Supporting Information

Dequéant et al. 10.1073/pnas.1517729112

SI Methods

Cell Culture. The five cell lines were derived as previously described in Simcox et al. (3) from primary cultures of embryos collected overnight at 17 °C from the cross *UAS-Ras*^{V12}, *UAS-GFP* × *Act5C/TM6*, *Tb*. The cell lines have been deposited at the Drosophila Genomics Resource Center.

RNA Isolation. Cells were lysed using buffer RLT (QIAGEN) or TRIzol (Thermo Fisher) for microarray and RNA-seq samples, respectively. RNA was extracted according to the manufacturer's directions. The quality of each RNA sample was confirmed using a Bioanalyzer (Agilent Technologies).

RNA Expression Measured on Expression Array. Each sample of the R3 and R7 cell line time series was prepared using the GeneChip 3′ IVT Express Kit (Affymetrix) and hybridized on whole-genome *Drosophila* Affymetrix arrays (Affymetrix GeneChip *Drosophila* Genome 2.0 Array), using standard procedures at the Microarray Core Facility of the Dana-Farber Cancer Institute. Array readouts were normalized using dChip (57) and were converted into $\log_2$ ratios of the expression level of each gene to the mean expression value of this gene at early passage time points to plot expression profiles. Time points for the R3 time series were P2, P2 (replicate), P5, P6, P7, P8, P16, P17, and P19; time points for the R7 time series were P2, P2 (replicate), P3, P4, P7, P8, P16, P17, and P19.

RNA Expression Measured by RNA-Seq. Five micrograms of long RNA species from the time series of the R1, R4, and R5 cell lines were depleted from rRNAs using the Ribo-Zero rRNA Removal Kit (Human/Mouse/Rat; Epicentre). RNA integrity and rRNA depletion efficiency were assessed on a Bioanalyzer (Agilent Technologies). RNA-Seq libraries were prepared starting with 50 ng of rRNA-depleted RNAs following the manufacturer's instructions (ScriptSeq v2 RNA-Seq; Epicentre). Thirty-two libraries were multiplexed, clustered, and sequenced on a Hi-Seq Illumina sequencer in paired-end sequencing runs of 76 cycles. Time points for the R1 time series were P2, P3, P4, P5, P7, P8, P10, P11, P16, P17, and P19; time points for the R4 time series were P2, P3, P4, P5, P6, P7, P9, P11, P12, P16, P17, and P19; and time points for the R5 time series were P2, P3, P4, P6, P7, P8, P16, P17, and P19.

Computational Analyses. Reads were mapped to the *Drosophila melanogaster* genome (Berkley Drosophila Genome Project release 5/dm3) and were annotated using TopHat (v1.4.1) and a custom Bowtie index (Bowtie v0.12.8) based on *Drosophila* annotated transcripts from FlyBase release dmel-r5.43. We determined read counts using Bedtools (58) and normalized counts with DESeq v1.8.2 (59) and the R package. To generate expression profiles, normalized counts were converted into $\log_2$ ratios of the expression level of each gene to the mean expression value of this gene at early-passage time points.

Analysis of Differential Expression. Differential expression analysis between early and late passage time points was performed using dChip and DESeq for microarray and RNA-Seq data, respectively. For technical reasons, biological replicates could not be generated. Therefore, to gain statistical significance, samples with consecutive passages in a short time window were treated as biological replicates. Early- and late-passage samples included samples from P2–P4 and P6–P19, respectively. This sample grouping was strongly supported by PCA performed using R and provided by the Harvard Chan Bioinformatics Core, Harvard T. H. Chan School of Public Health, Boston, MA (Fig. S1). Pairwise comparisons showed Pearson correlations higher than 0.9665 for all pairs of samples, thus providing a quality control for high consistency in the processing of the samples (Dataset S9).

Microarray data were filtered by mean expression value at late-passage time points (≥50) for reliably detectable expressed genes. Using this threshold, ∼44% of the genes detected on the array were expressed in the late time points (6,643 and 6,637 genes were detected in the late time points for the R3 and R7 cell lines, respectively). To select the up-regulated genes, we used a fold-change cutoff of 1.3 resulting from the ratio of the mean expression value of time points at late passages to the mean expression value of time points at early passages. As a result, 1,216 and 944 genes were up-regulated at late, compared with early, time points in the R3 and R7 cell lines, respectively (Dataset S1). Then two analyses were performed either using no Padj cutoff or Padj <0.15.

Similar to the microarray data analysis, normalized RNA-Seq data were filtered by mean expression at late passages (normalized counts ≥50, resulting in 7,159, 7,203, and 7,225 detectable genes expressed in the R1, R4, and R5 cell lines, respectively). The mean of the normalized values of early and late time points was calculated by pooling P2–P4 and P16–P19, respectively. Next, we identified the significantly up-regulated genes with a fold-change cutoff ≥1.3 using a two-way comparison. As a result, 1,061, 1,305, and 860 genes were up-regulated at late, compared with early, time points in the R1, R4, and R5 cell lines, respectively (Dataset S1). Then two analyses were performed using either no Padj cutoff or Padj <0.15.

Down-regulated genes were selected by mean expression value at early time points (≥50) and using a fold-change cutoff of −1.5 (Dataset S3).

Venn diagrams were generated using tools developed by Chris Seidel (www.pangloss.com/seidel/Protocols/venn.cgi).

GO term enrichment was performed using the DAVID analysis tool (https://david.ncifcrf.gov/) (60, 61) (Datasets S4 and S5). The most significant GO categories were selected that had *P* values less than 0.05 and a false-discovery rate (FDR) below 0.1. The GO heatmap in Fig. 3*B* was generated by removing GO redundancy manually and keeping only the most highly significant GO categories (*P* values $\leq 3.5 \times 10^{-7}$). Heatmaps were plotted using The Institute For Genomic Research (TIGR) MultiExperiment Viewer (TMEV) (62).

Permutation Test. To assess the significance of the genes found commonly up-regulated at late compared with early time points in the four time series (R1, R3, R4, and R7 cell lines), we performed a permutation experiment using a JAVA program developed in-house by randomly selecting the number of genes found up-regulated in each time series among expressed genes in each time series and then calculating the overlap for 10,000 times. The distribution of the number of overlapping genes was plotted. As a result, the *P* value associated with finding X number of genes in the overlap was calculated as followed: (number of times X number of genes are found in the overlap +1)/10,001.

Correlation Network Analysis. We first prioritized as seed genes the 121 genes (Dataset S2) that are commonly up-regulated at least 1.3-fold between early and late passages in the R1, R3, R4, and R7 cell lines. Because the R5 cell line reached a different late stable state, it was excluded from the first step of the analysis.

Next, using datasets for all five time series, we searched for all genes with an expression profile similar to that of each of the seed profiles with a Pearson correlation coefficient ≥0.8. The final correlation coefficient was calculated by averaging the coefficients of correlation computed separately for each of the five time-series datasets. This procedure allows us to give equal weight to each dataset, given that they have slight differences in the number of analyzed time points. This analysis resulted in the identification of 121 seed modules. Only the modules composed of at least 10 genes sharing similar expression profiles were selected, resulting in 51 relevant modules (Dataset S6). After building a matrix of seeds and correlated gene profiles, we performed unsupervised hierarchical clustering with TM4 (Pearson correlation, average linkage clustering) and identified eight clusters associated with *twi*, *Tsp66E*, *scf*, *Rps5b*, *tum*, *sty*, *sti*, and *Traf4* (the name of each cluster corresponds to the first gene of each cluster on the heatmap) (Dataset S6). We used Cytoscape version 3.1.0 to display the correlation network visually. A similar correlation network analysis was performed on the second list of genes generated using the criteria Padj <0.15. GO term enrichment on the eight clusters was performed using DAVID with the cutoffs $P < 0.01$ and FDR <10. Next, we selected the top eight GO terms and removed redundant terms (Dataset S8).

Comparison with the Cell Lines Analyzed by modENCODE. Reads per kilobase of transcript per million reads mapped (RPKM) values were computed from count files using TopHat and Cufflinks, which were set up on the orchestra computing cluster of Harvard Medical School Research Computing Group. Spearman rank correlation distance was used to compare the end points (P16–P17) of the Ras cell lines (R1, R4, and R5) with the RNA-Seq data cell lines from modENCODE (29). The expression level of the seed genes were calculated from expression data from ref. 63 with levels defined as very low expression (1–3), low expression (4–10), moderate expression (11–25), moderately high expression (26–50), high expression (51–100), very high expression (101–1,000), and extremely high expression (>1,000).

Western Blotting. Cell lysates were prepared using a standard cell extraction buffer supplemented with 1 mM PMSF, additional protease inhibitors (Sigma-Aldrich), and phosphatase inhibitors (Sigma-Aldrich). Lysates were centrifuged, and supernatants were used for analysis. Proteins were separated on polyacrylamide gels and were transferred to nitrocellulose membranes, which were blocked in Odyssey Blocking Buffer (LI-COR). Immunoblots were incubated with the antibodies rabbit anti-ERK and mouse anti-pERK (Cell Signaling Technology). We used the following secondary antibodies: IR dye 800 CW conjugated affinity-purified against rabbit IgG (goat) (Rockland) and Alexa Fluor 680 goat against mouse (A21057; Invitrogen). Blots were scanned using a LI-COR Odyssey Infrared.

Detection of Apoptotic Cells. Apoptosis was detected by TUNEL assay using the In Situ Cell Death Detection Kit (Roche) as described by the manufacturer. Apoptotic cells were visualized with a fluorescent microscope.

Immunohistochemistry on Wing Discs. Wing discs were dissected from third-instar larvae (72 h or greater) and were fixed in 4% (vol/vol) paraformaldehyde diluted in PBS (PBS pH 7.4) for 20 min, washed in PBS, blocked for 30 min in blocking buffer [PBS containing 0.2% Triton X-100 (PBT) containing 5% (wt/vol) BSA (Sigma-Aldrich)] and incubated overnight with the appropriate antibodies in blocking buffer at 4 °C. After overnight incubation, samples were rinsed three times for 15 min in 0.2% PBT and were incubated for 2 h with the appropriate Alexa-conjugated secondary antibody in blocking buffer. DAPI was used at a concentration of 1 μg/mL and was incubated with the secondary antibody. Tissues subsequently were rinsed three times for 15 min in PBT and were mounted in VECTASHIELD (Vector Laboratories). Antibodies used were anti-Twi (rat; 1:1,000), kindly provided by E. Wieschaus from the Department of Molecular Biology at Princeton University, and anti-PH3 (rabbit; 1:3,000; Millipore). Secondary antibodies (1:1,000) conjugated with Alexa Fluor 488 or Alexa Fluor 594 (Invitrogen) were used in all staining procedures. Zeiss confocal point scanning microscopy was used to acquire images, which were processed using Zeiss software.

Immunohistochemistry on Cells. Cells were grown on coverslips in six-well plates and were processed for antibody staining using myosin and tropomyosin antibodies (rat, 1:500) from the Babraham Institute. Cells were washed once in PBS and fixed for 20 min in 4% (vol/vol) paraformaldehyde in PBS, then were rinsed briefly in PBS and were washed three times for 5 min in PBS. PBT was used to permeabilize the cells. Cells were washed three times in PBS, blocked for 1 h, and then incubated with primary antibody in blocking buffer overnight at 4 °C. Cells then were washed three times in PBS and were incubated for 1 h at room temperature with secondary antibodies. After three washes in PBS, the cells were mounted in VECTASHIELD (Vector Laboratories). All secondary antibodies were from Invitrogen.

RNA Isolation, Reverse Transcription, and Real-Time qPCR. RNA was isolated by guanidinium thiocyanate-phenol-chloroform extraction using TRIzol (Life Technologies). Genomic DNA was eliminated through gDNA Eliminator spin columns (QIAGEN), and samples were processed for cleanup with an RNeasy Plus Kit (QIAGEN). One microgram of purified RNA was incubated with a mix of oligo(dT) and random hexamer primers and with iScript RT (iScript cDNA Synthesis Kit; Bio-Rad) for cDNA synthesis. cDNA was used as the template for amplification, using validated primers in iQ SYBR Green Supermix with a CFX96 real-time PCR detection system (Bio-Rad). Query gene expression was relative to a control sample, normalized to the expression of three reference genes: ribosomal protein L32, alpha-tubulin, and Gapdh1, using the $\Delta\Delta C(t)$ analysis method.

Drug and Ecdysone Treatments. Cells were seeded in six-well plates 24 h before drug treatments. The MEK inhibitor U0126 (Promega) was used at 10 μM and triggered cell death after 24 h. No apparent effect was observed in the cells treated with the Akt inhibitor 1L6-Hydroxymethyl-chiro-I, used at 20 μM, and the two PI3K inhibitors LY294002 (Calbiochem), used at 20 μM, and Wortmannin (Calbiochem), used at 100 nM. Inhibition of the Notch pathway was achieved by treatment with the Notch inhibitor DAPT (Sigma-Aldrich) at 10 μM. Ecdysone (20-Hydroxyecdysone; Sigma-Aldrich) treatment was at 1 μg/mL. Cells treated with drugs subsequently were analyzed by qPCR or were stained with antibody as described above.

***Drosophila* Strains.** The following *Drosophila* stocks were used: *tub-Gal80ts*; *Dmef2-Gal4* is described in ref. 64. The X-linked Gal4 driver *1151-Gal4* is active in adult muscle precursors (43). *UAS-RasV12* is described in ref. 65. RNAi stocks for disrupting *CG9650* (*9650-R1* and *9650-R4*) were from the National Institute of Genetics (NIG, Japan) *Drosophila* Stock Center.

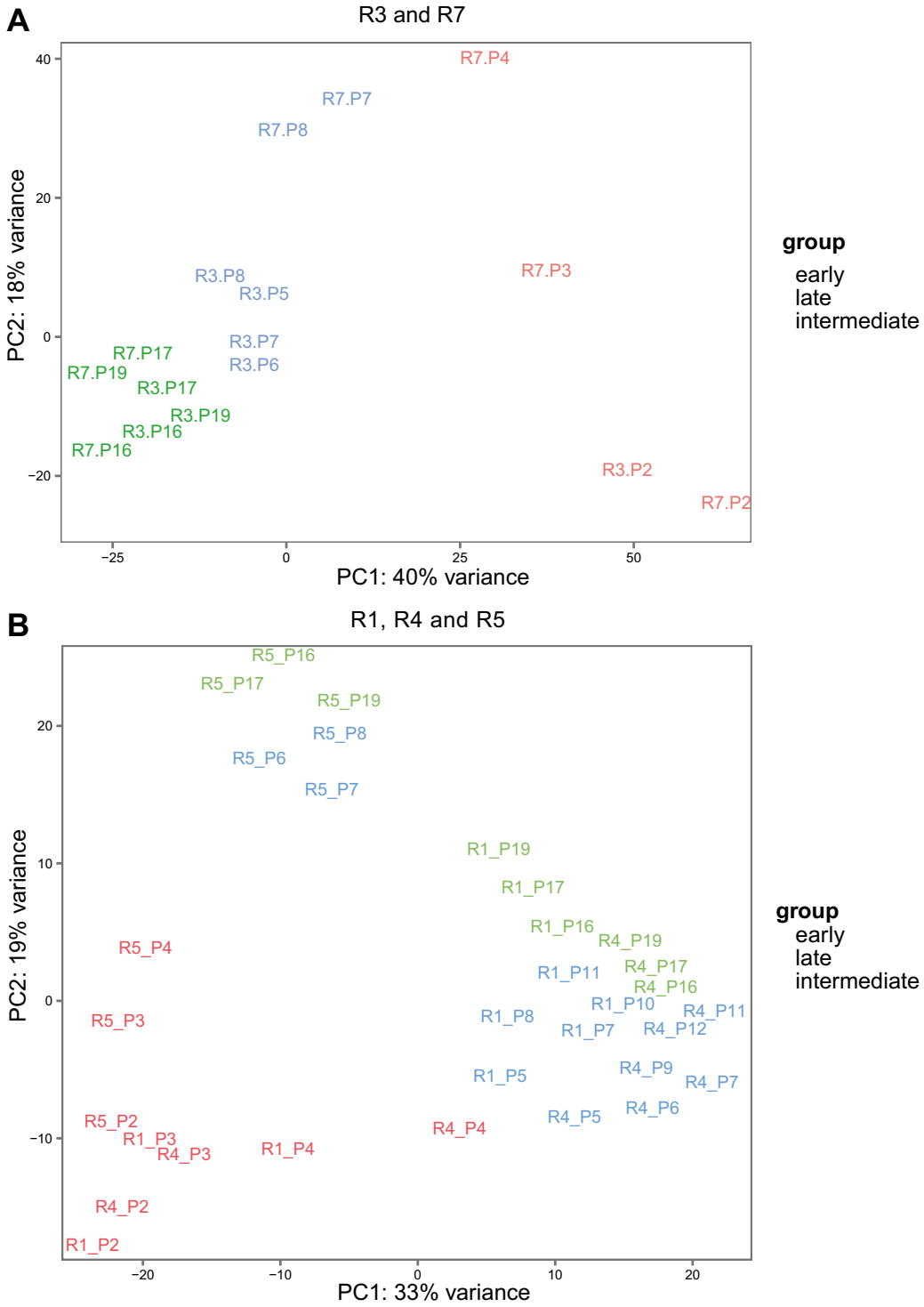
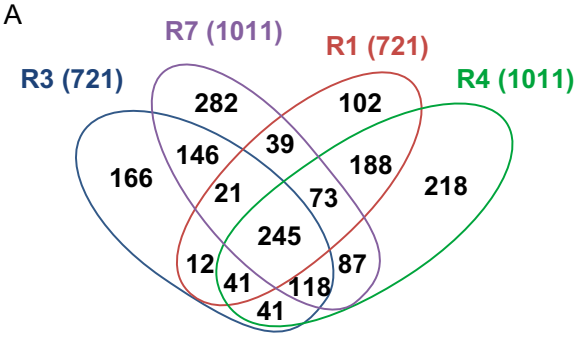


Fig. S1. PCA of time-series samples highlights similarities between most cell lines at early and late time points, respectively. (*A*) PCA for the R3 and R7 cell lines (analyzed by Affymetrix arrays). (*B*) PCA for the R1, R4, and R5 cell lines (analyzed by RNA-Seq). Time points corresponding to different stages of the cultures are color-coded: red, early; green, late; and blue, intermediate.

A



B

GO Term	p-Value	Matches
response to biotic stimulus [GO:0009607]	0.00508	21
response to external biotic stimulus [GO:0043207]	0.00927	21
response to other organism [GO:0051707]	0.00927	21
response to external stimulus [GO:0009605]	0.0114	37
aminoglycan catabolic process [GO:0006026]	0.01316	6
defense response [GO:0006952]	0.01631	22
amino sugar catabolic process [GO:0046348]	0.01652	5
glucosamine-containing compound catabolic process [GO:19010	0.01652	5
organonitrogen compound catabolic process [GO:1901565]	0.02852	9
carbohydrate derivative catabolic process [GO:1901136]	0.02885	7
response to stimulus [GO:0050896]	0.03001	78
detection of bacterium [GO:0016045]	0.03725	3
genital disc sexually dimorphic development [GO:0035263]	0.03725	3
detection of other organism [GO:0098543]	0.03725	3
regulation of biological quality [GO:0065008]	0.03843	38
cell surface receptor signaling pathway [GO:0007166]	0.04233	31
response to bacterium [GO:0009617]	0.04574	14
single-multicellular organism process [GO:0044707]	0.05472	90
single-organism process [GO:0044699]	0.05957	152
immune system process [GO:0002376]	0.05973	20
immune response [GO:0006955]	0.05977	16
regulation of melanization defense response [GO:0035007]	0.0691	3
detection of external biotic stimulus [GO:0098581]	0.0691	3
chitin catabolic process [GO:0006032]	0.07123	4
humoral immune response [GO:0006959]	0.07662	10
organic hydroxy compound metabolic process [GO:1901615]	0.08289	10
developmental process [GO:0032502]	0.08595	89
genital disc pattern formation [GO:0035221]	0.0888	3
genital disc anterior/posterior pattern formation [GO:0035224]	0.0888	3
fluid transport [GO:0042044]	0.09004	4

Fig. S2. Analysis of down-regulated genes at late time points compared with early time points reveals an early immune response in all cell lines. (*A*) Venn diagram showing the number of down-regulated genes (in parenthesis) at late, compared with early, passages in the R1, R3, R4, and R7 cell lines. (*B*) GO category enrichment performed on the common down-regulated genes in the R3, R7, R1, and R4 cell lines using FlyMine and Padj < 0.1.

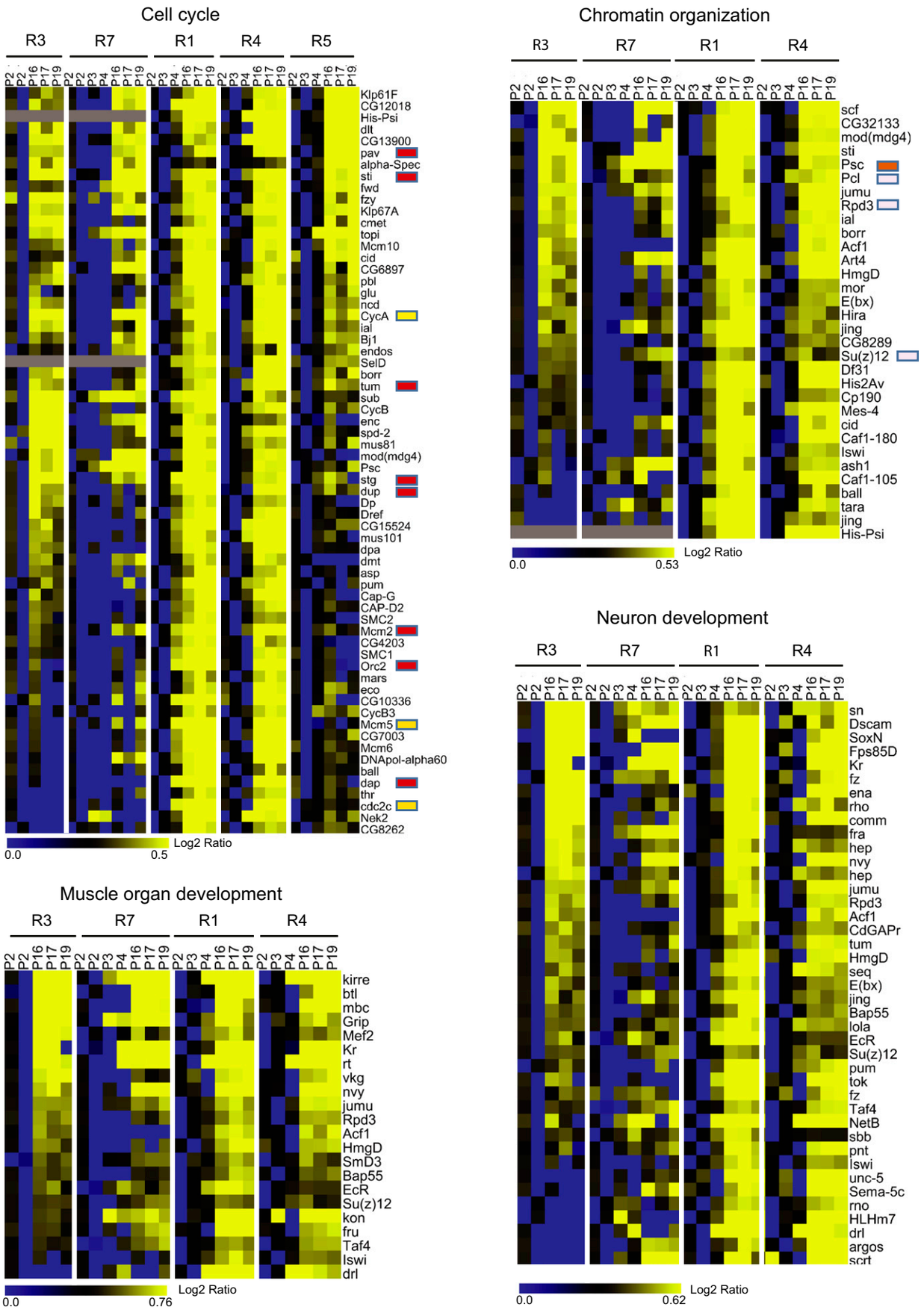


Fig. S3. The genes commonly up-regulated at late time points compared with early time points belong to four GO categories: cell cycle, chromatin organization, neuron, and muscle organ development. Log_2 ratios of the expression levels of genes up-regulated at late, compared with early, time points in each corresponding time series. Only genes up-regulated in at least both the R1 and R4 cell lines are displayed. In the heat maps, blue indicates a decrease in gene expression; yellow indicates an increase in gene expression; gray indicates no Affymetrix probeset. Color coding of gene targets: yellow squares indicate gene targets of the E2F pathway; red squares indicate gene targets of the E2F pathway that are rate-limiting for E2F-dependent proliferation; the orange square indicates a component of PRC1; the pink squares indicate components of PRC2.

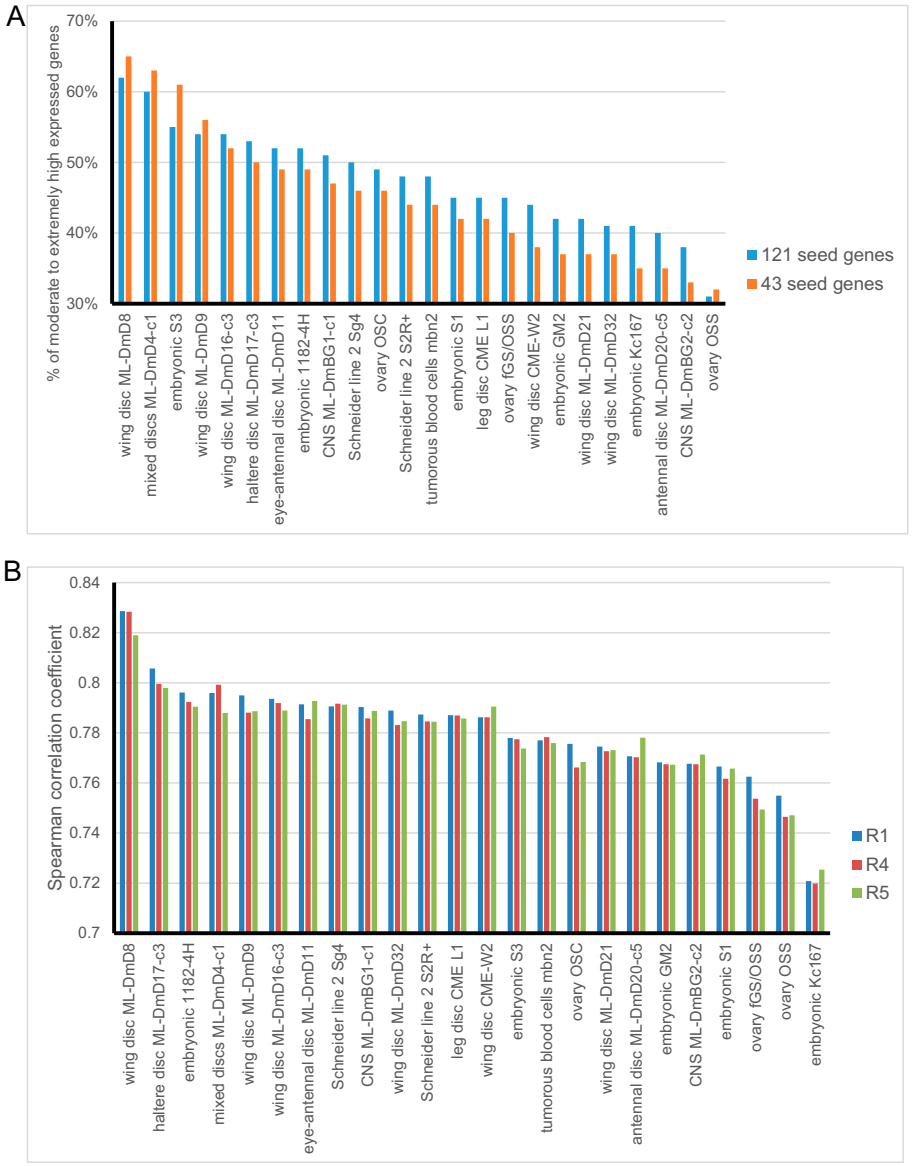


Fig. S4. Comparison of the *Ras*^{V12} cell lines with the modENCODE cell lines. (*A*) Percentage of common up-regulated genes expressed in the cell lines analyzed by modENCODE. (*B*) Spearman rank correlation analysis between the late time points of the Ras^{V12} (R1, R4, and R5) cell lines and the cell line RNA-Seq data from modENCODE.

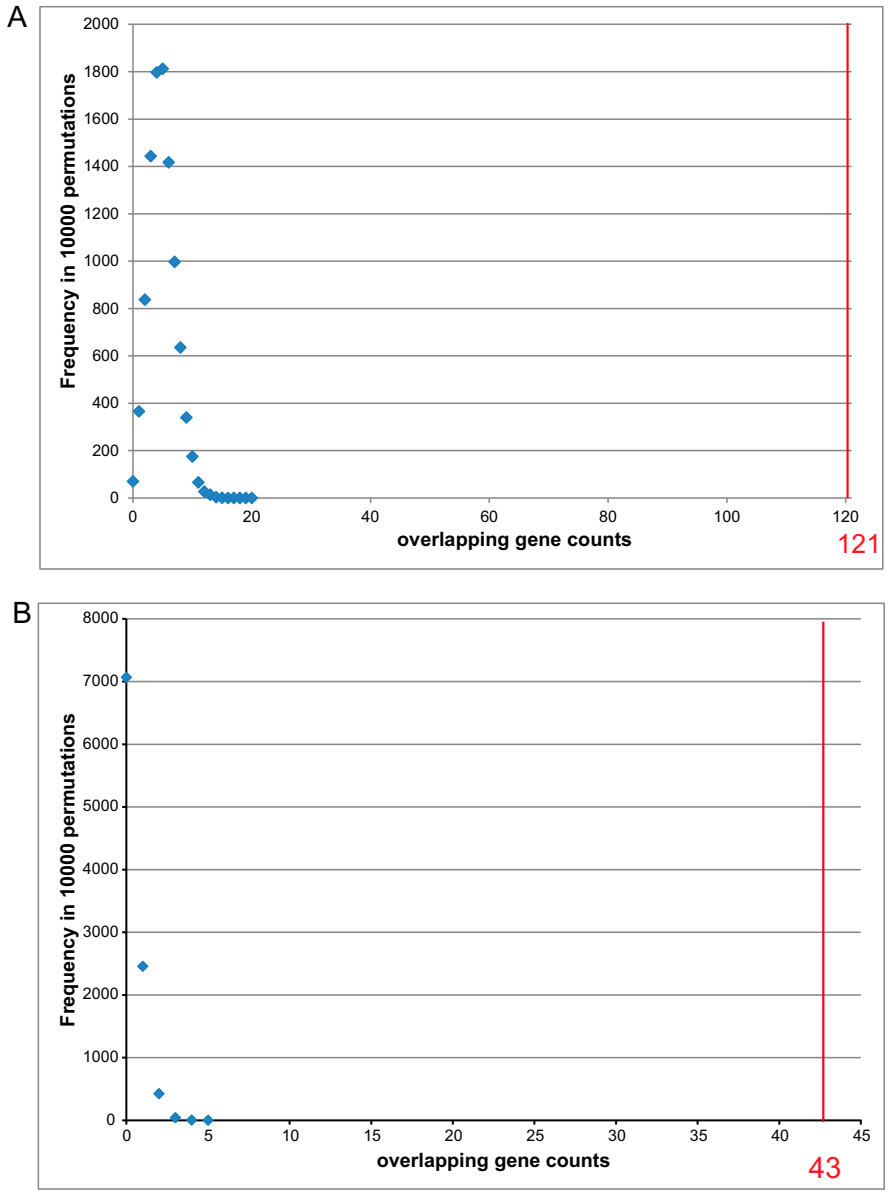


Fig. S5. Distribution plots of the number of genes common to the R1, R3, R4, and R7 cell lines obtained in permutation experiments performed as described in *SI Methods*. (*A*) Permutation test using parameters of the Venn diagram in Fig. 3. (*B*) Permutation test using parameters of the Venn diagram in Fig. S6*A*. The number of common up-regulated genes observed in the data is indicated in red.

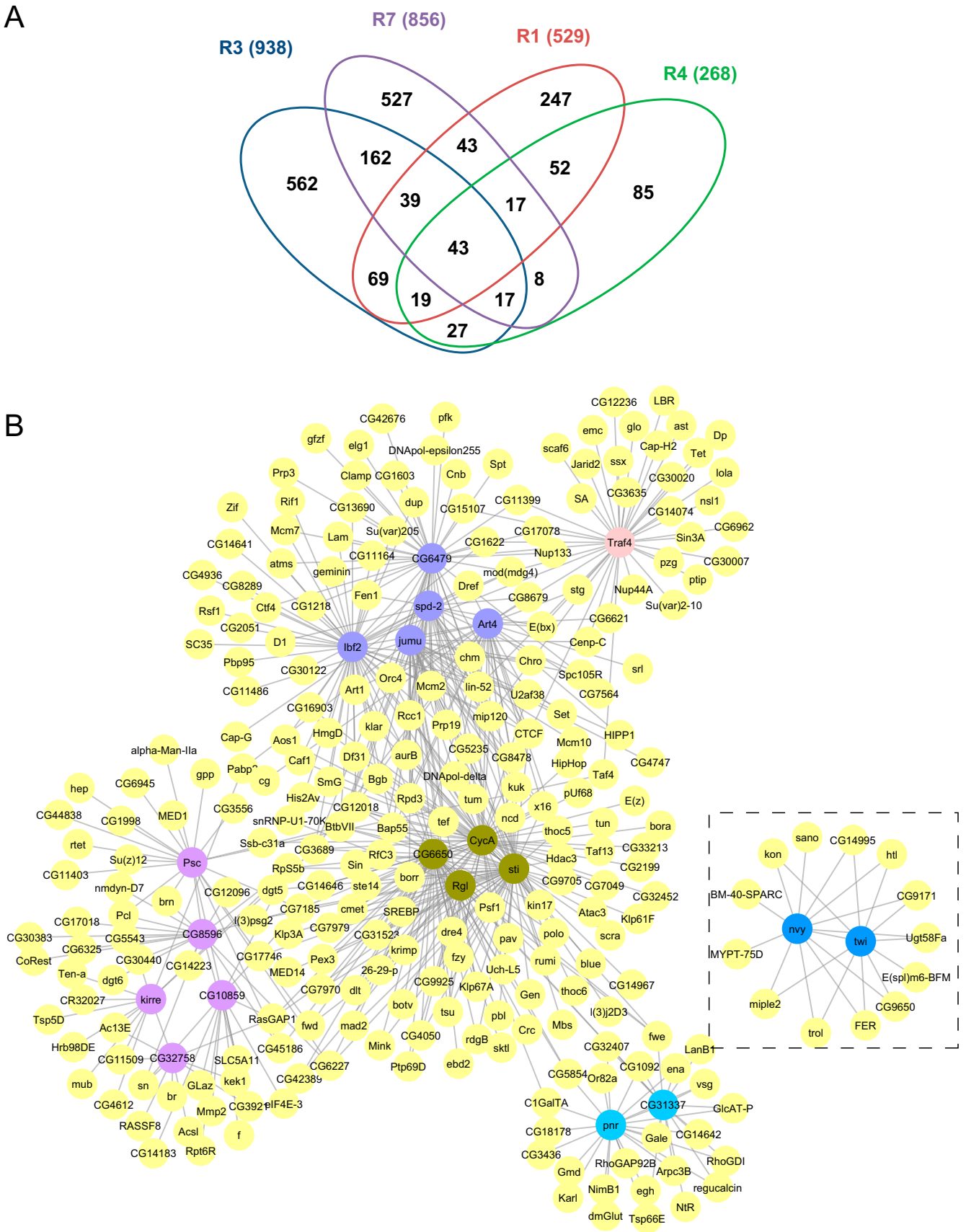


Fig. S6. Coexpression network analysis using more stringent criteria also identifies a specific cluster associated with the transcription factor *twi*. (*A*) A Venn diagram showing in parentheses the number of genes up-regulated at late, compared with early, passages in the R1, R3, R4, and R7 cell lines when using

Legend continued on following page

Padj < 0.15 (*SI Methods*). (*B*)The network is focused on commonly up-regulated genes (43 seeds) in the R1, R3, R4, and R7 cell lines. Genes in all datasets with expression profiles highly correlated (correlation coefficient ≥0.8) to the seeds are in yellow. Seeds with similar expression profiles are found in six clusters (Dataset S7). The *twi* cluster is highlighted in the dashed frame.

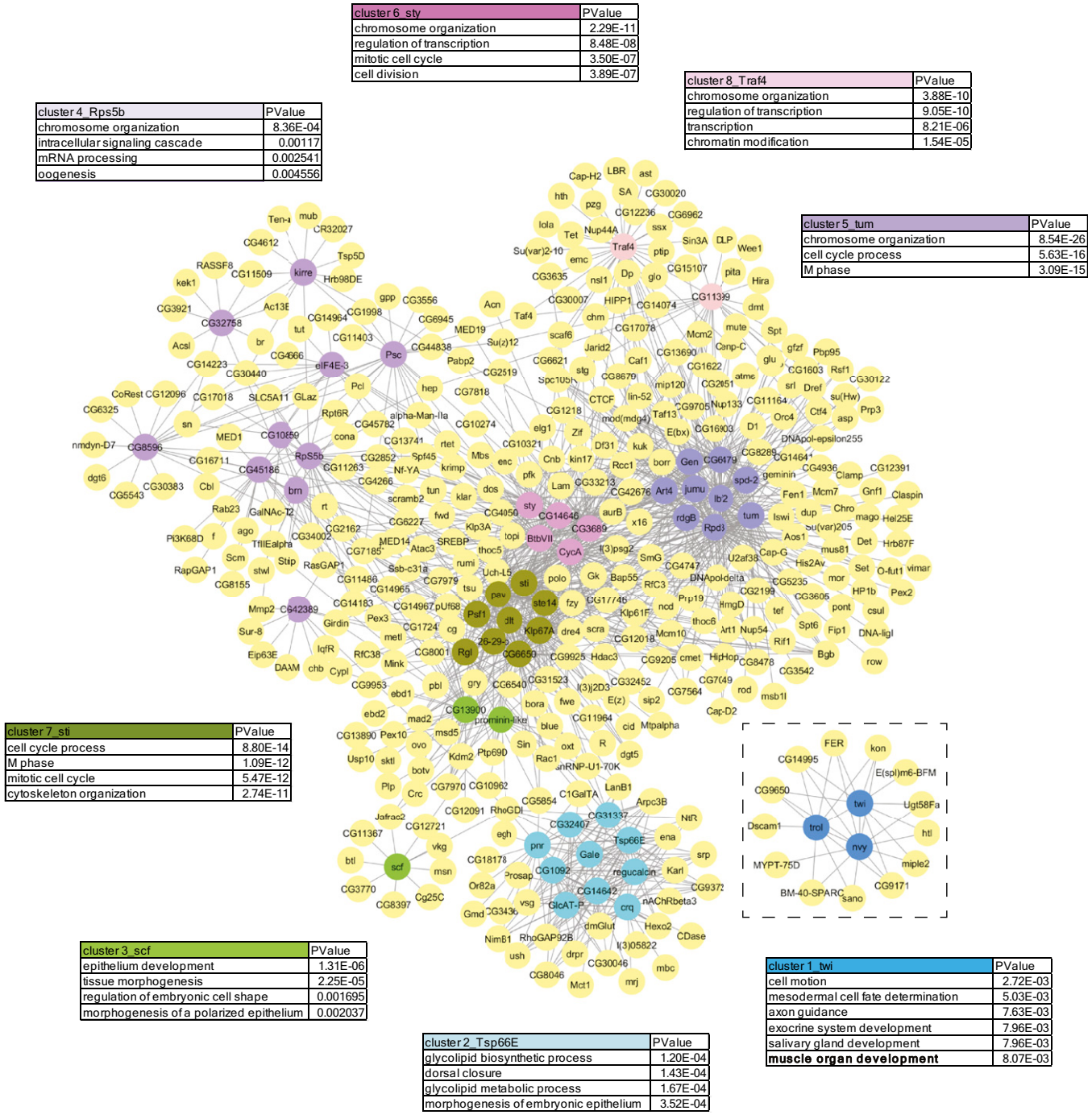


Fig. S7. The coexpression network identifies a specific cluster associated with the transcription factor *twi*. The network is focused on commonly up-regulated genes (seeds) in the R1, R3, R4, and R7 cell lines. Genes in all datasets with highly correlated expression profiles (correlation coefficient ≥0.8) to the seeds are in yellow. Seeds with similar expression profiles are found in eight clusters associated with *twi* (dark blue), *Tsp66E* (light blue), *scf* (light green), *Rps5b* (light purple), *tum* (purple), *sty* (dark pink), *sti* (dark green), and *Traf4* (pink). GO category enrichment is indicated next to each cluster. The *twi* cluster is highlighted in the dashed frame.

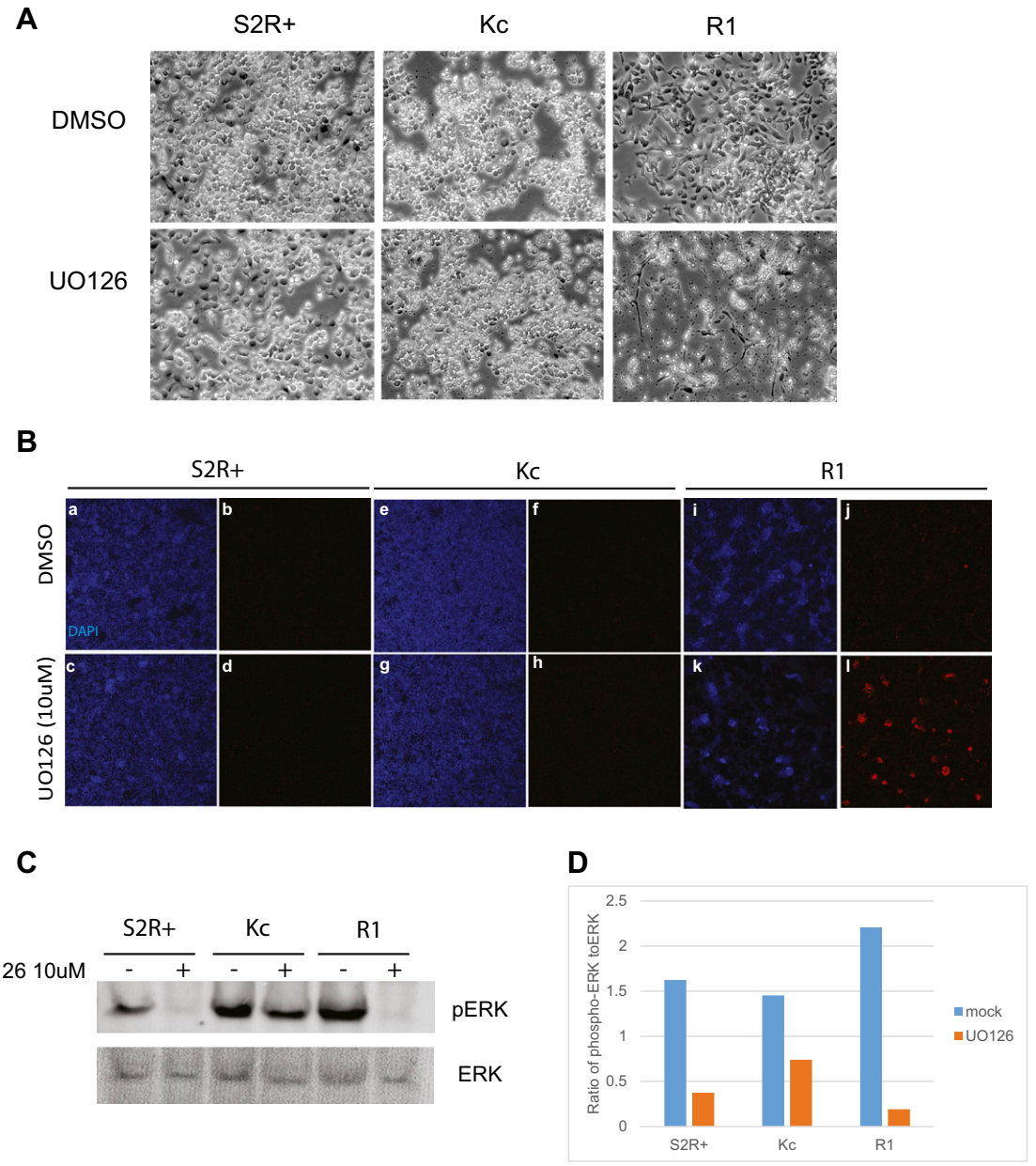


Fig. S8. Treatment of the *Ras*^{V12} cell lines with an inhibitor of the MAPK pathway induces apoptosis. The R1, S2R^{+}, and Kc cell lines were treated with either the MEK inhibitor UO126 at 10 μM or DMSO (mock treatment) after 24 h. (*A*) Microscope images. (*B*) A TUNEL assay was performed as described in *SI Methods*. (*B, a*, *c*, *e*, *g*, *i*, and *k*) DAPI staining is shown in blue. (*B, b*, *d*, *f*, *h*, *j*, and *l*) Apoptotic cells (red) were visualized with a fluorescent microscope. Representative samples are shown. (*C* and *D*) Controls for the effectiveness of the drug inhibiting MEK by Western blot analysis. (*C*) Anti-pERK and anti-ERK. (*D*) Ratios of pERK normalized to ERK.

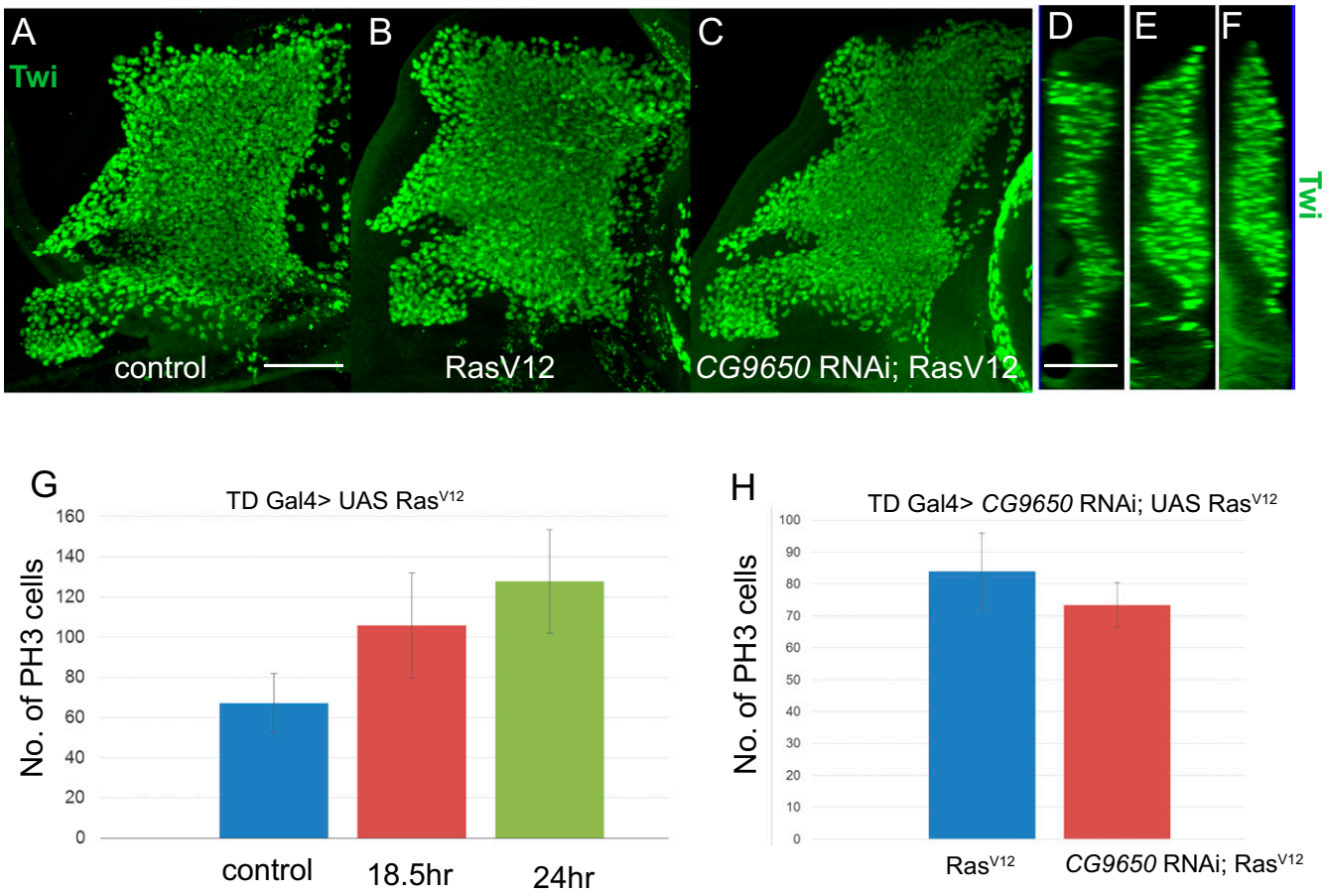


Fig. S9. *CG9650* is required for proliferative activity of Ras^{V12}-expressing AMPs. (*A–F*) Late third-instar discs stained for Twi (anti-Twist, green) in control (*A*), in *TubGal80ts; Dmef2-Gal4* > *UAS-*Ras^{V12} (*B*), and in *TubGal80ts; Dmef2-Gal4* > *UAS-CG9650 RNAi; UAS-*Ras^{V12} (*C*) cells, with an optical section of the wing disc shown in *D*, *E*, and *F*, respectively. (Scale bars, 50 μm.) (*G* and *H*) Quantification of the number of $PH3^{+}$ AMPs in the Ras^{V12} expression using *TubGal80ts; Dmef2-Gal4* > *UAS-*Ras^{V12} (*G*) and in the simultaneous Ras^{V12} expression and *CG9650* down-regulation using *TubGal80ts; Dmef2-Gal4* > *UAS-CG9650 RNAi; UAS-*Ras^{V12} (*H*). Gal80 repression was relieved from early third instar until late third instar by shifting from 18 °C to 29 °C. All plots show mean ± SD (n = 5).

Dataset S1. Genes up-regulated at late, compared with early, time points in the R1, R3, R4, R5, and R7 lines

Dataset S1

Dataset S2. List of genes in the Venn diagram (Fig. 3*A*) showing the number of up-regulated genes at late, compared with early, passages in the R1, R3, R4, and R7 lines

Dataset S2

Dataset S3. Genes down-regulated at late, compared with early, time points in the R1, R3, R4, R5, and R7 lines

Dataset S3

Dataset S4. GO category enrichment on genes that are up-regulated in the R1, R3, R4, R5, and R7 lines

Dataset S4

Dataset S5. GO category enrichment on genes that are down-regulated in the R1, R3, R4, R5, and R7 lines

Dataset S5

Dataset S6. List of gene components of the eight-cluster network analysis

Dataset S6

Dataset S7. List of gene components of the six-cluster network analysis

Dataset S7

Dataset S8. GO category enrichment on the eight clusters

Dataset S8

Dataset S9. Matrix of correlation coefficients between time-point samples in each time series

Dataset S9