PNAS
www.pnas.org

Supplementary Information for

***Drosophila* as a model for studying cystic fibrosis pathophysiology of the gastrointestinal system**

Kevin Kim[1,*], Elizabeth A. Lane[1], Aurelia Saftien[1], Haiyun Wang[3], Yue Xu[3], Frederik Wirtz-Peitz[1] and Norbert Perrimon[1,2,*]

[1]Department of Genetics, Harvard Medical School, Boston, MA, 02115, USA

[2]Howard Hughes Medical Institute, Harvard Medical School, Boston, MA, 02115, USA

[3]School of Life Sciences and Technology, Tongji University, Shanghai 200092, China

[*]Corresponding authors
Email: perrimon@receptor.med.harvard.edu (N.P.)

This PDF file includes:

Supplemental Materials and Methods

Drosophila Stocks and Genetics

The following *Drosophila* stocks were used in this study: *CG5789*^{NIG} RNAi (National Institute of Genetics Stock: 5789R-1 & 5789R-4); *CG7627*^{NIG} RNAi (National Institute of Genetics Stock: 7627R-1); *CG10505*^{NIG} RNAi (National Institute of Genetics Stock: 10505R-3); *CG31792*^{NIG} RNAi (National Institute of Genetics Stock: 10441R-3); UAS-miR-263a (1);*Muc68D*^{VDRC} RNAi (Vienna *Drosophila* RNAi Center Stock: 105461); *Muc68D*^{TRiP} RNAi (Bloomington *Drosophila* Stock Center: 56969) and *CG5789*^{VDRC} RNAi (Vienna *Drosophila* RNAi Center Stock: 1204). *puc-LacZ* (*puc*^{E69}); *10xSTAT-GFP*; *Dl-LacZ* (*Dl*^{05151}); *hsFlp; act>CD2>Gal4, UAS-GFP/CyO*; *Myo1A-Gal4*; *Escargot-Gal4*; and *Tachykinin-gut-Gal4* were from laboratory stocks. For flip-out clones, 2-3-day old adult flies were heat shocked for 1 hr at 37°C to induce clones and kept at 25°C for 7 days until dissection.

Immunostaining of the Midgut

Immunostaining of the midgut was performed as previously described (2). In brief, prior to dissection, flies were fed on 5% sucrose for 3 hr to remove food from the midgut. Guts from female flies were dissected in 1xPBS and fixed in 4% paraformaldehyde diluted with 1xPBS for 30 minutes. Samples were washed with 1xPBS, blocked for 30 minutes in 1xPBS, 5% Donkey Serum and 0.1% Triton X-100. Samples were incubated overnight at 4°C using the following antibodies: rabbit anti-β-galactosidase (1:1,000; Cappel), mouse anti-β-galactosidase (1:1,000; Promega), rabbit anti-phospho-Histone H3 (1:10,000; Millipore), and rabbit anti-P-p44/42 MAPK (dpERK) (1:200; Cell Signaling). Primary antibodies were detected using anti-mouse or anti-rabbit secondary antibodies conjugated to Alexa-Fluor 488 and 594 (1:1000; Invitrogen). Fluorescent images were acquired with a Leica TCS SP2 AOBS.

Generation of *hCFTR* Transgenic Flies

All cloning was done using Gateway technology. The *hCFTR* fragment was amplified from *pCMV-CFTR-pBQ6.2* (gift from P.J. Thomas) and cloned into *pENTR-D/TOPO* (Thermo Fisher). To generate mutant *hCFTR*s, *pENTR-hCFTR* was mutagenized using the QuikChange II XL Site-Directed Mutagenesis Kit (Stratagene). The entry vectors were then moved using the Gateway LR Clonase II Enzyme Mix (Thermo Fisher) into *pUASC-3xHA-attR* and *pUASp-EGFP-attR* destination vectors compatible with fly transgenesis. Transgenic flies were established by injecting *hCFTR* plasmids into flies carrying an attP docking site, attP2, located on the third chromosome.

Sequence Conservation Analysis

Candidate *Drosophila* orthologs of *hCFTR* were identified using the *Drosophila* RNAi Screening Center Integrative Ortholog Prediction Tool (DIOPT; http://www.flyrnai.org/diopt) (3). Sequence alignments were carried out using the Clustal Omega multiple sequence alignment tool (https://www.ebi.ac.uk/Tools/msa/clustalo/) (4). Pairwise sequence alignments were performed using the EMBOSS Stretcher alignment tool (https://www.ebi.ac.uk/Tools/psa/emboss_stretcher/) (5). To resolve the orthology of candidate *Dmel\CFTR* proteins, we used maximum-likelihood based phylogenetic reconstruction as implemented by RAxML BlackBox (https://embnet.vital-it.ch/raxml-bb/) (6). Phylogenic trees were visualized using FigTree v1.4.3 (http://tree.bio.ed.ac.uk/software/figtree/).

MQAE and Sodium Green Assay

MQAE and Sodium Green assays were performed as previously described (2). In brief, adult female flies, 14-16 days old, were fed 2 μM of cell-permeant Sodium Green tetraacetate Indicator (Thermo Fisher) or 2.5 mM of MQAE ((*N*-(Ethoxycarbonylmethyl)-6-Methoxyquinolinium Bromide) (Thermo Fisher) diluted in 5% sucrose overnight and dissected and fixed in 4% paraformaldehyde diluted with 1xPBS for 30 minutes. After a brief wash with 1xPBS, samples were mounted, and images were taken immediately using a Leica TCS SP2 AOBS (Fig. 1 and S2) or a W1 Yokogawa spinning disk, Nikon inverted Ti2 confocal microscope (Fig 3). For quantification of MQAE and Sodium Green fluorescence, the raw images were analyzed using

ImageJ (NIH).

Quantification of EC Volume

Mutant clones were generated using the flip-out system. To analyze cell volume, serial sections (~9-18 sections per individual EC) of individual ECs in flip-out clones and neighboring ECs were imaged using a confocal microscope and area of the cell in each section was measured in fiji. Total cell volume was calculated as follows:

$$\textit{cell volume } (\mu m^3) = \textit{total average cell surface area } (\mu m^2) \times \textit{thickness } (\mu m) \textit{ per section} \times \textit{number of sections}$$

Electron Microscopy

Electron microscopy was performed as previously described (2). In brief, dissected midguts were fixed overnight in 0.1 M sodium cacodylate buffer (pH 7.4) containing 2.5% glutaraldehyde, 2% paraformaldehyde. The fixed samples were then washed 3 times in distilled water, fixed again with 1% osmium tetroxide (OsO4) and 1.5% potassium ferrocyanide (KFeCN6) for 1 hr, and washed 3 times in distilled water. Next, the samples were washed in 1% maleate buffer and incubated in 1% aqueous uranyl acetate in 1% maleate buffer for 1 hr, followed by 2 washes in 1% maleate buffer and subsequent dehydration in grades of alcohol. The samples were put in propylene oxide for 1 hr and embedded with a solution of 50% propylene oxide and 50% TAAB Epon overnight. The next day, samples were embedded in fresh TAAB Epon and polymerized for 2 days at 60°C. Ultrathin sections (about 60 nm) were cut on a Reichert Ultracut-S microtome, picked up onto copper grids, and then stained with lead citrate. The sections were examined in a JEOL 1200EX Transmission electron microscope, and images were recorded with an AMT 2k CCD camera.

Bacterial Oral Infection Assays

Bacterial oral infection assays were performed as previously described (2). In brief, *P. aeruginosa* PA14 was grown on LB medium overnight. The following morning, 200 µl of overnight culture was added to 10 ml of LB and cultured for another 6-8 hr to reach OD_{600}=1.5. Bacteria/sucrose feeding solution was prepared by mixing 1 ml of the bacteria solution to 4 ml of sucrose solution to reach a final sucrose concentration of 4%. Bacterial infection assays were performed by placing ten 1-3-day old female flies into individual vials containing paper towels saturated with bacteria/sucrose feeding solution at 25°C. Three-four vials of each genotype with 10-15 flies were used per n. The number of dead flies was recorded every 24 hr. Flies were transferred to new vials with freshly prepared sucrose solution every other day and bacteria/sucrose solution every 3 days. *Ecc15* was grown on LB medium overnight at 30°C. Cells at OD_{600}=1.5 were then concentrated and mixed with sucrose solution to reach a final sucrose concentration of 5%. Bacterial infection assays were performed by placing 7-10-day old female flies into individual vials containing Kimwipes saturated with bacteria/sucrose feeding solution at 29°C. After 24 hr, 10 intestines from each genotype were dissected for qPCR analysis. Internal bacterial load was examined as previously described (2). For antibiotics feeding assay, standard laboratory fly food containing 30 µg/ml Kanamycin, 100 µg/ml Ampicillin, and 34 µg/ml Chloramphenicol was fed to newly eclosed female flies. Flies were transferred to fresh antibiotics food every 1–2 days. After 16 days, midguts were dissected and examined.

qPCR

Total RNA was prepared from 20-30 dissected adult intestines (after removal of the Malpighian tubules and crop) and RNA was extracted using TRIzol Reagent (Thermo Fisher). cDNA was prepared using iScript cDNA Synthesis Kit (Bio-Rad) and qPCR was performed using iQ SYBR Green Supermix (Bio-Rad). *RPL32* was used to normalize RNA levels. Relative quantification of mRNA levels was calculated using the comparative CT method. The following primer sequences were used:

<u>qPCR primers</u>

RPL32 F: 5'-AGCATACAGGCCCAAGATCG-3'

RPL32 R: 5'-TGTTGTCGATACCCTTGGGC-3'

Dpt F: 5'-CGTCGCCTTACTTTGCTGC-3'
Dpt R: 5'-CCCTGAAGATTGAGTGGGTACTG-3'
Crys F: 5'-ATGAAACGGACATACTTGTTGCT-3'
Crys R: 5'-CTGCTGAAGGTTGGAGGACTT-3'
Muc68D F: 5'-CTAGTCGTAAGAGTCCTATGGGC-3'
Muc68D R: 5'-GATTGGGGATCTTCGGAGTCG-3'

qPCR of *miR-263a*

Total RNA was extracted from 12 dissected adult intestines (after removal of the Malpighian tubules and crop) and RNA was extracted using Direct-zol RNA MicroPrep (Zymo Research) for each n. Micro-RNA levels were measured using the TaqMan™ MicroRNA Reverse Transcription Kit (Thermo-Fischer) and *miR-263a* was measured using assay ID number 000293 and levels were normalized to U14 (assay ID 001750) and U27 (assay ID 001752) micro-RNAs.

Processing of RNA-Seq Data

The quality of the reads was evaluated with FastQC (7). FastQ Screen(8) was used to screen if there were reads contamination due to Wolbachia and virus infection et al. We applied Trim Galore! (9) and Cutadapt (10) to remove the adaptors and overrepresented sequences, and used Trimmomatic to trim the bases with poor quality (11). The clean RNA-Seq reads were mapped onto the reference genome *Drosophila melanogaster* BDGP6 using STAR (v2.4.0j) (12). The abundance of each gene was quantified as TPM (Transcripts per million) value, which was evaluated by a statistical method RSEM (RNA-Seq by Expectation Maximization). RSEM uses a generative model of RNA-Seq reads and the EM algorithm, taking read mapping uncertainty into account and achieving the most accurate abundance estimates (13). Pairwise Pearson correlation coefficient (R) between four samples were calculated using log2(TPM) of genes. We employed DESeq2 to call the differentially expressed genes (DEGs) between two conditions. Sequencing results between S1 and S2 were paired while S3 and S4 were paired for this analysis to exclude the bias from sequencing depth that may have resulted from batch differences (14). DEGs with P-value <0.05 and fold-change ≥1.2 were further mapped onto the Gene Ontology (GO) and KEGG, and the P-value indicating whether a function was enriched by DEGs was calculated using hypergeometric distribution. Furthermore, we applied Benjamini-Hochberg procedure to control the false discovery rate from the multiple testing. DEGs with adjusted P-value < 0.1 and fold-change≥1.2 were also used to have the same analysis.

DRSC Integrative Ortholog Prediction Tool (DIOPT, https://www.flyrnai.org/diopt) and NCBI HomoloGene database (https://www.ncbi.nlm.nih.gov/homologene/) were applied to annotate the orthologous genes in Human. A tool Gene List Annotation for Drosophila (GLAD, https://www.flyrnai.org/tools/glad/web/) was used to annotate Drosophila genes in terms of the information including transcription factor, transmembrane protein and CPCR. The information of function summary for genes was downloaded from FlyBase database (http://flybase.org).

Public Datasets in GEO Database and DEGs Analysis

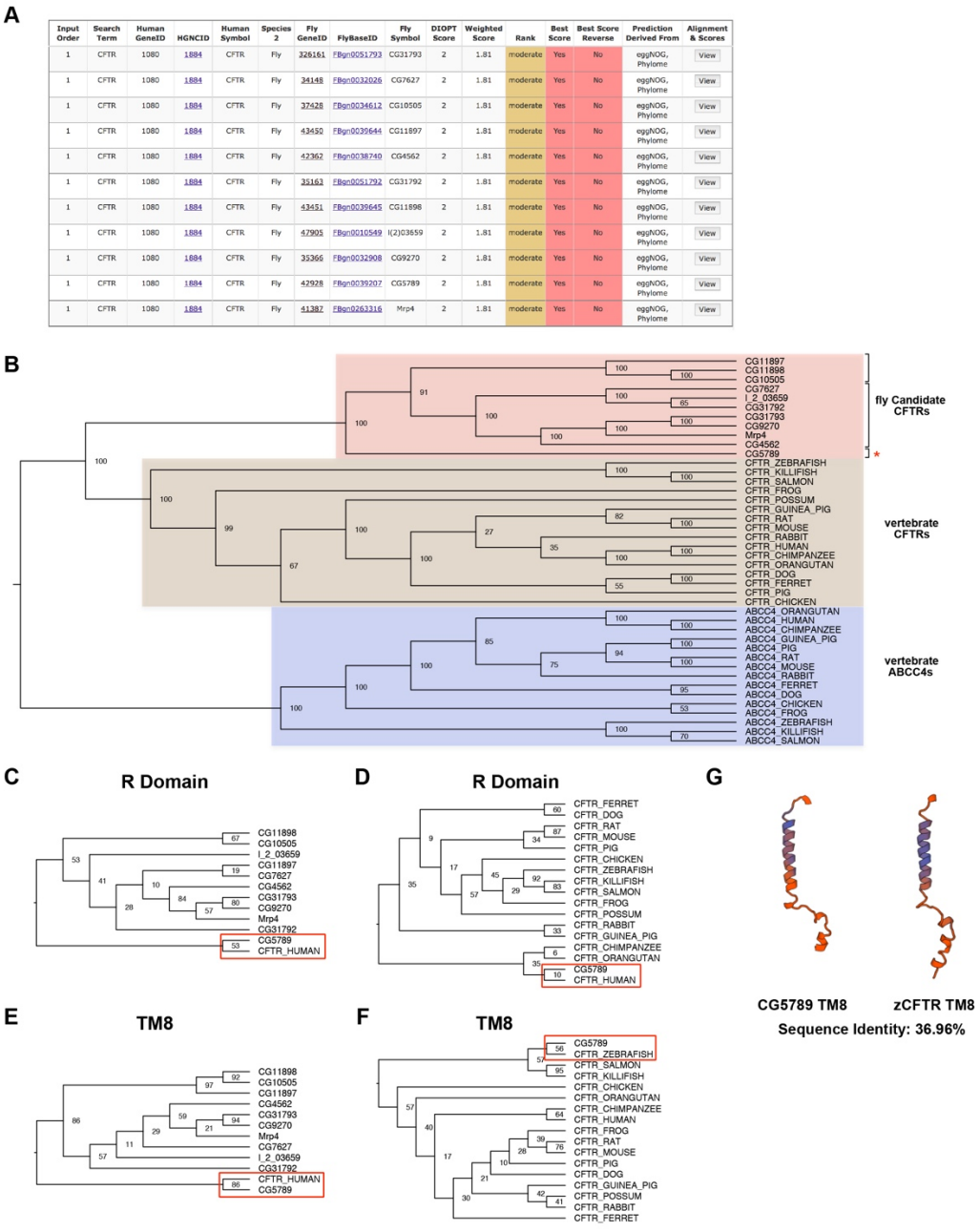
To compare the similarity of the transcriptional changes between the fruit fly model and other mammalian models, three gene expression datasets in Gene Expression Omnibus (GEO, http://www.ncbi.nlm.nih.gov/geo/) database (15), with RNAs from a CFTR knockout mammalian intestine, were used. They included two microarray datasets GSE765 and GSE5715 in mouse, and a RNA-Seq dataset GSE81114 in rat.

For the microarray datasets GSE765 and GSE5715, the raw data with CEL format were downloaded from GEO database. Next, Robust Multichip Average (RMA) method was used for background adjustment and expression index calculation (16). Then, DEGs at a significant level of P-value <0.05 and a fold-change ≥1.2 were detected by Significant Analysis of Microarrays (SAM) tool (17). Probe IDs were mapped to gene symbols using the "biomaRt" package. For the RNA-Seq data GSE81114, the raw counts were downloaded from GEO database and DEGs with P-value <0.05 and fold-change ≥1.2 were detected by DESeq2 (18).

Identification of Overlapping Differentially Expressed Genes in Fruit Fly Model and Mammalian Models

We calculated the number of DEGs in fruit fly (N_{DEGs}) whose orthologs in rat or mouse were also differentially expressed in the same direction. The orthologous genes of *Drosophila* genes in mouse and rat genome were annotated using DIOPT. To evaluation if N_{DEGs} are statistical significant, we applied a test for N_{DEGs} with random sampling techniques (19). For this hypothesis, we tested whether N_{DEGs} is a random sample from the background distribution while randomly selecting the same number of genes as DEGs. The test procedure is as follows:

1. Randomly select genes from fruit fly, with the number of them same as that of DEGs in fruit fly. Randomly label them as upregulated or downregulated genes, with the same number as that of DEGs.
2. Calculate the number of overlapping genes of randomly selected genes and DEGs in rat or mouse.
3. Repeat steps 1–2 for 200 times.
4. Create histogram of H_0 (null distribution).
5. Calculate $p(N_{DEGs} \mid H_0)$, if $p < 0.05$ then reject H_0.



Input Order	Search Term	Human GeneID	HGNCID	Human Symbol	Species 2	Fly GeneID	FlyBaseID	Fly Symbol	DIOPT Score	Weighted Score	Rank	Best Score	Best Score Reverse	Prediction Derived From	Alignment & Scores
1	CFTR	1080	1884	CFTR	Fly	326161	FBgn0051793	CG31793	2	1.81	moderate	Yes	No	eggNOG, Phylome	View
1	CFTR	1080	1884	CFTR	Fly	34148	FBgn0032026	CG7627	2	1.81	moderate	Yes	No	eggNOG, Phylome	View
1	CFTR	1080	1884	CFTR	Fly	37428	FBgn0034612	CG10505	2	1.81	moderate	Yes	No	eggNOG, Phylome	View
1	CFTR	1080	1884	CFTR	Fly	43450	FBgn0039644	CG11897	2	1.81	moderate	Yes	No	eggNOG, Phylome	View
1	CFTR	1080	1884	CFTR	Fly	42362	FBgn0038740	CG4562	2	1.81	moderate	Yes	No	eggNOG, Phylome	View
1	CFTR	1080	1884	CFTR	Fly	35163	FBgn0051792	CG31792	2	1.81	moderate	Yes	No	eggNOG, Phylome	View
1	CFTR	1080	1884	CFTR	Fly	43451	FBgn0039645	CG11898	2	1.81	moderate	Yes	No	eggNOG, Phylome	View
1	CFTR	1080	1884	CFTR	Fly	47905	FBgn0010549	l(2)03659	2	1.81	moderate	Yes	No	eggNOG, Phylome	View
1	CFTR	1080	1884	CFTR	Fly	35366	FBgn0032908	CG9270	2	1.81	moderate	Yes	No	eggNOG, Phylome	View
1	CFTR	1080	1884	CFTR	Fly	42928	FBgn0039207	CG5789	2	1.81	moderate	Yes	No	eggNOG, Phylome	View
1	CFTR	1080	1884	CFTR	Fly	41387	FBgn0263316	Mrp4	2	1.81	moderate	Yes	No	eggNOG, Phylome	View

Fig. S1. Predicted Orthologs of *hCFTR* in the *Drosophila* Genome.
(A) Snapshot of DIOPT prediction for the predicted orthologs of *hCFTR* in the *Drosophila* genome. **(B)** Phylogenetic relationship between 11 Dmel\CFTR candidates and 16 functionally characterized vertebrate CFTR and 15 functionally ABCC4 orthologs. Red asterisk marks *CG5789* that branches basally from all other *Dmel\CFTR* candidates. **(C)** Phylogenetic relationship between 11 Dmel\CFTR candidate R Domain and hCFTR R Domain. **(D)** Phylogenetic relationship between 16 functionally characterized CFTR R Domain and putative R Domain of CG5789. **(E)** Phylogenetic relationship between 11 Dmel\CFTR candidate TM8 and hCFTR TM8. **(F)** Phylogenetic relationship between 16 functionally characterized CFTR TM8 and putative TM8 of CG5789. **(B-F)** The numbers at each node indicate the bootstrap support values. **(G)** Predicted structure of CG5789 TM8 and previously solved structure of zCFTR TM8 (PDB ID: 5w81.1).

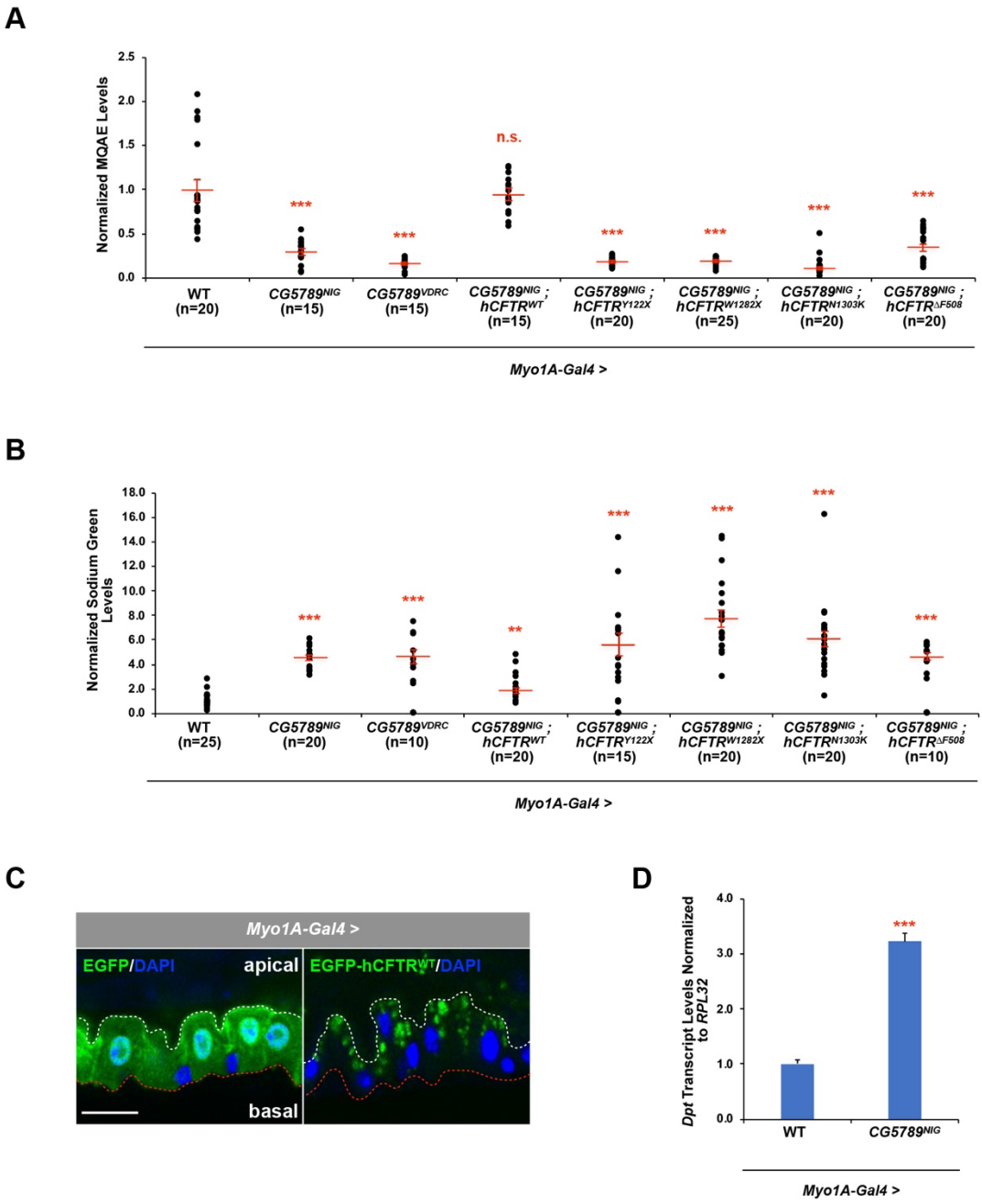


Fig. S2. Distribution of Cl^- and Na^+ in the absence of CG5789
(A) Quantification of MQAE flourescence. **(B)** Quantification of Sodium Green fluorescence. **(A and B)** "n" denotes the number of posterior midguts examined for each genotype. **(C)** Cross-section view of the posterior midgut expressing EGFP alone or EGFP tagged $hCFTR^{WT}$. White dotted lines mark the outline of the apical membrane and red dotted lines mark the outline of the basal membrane of the intestinal epithelium. The scale bar represents 25 μm. **(D)** qPCR analysis of *Dpt* using total RNA from dissected midguts of indicated genotypes. Error bars indicate SEM. **P <0.05 and ***P <0.001 (two-tailed t-test).

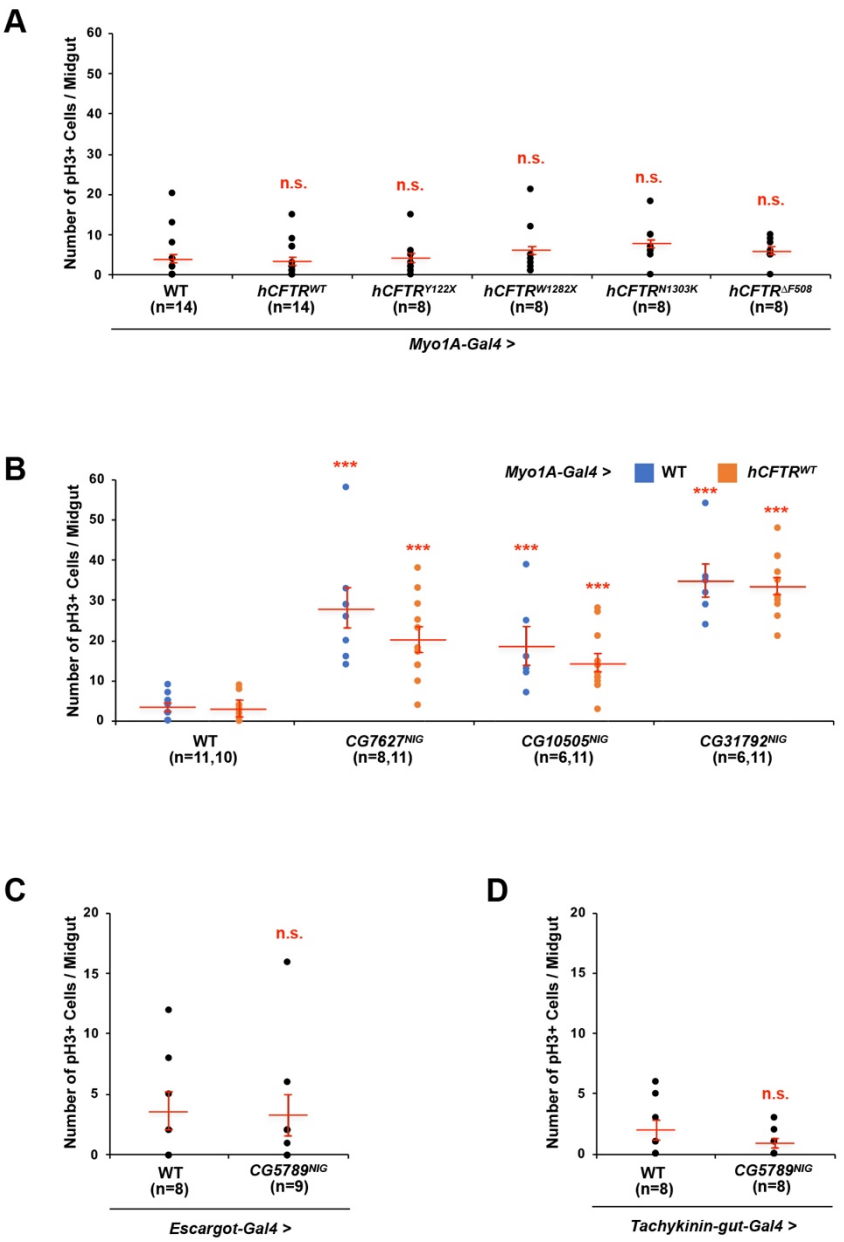


Fig. S3. Intestinal stem cell phenotypes of hCFTRs and candidate Dmel\CFTRs
(A) The average number of pH3+ cells in the posterior midguts expressing wild-type or mutant *hCFTR*s in wild-type background. **(B)** The average number of pH3+ cells in the posterior midguts expressing RNAi against *CG7627*, *CG10505*, and *CG31792*, and co-expressing wild-type *hCFTR*. **(C)** The average number of pH3+ cells in the posterior midguts expressing RNAi against *CG5789* in the precursor cells using precursor specific *Escargot-Gal4* driver. **(D)** The average number of pH3+ cells in the posterior midguts expressing RNAi against *CG5789* in the EEs using EE specific *Tachykinin-gut-Gal4* driver. **(A-D)** "n" denotes the number of posterior midguts examined for each genotype. Error bars indicate SEM. ***P <0.001 (two-tailed t-test).

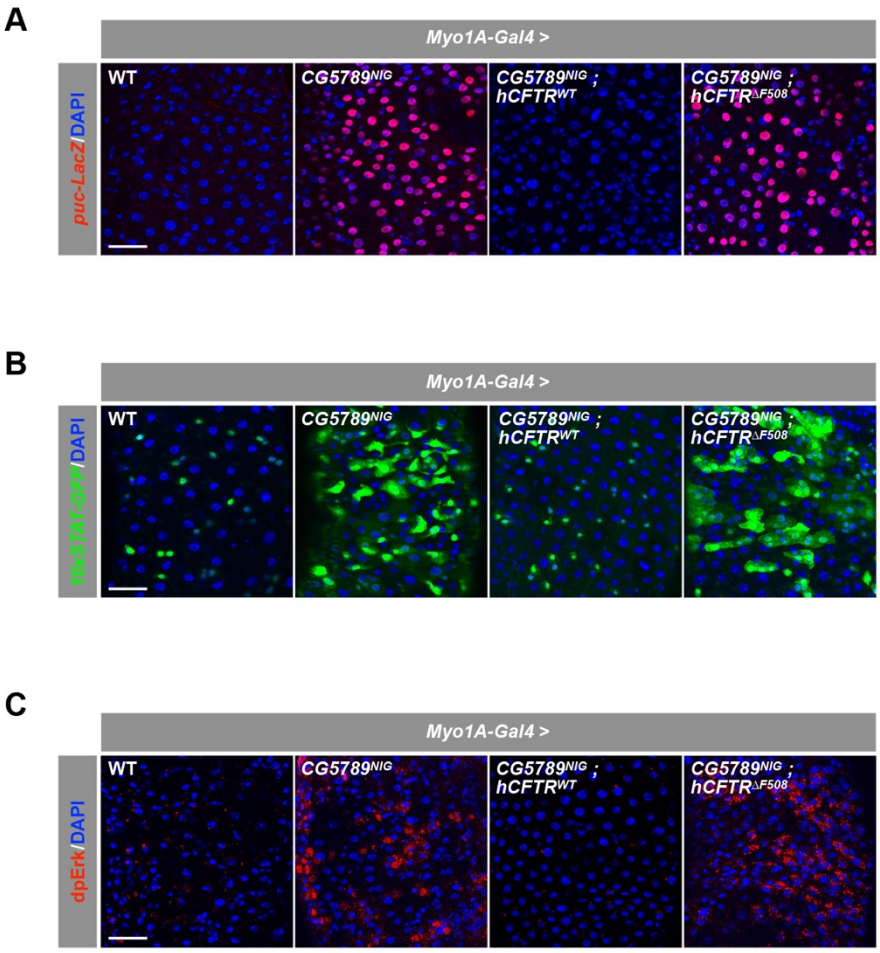


Fig. S4. *CG5789* RNAi activates stress and developmental signaling pathways
(A) JNK pathway activity visualized by *puc-LacZ*. **(B)** JAK/STAT pathway activity visualized by *10xSTAT-GFP* reporter. **(C)** EGFR pathway activity visualized by dpERK expression. **(A-C)** The scale bar represents 50 μm.

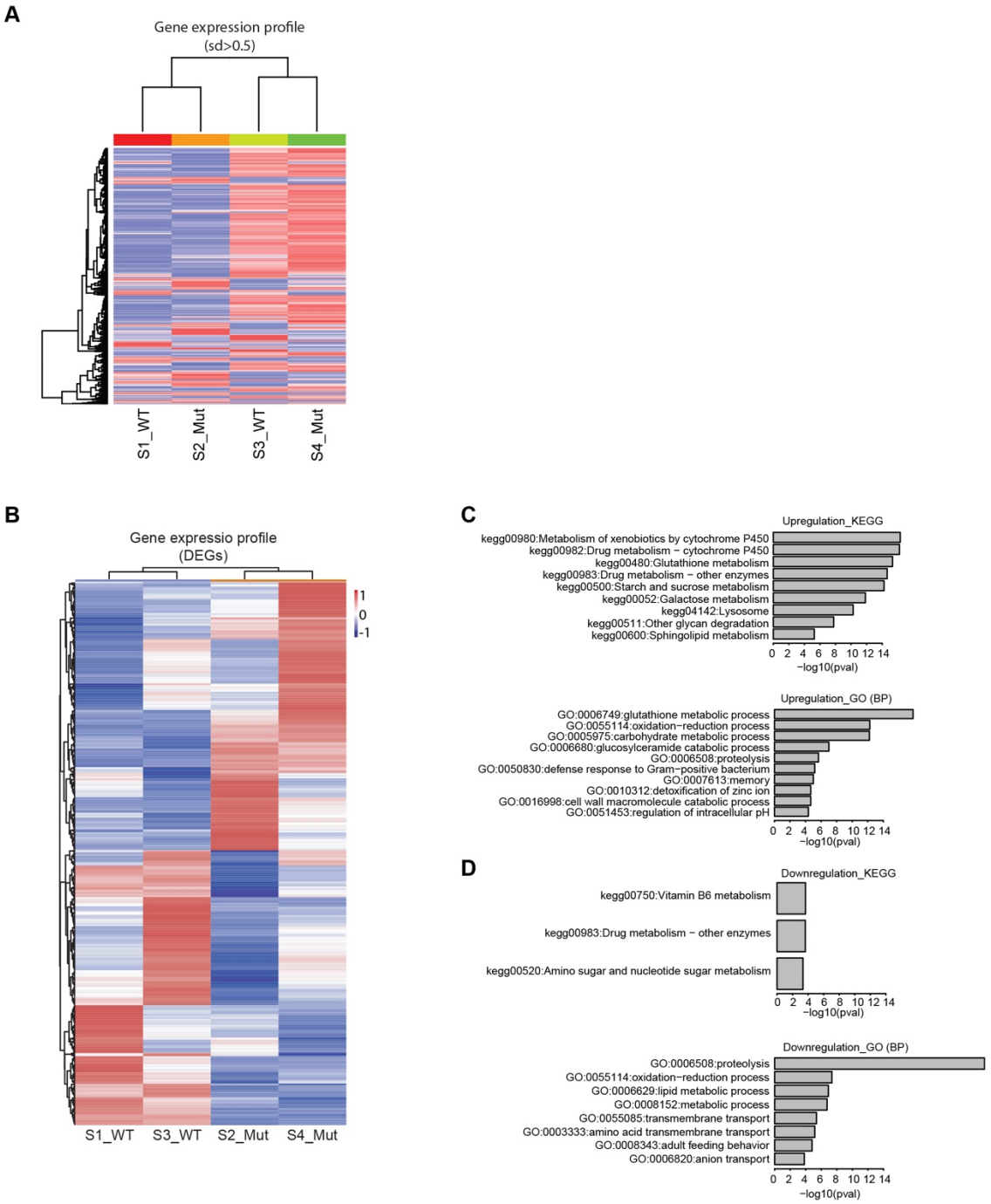


Fig. S5. Gene enrichment analysis of *Dmel\CFTR* intestine
(A) Heat map of RNA-Seq transcriptome analysis from samples S1-S4. The results showed that the expression profile of S1 was more similar with that of S2, and the expression profile of S3 was more similar with that of S4. **(B)** Heat map generated from 451 upregulated and 463 downregulated genes from samples S1-S4. **(C)** Gene Ontology (GO) and KEGG enrichment analyses for 451 upregulated genes identified from 914 DEGs. **(D)** Gene Ontology (GO) and KEGG enrichment analyses for 463 downregulated genes from 914 DEGs.

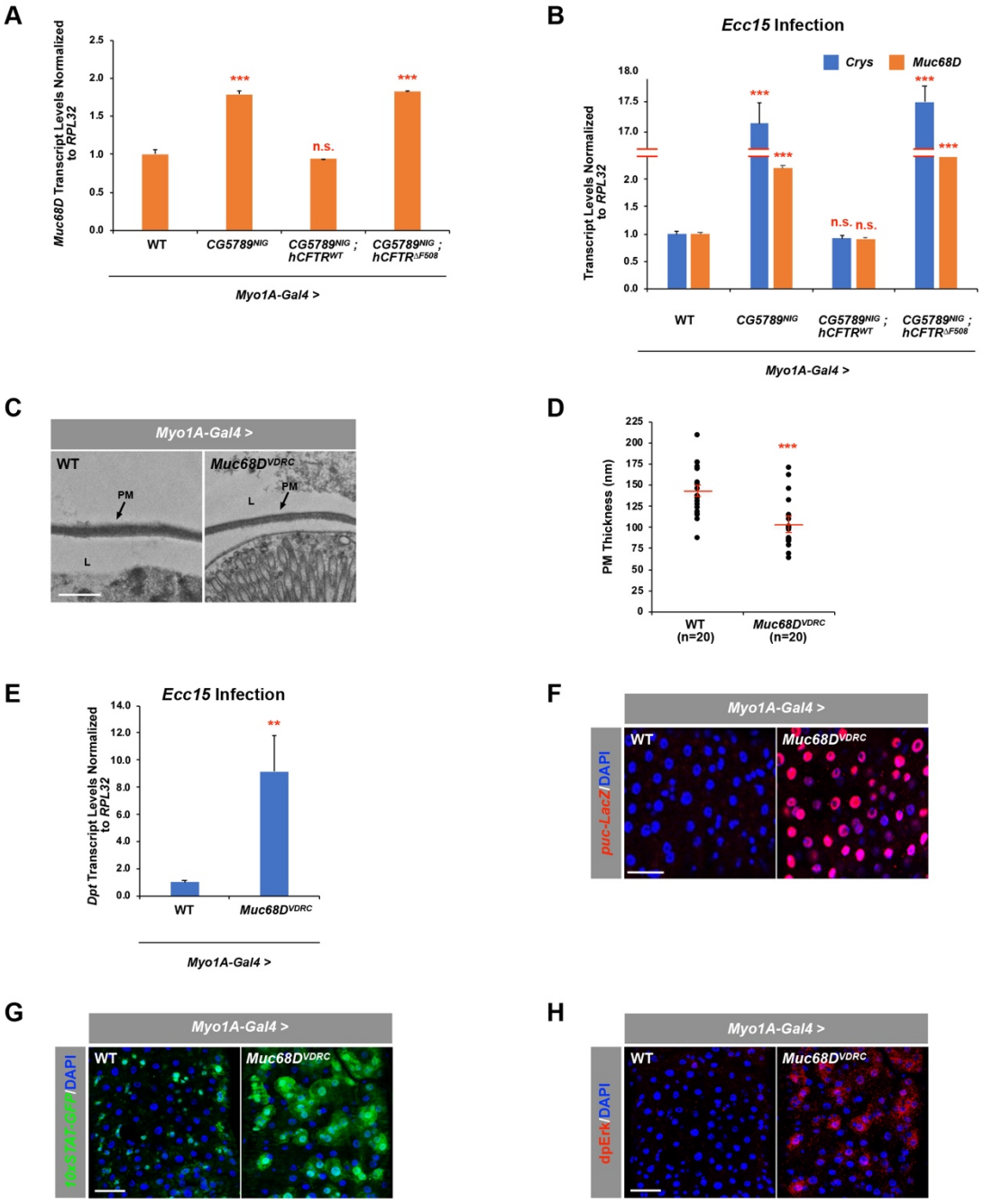


Fig. S6. Intestinal phenotypes of *Muc68D* RNAi
(A) qPCR analysis of *Muc68D* using total RNA from dissected midguts of indicated genotypes. **(B)** qPCR analyses of *Crys* and *Muc68D* using total RNA from dissected midguts of indicated genotypes 24 hr after *Ecc15* oral infection. **(C)** EM cross-sections of posterior midguts. Arrows indicate the PM (peritrophic matrix) and L (lumen). The scale bar represents 800 nm. **(D)** Quantitative measurements of the PM thickness. "n" denotes the number of PM thickness measurements for each genotype. **(E)** qPCR analysis of *Dpt* using total RNA from dissected midguts of indicated genotypes. **(A-B and D-E)** Error bars indicate SEM. **P <0.05 and ***P <0.001 (two-tailed t-test). **(F)** JNK pathway activity visualized by *puc-LacZ*. **(G)** JAK/STAT pathway activity visualized by *10xSTAT-GFP* reporter. **(H)** EGFR pathway activity visualized by dpERK expression. **(F-H)** The scale bar represents 50 μm.

Supplementary File S1: Pairwise Pearson correlation coefficient (R) of all samples

Supplementary File S2: DEGs associated with cystic fibrosis in the *dCFTR* intestine

Supplementary File S3: Upregulated DEGs

Supplementary File S4: Downregulated DEGs

SI References

1. Bejarano F*, et al.* (2012) A genome-wide transgenic resource for conditional expression of Drosophila microRNAs. *Development* 139(15):2821-2831.
2. Kim K, Hung RJ, & Perrimon N (2017) miR-263a Regulates ENaC to Maintain Osmotic and Intestinal Stem Cell Homeostasis in Drosophila. *Dev Cell* 40(1):23-36.
3. Hu Y*, et al.* (2011) An integrative approach to ortholog prediction for disease-focused and other functional studies. *BMC bioinformatics* 12:357.
4. Sievers F*, et al.* (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* 7:539.
5. Li W*, et al.* (2015) The EMBL-EBI bioinformatics web and programmatic tools framework. *Nucleic Acids Res* 43(W1):W580-584.
6. Stamatakis A, Hoover P, & Rougemont J (2008) A rapid bootstrap algorithm for the RAxML Web servers. *Syst Biol* 57(5):758-771.
7. Andrews S (2010) Babraham Bioinformatics: FastQC.
8. Wingett SW & Andrews S (2018) FastQ Screen: A tool for multi-genome mapping and quality control. *F1000Res* 7:1338.
9. Krueger F (2012) Babraham Bioinformatics: Trim Galore!
10. Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* 17(1):10.
11. Bolger AM, Lohse M, & Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30(15):2114-2120.
12. Dobin A*, et al.* (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29(1):15-21.
13. Li B, Ruotti V, Stewart RM, Thomson JA, & Dewey CN (2010) RNA-Seq gene expression estimation with read mapping uncertainty. *Bioinformatics* 26(4):493-500.
14. Love MI, Huber W, & Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 15(12):550.
15. Tanya B*, et al.* (2011) NCBI GEO: archive for functional genomics data sets--10 years on. *Nucleic Acids Research* 39(Database issue):D1005.
16. Irizarry RA*, et al.* (2003) Summaries of Affymetrix GeneChip probe level data. *Nucleic Acids Res* 31(4):e15.
17. Tusher VG, Tibshirani R, & Chu G (2001) Significance analysis of microarrays applied to the ionizing radiation response. *Proceedings of the National Academy of Sciences of the United States of America* 98(9):5116-5121.
18. Anders S & Huber W (2010) Differential expression analysis for sequence count data. *Genome Biology* 11(10):R106.
19. Niknian M (1994) Permutation Tests: A Practical Guide to Resampling Methods for Testing Hypotheses. *Technometrics* 37(3):341-342.