterisk in **e**) that does not exist in the previous (**d**) or next time frame (**f**). This oversegmentation is visible in the segmentation mask (asterisk in **b**); it can be fixed as described in Subheading 3.4, **step 3**. The arrow in (**e**) indicates two cell tracks that have been swapped (compare with **d**). Note that this cell swap generates a (darkening) strong environmental change (arrow in **g**). Cell swapping can be fixed as described in Subheading 3.6, **step 2**. Finally, the dark color-coded cell (labeled with a + in **h**) shows a strong environmental change caused by a tracking failure. Compare the identity given to the cell in (**e**) and (**f**) (o and + symbols in **e** and **f**). This error can be fixed as described in Subheading 3.6, **step 3**

segmentation or a tracking error and calls for a careful user inspection. In contrast, a high score (a bright color-coded cell) indicates environmental preservation and does not require user attention.

1. Visually identifying tracking or segmentation errors:
   (a) Select the "Advanced segmentation and tracking correction" tab (*see* Fig. 1i).
   (b) Click on the "Fix masks and tracks dynamically" button, and wait until a window with three rows appear (*see* Fig. 3).
   (c) Visually identify a dark cell (*see* Fig. 3g, h) and zoom ("Ctrl/Cmd +") on it and move the mouse to it.
   (d) Using the synchronization of the mouse pointer, search for associated segmentation errors (*see* Fig. 3a–c) and fix them as described in Subheading 3.4. If there is no segmentation error, check the tracks (*see* Fig. 3d–f) for errors (swapping error or lost track) and fix them (*see* **steps 2** and **3**).
2. Correction of a cell swapping error:
   Please correct all the segmentation errors and repeat the tracking procedure (*see* Subheading 3.5) before attempting to correct for tracking errors. Indeed, as most tracking errors are linked to segmentation errors, fixing the latter is often sufficient to fix the former. However, if tracking errors persist, please do as follows:
   (a) Move the mouse cursor over the image where the swapping is first detected (*see* arrow in Fig. 3e).
   (b) Left-click the centroid of the two swapped cells.
   (c) Press "Enter" while keeping the mouse pointer on this image to fix the error for the current and following frames.
3. Correction of an interrupted track error:
   (a) Place the mouse cursor on the last image where the cell is still tracked properly (i.e., shows the right color/identity in Fig. 3e).
   (b) Left-click the cell that is about to lose its identity/color (o).
   (c) Place the mouse cursor on the following image (*see* Fig. 3f) and left-click the cell with the wrong identity/color (+).
   (d) Press "Enter" while keeping the mouse on this image; this will fix the tracking error for the current and following frames.

3.7 Data Analysis and Quantifications

TA provides the user with a thorough description of cells including area, perimeter, elongation, planar distribution of proteins, and cell neighborhood (number and identity of neighbors) [1]. TA additionally provides information on junction length, orientation, and intensity. Most cell metrics can also be corrected for tissue bending as if they would have been obtained directly from a 3D segmentation when a height map is provided (also *see* Subheading 3.2).

Data extracted by TA are stored in an internal SQLite database (*see* **Notes 17** and **18**). These data can also be exported to facilitate plotting using third-party software as follows:

1. Click on the "Export (CSV files)" tab (*see* Fig. 1j).
2. To export cell data, press the "Export cell data" button.
3. To export junction data, press the "Export bond data" button.

3.8 Data Presentation

Because visual presentation of quantitative data is often key to human understanding of tissue dynamics, TA offers the possibility to overlay/blend virtually any data over the original epithelial images.

1. Select the "Preview" tab (*see* Fig. 1k).
2. Select the data to be visualized from the "Image/Data" drop-down list.
3. Enable "Color coding."
4. Select the look-up table to be used from the "LUT" drop-down list.
5. Enable "Overlay/Blend" and select the channel of interest for the background image.
6. Set the "foreground opacity" value (0 totally transparent; 1 totally opaque).
7. The view can be exported as a TIF for use with third-party tools or for publication by pressing the "Single image" or the "Stack" button.

3.9 Going Further with Tissue Analyzer

Here, we have covered the basic TA functionalities from surface extraction to segmentation and data presentation. However, due to space limitation and for the sake of simplicity, we have left out several useful TA functionalities. We leave it to you to try by yourselves more advanced functionalities that have not been documented here. To assist you in that process, we offer an online tutorial (https://github.com/baigouy/tissue_analyzer). Alternatively, place your mouse cursor over the TA component of interest and wait for a few seconds until a small tooltip text appears.

4 Notes

1. If Python is not installed in your system or if the installed version is <3.7 (*see* **Note 2**), then please download and install Miniconda from https://docs.conda.io/en/latest/miniconda.html
2. To determine the version of Python running on a computer:
   (a) Open a terminal (*see* **Note 3**).
   (b) Type in the terminal "python -V" without the quotes.
3. Open a terminal:
   (a) On a Mac: open the command line terminal by pressing together the Command and Space keys, and then type in the new window "terminal" without the quotes.
   (b) On Windows: press the "Windows" key and the "R" letter, and then type "cmd" without the quotes.
   (c) On Linux (Ubuntu): press Ctrl+Alt+T.
4. TA currently relies on deep learning for segmentation and image pre-processing. Even though a graphic card is not mandatory to use trained deep learning models, being equipped with a CUDA®-compatible graphic card will significantly speed up the analysis. For example, we recommend using, whenever possible, Nvidia® graphic cards with, at least, Turing or Ampere microarchitectures.
5. The JPEG compression is deeply modifying pixel data and should not be used for quantifications. Instead, we recommend using TIF images with non-lossy or no compression.
6. When a file format cannot be read by TA, open the image using a third-party tool such as ImageJ or FIJI [21, 22], and then export it as a TIF file. TIF files can be read flawlessly by TA.
7. Deep learning segmentation models can be trained using images having a low signal-to-noise ratio; such models can then achieve good segmentation even for tissues showing weak fluorescence.
8. Deep learning models can be used to reconstruct/denoise images prior to segmentation [7]. Note, however, that if intensity measurements are key for your analysis, denoising algorithms should be used with a lot of care, as they heavily modify pixel intensities and are shown to statistically guarantee a good quality and a biologically relevant denoising only when the images used are of the same tissue and acquired under the exact same conditions as the ones used for training the model. This will almost inevitably not be the case for your tissue unless you do train the model on your own data, a step that is often difficult and time-consuming.

9. In our hands, surface extraction is best achieved using deep learning.
10. TA output might be channel-dependent, and the software will warn the user when channel selection is required.
11. Most often, surface extraction works well, but it is rarely error-free; very importantly, any error in the height map will impair the 3D correction of 2D measurements such as area and perimeter. It is, therefore, important to check for the quality of the height map. To help you visually assess height map quality, TA generates a file called "height_map_quality_test.tif" for each surface extraction it performs. If there are discontinuities/holes in the cortical signal of the region of interest, then the height map is not of good quality and the 3D corrections applied to the 2D cell metrics will be locally erroneous. In such case, and if the epithelium studied is flat, it is probably a good idea to rely on raw 2D measurements rather than using the 3D corrected ones, or, alternatively, to correct the height map. Finally, when using a third-party tool to create the 2D surface projections from a 3D epithelial stack, we recommend that you set the projection tool in a manner so that only one frame is used for building the final 2D projection; this allows to see the height map errors better. Indeed, height map errors are hard to visualize when several frames are summed or averaged to create a 2D projection, and most tools use several frames to compensate for their low-quality height map extraction.
12. At this step, TA must have extracted the voxel depth/voxel width ratio. Given the importance of this ratio for the 3D correction of 2D cell metrics, it is necessary to verify its correctness. This can be done in the "Properties" tab of TA (*see* Fig. 1d). If there is an error, it is possible to edit the value stored in the table and to save those changes.
13. To enter the full screen mode and see cells better, press "F12."
14. Some segmentation errors may be hard to see. It is often useful to switch back and forth between the original image and the image overlaid with its segmentation mask to detect errors. This can be achieved by pressing the "M" key repeatedly, to show or hide the segmentation mask.
15. When tracked, each cell is given a unique pseudorandom color, defining its identity. A cell keeps this identity as long as it is successfully tracked. It is attributed a new identity when lost.
16. Please note that a low score is not necessarily due to an error in segmentation or tracking; it can instead indicate that a cell underwent a rapid, yet natural, environmental change (e.g., such a change can result from a division or a cell delamination).

17. For each image, Tissue Analyzer stores the segmentation mask and various output files in a folder. To open this folder, double-click the file of interest in the list (*see* Fig. 1b).
18. Those databases can be accessed directly by users having experience with manipulating Structured Query Language (SQL) databases. Since SQL databases are heavily used by TA, it is important that all TA-generated tables remain unaltered. Experienced users can nevertheless create extra tables or duplicate the existing ones and modify them at will. We recommend closing TA before accessing databases using an external software to avoid database locking problems and, similarly, we recommend closing external database browsers before launching TA.

Acknowledgments

This work was supported by the CNRS.

References

1. Aigouy B, Farhadifar R, Staple DB, Sagner A, Röper J-C, Jülicher F, Eaton S (2010) Cell flow reorients the axis of planar polarity in the wing epithelium of Drosophila. Cell 142(5): 773–786. https://doi.org/10.1016/j.cell.2010.07.042
2. Bosveld F, Bonnet I, Guirao B, Tlili S, Wang Z, Petitalot A, Marchand R, Bardet P-L, Marcq P, Graner F, Bellaïche Y (2012) Mechanical control of morphogenesis by Fat/Dachsous/four-jointed planar cell polarity pathway. Science 336(6082):724–727. https://doi.org/10.1126/science.1221071
3. Etournay R, Popović M, Merkel M, Nandi A, Blasse C, Aigouy B, Brandl H, Myers G, Salbreux G, Jülicher F, Eaton S (2015) Interplay of cell dynamics and epithelial tension during morphogenesis of the Drosophila pupal wing. eLife 4:e07090. https://doi.org/10.7554/eLife.07090
4. Saadaoui M, Rocancourt D, Roussel J, Corson F, Gros J (2020) A tensile ring drives tissue flows to shape the gastrulating amniote embryo. Science 367(6476):453–458. https://doi.org/10.1126/science.aaw1965
5. Umetsu D, Aigouy B, Aliee M, Sui L, Eaton S, Jülicher F, Dahmann C (2014) Local increases in mechanical tension shape compartment boundaries by biasing cell intercalations. Curr Biol 24(15):1798–1805. https://doi.org/10.1016/j.cub.2014.06.052
6. Valon L, Davidović A, Levillayer F, Villars A, Chouly M, Cerqueira-Campos F, Levayer R (2021) Robustness of epithelial sealing is an emerging property of local ERK feedback driven by cell elimination. Dev Cell 56(12): 1700–1711, e1708. https://doi.org/10.1016/j.devcel.2021.05.006
7. Weigert M, Schmidt U, Boothe T, Müller A, Dibrov A, Jain A, Wilhelm B, Schmidt D, Broaddus C, Culley S, Rocha-Martins M, Segovia-Miranda F, Norden C, Henriques R, Zerial M, Solimena M, Rink J, Tomancak P, Royer L, Jug F, Myers EW (2018) Content-aware image restoration: pushing the limits of fluorescence microscopy. Nat Methods 15(12): 1090–1097. https://doi.org/10.1038/s41592-018-0216-7
8. Basu S, Rexhepaj E, Spassky N, Genovesio A, Paulsen RR, FastSME SA (2018) Faster and smoother manifold extraction from 3D stack. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2018-06, pp 2362–23628. https://doi.org/10.1109/CVPRW.2018.00305
9. Blasse C, Saalfeld S, Etournay R, Sagner A, Eaton S, Myers EW (2017) PreMosa: extracting 2D surfaces from 3D microscopy mosaics. Bioinformatics 33(16):2563–2569. https://doi.org/10.1093/bioinformatics/btx195

10. Erguvan Ö, Louveaux M, Hamant O, Verger S (2019) ImageJ SurfCut: a user-friendly pipeline for high-throughput extraction of cell contours from 3D image stacks. BMC Biol 17(1): 38. https://doi.org/10.1186/s12915-019-0657-1
11. Forster B, Ville DVD, Berent J, Sage D, Unser M (2004) Complex wavelets for extended depth-of-field: a new method for the fusion of multichannel microscopy images. Microsc Res Tech 65(1-2):33–42. https://doi.org/10.1002/jemt.20092
12. Herbert S, Valon L, Mancini L, Dray N, Caldarelli P, Gros J, Esposito E, Shorte SL, Bally-Cuif L, Aulner N, Levayer R, Tinevez J-Y (2021) LocalZProjector and DeProj: a toolbox for local 2D projection and accurate morphometrics of large 3D microscopy images. BMC Biol 19(1):136. https://doi.org/10.1186/s12915-021-01037-w
13. Li K, Wu X, Chen DZ, Sonka M (2006) Optimal surface segmentation in volumetric images-a graph-theoretic approach. IEEE Trans Pattern Anal Mach Intell 28(1): 119–134. https://doi.org/10.1109/TPAMI.2006.19
14. Lombardot B (2017) Minimum cost Z surface projection. https://imagej.net/Minimum_Cost_Z_surface_Projection
15. Shihavuddin A, Basu S, Rexhepaj E, Delestro F, Menezes N, Sigoillot SM, Del Nery E, Selimi F, Spassky N, Genovesio A (2017) Smooth 2D manifold extraction from 3D image stack. Nat Commun 8(1):15554. https://doi.org/10.1038/ncomms15554
16. Umorin M (2006) Stack focuser. https://imagej.nih.gov/ij/plugins/stack-focuser.html
17. Wu X, Chen DZ (2002) Optimal net surface problems with applications. In: Widmayer P, Eidenbenz S, Triguero F, Morales R, Conejo R, Hennessy M (eds) Lecture notes in computer science. Springer, Berlin, Heidelberg, pp 1029–1042. https://doi.org/10.1007/3-540-45465-9_88
18. Barbier de Reuille P, Routier-Kierzkowska A-L, Kierzkowski D, Bassel GW, Schüpbach T, Tauriello G, Bajpai N, Strauss S, Weber A, Kiss A, Burian A, Hofhuis H, Sapala A, Lipowczan M, Heimlicher MB, Robinson S, Bayer EM, Basler K, Koumoutsakos P, Roeder AH, Aegerter-Wilmsen T, Nakayama N, Tsiantis M, Hay A, Kwiatkowska D, Xenarios I, Kuhlemeier C, Smith RS (2015) MorphoGraphX: a platform for quantifying morphogenesis in 4D. eLife 4:e05864. https://doi.org/10.7554/eLife.05864
19. Aigouy B, Cortes C, Liu S, Prud'Homme B (2020) EPySeg: a coding-free solution for automated segmentation of epithelia using deep learning. Development 147(24): dev194589. https://doi.org/10.1242/dev.194589
20. Vincent L, Soille P (1991) Watersheds in digital spaces: an efficient algorithm based on immersion simulations. IEEE Trans Pattern Anal Mach Intell 13(6):583–598. https://doi.org/10.1109/34.87344
21. Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, Preibisch S, Rueden C, Saalfeld S, Schmid B, Tinevez J-Y, White DJ, Hartenstein V, Eliceiri K, Tomancak P, Cardona A (2012) Fiji: an open-source platform for biological-image analysis. Nat Methods 9(7):676–682. https://doi.org/10.1038/nmeth.2019
22. Schneider CA, Rasband WS, Eliceiri KW (2012) NIH Image to ImageJ: 25 years of image analysis. Nat Methods 9(7):671–675. https://doi.org/10.1038/nmeth.2089

Chapter 21

Genetically Encoded Sensors to Study Metabolism in *Drosophila*

Ellen McMullen, Helen Hertenstein, Stephan Müller, and Stefanie Schirmeier

Abstract

The rather recent development of genetically encoded metabolite sensors has changed the way we can study metabolism in living cells, ex vivo tissues, and in vivo immensely. In recent years, these sensors have also been adapted for use in *Drosophila* tissues. Here, we describe a standard protocol to image such sensors in ex vivo *Drosophila* larval brains using the glucose sensor FLII12Pglu-700μδ6. The protocol, however, can be adapted for the use of other sensors, tissues, and can even be used in vivo.

Key words Genetically encoded metabolite sensors, Live imaging, Metabolism, FRET, Carbohydrate transport

1 Introduction

For decades, cellular integrity was one of the main obstacles precluding the discovery of what lies within a cell, as well as limiting knowledge of the intracellular processes that occur. This is especially true when it comes to the question of how metabolites enter cells, travel through cells, and are metabolized within cells. Therefore, the development of tools that enable monitoring of intracellular processes, while allowing the tissue to remain intact, was a major scientific breakthrough of the late twentieth century. Since the development of the first genetically encoded calcium indicator [1], this ever-expanding field has had a tremendous impact on noninvasive, target-oriented biological research.

In general, any genetically encoded fluorescent indicator (GEFI) consists of three major components: one or multiple

Supplementary Information The online version contains supplementary material available at [https://doi.org/10.1007/978-1-0716-2541-5_21].

Christian Dahmann (ed.), *Drosophila: Methods and Protocols*,
Methods in Molecular Biology, vol. 2540, https://doi.org/10.1007/978-1-0716-2541-5_21,

fluorescent proteins, a target-binding domain, and a protein scaffold. In theory, this basic layout can give rise to as many biosensors as there are targets in a cell. However, after the first nongenetically encoded Förster resonance energy transfer (FRET)-based biosensor for cAMP (3′2,5′2-cyclic adenosine monophosphate; [2]), it took another 15 years until the development of the first single-fluorophore (SF) small molecule-binding GEFI (H_2O_2, [3]). Although many advancements have been made over the last two decades (reviewed in ref. (4–7)), there is still a lack of concise and comprehensive hands-on manuals on how to best approach GEFIs as well as the considerations that should be made in experimental design when working with them.

This chapter aims to give researchers an overview of the advantages and disadvantages of metabolite sensors as well as what to look out for when using such a sensor as a drosophilist. At present, there are two main classes of GEFIs: FRET (Förster resonance energy transfer)-based biosensors and single-fluorophore (SF) biosensors, which we will describe in more detail.

FRET Sensors

Among groups of biosensors with two (or more) fluorophores, FRET probes are the most prevalent. FRET-based sensors were among the first biosensors to be available [2]. The underlying principle is based on a spectral overlap in the emission of the donor fluorophore and absorption of the acceptor, thus enabling the direct transfer of energy between the two fluorophores if in close proximity (reviewed in ref. 8). Depending on the sensor design, binding of the target molecule causes a conformational change that either brings the FRET pair closer together, resulting in energy transfer from one fluorophore to the other, or reduces the FRET efficiency as the distance between the fluorophores increases. Advantages of FRET sensors include low pH sensitivity, which results from relatively similar pH sensitivities of the two fluorophores used (leading to very low pH sensitivity of the relative fluorescent change (fluorophore A/fluorophore B)). Additionally, focus changes and z-drift affect both fluorophores simultaneously; therefore, the fluorescence ratio is maintained. Additionally, both fluorophores are translated from the same mRNA transcript, resulting in an equimolar ratio. Disadvantages include the relatively low dynamic range and error-prone folding of such large proteins. The latter can lead to oligomerization, which can affect the experimental outcome (reviewed in ref. 6, 9).

Single-Fluorophore (SF) Sensors

SF sensors are derived from a circular permutation and genetic engineering of just one fluorescent protein, fused with a sensing domain (for a detailed review about circular permutation, *see* ref. 10). In general, target binding leads to a conformational change of

the protein that in turn induces a change of the chromophore's degree of protonation or enables correct chromophore folding. Nowadays, these sensors are designed to have a broad range of emission spectra as well as different K_Ds.

Advantages of SF sensors often include the very high dynamic range. In addition, the entire protein structure is smaller and, thus, the expression rates are higher and the risk of misfolding is reduced. Given the narrower emission spectrum, SF-GEFIs are most suitable for multiparametric imaging, i.e., measuring different metabolites with different sensors at the same time [11]. Disadvantages include the notoriously high pH sensitivity, susceptibility to small ions (e.g., sensors often have a second Ca^{2+}-binding site that has been overlooked), and the issue of being able to reliably quantify fluorescence if the sensor is intensiometric. In addition, focus shift and z-drift can be problematic; however, this can be overcome by the addition of a second, insensitive fluorophore. This requires the expression of the second fluorophore using another transgene (pseudo-ratiometric) or, more elegantly, by using the unique properties of 2A sequences [12] allowing the equimolar expression of two fluorophores from a single transcript (ratiometric).

The Materials and Methods sections that follow detail a well-established protocol utilizing the FRET glucose sensor (*UAS-FLII12Pglu-700μδ6*; [13–15]) expressed in the astrocyte-like glia of a third instar larval brain. In this example experiment, the baseline, maximal glucose levels, and glucose uptake rate are measured. In the Notes section, we will address how the described experimental design can be adapted for the use of different sensors, tissue types, or buffers.

2 Materials

2.1 Equipment for Microscope Setup

1. Custom-made, three-dimensional (3D) printed flow-through chamber (*see* Fig. 1, Supplemental file 1; *see* **Note 1**).
2. Mini-peristaltic pump.
3. Silicone tubing to connect the pump, chamber, and waste (a pipette tip can be used to secure the tubing to the chamber) (*see* **Note 2**).
4. Laser scanning microscope with water emersion objectives (*see* **Note 3**).
5. Poly-L-lysine-coated coverslip (*see* **Note 4**).
6. Silicone to adhere the coverslip to the chamber.

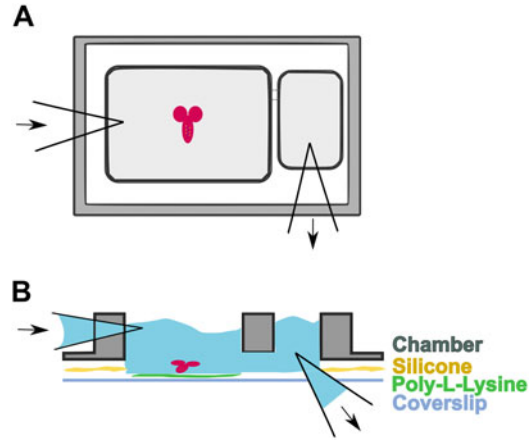


Fig. 1 Schematic of the custom-made flow-through chamber. (**a**) Top view of the chamber. The chamber has a larger and a smaller compartment, which are interconnected, indicated by the dotted lines. In the larger compartment, the tissue of interest, e.g., a larval brain (red), is placed. Here, a pipette tip filling in fresh buffer from the pump is inserted. A second pipette tip inserted into the smaller compartment aspirates the buffer. Arrows indicate the direction of the buffer flow. (**b**) Transverse section of the chamber. The chamber is adhered to a poly-L-lysine (green)-coated coverslip (blue) via silicone (yellow). The smaller compartment of the chamber is connected to the larger part via a small channel. Fresh buffer flows into the first compartment in which the sample is placed (red). The buffer is removed from the connected second compartment via suction from the pump, as indicated by the arrows

2.2 Fly Stocks and Dissection Equipment

1. Third instar larvae expressing *alrm-Gal4* ≫ *FLII12Pglu-700μδ6*; [16].
2. A pair of forceps.
3. Dissection pad or dish.
4. Binocular dissecting microscope.

2.3 Buffers (See Note 5)

1. Hemolymph-like buffer 3 (HL3): 70 mM NaCl, 5 mM KCl, 20 mM $MgCl_2$, 10 mM $NaHCO_3$, 115 mM sucrose, 5 mM trehalose, 5 mM HEPES; pH 7.2; ca. 350 mOsm; [17].
2. Glucose buffer: HL3 supplemented with glucose (10 mM; pH 7.2).

3 Methods (*See* Note 6)

3.1 Sample Preparation

1. Dissect the brains of third instar larvae expressing the FRET glucose sensor (*UAS-FLII12Pglu-700μδ6;* for sensor specifics, *see* **Note 7**) in astrocyte-like glia (*alrm-Gal4*) in the HL3 buffer (*see* **Note 8**).

2. Adhere freshly dissected brains onto a poly-L-lysine-coated coverslip, and submerge sample in HL3 (*see* **Note 9**).
3. Attach the coverslip to the chamber using silicone to avert any leaks (*see* Fig. 1).

3.2 Pump Setup

1. Attach the silicone tubing to the chamber.
2. Place one end of the first length of the tubing into the buffer and the other end into the first compartment of the chamber (*see* Fig. 1).
3. Place one end of the second length of the tubing into the outflow, second chamber compartment, and the other end into an empty Schott bottle for waste (*see* Fig. 1).
4. Connect both lengths of the tubing to a mini-peristaltic pump. Alternatively, the buffer flow can be created using a "variable speed pump medium flow" (neoLab Migge GmBH, Heidelberg), and waste buffer removed by a modified commercial aquarium pump (Tetra Aquarium Air Pump, APS 50).
5. Run the buffer through the system before mounting it onto the microscope to check for leaks, and adjust the flow rate as desired.

3.3 Buffer Exchange

1. Mount the chamber onto the microscope, and fill the first compartment of the chamber with HL3 buffer.
2. Lower the emersion objective into the buffer, and locate the sample (*see* **Note 10**).
3. Run the HL3 buffer through the system for 2 min (*see* **Note 11**; start imaging (*see* below) at this step).
4. Some metabolites are not transported over cellular membranes; in this case, permeabilization of the tissue is needed prior to washing such metabolites into a tissue. Please refer to **Note 12** for comments on permeabilization. Permeabilization is not needed in the experiment described here.
5. Exchange with glucose buffer by manually moving the tubing from one buffer to another and leave for 6 min (*see* **Note 13**).
6. Return the tubing to the HL3 buffer until the end of the experiment.

3.4 Image Acquisition (See Note 14)

1. Set image acquisition to 5 s at a resolution of 1024*1024 pixels.
2. Set excitation of 436/25 nm, beam splitter 455 nm, emission 480/40 nm Cyan fluorescent protein (CFP channel) and excitation 436/25 nm, beam splitter 455 nm, emission 535/30 nm yellow fluorescent protein (YFP channel).

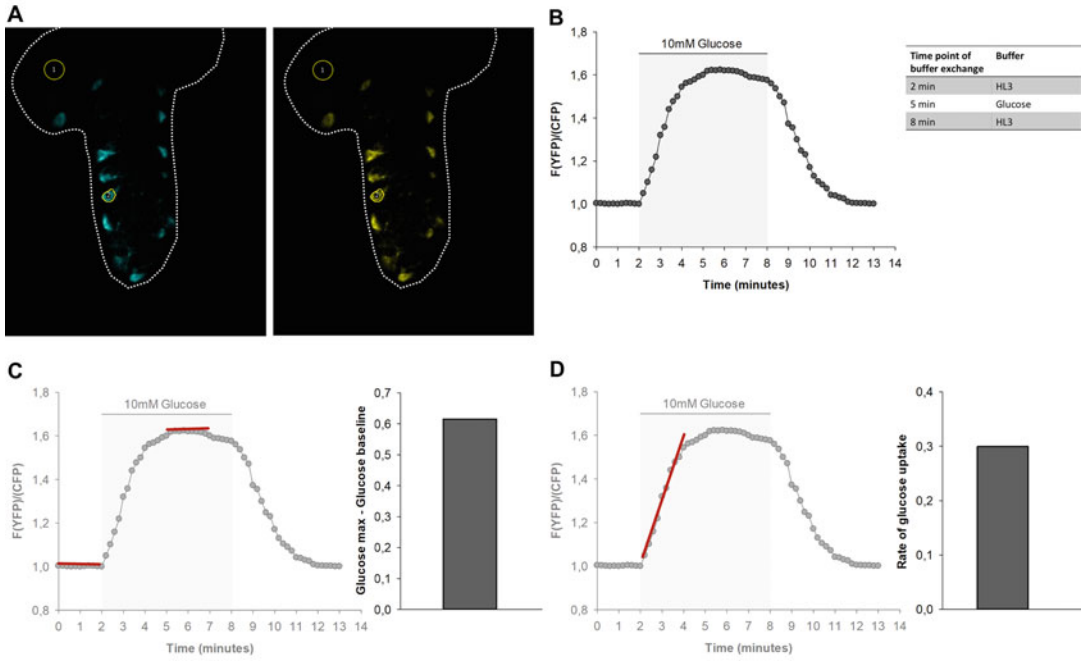


Fig. 2 Data analysis. (**a**) Confocal image of the ventral nerve cord of a third instar larval brain expressing the FRET glucose sensor (*UAS-FLII12Pglu-700μδ6*) in astrocyte-like glia. ROIs have been created in the background of the image (ROI 1) and around an individual astrocyte-like glial cell body (ROI 2) so that the background fluorescence can be subtracted. Analysis can be performed on single or multiple cells or on the entire image. In this experiment, the FRET ratio was determined for the single cell indicated. (**b**) Scatter graph and table indicating the buffer exchange during the experiment. HL3 was exchanged for 10 mM glucose buffer after 2 min and then exchanged back to the HL3 buffer 6 min later, after a plateau was reached. Images were captured every 5 s. (**c**) The baseline can be subtracted from the maximum to determine the overall change in the FRET signal and give an indication of the amount of glucose entering the cell type of interest during the experiment. (**d**) The rate of glucose uptake was determined by performing regression analysis of the first 10 time points at the beginning of the slope (red line). This indicates the rate of change in the FRET signal and therefore the rate of glucose uptake into the cell

3.5 Data Analysis (See Note 15)

1. In Fiji, select a region of interest (ROI) to measure the background.
2. Create an additional ROI (or ROIs) around the exact expression of the sensor (*see* Fig. 2a and **Note 16**).
3. Determine the mean gray values for the background and each channel separately, for each time frame (*see* **Note 17**).
4. Subtract the background from the gray values of each channel.
5. Calculate the ratio of the two channels to negate any focus change that may have occurred during the experiment (*see* **Note 18**).

6. For further analysis and graphical illustrations, transfer data to SigmaPlot (Jandel) (*see* **Note 19**).
7. Graph the change in fluorescent intensities of the biosensors fluorophore over time (*see* Fig. 2b).
8. Calculate the overall change in fluorescence by calculating the mean of 10 consecutive time points of the maximal plateau and subtracting the mean of 10 consecutive time points of the baseline (*see* Fig. 2c and **Note 20**).
9. Calculate the glucose uptake rate by performing a linear regression on 10 consecutive time points from when the slope begins (*see* Fig. 2d and **Note 21**).

4 Notes

1. A big advantage of closed chambers is that a smaller volume of buffer is required per experiment; it is also beneficial when working with harmful or toxic substances. However, the air bubble or pressure changes caused by buffer exchanges can lead to a significant focus change. This issue can be reduced if an automated buffer exchange system is used. Setups that use an epifluorescence microscope are less susceptible to such problems than setups using a confocal microscope. Open chambers require a higher volume of buffer and have the added risk of overflowing, which can potentially damage the microscope. A big advantage of an open chamber is that air bubbles or pressure changes do not cause significant focus drift. Both open and closed chambers can be used in inverted and upright microscopes. Which setup is best to use depends on the individual experimental design. Imaging biosensors in an intact larval filet/tissue are best in an open chamber with an upright microscope to allow direct access to the tissue. However, using an upright microscope with an open chamber requires a dipping objective to receive high-quality images.

   The two-chamber design of the imaging chamber used here reduces the movement of the sample caused by the peristaltic movement produced by the pump. However, when imaging cells, any standard flow-through chamber is suitable.
2. When exchanging buffers, the new buffer takes some time (depending on the flow rate and the length and diameter of the tubing) to reach the chamber. This leads to a delayed reaction in the tissue. A small air bubble is formed when buffers are being exchanged; this air bubble can be followed through the tubing. Thereby, the time the new buffer needs to reach the chamber can be determined.

3. Different microscopes can be used for the system described here. The best suited for these fluorescent sensors are epifluorescence microscopes, but confocal LSMs (laser scanning microscopes) also work well. Sensors with a weaker expression and high dynamics are more difficult to image using a confocal microscope and to benefit from a camera system used by an epifluorescence microscope.
4. Coverslips coated with poly-L-lysine are autofluorescent. However, we found that poly-L-lysine worked best for the described setup. The background fluorescence is subtracted during analysis. We received the best results using poly-L-lysine-coated coverslips, when 200 µl of poly-L-lysine was added to a 22 × 40 mm coverslip and incubated for 2 h at room temperature (RT). The residual solution was removed, and the coverslips were allowed to dry completely. These coverslips are best used the next day but can also be stored at room temperature and used within 2 weeks. Poly-L-lysine-coated coverslips can lose adherence during long-term imaging and with some permeabilization treatments.
5. pH varies in different cellular compartments; therefore, buffers should be adjusted to the predominant pH of your tissue. In the experiment described here, the pH was adjusted to 7.2, which is said to be the pH of *Drosophila* hemolymph [17]. All buffers should also have comparable temperatures. If effects are studied in, for example, an ex vivo larval brain, the temperature influences the living organ in several ways, including metabolic rate. Due to the small volume of the tubing, buffers can rapidly cool down or heat up in the time it takes to reach the chamber. Therefore, the temperature at which the buffers are stored should be adjusted in a way that they are at the desired temperature when the chamber is reached. This can be achieved by keeping the buffers on ice or with the use of a heat block. If needed, the temperature can be measured in the chamber by adding a suitable digital thermometer to the setup.

   Aliquots of each element of the HL3 (or other) buffers can be made ahead of time and stored at RT, 4 °C, or −20 °C; however, buffers should be brought to RT (or the desired experimental temperature) prior to use.
6. Explore the dynamic range and sensitivity of the sensor in the cell or tissue of interest to optimize your experiments. Biosensors can be used to analyze a variety of parameters, from metabolite levels to pH, but, to do so, other variables need to be controlled, meaning only one variable should change at a time. For example, if one wants to compare different concentrations of a metabolite, e.g., glucose, all buffer solutions used must have the same pH. The fluorescence of some biosensors is affected by pH; this should be monitored to prevent

quenching. In case the metabolite of interest or its transport may alter the local pH, a pH sensor should first be used to determine any potential pH changes. If indeed a pH change occurs, this should be considered when interpreting the results.

7. When using an SF biosensor, the sensor's fluorescence should ideally be normalized to a fluorophore unaffected by the treatment to ensure that fluorescence change is due to the treatment and not any other factors. This second fluorophore must remain constant throughout the experiment (besides bleaching and changes due to tissue movement) and should be expressed in the same cellular compartment. When using a green SF sensor, we express tdTomato or mCherry in the same cellular compartment as the circularly permuted green fluorescent protein (cpGFP)-coupled biosensor. FRET biosensors harbor two fluorophores. Thus, a focus drift, for example, affects both fluorophores equally. Taking the ratio of both fluorescence intensities abolishes changes caused by a focus drift. However, over time, the two fluorophores can bleach at different rates.

   Since biosensors are proteins, they exhibit characteristic equilibrium dissociation constants determined by their respective K_D values. Thus, the sensor might already be saturated at low metabolite concentrations and further changes are no longer detectable. The variance in fluorescence intensity correlates with the K_D value. In other words, a biosensor can show a much greater range in fluorescence the less saturated the target site is. Therefore, careful normalization of baseline fluorescence is necessary.

8. To minimize stress reactions, dissections are performed in the HL3 buffer and the tissue is transferred to the chamber immediately.

9. Leaving the surrounding tissue, for example, imaginal discs, attached to larval brain, help stick the tissue to the poly-L-lysine-coated coverslips. Once attached to the coated coverslips, the tissue cannot be easily removed and will not stick well elsewhere.

10. Specific microscope objectives are available to allow better image quality, for example, a dipping objective for use with an open chamber in an upright microscope. Furthermore, there are water immersion objectives for inverse microscopes available. Here, the refraction index of the immersion medium is similar to that of the buffer, which is beneficial for image quality.

11. The flow rate can be calculated prior to starting the experiment by measuring the time it takes for the buffer to run through the empty system. It is advised not to alter the flow rate once the experiment has started to reduce the variables.

12. Most available GEFIs have targets that can naturally enter a given tissue via ion channels or transporters (glucose, pyruvate, lactate). Thus, fluorescence can be manipulated by titrating different concentrations of the target to get an estimate of the sensor's fluorescence behavior. However, some molecules are not transported over cell membranes (reduced nicotinamide adenine dinucleotide phosphate (NADPH), H_2O_2; even though in the case of the latter, this is still a matter of debate). In order to assess the fluorescent range of a given sensor for such a molecule, cell membranes need to be permeabilized. Fig. 3 summarizes the common molecules used for permeabilization, with some characteristic features. In general, the size of the metabolite to titrate determines the choice and concentration of the permeabilizing agent and in turn the length of the treatment. The bigger the target, the later it should be washed into the sample, as permeabilization itself causes damage to the tissue. One specialty of metabolite live imaging lies in the complexity of metabolic interactions. Hence, the choice of molecules to normalize the baseline

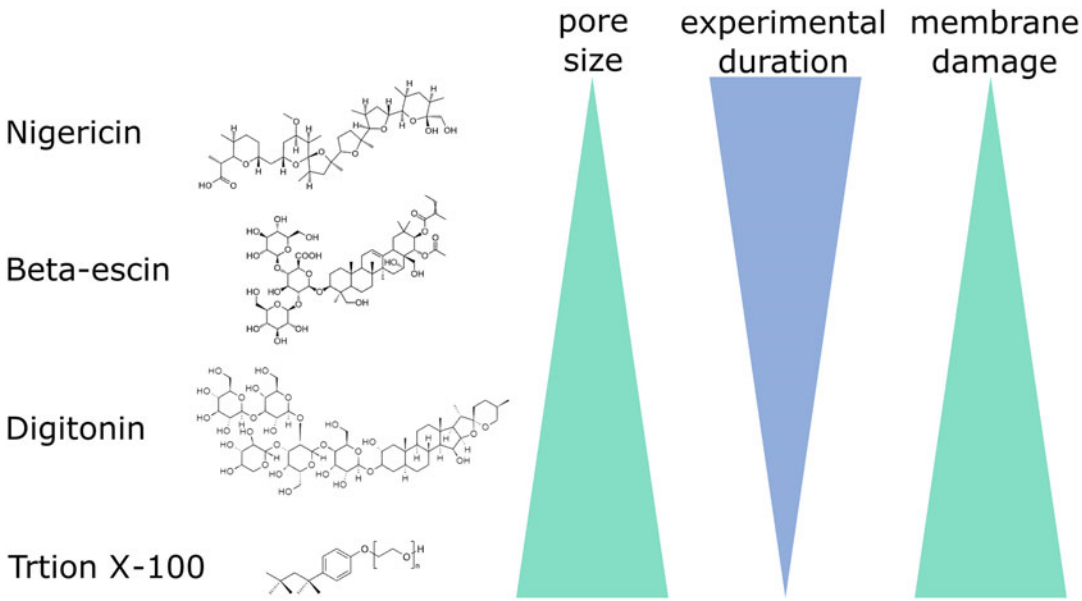


Fig. 3 Permeabilizing agents. The degree of damage to the cell membrane varies with the choice of the permeabilizing agent. The smaller the resulting pore size and the more reversible the interaction with the cell membrane is, the longer the samples can be imaged. Nigericin creates the smallest pores but can only be used in pH measurements. Beta-Escin, together with other saponins, is weakly permeabilizing. Digitonin can create large, irreversible pores, of which the size is concentration-dependent. It is useful for big metabolites and also causes the most damage. Metabolites that require digitonin to enter the cell should be washed last. Triton X-100 is a brute force attack to the cell membrane; it has the worst trade-off between permeabilization and damage

fluorescence must be carefully evaluated for each sensor to be used. For instance, the effects of peroxides can be counteracted via the addition of NADPH. However, since NADPH is the larger molecule, it needs to be washed last, since it requires stronger permeabilization to be titrated successfully.

Since the buffer enters the cell after permeabilization (*see* below), a different, intracellular buffer (HL6) should be used.

Hemolymph-like buffer 6 (HL6): 15 mM $MgCl_2$, 24.8 mM KCl, 23.7 NaCl, 20.mM isenthionic acid, 5 mM *N*, *N*-bis(2-hydroxyethyl)-2-aminoethanesulfonic acid (BES), 80 mM trehalose, 5.7 mM L-alanine, 2.0 mM L-arginine, 14.5 mM glycine, 11 mM L-histidine, 1.7 mM L-methionine, 13 mM L-proline, 2.3 mM L-serine, 2.5 mM L-threonine, 1.4 mM L-tyrosine, 1 mM L-valine (pH 7.2) [18].

13. Using an automated buffer exchange system can avoid the production of air bubbles and thereby disturbances of the tissue.

14. Sometimes a drift of the tissue is unavoidable. This can be corrected for with a second fluorophore or landmarks that do not change their relative position during imaging.

    The length of imaging depends on the tissue, sensor, protocol, and treatment being used. A protocol without the addition of any harsh chemicals can run for upward of an hour. However, the specific properties of the sensor used need to be considered for each individual experimental setup. The kinetics of the metabolite measured can be highly diverse and thereby affect the endurance of the sensor. Furthermore, some treatments damage the tissue or the cellular properties irreversible, which decreases maximal experimental duration. In the case of ex vivo samples, the tissue can become less reactive to changing conditions over time.

    For the biosensor and experimental setup used here, 12 frames per min were used to analyze the uptake rate. However, this is dependent on the biophysical properties of a given sensor and the kinetics of the metabolite of interest in the cell. Some metabolites show extremely fast concentration changes and thus require higher frame rates, e.g., Ca^{2+}.

    Laser intensities can be adjusted to the individual sample; however, a medium range is desired so that fluctuating fluorescent intensities can be detected in both directions. Acquisition rate and resolution can be modified to fit the needs of the experiment. When using a FRET sensor or when expressing two fluorophores and working with a confocal microscope, the line acquisition mode must be employed to detect both channels; a sequential acquisition mode can be used with single-fluorophore sensors.

At which time point a buffer exchange needs to occur depends on the individual's research interest. When studying an uptake rate, or relative metabolite concentrations, the fluorescence should always reach a stable plateau before a different condition is applied. Otherwise, the effect of the treatment cannot be clearly analyzed. However, this is not the case when studying metabolite consumption rates, for example.

The uptake rate and time taken to reach the plateau differs between tissues and genotypes. It is advised to perform a trial to understand how the tissue of interest will react and adjust the experiment accordingly.

Depending on the experiment and setup used, multiple tissues or cells can be imaged simultaneously; this increases the *n* of the experiment without increasing the time and additionally ensures that the experimental conditions are the same for these samples/cells.

15. Interpretation of the data highly depends on the system used, taking into consideration the characteristics of transporters, cellular metabolism, external influences, etc. The diet on which the flies/larvae are raised can have a significant impact on their metabolic requirements; therefore, this should be considered when designing the experiment and interpreting data. The data obtained can be used to compare different genotypes, for example, to determine the effect of the loss of a specific transporter or animals raised on different food conditions [19, 20].
16. In order to analyze the fluorescence emission from a sensor expressed differently in subcellular compartments, the region of interest (ROI) should always be solely based on the pixels comprising the actual signal. A mask can be used to achieve a precise cutout of the area expressing the sensor; this is particularly useful when the sensor is expressed in unconnected regions, as is the case e.g., for mitochondria. When creating a mask, a duplicate of the image should be rendered binary. The mask should cover the area of interest in which the signal will be quantified. If there are several unconnected areas with sensor expression, the resulting ROI has to be adapted accordingly; combining ROIs (Fiji>ROI manager>more>OR) or using an exclusive OR function (Fiji>ROI manager>more>XOR) will work in this situation.
17. Different channels have different levels of autofluorescence, so a background subtraction needs to be performed for both channels.
18. If the experimental design allows, linear baseline condition phases can be used to correct for bleaching. Therefore, a linear function is calculated in order to determine the bleaching rate.

The actual experimental values can then be corrected accordingly.

19. The initial image processing can be carried out using ImageJ or the accompanying software of the microscope used for the experiment. Further calculations can be performed using a statistical software of choice, for example, SigmaPlot (Jandel), R Studio (https://www.rstudio.com/), or Anaconda (https://www.anaconda.com/).
20. When the metabolite is taken away after a certain incubation time, the observed decay in fluorescence is not only the efflux of the metabolite but also its consumption.
21. When analyzing the uptake rate of a metabolite, it must be kept in mind that this does not reflect the net uptake but rather the metabolites uptake minus its cellular consumption.

Acknowledgments

We thank Luis Garcia for designing the flow-through chamber and for providing the specifications. Work in the laboratory has been funded by grants from the DFG to S.S. (SFB1009, SCHI 1380/2-1).

References

1. Miyawaki A, Llopis J, Heim R et al (1997) Fluorescent indicators for Ca2+based on green fluorescent proteins and calmodulin. Nature 388(6645):882–887
2. Adams SR, Harootunian AT, Buechler YJ et al (1991) Fluorescence ratio imaging of cyclic AMP in single cells. Nature 349(6311): 694–697
3. Belousov VV, Fradkov AF, Lukyanov KA et al (2006) Genetically encoded fluorescent indicator for intracellular hydrogen peroxide. Nat Methods 34(3):281–286
4. Sanford L, Palmer A (2017) Recent advances in development of genetically encoded fluorescent sensors. In: Methods in enzymology, pp 1–49
5. Chudakov DM, Matz MV, Lukyanov S et al (2010) Fluorescent proteins and their applications in imaging living cells and tissues. Physiol Rev 90(3):1103–1163. https://pubmed.ncbi.nlm.nih.gov/20664080/
6. Palmer AE, Qin Y, Park JG et al (2011) Design and application of genetically encoded biosensors. Trends Biotechnol 29:144
7. San Martín A, Sotelo-Hitschfeld T, Lerchundi R et al (2014) Single-cell imaging tools for brain energy metabolism: a review. Neurophotonics 1(1):11004
8. Zaccolo M (2004) Use of chimeric fluorescent proteins and fluorescence resonance energy transfer to monitor cellular responses. Circ Res 94(7):866–873, https://pubmed.ncbi.nlm.nih.gov/15087426/
9. Bajar BT, Wang ES, Zhang S et al (2016) A guide to fluorescent protein FRET pairs. Sensors (Basel) 16(9):1488
10. Kostyuk AI, Demidovich AD, Kotova DA et al (2019) Circularly permuted fluorescent protein-based indicators: history, principles, and classification. Int J Mol Sci 20:4200
11. Mehta S, Zhang Y, Roth RH et al (2018) Single-fluorophore biosensors for sensitive and multiplexed detection of signalling activities. Nat Cell Biol 2010(20):1215–1225
12. Daniels RW, Rossano AJ, Macleod GT et al (2014) Expression of multiple transgenes from a single construct using viral 2A peptides in drosophila. PLoS One 9:e100637

13. Fehr M, Lalonde S, Lager I et al (2003) In vivo imaging of the dynamics of glucose uptake in the cytosol of COS-7 cells by fluorescent nanosensors. J Biol Chem 278:19127–19133
14. Takanaga H, Chaudhuri B, Frommer WB (2008) GLUT1 and GLUT9 as major contributors to glucose influx in HepG2 cells identified by a high sensitivity intramolecular FRET glucose sensor. Biochim Biophys Acta Biomembr 1778:1091–1099
15. Volkenhoff A, Hirrlinger J, Kappel JM et al (2018) Live imaging using a FRET glucose sensor reveals glucose delivery to all cell types in the Drosophila brain. J Insect Physiol 106: 55–64
16. Doherty J, Logan MA, Taşdemir ÖE et al (2009) Ensheathing glia function as phagocytes in the adult *Drosophila* brain. J Neurosci 29:4768–4781
17. Stewart BA, Atwood HL, Renger JJ et al (1994) Improved stability of Drosophila larval neuromuscular preparations in haemolymph-like physiological solutions. J Comp Physiol A 175:179–191
18. Macleod GT, Hegström-Wojtowicz M, Charlton MP et al (2002) Fast calcium signals in drosophila motor neuron terminals. J Neurophysiol 88:2659–2663
19. McMullen E, Weiler A, Becker HM et al (2021) Plasticity of carbohydrate transport at the blood-brain barrier. Front Behav Neurosci 14:271
20. Hertenstein H, McMullen E, Weiler A et al (2021) Starvation-induced regulation of carbohydrate transport at the blood–brain barrier is TGF-β-signaling dependent. elife 10:e62503

INDEX

Christian Dahmann (ed.), *Drosophila: Methods and Protocols*,
Methods in Molecular Biology, vol. 2540, https://doi.org/10.1007/978-1-0716-2541-5,

D

E

F

G

H

I

J

L

M

N

O

P

Q

R

S

T

U

V

W

Y

Z