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

Mary-Lee Dequéant, Delphine Fagegaltie, Yanhui Hu, Kerstin Spirohn, Amanda Simcox, Gregory J. Hannon, and Norbert Perrimon

*Department of Genetics, Harvard Medical School, Boston, MA 02115,* *Cold Spring Harbor Laboratories, Cold Spring Harbor, NY 11724;* *Department of Molecular Genetics, The Ohio State University, Columbus, OH 43210;* *Howard Hughes Medical Institute, Cold Spring Harbor Laboratories, Cold Spring Harbor, NY 11724;* and *Howard Hughes Medical Institute, Harvard Medical School, Boston, MA 02115*

Contribution 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 RasV12 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 Bdi1a/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.

**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 (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 Bdi1a/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-RasV12*, *UAS-GFP* embryos in which ubiquitously expressed Gal4 drives the expression of both *RasV12* 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 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. S1B).

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. 3A, Fig. S2B, and Datasets S4 and S5). Gene ontology (GO) category enrichment was performed on each gene list using the DAVID analysis tool (Fig. 3B, Fig. S2B, 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. S2B). 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. 3B), 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 RasV12 lines (Fig. 3B and Fig. S3). Interestingly, this set contains many known gene targets of the E2 promoter binding factor/retnoblastoma 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* (sg) (13), *dacapo* (dapo) (14), and *Cyclin-dependent kinase 2* (cdk2) 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 (sg) (13) transitions or both (cdk2) (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* (*dop*) in the R1, R3, and R4 cell lines; and *dop*, *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. 3B 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, *prcl* was up-regulated in all cell lines; *Set2* (*set*) 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). *Pc* 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 undifferentiated and proliferative states.
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 profiles 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. 3A 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. S6B), 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 [nerve (nry), CG14995, CG2650, trol, FER ortholog (Homo sapiens) (FpS95D), Ugt58Fa, CG9171, midkine and plasminogen 2 (Miple2), BM-40-secreted protein acidic and rich in cysteine (SPARC), semano (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...
finding, four of these genes (CG9650, BM-40-SPARC, Ugt58Fa, and \(t\)ol) are expressed in AMPs (32, 33). Two of them have been associated with regulation of proliferation. The heparan sulfate proteoglycan (HSPG) \(t\)ol can be secreted from EGF receptor (Ergr)-overexpressing wing imaginal disc epithelia and drive the over-proliferation 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\(^{12}\) cells after 24 h by visual inspection and TUNEL assay, but no effect was observed in S2\(^{R}\) and Kc cell lines (Fig. S8A 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\(^{12}\) 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. 6C). Expression of \(Mez\), a critical component of adult muscle differentiation and a target of ecdysone (8), also increased (2.5-fold) (Fig. 6C). 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\(^{12}\) immortalization phenotype.

**CG9650 Is a Regulator of AMP Proliferation.** To test the role of a previously unidentifed 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 \(Mez\)-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. 7A–D). An assay for cell mitosis showed that the number of AMPs proliferating in the knockdown animals was half that in control individuals (Fig. 7E), indicating that CG9650 knockdown affects the number of proliferating AMPs.

To investigate the role of CG9650 in Ras\(^{12}\)–induced overproliferation further (Fig. S9), we first established an assay for this phenomenon in AMPs. We expressed Ras\(^{12}\) in AMPs during the larval proliferation phase using \(Dmez\)-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\(^{12}\)-expressing AMPs (Fig. S9G). The longer the induction of Ras\(^{12}\) expression, the higher was the number of PH3\(^{+}\) cells (more than two-fold after 24 h of induction) compared with controls (Fig. S9G). 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\(^{12}\) overproliferation, we drove the expression of Ras\(^{12}\) and CG9650 RNAi together in AMPs. The coexpression led to a less severe overgrowth than seen with Ras\(^{12}\) 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\(^{12}\)-induced overproliferation phenotype.

Strikingly, CG9650 is coexpressed with targets of the Notch pathway (\(twi\), \(mhc\), and \(him\)) (Fig. 5), suggesting that CG9650 might be regulated by Notch signaling. Treatment of Ras\(^{12}\)-expressing AMPs with the Notch pathway inhibitor N-[\(N\)-(3,5-Difluorophenacyl)-L-alamyl]-S-phenylglycine t-buty ester (DAPT) for 24 h down-regulated CG9650 expression by 40\% (Fig. 7F), similar to its effect on other known Notch targets including \(twi\), \(him\), and \(m6\). This result
**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*AR* 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, *hvy/ETO* and *CG9650/Bcl11ab* have been implicated in leukemia (49–51). Finally, *troll/pericca* expression is up-regulated in Ras*V12* tumors and is associated with promoting tumor cell proliferation (34, 52).

**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 edcsyne treatments, Western blotting, and TUNEL assay can be found in SI Methods.

**ACKNOWLEDGMENTS.** We thank R. Binari, M. Kuroda, C. Pitsoulis, B. Matthey-Prevot, R. Sopko, and R. Sugimura for critical reading of the manuscript; R. Gunase for advice on AMP dissection; 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.