Discovery of progenitor cell signatures by time-series synexpression analysis during *Drosophila* embryonic cell immortalization

Mary-Lee Dequéant[a,1], Delphine Fagegaltier[b], Yanhui Hu[a], Kerstin Spirohn[a], Amanda Simcox[c], Gregory J. Hannon[d], and Norbert Perrimon[a,e,1]

[a]Department of Genetics, Harvard Medical School, Boston, MA 02115, [b]Cold Spring Harbor Laboratories, Cold Spring Harbor, NY 11724; [c]Department of Molecular Genetics, The Ohio State University, Columbus, OH 43210; [d]Howard Hughes Medical Institute, Cold Spring Harbor Laboratories, Cold Spring Harbor, NY 11724; and [e]Howard Hughes Medical Institute, Harvard Medical School, Boston, MA 02115

Contributed by Norbert Perrimon, September 10, 2015 (sent for review May 18, 2015; reviewed by Peter Cherbas, Gary Karpen, and Renato Paro)

The use of time series profiling to identify groups of functionally related genes (synexpression groups) is a powerful approach for the discovery of gene function. Here we apply this strategy during Ras^{V12} immortalization of *Drosophila* embryonic cells, a phenomenon not well characterized. Using high-resolution transcriptional time-series datasets, we generated a gene network based on temporal expression profile similarities. This analysis revealed that common immortalized cells are related to adult muscle precursors (AMPs), a stem cell-like population contributing to adult muscles and sharing properties with vertebrate satellite cells. Remarkably, the immortalized cells retained the capacity for myogenic differentiation when treated with the steroid hormone ecdysone. Further, we validated in vivo the transcription factor *CG9650*, the ortholog of mammalian *Bcl11a/b*, as a regulator of AMP proliferation predicted by our analysis. Our study demonstrates the power of time series synexpression analysis to characterize *Drosophila* embryonic progenitor lines and identify stem/progenitor cell regulators.

time series | RNA-Seq | immortalization | *Drosophila*

The highly coordinated expression of genes functioning in common processes is a widespread phenomenon from bacterial operons (1) to eukaryotic synexpression groups (2). Because of the strong correlation between cofunction and coexpression (2), inferring gene function based on covariation of expression profiles (called "guilt by association") is a powerful approach in functional genomics. Importantly, because biological systems are dynamic, recording gene expression over a time series rather than determining a static single measurement can greatly facilitate the characterization of coregulated genes in a particular process.

Using a *Drosophila* embryonic culture system, cell lines can be derived efficiently from primary cell cultures established from embryos expressing constitutively active Ras^{V12} (3). This process is progressive, with the cell lines reaching a stable state within approximately 6 mo. This system provides a unique opportunity to apply a time-series profiling approach to discover synexpression groups in essential biological processes involved in cell immortalization such as cell-cycle regulation, epigenetic regulation, and cellular differentiation. Furthermore, this unbiased transcriptomic approach can provide insights into the unknown origins, regulators, and properties of the immortalized cells.

Here, we perform the first, to our knowledge, in-depth genomic and temporal characterization of five *Drosophila* embryonic cell lines during their establishment. Analysis of differential expression between early and late time points of the cultures indicated that most cell lines reached a similar stable state reminiscent of neurogenic and myogenic progenitor types. To uncover groups of functionally related genes, we applied systematic synexpression network analyses clustering genes with correlated expression profile dynamics, using high-resolution time-series profiling datasets. By analyzing the transcriptional signature of a module associated with the transcription factor *twist* (*twi*), we found that the immortalized cells are related to adult muscle precursors (AMPs), a stem cell-like population contributing to adult muscles (4–7). Consistently, treatment of the immortalized cells with ecdysone, a steroid hormone triggering adult muscle differentiation (8), induced in vitro terminal myogenic differentiation. Our coexpression network analysis further predicted a number of candidate regulators of AMPs, allowing us to validate the transcription factor *CG9650*, an ortholog of the mammalian genes *Bcl11a/b*, as a previously unidentified regulator of AMP proliferation. In conclusion, we show the value of this newly characterized *Drosophila* embryonic progenitor culture system for time-series genomic approaches to identify stem/progenitor cell regulators.

Results

Generation of Immortalized Cell Lines. We established primary cultures from *Act5C* > *UAS-*Ras^{V12}, *UAS-GFP* embryos in which ubiquitously expressed Gal4 drives the expression of both Ras^{V12} and *GFP* (Fig. 1) (3). In early passages, the cultures showed heterogeneous cell morphologies with different levels of GFP and exhibited nonuniform growth across the flasks, suggesting that some cells proliferated more rapidly than others. However, in later passages the cells appeared more homogeneous, suggesting that a single or a few cell types predominated. We successfully derived seven cell lines (R1–R7) from independent primary cultures. All seven lines showed similar passage kinetics before they reached a stable state, and different phases could be distinguished based on

Significance

Cell line derivation is a complex process and a major challenge outside of mammalian systems. Here we take an unbiased genomic approach to characterize a *Drosophila* embryonic culture system allowing the efficient derivation of cell lines of still unknown origins. We reveal that commonly immortalized cells are related to adult muscle precursors (AMPs), a stem cell-like population contributing to adult muscles and sharing properties with vertebrate satellite cells. Remarkably, the cells retain the ability to differentiate terminally in vitro. We also confirm in vivo a previously unidentified regulator of AMP proliferation predicted by our analysis. Our study reveals the potential of this genetically controlled progenitor culture system to provide key missing resources to the *Drosophila* toolbox for cell-based assays.

Author contributions: M.-L.D., D.F., A.S., G.J.H., and N.P. designed research; M.-L.D., D.F., K.S., and A.S. performed research; M.-L.D., D.F., and Y.H. analyzed data; and M.-L.D. and N.P. wrote the paper with contributions from D.F., A.S., and G.J.H.

Reviewers: P.C., Indiana University; G.K., Lawrence Berkeley National Laboratory; and R.P., Swiss Federal Institute of Technology in Zurich.

The authors declare no conflict of interest.

Data deposition: Data have been deposited in the National Center for Biotechnology Information Gene Expression Omnibus (GEO), www.ncbi.nlm.nih.gov/geo/info/submission.html (accession no. GSE73354).

[1]To whom correspondence may be addressed. Email: mdequeant@genetics.med.harvard.edu or perrimon@receptor.med.harvard.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1517729112/-/DCSupplemental.

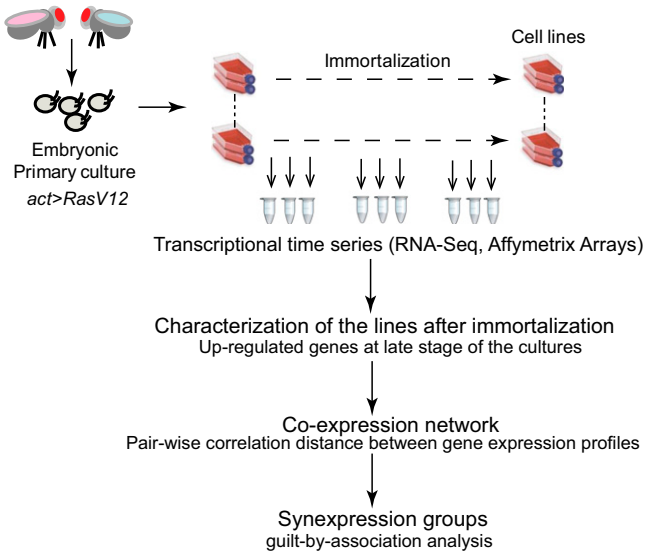

Fig. 1. A time-series profiling approach to characterize *Ras*^{V12} embryonic cell lines by identifying synexpression groups. Seven independent primary cultures were established from *Act5C* > *UAS-Ras*^{V12}, *UAS-GFP* embryos, in which ubiquitously expressed Gal4 drives the expression of both *Ras*^{V12} and *GFP*. High-resolution time-series profiling using RNA-Seq and Affymetrix arrays were generated by sampling five cultures at different passages during cell line derivation over a period of 6 mo. Differential expression of up-regulated genes at late, compared with early, time points, was performed to identify genes associated with the establishment of the final stage of the cultures. Prioritizing these genes, we generated a network of genes with highly correlated temporal expression profiles in the time-series datasets, leading to the identification of synexpression groups.

the shortening of passage times over a 6-mo period: 3–4 wk for passage 1 (P1), 5–20 d for P2–P12, and less than 7 d after P15 (Fig. 2). Most cell lines now have undergone more than 60 passages, equivalent to 120–240 population doublings.

Most Cell Lines Reach a Similar State. To characterize the sequence of events associated with cell line establishment, we generated transcriptional time series from five cell lines by sampling the cultures at successive stages, early (P2–4), intermediate (P4–11), and late (P16–19), characterized by different passage times (Fig. 2). The time series for the R3 and R7 cell lines were analyzed using Affymetrix arrays, and those for R1, R4, and R5 were analyzed using next-generation sequencing (*SI Methods*). Because the cell lines were derived using similar conditions but from independent primary cultures, we first asked whether they progressed similarly during establishment by looking at broad patterns of expression. Principal component analyses (PCA) (Fig. S1) showed for all cell lines a similar clustering pattern of the samples ordered according to the stage of the cultures (early, intermediate, and late) along the first axis, accounting for 33% and 40% of the variance in RNA-Seq and Affymetrix array datasets, respectively (Fig. S1). Furthermore, the plots highlighted the similarity between the early time points of all cell lines and also suggested that the R3 and R7 cell lines reached a similar final stage, as did the R1 and R4 cell lines. However, despite similar initial conditions, the R5 cell line behaved differently from the R1 and R4 lines, rapidly reaching a different late state (Fig. S1*B*).

To characterize the states reached by the cell lines, we analyzed their transcriptomes at the latest stages of the cultures. Genes that were up- or down-regulated at late, compared with early, time points for each culture were identified (*SI Methods*) (Fig. 3*A*, Fig. S2, and Datasets S1, S2, and S3). Gene ontology (GO) category enrichment was performed on each gene list using the DAVID analysis tool (Fig. 3*B*, Fig. S2*B*, and Datasets S4 and S5). Analysis of down-regulated gene expression revealed a common immune response at early time points in all cell lines, potentially reflecting the stress associated with the establishment of the primary cultures (Fig. S2*B*). Despite some differences, comparisons of the up-regulated genes among the lines revealed global similarities, based on enrichments of GO categories (described in the next sections), in the R1, R3, R4, and R7 cell lines, with partial differences with the R5 cell line (Fig. 3*B*), as is consistent with the PCA analysis.

A Proliferative State Associated with the E2 Promoter Binding Factor/Retinoblastoma Protein Pathway. GO category enrichment revealed that cell-cycle and cell-division genes represent the most common significantly up-regulated genes in all Ras^{V12} lines (Fig. 3*B* and Fig. S3). Interestingly, this set contains many known gene targets of the E2 promoter binding factor/retinoblastoma protein (E2F/RB) pathway (highlighted in Fig. S3) (9–11) that plays a central role in cell proliferation and cell growth and which is disrupted in virtually all human cancers (12). It includes regulators of cell-cycle progression, such as *Cyclin A* (*Cyc* A) in all cell lines (11); *string* (*stg*) (13), *dacapo* (*dap*) (14), and *Cyclin-dependent kinase 2* (*cdc2c*) in the R1 and R4 cell lines (9); and *Cyclin E* (*Cyc E*) in the R4 cell line (15). Some of them promote the G1/S (*Cyc A*, *Cyc E*) (15, 16) or G2/M (*stg*) (13) transitions or both (*cdc2c*) (9). Importantly, genes that have been shown to be rate-limiting for E2F1-dependent cell proliferation (17) also are up-regulated in most cell lines. These include *tumbleweed* (*tum*), *sticky* (*sti*), and *pavarotti* (*pav*) in all cell lines; *stg* and *double parked* (*dup*) in the R1, R3, and R4 cell lines; and *dap*, *Origin recognition complex subunit 2* (*Orc2*), and *Minichromosome maintenance 2* (*Mcm2*) in the R1 and R4 cell lines (17). Taken together, these data suggest that the enhanced proliferation of these cells reflects increased E2F activity.

Increasing Levels of Polycomb Group Expression in Established Cell Lines Suggest an Undifferentiated State. Significantly for the R1, R3, R4, and R7 cell lines, the chromatin organization category indicated that the lines were characterized by an epigenetic state associated with increasing levels of Polycomb Group (PcG) transcripts during the immortalization process (Fig. 3*B* and Fig. S3). These transcripts corresponded to components of two cooperating protein complexes: Pc-repressive complexes 1 (PRC1) and 2 (PRC2). Encoding a PRC1 core component, *Posterior sex combs* (*Psc*) was up-regulated in all cell lines. Of the genes encoding PRC2 components, *rpd3* was up-regulated in all cell lines; *Su(z)12* was up-regulated in the R1, R4, and R7 cell lines; *Polycomb-like* (*Pcl*) was up-regulated in the R1, R3, and R4 cell lines; and *Enhancer of zeste* [*E(z)*] was up-regulated in the R4 and R7 cell lines (highlighted in Fig. S3 and Dataset S4). PcG proteins are transcriptional repressors of developmental programs (18) and are expressed at high levels in stem and progenitor cells (19). They are important regulators of stem-cell maintenance in both the

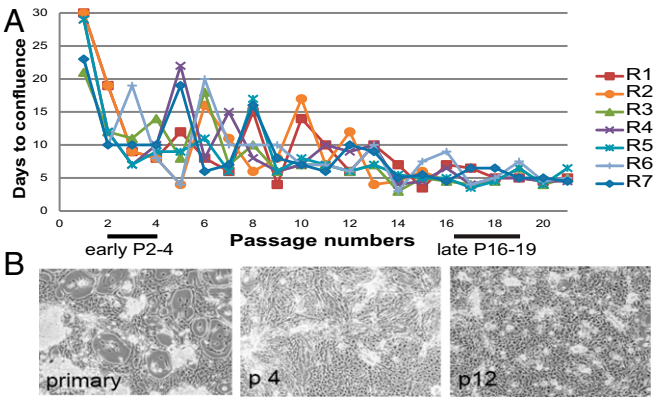

Fig. 2. Primary culture development is characterized by a progressive shortening and stabilization of the passage time. (*A*) The number of days to confluence at each passage number is shown for the seven cultures R1–R7. Cultures were diluted 1:2 from P1 to P13 and 1:4 after P14 (data were normalized for comparison with the 1:2 dilutions). (*B*) Morphology of cultured cells at P0, P4, and P12. Images shown are representative of all the lines.

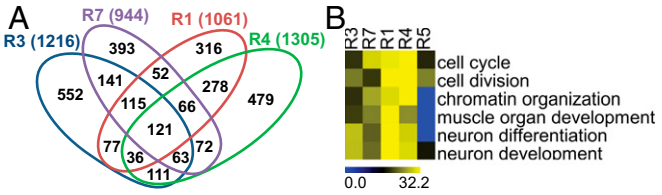

Fig. 3. Time-series transcriptome analyses reveal that most Ras^{V12} cell lines reach a similar state. (*A*) Venn diagram showing the number of up-regulated genes (in parenthesis) at late, compared with early, passages in the R1, R3, R4, and R7 cell lines. (*B*) Heat map showing the results of GO category enrichment performed on the genes up-regulated at late time points compared with early time points in each cell line. Scale is in negative $\log_2 P$ value. Yellow, highly significant; blue, not significant.

undifferentiated and proliferative states (20–22). Therefore, the increased levels of PcG transcripts during cell line establishment indicate a proliferative undifferentiated progenitor-like state.

Immortalized Cell Lines Have Common Tissue Type Origins. Similar GO category enrichment in the R1, R3, R4, and R7 cell lines revealed that they shared similar origins, including neuronal and myogenic tissue types (Fig. 3*B* and Fig. S3). The statistically significant GO categories include "neuron development," "neuron differentiation," and "muscle organ development." They included genes involved in the regulation of neuronal precursors/neuroblasts, such as *Kruppel* (*Kr*) (23) and *SoxNeuro* (*SoxN*) (24), and genes expressed in muscle progenitors, such as *Mef2*, *Kr*, *Kirre*, and *kontiki* (*kon*) (25–28). *Mef2* expression in R1, R3, R7, and R4 cells was confirmed by immunofluorescence staining (3). No significant GO terms associated with tissue type were found for the R5 cell line, so that the origin of this cell population is unclear. Taken together, these data suggest that the cell lines contain proliferative populations of undifferentiated cells from different lineages, among which neuron and muscle lineages are selected consistently. Thereafter, we focused our analysis on the most similar lines.

To determine the state of the cell lines further, we compared the data with modENCODE datasets on available *Drosophila* cell lines (Fig. S4) (29). First we checked the percentage of common up-regulated genes in the generated lines that were expressed in each of the modENCODE cell lines (*SI Methods*). Of note, a standard adjusted *P* value (Padj) cutoff could not be used to select the up-regulated genes in our study because of the lack of true biological replicates and because we were comparing datasets from two different technology platforms, RNA-Seq and Affymetrix arrays. The analyses performed on the R1, R3, R4, and R7 cell lines to find common up-regulated genes yielded 121 genes without using a Padj cutoff (Fig. 3 and Fig. S5*A*) and 43 genes using the cutoff Padj < 0.15 (Figs. S5*B* and S6*A*). To address the issue of false positives, we performed a permutation test (*SI Methods* and Fig. S5) giving statistical confidence (*P* value = 1E-04) to both gene lists. In these lists, the highest percentages of expressed genes were found mostly in disc cell lines and specifically in the Dmd8 line (62 and 65%, respectively) (Fig. S4*A*). Similarly, comparison of the late time points of the R1, R4, and R5 cell lines analyzed by RNA-Seq with modENCODE datasets by Spearman rank correlation analysis (Fig. S4*B*) suggested that the generated cell lines are most similar to the Dmd8 line, a line derived from wing discs with AMP characteristics (30).

Coexpression Network Analysis Identifies a Transcriptional Signature Reminiscent of AMPs. To isolate clusters of coregulated genes revealing tissue-specific transcriptional signatures, we generated a gene coexpression network grouping genes with similar expression profile dynamics across time series (Fig. 4 and Figs. S6 and S7). First, we prioritized as seeds the 121 genes that are commonly up-regulated by at least 1.3-fold at late, compared with early, time points in most cell lines (including the R1, R3, R4, and R7 cell lines but excluding the R5 cell line from the first step of the analysis because it reached a different late stable state) (Fig. 3*A* and Dataset S2). Second, using the five time-series datasets, we searched for all the genes with an expression profile highly correlated (coefficient correlation ≥0.8) with each of the seed profiles; after additional filtration (*SI Methods*), this search identified eight clusters (Dataset S6). A second network analysis using more stringent criteria (selection of gene seeds with Padj < 0.15, yielding 43 genes) was performed also (Fig. S6 and Dataset S7).

In both analyses, the network topology exhibited an isolated cluster associated with the transcription factor *twi* (Fig. 4 and highlighted in Fig. S6*B*), reflecting the high specificity of the correlated expression profiles. Although GO category enrichment performed on the eight clusters indicated broad categories such as cell cycle and chromatin modification (Fig. S7 and Dataset S8), the *twi* cluster was associated with the specific significant category "muscle organ development." Interestingly, many genes of this cluster have been reported to be expressed in AMPs, which are stem cell-like cells set aside during embryogenesis that contribute to adult muscles (4–7). These genes include the transcription factor Twi, whose expression is retained in AMPs until they differentiate (4, 31) and also *kon*, *heartless* (*htl*), *trol*, *ugt58Fa*, *Enhancer of split m6*, *Bearded family member* (*m6*), and *CG9650*) (Figs. 4 and 5 and Fig. S7) (27, 32). Additional genes reported to be expressed in AMPs did not pass the stringent criteria of the correlation analysis but still are up-regulated during most cell line establishment (Fig. 5). Taken together, the global combinatorial coexpression signature strongly points toward an AMP origin of the cell lines.

In addition to revealing an identity of the immortalized cells, the clustering analysis identifies candidate regulators of AMPs. Importantly, 11 genes [*nervy* (*nvy*), *CG14995*, *CG9650*, *trol*, *FER ortholog* (*Homo sapiens*) (*FpS95D*), *Ugt58Fa*, *CG9171*, *midkine and pleiotrophin 2* (*Miple2*), *BM-40-secreted protein acidic and rich in cysteine* (*SPARC*), *serrano* (*sano*), and *MYPT-75D*] exhibited an expression profile similar to that of *twi*, with correlation coefficients ≥0.75 and with unknown function in AMPs (Figs. 4 and 5). Consistent with this

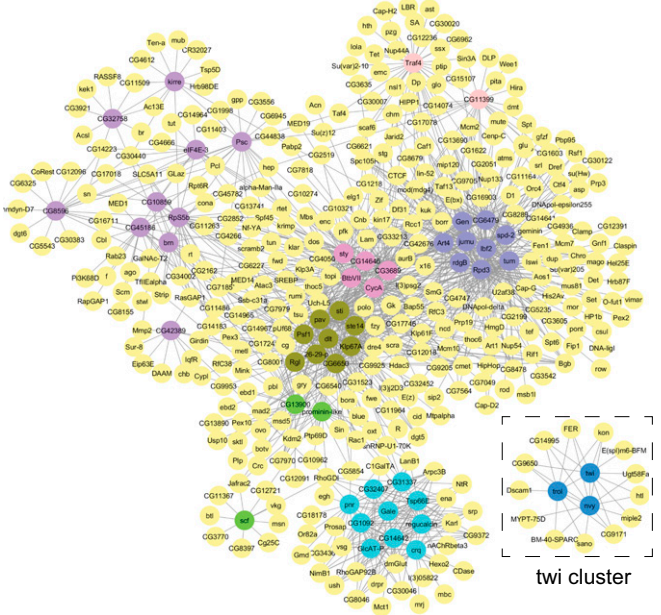

Fig. 4. Coexpression network topology identifies a specific cluster associated with the transcription factor *twi*. The network is focused on commonly up-regulated genes (called "seeds") in the R1, R3, R4, and R7 lines. Genes in all datasets with expression profiles highly correlated (correlation coefficient ≥0.8) to the seeds are in yellow. Seeds with similar expression profiles are found in eight clusters shown in different colors: dark or light blue, green, purple, and pink. The *twi* cluster is highlighted in the dashed frame. The network is shown at higher magnification in Fig. S7.

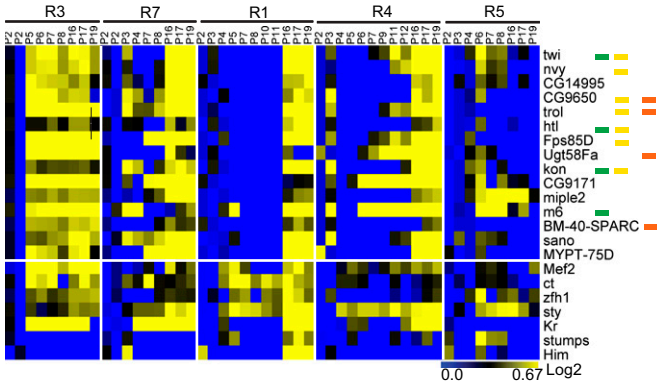

Fig. 5. The transcriptional signature of the *twi* module is reminiscent of that of AMPs. The heat map shows gene expression of components of the *twi* module (correlation coefficient ≥0.75). Green squares: markers of AMPs; orange squares: expressed in AMPs but role not characterized; yellow squares: associated with cancer. Genes below the white line have a correlation coefficient below 0.75 but were found to be up-regulated during the time course and are expressed with a known function in AMPs. Log_2 ratios of the expression levels are represented. Blue indicates a decrease in gene expression; yellow indicates an increase in gene expression.

finding, four of these genes (*CG9650*, *BM-40-SPARC*, *Ugt58Fa*, and *trol*) are expressed in AMPs (32, 33). Two of them have been associated with regulation of proliferation. The heparan sulfate proteoglycan (HSPG) *trol* can be secreted from *EGF receptor* (*Egfr*)-overexpressing wing imaginal disc epithelia and drive the overproliferation of AMPs (34). Although the role of *BM-40-SPARC* is not known in AMPs, this multifunctional calcium-binding glycoprotein associated with the ECM (35) is a direct modulator of several mitotic factors (36) and is up-regulated during skeletal muscle regeneration involving the activation/proliferation of satellite cells (37).

Immortalized AMP-Like Cells Differentiate into Muscle Cells in Response to the Steroid Hormone Ecdysone. Because the immortalized cells exhibited a transcriptional signature reminiscent of progenitor cells, we attempted to differentiate them to reveal their tissue of origin. Reasoning that continuous MAPK activity inducing a proliferative state could antagonize differentiation, we first attempted to inhibit the MAPK pathway using treatment with the MAPK/ERK kinase (MEK) inhibitor U0126. As a result, we observed cell death of the Ras^{V12} cells after 24 h by visual inspection and TUNEL assay, but no effect was observed in $S2R^+$ and Kc cell lines (Fig. S8 *A* and *B*). Although the drug inhibited MEK (as checked by readout of decreased levels of phospho-ERK in Fig. S8 *C* and *D*) in all cell lines, the effect was very diverse in terms of survival/cell death, with the Ras^{V12} lines being particularly sensitive to MAPK activity. However, no cell differentiation was observed. No apparent effect was observed when the cells were treated with Akt or PI3K) inhibitors (*SI Methods*).

Using our predictive analysis on AMPs, we devised a hormonal treatment (*SI Methods*) mimicking adult muscle differentiation during metamorphosis. Remarkably, treatment of the R1 cell line with ecdysone induced dramatic changes in cell morphology within 24 h. As cells elongated, they started to express the muscle terminal differentiation marker *myosin heavy chain* (*mhc*) (Fig. 6 *A* and *B*), whereas Twi expression was down-regulated (Fig. 6 *A* and *B*). Quantitative PCR (qPCR) analysis quantified the increase in expression of *mhc* (threefold) and *Tropomyosin* (*Tm*) (2.5-fold) (Fig. 6*C*). Expression of *Mef2*, a critical component of adult muscle differentiation and a target of ecdysone (8), also was increased (2.5-fold) (Fig. 6*C*). We also observed similar effects in the R3, R4, and R7 cell lines, although with variable efficiency. Occasionally, contractile muscle cells were found in culture, suggesting terminal myogenic differentiation of the cells. This treatment was ineffective in the R5 cell line; no R5 cells stained for *mhc*, as is consistent with the disappearance of the AMP transcriptional signature from this cell line at late passages. Taken together, these results confirm a muscle progenitor origin for most cells and reveal the reversibility of the Ras^{V12} immortalization phenotype.

***CG9650* Is a Regulator of AMP Proliferation.** To test the role of a previously unidentified factor in AMP regulation, we examined the role of the zinc-finger–containing putative transcription factor *CG9650*, which is orthologous to the mammalian genes *Bcl11a/b*. Highly correlated (coefficient correlation >0.9) with the *twi* profile in all cell lines, *CG9650* is a strong candidate for a role in AMPs. It is expressed in the mesoderm (38), in the embryonic nervous system where it has been implicated in axon guidance (39), and in AMPs (33). Previous overexpression experiments have suggested that *CG9650* influences Notch signaling in sensory organ and eye development and/or cell viability (40), FGF signaling (41), and growth or cell-cycle progression in the developing eye (42).

To characterize the role of *CG9650* in AMPs, we depleted *CG9650* in AMPs by RNAi using the AMP-specific Gal4 drivers *1151*-Gal4 or *Mef2*-Gal4 (7, 43). During larval development, AMPs for adult flight muscles are found in the ventral region of the wing imaginal disc below the epithelial cells that give rise to the body wall and are labeled by Twi antibody. They proliferate during the L2 and L3 stages (7). Knockdown of *CG9650* by RNAi during the AMP proliferation stage resulted in a reduction of both the number and layers of the Twi^+ cells in the late L3 wing disc (Fig. 7 *A*–*D*). An assay for cell mitosis showed that the number of AMPs proliferating in the knockdown animals was half that in control individuals (Fig. 7*E*), indicating that *CG9650* knockdown affects the number of proliferating AMPs.

To investigate the role of *CG9650* in Ras^{V12}-induced overproliferation further (Fig. S9), we first established an assay for this phenomenon in AMPs. We expressed Ras^{V12} in AMPs during the larval proliferation phase using *Dmef2-Gal4* and then assayed mitotic activity at late third instar using the phospho-histone 3 (PH3) antibody (*SI Methods*). Strikingly, a marked increase in the number of mitotically active cells was observed in Ras^{V12}-expressing AMPs (Fig. S9*G*). The longer the induction of Ras^{V12} expression, the higher was the number of $PH3^+$ cells (more than two-fold after 24 h of induction) compared with controls (Fig. S9*G*). Furthermore, the number and layers of Twi^+ cells were increased (Fig. S9 *A*, *B*, *D*, and *E*). To test the role of *CG9650* in the context of Ras^{V12} overproliferation, we drove the expression of Ras^{V12} and *CG9650* RNAi together in AMPs. The coexpression led to a less severe overgrowth than seen with Ras^{V12} expression alone (Fig. S9 *B*, *C*, *E*, *F*, and *H*), with fewer proliferative cells and a decrease in the layers of Twi-labeled cells, although not to the level seen in controls. These results are consistent with a requirement for *CG9650* in the Ras^{V12}-induced overproliferation phenotype.

Strikingly, *CG9650* is coexpressed with targets of the Notch pathway (*twi*, *m6*, and *him*) (Fig. 5), suggesting that *CG9650* might be regulated by Notch signaling. Treatment of Ras^{V12} cells with the Notch pathway inhibitor N-[N-(3,5-Difluorophenacetyl)-L-alanyl]-S-phenylglycine t-butyl ester (DAPT) for 24 h down-regulated *CG9650* expression by 40% (Fig. 7*F*), similar to its effect on other known Notch targets including *twi*, *him*, and *m6*. This result

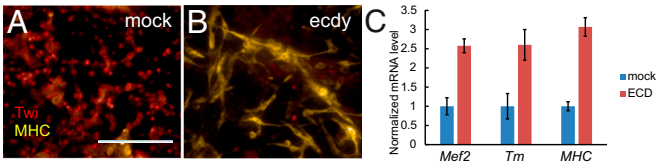

Fig. 6. Immortalized AMP-like cells differentiate in vitro into muscle cells in response to ecdysone treatment. (*A* and *B*) Immunofluorescence staining showing Twi (anti-Twi, red) and MHC expression (anti-MHC, yellow) in cells that were mock treated (*A*) or treated with ecdysone (*B*) after 24 h. (Scale bar, 100 μm.) (*C*) Relative quantification of known muscle differentiation markers by qPCR, comparing ecdysone and mock-treated R1 cells. Data are shown as mean ± SE ($n = 3$).

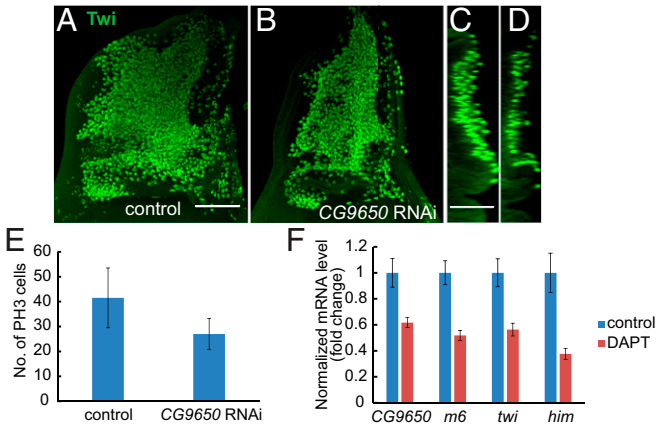

Fig. 7. *CG9650* is required for the proliferative activity of AMPs and is regulated by the Notch pathway. Two different *CG9650* RNAi lines expressed using the AMP-specific Gal4 drivers *1151-Gal4* or *TubGal80ts; dMef2-Gal4* (TD-Gal4) resulted in pupal lethality when shifted from 18 °C to 29 °C at early second instar. Further validation was performed using the *CG9650-R1 Drosophila* transgenic RNAi line and the TD-Gal4 driver. (*A–D*) Late third-instar discs stained for Twi (anti-Twist, green) in control (*A*) and *TD-Gal4 > UAS-CG9650* RNAi (*B*), with optical sections of the wing discs in *C* and *D*, respectively. (*E*) Quantification of number of $PH3^+$ AMPs following *CG9650* down-regulation using *TubGal80ts; Dmef2-Gal4 > UAS-CG9650* RNAi. Gal80 repression was relieved from early second instar until late third instar by shifting from 18 °C to 29 °C. The data are shown as mean ± SD. (n = 5). (*F*) Relative quantification of *CG9650* and other known Notch targets by qPCR, comparing R1 cells treated with the Notch pathway inhibitor DAPT and mock-treated cells. Data are shown as mean ± SE (n = 3).

suggests that *CG9650* is a bona fide Notch target in the immortalized cells, is consistent with a previous report indicating that *CG9650* is a Notch target in Dmd8 cells (33), and also is consistent with the activity of Notch signaling in maintaining AMPs in a proliferative state during larval stages (7, 44).

Discussion

To uncover synexpression groups during immortalization of *Drosophila* embryonic cells, we generated high-resolution time-series profiling during the establishment of five cell lines. Analysis of temporal coexpression profiles identified transcriptional signatures suggesting an AMP origin for the cells. We revealed that the immortalized cells can be differentiated in vitro. Finally, we predicted by guilt-by-association analysis that the transcription factor *CG9650* is a previously unidentified regulator of AMP proliferation and then validated the prediction.

Although the cultures were derived from whole embryos, the most frequently immortalized cells were related to AMPs, a stem cell-like population contributing to adult muscles (6, 7) that is specified during early embryogenesis (45). During embryogenesis, the MAPK pathway is a key regulator of the specification and survival of AMPs (27). Here we show that driving Ras^{V12} expression in AMPs during larval proliferation induces an overproliferation phenotype. Taken together these in vivo observations are consistent with the common immortalization of AMP-like cells from embryonic cultures expressing Ras^{V12}. The generation of other progenitor-type cell lines such as gut progenitor cells might require a combination of oncogenes/tumor suppressors that would support epithelial proliferation. For example, expressing both Ras^{V12} and wts^{dsRNA} successfully generated epithelial cell lines (46). Furthermore, different culture conditions (e.g., the addition of growth factors, insulin, or fly extract) might be needed for different cell types.

Despite sharing many markers and properties with AMPs, the immortalized cells are highly proliferative and continuously express Ras^{V12}, making them similar to a cancer stem cell-like state. Interestingly, the *twi* transcriptional module contains many genes that have mammalian orthologs associated with cancers, notably rhabdomyosarcoma (RMS), a childhood muscle cancer. For example, *Twi* and *htl/FGFR1* are overexpressed in primary RMS tumors (47), and *kon/CSPG4* is expressed in RMS cell lines and patient material (48). Furthermore, *nvy/ETO* and *CG9650/Bcl11a/b* have been implicated in leukemia (49–51). Finally, *trol/perlecan* expression is up-regulated in Ras^{V12} tumors and is associated with promoting tumor cell proliferation (34, 52).

Clustering analysis with the *twi* expression profile revealed many coexpressed genes with unknown function in AMPs. We show that one of them, *CG9650*, a zinc-finger–containing putative transcription factor expressed in AMPs (33), is required for AMP proliferation and is regulated by Notch signaling, which also is involved in AMP proliferation (7). In addition, the high correlation of *CG9650* expression with *twi* (coefficient correlation >0.9) in all cell lines suggests that the transcription factor Twi might regulate *CG9650* expression directly. This finding is consistent with previous ChIP-on-chip analyses identifying Twi *cis*-regulatory modules in the vicinity of the *CG9650* promoter (38, 53) during embryonic mesoderm expression. The mammalian orthologs of *CG9650*, *Bcl11a* and *Bcl11b*, are Kr-like transcription factors that have been associated with the maintenance of lymphoid and ameloblast progenitors, respectively (54, 55). *Bcl11b* also is expressed in murine myogenic progenitors and disappears during differentiation, as is consistent with a possible conserved role in vertebrate muscle progenitor proliferation (56).

We show, for the first time to our knowledge, that immortalized *Drosophila* cells can be terminally differentiated in vitro into the myogenic lineage by treatment with the steroid hormone ecdysone, which is known to induce AMP differentiation in vivo (8). Despite the complexity of adult muscle differentiation (5), we show that the differentiated cells express markers of terminal muscle differentiation, such as *mhc* and *Tm*. Furthermore, contractile muscle cells were observed occasionally in the differentiated cultures. Consistently, our in vitro differentiation system recapitulates an in vivo regulation of adult muscle differentiation. For example, *Mef2*, an ecdysone target that plays a critical role in adult muscle differentiation (8), is also up-regulated by ecdysone in vitro.

Finally, in contrast to previously existing *Drosophila* cell lines obtained by spontaneous immortalization, the cell lines characterized in this study have been derived using a genetic method (3). Therefore, several manipulations can be implemented to improve the system. Using Gal4 lines driving expression in the population cell type of interest would make the process tissue specific and potentially faster, because it would direct the selection of the culture toward the desired final stage. At early stages of the culture, selecting cells expressing specific levels of GFP (correlated to Ras^{V12} expression levels) also could stabilize the culture faster. In addition, the increased expression of PcG genes during immortalization suggests that affecting the epigenetic cell state, for example by expressing high levels of PcG in combination with Ras^{V12}, might facilitate the immortalization process. Finally, the use of an inducible system to control the expression of the oncogene would make it possible to limit the potential impact of continuous Ras^{V12} expression on cell behavior and properties. Combined with the powerful *Drosophila* genetic tools, this newly characterized culture system opens the door for the establishment of progenitor lines of a desired genotype, amenable to cell-based assays to shed light on a variety of biological processes.

Methods

Details on sample generation and analysis (cell culture, preparation of Affymetrix array and RNA-Seq samples, data analysis including differential expression, PCA, permutation test, correlation network, clustering, GO enrichment), the fly strains used in this study, and protocols used for antibody staining, real-time qPCR, drug and ecdysone treatments, Western blotting, and TUNEL assay can be found in *SI Methods*.

ACKNOWLEDGMENTS. We thank R. Binari, M. Kuroda, C. Pitsouli, B. Matthey-Prevot, R. Sopko, and R. Sugimura for critical reading of the manuscript; R. Gunage for advice on AMP dissections; the Microarray Core Facility of the Dana-Farber Cancer Institute; E. Wieschaus for providing anti-Twi antibody; and L. Pantano, A. Sjödin, and J. Hutchinson of the Harvard Chan Bioinformatics Core for assistance with analysis and suggestions. This work was supported in part by funding from the Starr Cancer Consortium. N.P. and G.J.H. are Investigators of the Howard Hughes Medical Institute.

1. Jacob F, Perrin D, Sanchez C, Monod J (1960) [Operon: A group of genes with the expression coordinated by an operator]. *C R Hebd Seances Acad Sci* 250:1727–1729.
2. Niehrs C, Pollet N (1999) Synexpression groups in eukaryotes. *Nature* 402(6761):483–487.
3. Simcox A, et al. (2008) Efficient genetic method for establishing Drosophila cell lines unlocks the potential to create lines of specific genotypes. *PLoS Genet* 4(8):e1000142.
4. Currie DA, Bate M (1991) The development of adult abdominal muscles in Drosophila: Myoblasts express twist and are associated with nerves. *Development* 113(1):91–102.
5. Fernandes J, Bate M, Vijayraghavan K (1991) Development of the indirect flight muscles of Drosophila. *Development* 113(1):67–77.
6. Roy S, VijayRaghavan K (1999) Muscle pattern diversification in Drosophila: The story of imaginal myogenesis. *Bioessays* 21(6):486–498.
7. Gunage RD, Reichert H, VijayRaghavan K (2014) Identification of a new stem cell population that generates Drosophila flight muscles. *eLife* 3:e03126.
8. Lovato TL, Benjamin AR, Cripps RM (2005) Transcription of Myocyte enhancer factor-2 in adult Drosophila myoblasts is induced by the steroid hormone ecdysone. *Dev Biol* 288(2):612–621.
9. Dimova DK, Stevaux O, Frolov MV, Dyson NJ (2003) Cell cycle-dependent and cell cycle-independent control of transcription by the Drosophila E2F/RB pathway. *Genes Dev* 17(18):2308–2320.
10. Neufeld TP, de la Cruz AF, Johnston LA, Edgar BA (1998) Coordination of growth and cell division in the Drosophila wing. *Cell* 93(7):1183–1193.
11. Knudsen KE, Fribourg AF, Strobeck MW, Blanchard JM, Knudsen ES (1999) Cyclin A is a functional target of retinoblastoma tumor suppressor protein-mediated cell cycle arrest. *J Biol Chem* 274(39):27632–27641.
12. Nevins JR (2001) The Rb/E2F pathway and cancer. *Hum Mol Genet* 10(7):699–703.
13. Edgar BA, Lehman DA, O'Farrell PH (1994) Transcriptional regulation of string (cdc25): A link between developmental programming and the cell cycle. *Development* 120(11):3131–3143.
14. Lane ME, et al. (1996) Dacapo, a cyclin-dependent kinase inhibitor, stops cell proliferation during Drosophila development. *Cell* 87(7):1225–1235.
15. Duronio RJ, O'Farrell PH (1995) Developmental control of the G1 to S transition in Drosophila: Cyclin E is a limiting downstream target of E2F. *Genes Dev* 9(12):1456–1468.
16. Sprenger F, Yakubovich N, O'Farrell PH (1997) S-phase function of Drosophila cyclin A and its downregulation in G1 phase. *Curr Biol* 7(7):488–499.
17. Herr A, et al. (2012) Identification of E2F target genes that are rate limiting for dE2F1-dependent cell proliferation. *Dev Dyn* 241(11):1695–1707.
18. Sparmann A, van Lohuizen M (2006) Polycomb silencers control cell fate, development and cancer. *Nat Rev Cancer* 6(11):846–856.
19. Rajasekhar VK (2009) *Regulatory Networks in Stem Cells* (Springer, New York).
20. Bello B, Holbro N, Reichert H (2007) Polycomb group genes are required for neural stem cell survival in postembryonic neurogenesis of Drosophila. *Development* 134(6):1091–1099.
21. Lessard J, Sauvageau G (2003) Bmi-1 determines the proliferative capacity of normal and leukaemic stem cells. *Nature* 423(6937):255–260.
22. Valk-Lingbeek ME, Bruggeman SW, van Lohuizen M (2004) Stem cells and cancer; the polycomb connection. *Cell* 118(4):409–418.
23. Li X, et al. (2013) Temporal patterning of Drosophila medulla neuroblasts controls neural fates. *Nature* 498(7455):456–462.
24. Buescher M, Hing FS, Chia W (2002) Formation of neuroblasts in the embryonic central nervous system of Drosophila melanogaster is controlled by SoxNeuro. *Development* 129(18):4193–4203.
25. Bour BA, et al. (1995) Drosophila MEF2, a transcription factor that is essential for myogenesis. *Genes Dev* 9(6):730–741.
26. Isshiki T, Takeichi M, Nose A (1997) The role of the msh homeobox gene during Drosophila neurogenesis: Implication for the dorsoventral specification of the neuroectoderm. *Development* 124(16):3099–3109.
27. Figeac N, Jagla T, Aradhya R, Da Ponte JP, Jagla K (2010) Drosophila adult muscle precursors form a network of interconnected cells and are specified by the rhomboid-triggered EGF pathway. *Development* 137(12):1965–1973.
28. Ranganayakulu G, et al. (1995) A series of mutations in the D-MEF2 transcription factor reveal multiple functions in larval and adult myogenesis in Drosophila. *Dev Biol* 171(1):169–181.
29. Cherbas L, et al. (2010) The transcriptional diversity of 25 Drosophila cell lines. *Genome Res* 21(2):301–314.
30. Ui K, Ueda R, Miyake T (1987) Cell lines from imaginal discs of Drosophila melanogaster. *In Vitro Cellular Dev Biol Anim* 23(10):707–711.
31. Bate M, Rushton E, Currie DA (1991) Cells with persistent twist expression are the embryonic precursors of adult muscles in Drosophila. *Development* 113(1):79–89.
32. Butler MJ, et al. (2003) Discovery of genes with highly restricted expression patterns in the Drosophila wing disc using DNA oligonucleotide microarrays. *Development* 130(4):659–670.
33. Krejcí A, Bernard F, Housden BE, Collins S, Bray SJ (2009) Direct response to Notch activation: Signaling crosstalk and incoherent logic. *Sci Signal* 2(55):ra1.
34. Herranz H, Weng R, Cohen SM (2014) Crosstalk between epithelial and mesenchymal tissues in tumorigenesis and imaginal disc development. *Curr Biol* 24(13):1476–1484.
35. Lane TF, Sage EH (1994) The biology of SPARC, a protein that modulates cell-matrix interactions. *FASEB J* 8(2):163–173.
36. Motamed K, et al. (2003) Fibroblast growth factor receptor-1 mediates the inhibition of endothelial cell proliferation and the promotion of skeletal myoblast differentiation by SPARC: A role for protein kinase A. *J Cell Biochem* 90(2):408–423.
37. Petersson SJ, et al. (2013) SPARC is up-regulated during skeletal muscle regeneration and inhibits myoblast differentiation. *Histol Histopathol* 28(11):1451–1460.
38. Ciglar L (2010) Drosophila myogenesis as a model for studying cis-regulatory networks: Identifying novel players and dissecting the role of transcriptional repression. Ph.D. thesis. (The Ruperto-Carola University of Heidelberg, Heidelberg).
39. McGovern VL, Pacak CA, Sewell ST, Turski ML, Seeger MA (2003) A targeted gain of function screen in the embryonic CNS of Drosophila. *Mech Dev* 120(10):1193–1207.
40. Shalaby NA, et al. (2009) A screen for modifiers of notch signaling uncovers Amun, a protein with a critical role in sensory organ development. *Genetics* 182(4):1061–1076.
41. Zhu MY, Wilson R, Leptin M (2005) A screen for genes that influence fibroblast growth factor signal transduction in Drosophila. *Genetics* 170(2):767–777.
42. Tseng AS, Hariharan IK (2002) An overexpression screen in Drosophila for genes that restrict growth or cell-cycle progression in the developing eye. *Genetics* 162(1):229–243.
43. Roy S, VijayRaghavan K (1998) Patterning muscles using organizers: Larval muscle templates and adult myoblasts actively interact to pattern the dorsal longitudinal flight muscles of Drosophila. *J Cell Biol* 141(5):1135–1145.
44. Anant S, Roy S, VijayRaghavan K (1998) Twist and Notch negatively regulate adult muscle differentiation in Drosophila. *Development* 125(8):1361–1369.
45. Ruiz Gómez M, Bate M (1997) Segregation of myogenic lineages in Drosophila requires numb. *Development* 124(23):4857–4866.
46. Simcox A (2013) Progress towards Drosophila epithelial cell culture. *Methods Mol Biol* 945:1–11.
47. Goldstein M, Meller I, Orr-Urtreger A (2007) FGFR1 over-expression in primary rhabdomyosarcoma tumors is associated with hypomethylation of a 5′ CpG island and abnormal expression of the AKT1, NOG, and BMP4 genes. *Genes Chromosomes Cancer* 46(11):1028–1038.
48. Brehm H, et al. (2014) A CSPG4-specific immunotoxin kills rhabdomyosarcoma cells and binds to primary tumor tissues. *Cancer Lett* 352(2):228–235.
49. Weniger MA, et al. (2006) Gains of the proto-oncogene BCL11A and nuclear accumulation of BCL11A(XL) protein are frequent in primary mediastinal B-cell lymphoma. *Leukemia* 20(10):1880–1882.
50. Huang X, Du X, Li Y (2012) The role of BCL11B in hematological malignancy. *Exp Hematol Oncol* 1(1):1–22.
51. Nimer SD, Moore MA (2004) Effects of the leukemia-associated AML1-ETO protein on hematopoietic stem and progenitor cells. *Oncogene* 23(24):4249–4254.
52. Jiang X, Couchman JR (2003) Perlecan and tumor angiogenesis. J Histochem Cytochem 51(11):1393–1410.
53. Zinzen RP, Girardot C, Gagneur J, Braun M, Furlong EE (2009) Combinatorial binding predicts spatio-temporal cis-regulatory activity. *Nature* 462(7269):65–70.
54. Yu Y, et al. (2012) Bcl11a is essential for lymphoid development and negatively regulates p53. *J Exp Med* 209(13):2467–2483.
55. Katsuragi Y, et al. (2013) Bcl11b transcription factor plays a role in the maintenance of the ameloblast-progenitors in mouse adult maxillary incisors. *Mech Dev* 130(9-10):482–492.
56. Duong BB (2014) Bcl11a and Bcl11b Regulation of the Decision for Murine Cells to Proliferate or Differentiate During Skeletal Muscle Development and Repair. Thesis (Barrett, The Honors College at Arizona State University, Tempe, AZ).
57. Li C, Wong WH (2001) Model-based analysis of oligonucleotide arrays: Expression index computation and outlier detection. *Proc Natl Acad Sci USA* 98(1):31–36.
58. Quinlan AR, Hall IM (2010) BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* 26(6):841–842.
59. Anders S, Huber W (2010) Differential expression analysis for sequence count data. *Genome Biol* 11(10):R106.
60. Huang da W, Sherman BT, Lempicki RA (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* 4(1):44–57.
61. Huang da W, Sherman BT, Lempicki RA (2009) Bioinformatics enrichment tools: Paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res* 37(1):1–13.
62. Saeed AI, et al. (2003) TM4: A free, open-source system for microarray data management and analysis. *Biotechniques* 34(2):374–378.
63. St Pierre SE, Ponting L, Stefancsik R, McQuilton P; FlyBase Consortium (2014) FlyBase 102–advanced approaches to interrogating FlyBase. *Nucleic Acids Res* 42(Database issue):D780–D788.
64. Owusu-Ansah E, Song W, Perrimon N (2013) Muscle mitohormesis promotes longevity via systemic repression of insulin signaling. *Cell* 155(3):699–712.
65. Karim FD, Rubin GM (1998) Ectopic expression of activated Ras1 induces hyperplastic growth and increased cell death in Drosophila imaginal tissues. *Development* 125(1):1–9.