

# Temporal waves of coherent gene expression during *Drosophila* embryogenesis

Ilya Papatsenko<sup>1</sup>, Mike Levine<sup>1</sup> and Dmitri Papatsenko<sup>2,\*</sup>

<sup>1</sup>University of California Berkeley, Department of Molecular and cell Biology and <sup>2</sup>Black Family Stem Cell Institute, Mount Sinai School of Medicine

Associate Editor: Martin Bishop

## ABSTRACT

**Motivation:** Animal development depends on localized patterns of gene expression. Whole-genome methods permit the global identification of differential expression patterns. However, most gene-expression-clustering methods focus on the analysis of entire expression profiles, rather than temporal segments or time windows.

**Results:** In the current study, local clustering of temporal time windows was applied to developing embryos of the fruitfly, *Drosophila melanogaster*. Large-scale developmental events, involving temporal activation of hundreds of genes, were identified as discrete gene clusters. The time-duration analysis revealed six temporal waves of coherent gene expression during *Drosophila* embryogenesis. The most powerful expression waves preceded major morphogenetic movements, such as germ band elongation and dorsal closure. These waves of gene expression coincide with the inhibition of maternal transcripts during early development, the specification of ectoderm, differentiation of the nervous system, differentiation of the digestive tract, deposition of the larval cuticle and the reorganization of the cytoskeleton during global morphogenetic events. We discuss the implications of these findings with respect to the gene regulatory networks governing *Drosophila* development.

**Availability:** Data and software are available from the UC Berkeley web resource [http://flydev.berkeley.edu/cgi-bin/GTEM/dmap\\_dm-ag/index\\_dmap.htm](http://flydev.berkeley.edu/cgi-bin/GTEM/dmap_dm-ag/index_dmap.htm)

**Contact:** dimitri.papatsenko@mssm.edu

**Supplementary information:** Supplementary data are available at *Bioinformatics* online.

Received on June 23, 2010; revised on August 30, 2010; accepted on August 31, 2010

## 1 INTRODUCTION

*Drosophila* embryogenesis has been divided into 17 separate stages, each stage marked by global changes of embryo morphology, such as cellularization, gastrulation, germ band elongation (GBE), germ band retraction (GBR) or dorsal closure (DC). Different parts of an embryo might undergo simultaneous morphological changes. Parallel developmental events, such as DC and head involution, might begin at the same time (stages 14–15), but terminate at different stages. Based on the time and the duration parameters,

the parallel events can be mapped to a 2D parameter space and separated. Thus, parallel and temporally overlapping events can be separated and analyzed (locally) with respect to distinct time and duration parameters. These parameters are convenient for the classification of peaks of gene expression observed in the global time series data. Presence of temporal peaks of expression with similar time/duration properties might indicate a functional link between genes, even if the genes' expression profiles are not similar.

Complete expression data have been obtained for the entire developmental time course of *Drosophila melanogaster* using both expression (Arbeitman *et al.*, 2002; Hooper *et al.*, 2007; Stolc *et al.*, 2004; White *et al.*, 1999) and tiling microarrays (Biemar *et al.*, 2005). Previous analyses of these datasets have shown that the vast majority of genes exhibit at least 4-fold expression changes during embryogenesis (Arbeitman *et al.*, 2002; White *et al.*, 1999) and most genes have two or more major temporal peaks of expression during the life cycle. Such reengagement of genes and gene networks has recently been demonstrated during the development of the mosquito embryo, *Anopheles gambiae* (D.Papatsenko, M.Levine and U.Goltsev, personal communication). These studies suggest that morphologically distinct and temporally separated developmental events, such as formation of the serosa and cuticle, depend on the deployment of coherent gene batteries. A recent study by Hooper and co-workers (2007) identified a number of time points during fly development when clusters of genes coherently change their expression levels. Using an approach initially introduced for the analysis of development in *Dictyostelium* (Sasik *et al.*, 2002), Hooper and co-workers have identified several coherently switching classes of genes in *Drosophila*, including maternal (off switch), activated (on switch) and transient (peak) expression types. This clustering strategy (Hooper *et al.*, 2007; Sasik *et al.*, 2002) relies on local similarities (coherent on/off switches) in the global expression data.

The time-duration analysis of fly development described in this study represents an unsupervised learning algorithm [a supervised method was introduced earlier (Samsonova *et al.*, 2007)], similar to the temporal-clustering method implemented by Hooper and co-workers (2007). Technically, the method can be classified as a query-driven, local clustering of time series expression data (Balasubramanian *et al.*, 2005; Dhollander *et al.*, 2007). In this study, the search patterns (queries) were temporal sequences corresponding to two major types of gene activity: 'peak' (transient type) and 'switch' (activated type). The search patterns were matched to the expression profiles locally (time windows), in order

\*To whom correspondence should be addressed.

to retrieve genes with locally coherent patterns of expression. The duration parameter was incorporated by scaling each search pattern to a range of time windows. This method identified nearly a dozen distinct, large-scale developmental events during *Drosophila* embryogenesis, each involving coherent expression of hundreds of genes. Based on gene annotation enrichment, many of the identified events were attributed to known morphological changes, such as formation of the nervous system, digestive tract, deposition of cuticle or reorganization of the cytoskeleton during morphogenetic movements.

This method successfully identified discrete temporal events from global time-series expression data. The approach can be applied to the analysis of a variety of dynamic biological processes.

## 2 METHODS

### 2.1 Microarray data and data treatment

Time-series data for the *Drosophila* developmental course was taken from the Hooper *et al.*'s (2007) study. Low-level microarray data treatment involved standard quintile normalization of microarrays (Bolstad *et al.*, 2003) and Z-score normalization of time points (Eisen *et al.*, 1998). Microarray data were filtered based on variation between biological triplicate's points using standard ANOVA analysis with very mild thresholds ( $P < 0.05$ ). These procedures produced 7052 gene expression profiles. Finally, the data has been upsampled to 100 points and smoothed using Gaussian filter with standard deviation equal to 5 points (after upsampling).

### 2.2 Local clustering algorithm

The algorithm inputs included a query pattern (sequence of values) and a time-series data. The query patterns were rescaled (resampled and smoothed) to a range 5–50% with respect to the number of time points in the input time series data (duration parameter  $j$ ) and then matched to every time window of the microarray data set  $A$ . Pearson correlation  $r$  and standard deviation  $\sigma$  were calculated for every window position in every gene expression profile. In the next step, for each expression profile and each combination of time  $i$  and duration  $j$  parameters, a matching score  $S$  was calculated as:

$$S(i, j) = \begin{cases} -\log(p(r_{ij}|A) + \alpha), & p(\sigma_j|A) < h \\ 0, & p(\sigma_j|A) \geq h \end{cases} \quad (1)$$

In this formula, the parameter  $h$  captures relative amplitude of the signals (i.e. peaks). For  $h=0.5$  (selected in the case of peak matching), only the highest 50% of all peaks were considered for a given window size  $j$ . The  $h=0.9$  (highest 10%) has been selected for the 'switch' query pattern. Correlation values ( $r$ ) were translated to conditional  $P$ -values in order to obtain scores, which were comparable for different values of the duration  $j$  (window sizes). Parameter  $\alpha$  is a pseudocount limiting maximal value of the score  $S$  ( $\alpha=0.0001$  in this study). Clusters were identified as local peaks on the time/duration parameter space, in other words—parameter combinations that produced a large number of genes with scores  $S$  above an arbitrarily chosen score cutoff ( $S=1$ ). Genes, producing  $S>1$  were extracted from each cluster and annotated. Notice that this procedure returns overlapping clusters as any part of a given temporal expression profile can be associated with different clusters.

### 2.3 Annotation term enrichment analysis

GO slim version of gene annotations (Adams *et al.*, 2000), containing terms related to protein function and biochemistry and so-called 'controlled vocabulary (CV)' annotations describing morphological parts of fly embryo were used in this study (Tomancak *et al.*, 2002). The enrichment analysis was performed using custom software 'GO-piper' (available upon request),

based on using binomial statistics to calculate term enrichment. The  $P$ -values were adjusted using multiple test correction (correction Bonferroni).

## 3 RESULTS

### 3.1 Biological assumptions and selection of input search patterns

Previous analyses of fly and mosquito embryogenesis pointed to the presence of temporally local characteristic patterns of gene activity, such as on/off switches (activation/repression type) (Hooper *et al.*, 2007) and peaks of expression (transient type) (Goltsev *et al.*, 2009). These features are shared by many genes that are coherently expressed at certain time intervals. It is reasonable to assume that such coherent expression profiles might precede or coincide with major morphogenetic events. Perhaps, such a straightforward 1D interpretation of the time-series data is too simplistic as there are unrelated, parallel developmental events that overlap in time (e.g. DC and head involution). The application of both time and duration parameters should help resolve such developmental events (including parallel events) and provide a clear foundation for functional analysis.

In theory, any type of temporal expression pattern can be of interest; in this study, however, the search patterns for the time-duration analysis were limited by the two previously described types: local on/off switches and transient peaks of gene expression. The peaks and the switches also represent one of the simplest possible and biologically reasonable temporal expression patterns. The particular shapes of the search patterns were adapted to the format of the time series data. In this study, the original data have been upsampled and smoothed using a Gaussian filter, in order to minimize measurement errors for the neighboring time points (see Section 2). Therefore, Gaussian shapes were accepted for the selected input search patterns (peaks and switches) and all their scaling variants (Supplementary Fig. S1). The upsampling step (linear interpolation) changed the number of time points in the dataset by inserting additional time points between the original ones (20:100), thus preventing information loss during the consequent data treatment.

### 3.2 Resolution limits and selection of input time-series data

Developmental gene networks can include hundreds of genes, comprising 1% or more of the entire *Drosophila* genome. These considerations dictate sensitivity limits for the time-duration analysis. To test whether these limits are met, an artificial temporal data set was generated, containing 10 000 random temporal profiles. Coherent signals (peak type) with defined start time and duration time were introduced in some of the profiles from the simulated dataset. In this benchmarking test, the method demonstrated reliable detection of signals, introduced into 100 profiles (1% of total, see Supplementary Fig. S2). It was possible to achieve even better sensitivity when the introduced signals were highly coherent.

Sensitivity of the time-duration analysis also depends on the type of input data. There are several global time-series expression datasets currently available for *Drosophila* embryogenesis. One such dataset was generated using expression arrays (Hooper *et al.*, 2007). It has rather high levels of noise (higher ANOVA  $P$ -values, see Section 2), but contained more time points (30 time points for 24 h

of embryogenesis) than other similar datasets (Biemar *et al.*, 2005). Whole-genome tiling array datasets have much lower noise levels, but include fewer time points (12 time points for 24 h) (J.R. Manak, time-series tiling array expression data, personal communication, 2005). Comparison of the datasets shows that time resolution is critical for time-duration analyses (Supplementary Fig. S3). The unreliable profiles can be removed based on their ANOVA *P*-values, while the information in the missing time points could not be recovered even by the use of data interpolation methods. Side by side comparisons of gene expression profiles from the expression and tiling array data indicate that the tiling arrays capture no more than 1 data point per event (peak) lasting 2–4 h, while the expression array data captures up to 3–4 points for the same event.

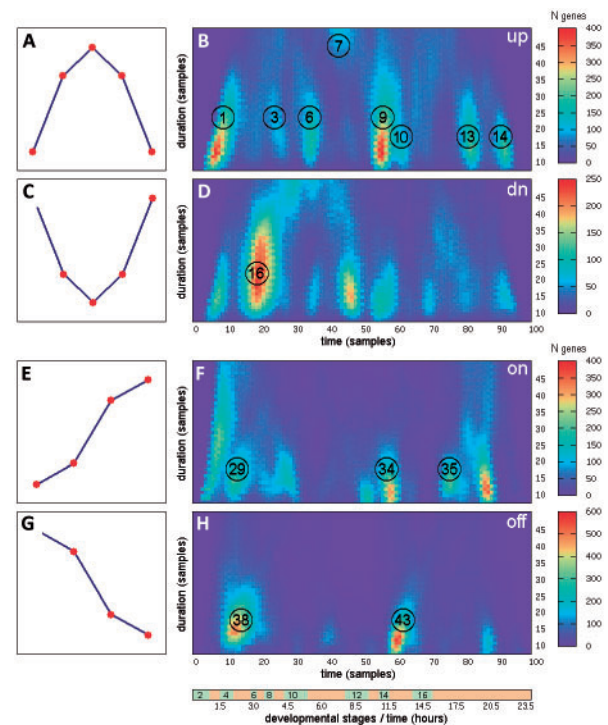
Alternative methods of handling the input data were also investigated. Performance of the time-duration analysis has been tested on both normalized data and smoothed data. Upsampling (linear interpolation, see Section 2) and Gaussian smoothing parameters were selected in order to remove noise, corresponding to point-by-point variation. This noise filtering increased the resolution and revealed more events (clusters) in the expression array data (Supplementary Fig. S3).

### 3.3 Identified clusters reflect temporal sequence of known developmental events

Figure 1 shows the results of the time-duration analysis applied to the normalized and upsampled/smoothed *Drosophila* expression array data (Hooper *et al.*, 2007). Altogether, the four search patterns (peak, switch and their reversed variants) identified 45 distinctive temporal clusters of genes, with the largest clusters containing up to 1300 genes and the smallest clusters containing ~100 genes; the majority of clusters contain 400–500 genes. Certain clusters exhibit extensive overlaps (Supplementary Table S3), and are presumably engaged in the same developmental processes (Supplementary Fig. S5D and E). For this reason, the actual number of the identified events was estimated to be lower (15–20 events) than the total number of distinctive clusters. Out of 7052 profiles in the input dataset, 3128 (44%) were shared by three or more clusters, 1295 (18%) were shared by two clusters, 1239 (17.5%) profiles were not shared (unique) and 1390 (20%) profiles were not present in any cluster.

As expected, the time-duration analysis allowed clear discrimination of slow and fast events (compare clusters #6 and #7 in Fig. 1B). Annotation term enrichment analysis was used to determine the biological function shared by genes in each cluster (van den Berg *et al.*, 2009). Two types of annotations were used: Gene Ontology (GO-slim) (Ashburner *et al.*, 2000) and *Drosophila* CV for morphological parts of the embryo (Tomancak *et al.*, 2007). GO-slim terms associated with genes indicate their biochemical functions, while CV terms identify embryonic tissues in which the genes are expressed (expression patterns).

Most of the gene clusters contain specific classes of enrichment in the GO and CV terms (Supplementary Table S1). A strong agreement has been found between the temporal sequence of known developmental events and the timing of events associated with the identified clusters. For instance, the CV term ‘maternal’ has been enriched in cluster #1 (Fig. 1B), containing genes expressed at the beginning of embryogenesis. CV terms ‘Dorsal ectoderm’ and ‘Pair-rule’ were enriched in cluster #4, with the peak time of expression at



**Fig. 1.** Temporally local clusters identified using two types of patterns. 2D histograms (event maps) are shown for the input patterns ‘peak’ (A and B), ‘switch’ (C and D) and their reverse variants (E and F and G and H correspondingly). The hotspots on the event maps reveal coherent switches of gene activity during embryogenesis (local clusters). Both types of patterns were able to detect shutdown of maternal genes (clusters #1, #16 and #38), differentiation of ectoderm (clusters #6 and #29), synthesis of cuticle (clusters #13 and 35) and other events. Massive shutdown/restart of gene expression occurs early in embryogenesis (stage 4) and at the time of DC (stage 15, clusters #34 and #43); see panels (F and H).

~3.8 h, while GO term ‘cuticle’ and CV terms including ‘epidermis’ were enriched in cluster #13 with a peak of expression at 18 h. The relevant temporal sequence of events predicted by the analysis, along with the enriched annotation terms, suggested that the identified clusters coincide with many of the key developmental processes taking place during fly embryogenesis. Examples of expression profiles and the corresponding gene expression patterns are shown in Supplementary Figure S4. Detailed description of each cluster and lists of genes are available from the UCB online resource.

### 3.4 Time-duration analysis and conventional clustering methods

The results of local time-duration cluster analysis were compared with more conventional clustering methods—hierarchical clustering, *k*-means clustering (Eisen *et al.*, 1998) and Principal Component Analysis (PCA) (Katagiri and Glazebrook, 2009). Conventional clustering was carried out using Cluster 3.0 program with the default parameter settings (de Hoon *et al.*, 2004). Supplementary Figure S6 shows that hierarchical, *k*-means (Cluster 3.0) and local clustering (this work) methods produce similar, but not identical results. Most of the tested methods identify clusters engaged in



common processes, such as maternal genes and cuticle/epidermis genes. In some cases, the different methods produce gene clusters that share >50% of the expression profiles (Supplementary Fig. S5). Annotation enrichment levels were comparable for all three clustering methods using GO and CV annotation (Supplementary Fig. S6). However, a distinctive feature of the local method was its ability to retrieve pure local components of gene expression, matching the search patterns. For example, a major component of cluster #9 (see the arrow in Supplementary Fig. S6C) is also present in clusters produced by the other methods (clusters #1 and #6, Supplementary Fig. S6B), but in these latter clusters contain additional signals as well.

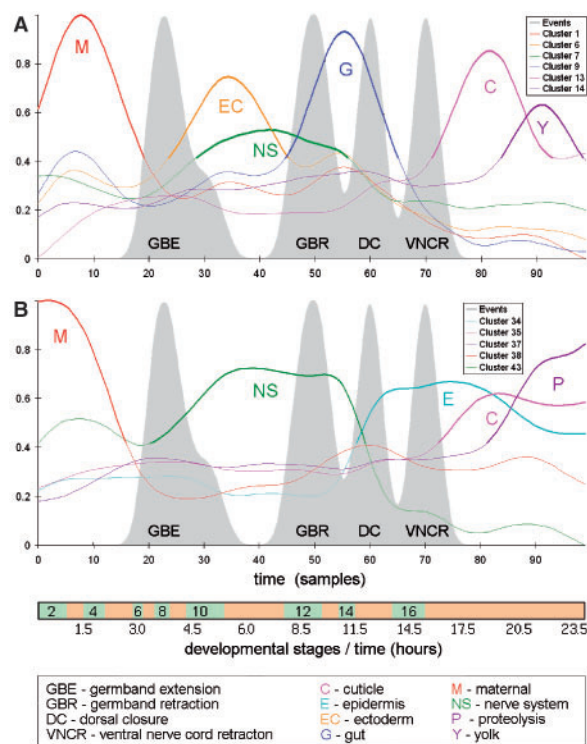
Sensitivity and selectivity of different clustering methods were compared based on their receiver operating characteristics (ROC space, see Supplementary Fig. S7). The local time duration clustering method consistently produced lower false-positive rates and somewhat lower true positive rates than the other two methods. The most striking difference is observed for ‘maternal’ gene clusters (Supplementary Fig. S7E). 97% of all genes in cluster #17 (Supplementary Fig. S6C) are associated with this term, although a large fraction was lost due to the lower sensitivity of this method. Each clustering method produced a list of enriched annotation terms not found or poorly enriched by the other two methods. Some interesting terms, such as ‘pair rule’ (CV, id=262) or ‘neuroblasts of ventral nervous system’ (CV, id=172) were selectively enriched by the local clustering method (Supplementary Table S2).

The results of local clustering were also compared with the results of PCA (Supplementary Fig. S8). Interestingly, the first four principal components (PC1–PC4) were quite similar to some of the local clusters, but the fifth component and additional components are quite distinct (Supplementary Fig. S8E). As in the case with the hierarchical and the *k*-means clustering, PCA failed to retrieve the local signals, such as the signal in the largest cluster #9, produced by local clustering (Supplementary Fig. S6C).

Benchmarking tests based on annotation term enrichment have shown that the local time-duration clustering method performed nearly as well as the conventional clustering methods. A distinctive feature of the local clustering was its ability to retrieve local signals, matching the search patterns, such as the peaks corresponding to clusters #6, #7 or #9. Technical differences among the different clustering methods are summarized in the Supplementary Table S3.

### 3.5 Coherent waves of gene expression often precede major morphogenetic events

Temporal clusters of gene expression might be linked to changes in embryo morphology. The most dramatic morphological changes include gastrulation, germ band extension (GBE), GBR and DC. To investigate the possible connection between these processes and coherent waves of gene expression, the average values of expression for the largest local clusters (average profiles) were superimposed on a timescale of major morphogenetic events (Fig. 2). The most extensive gene clusters do not coincide with morphogenesis, but rather, they occur either before or after these events. For instance, the maternal genes (cluster #1) have a peak of expression before gastrulation and GBE, ectodermal genes (cluster #6) peak between GBE and GBR, and brain-related genes between GBE and DC

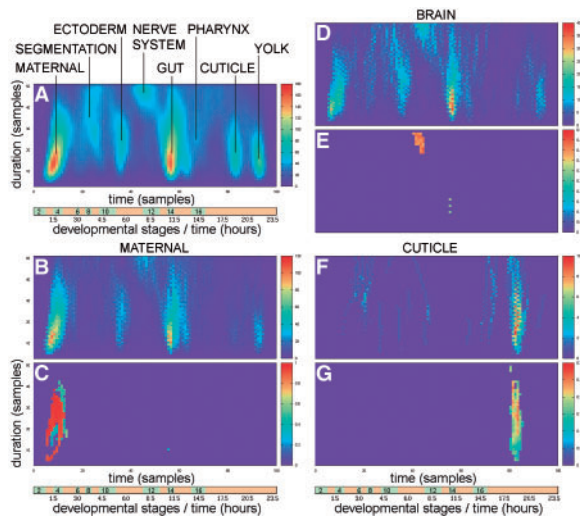


**Fig. 2.** Bursts of gene activity and morphogenetic movements in fly embryo. Figure compares the timing of major switches in the gene activity and the timing of major morphogenetic events. Average expression for each local cluster is shown; the amplitude is proportional to the number of genes in each local cluster. (A) The largest clusters produced by the ‘peak on’ pattern. (B) The largest clusters produced by the ‘switch’ patterns. Morphogenetic movements often separate the ‘bursts’ of gene activity (see clusters #34 and #43 on B, separated by DC).

(cluster #7) (Fig. 2A). Thus, gene expression waves appear to be linked to major morphogenetic events. For instance, GBE separates the maternal wave of expression from the wave associated with the specification of the nervous system (Fig. 2A). Similarly, DC separates the gene cluster associated with the nervous system from those associated with the differentiation of the epidermis (Fig. 2B).

Many of the coherent gene expression waves identified in this study have been observed previously. Thus, Class IIa transcripts reported by Hooper and co-workers (2007) correspond to clusters #7 and #43 in this study. Shutdown of class IIa expression coincides with the activation of classes IIb and IIIa, which is consistent with co-temporal shutdown of the neural genes (cluster #43) and activation of epidermal genes (cluster #34) (Fig. 2B). Classes of transcripts corresponding to local clusters #6 (ectoderm) and #9 (digestive tract) identified in this study were not previously reported.

There is a lag between the timing of coherent expression waves and morphogenetic movements. This lag is consistent with a step-wise mechanism of embryogenesis, whereby tissue specification precedes morphogenetic movements. Specific examples have been extensively documented in the literature, including mesoderm specification followed by invagination of the ventral furrow during gastrulation (Hong *et al.*, 2008). Tissue specification coincides with the deployment of extensive gene networks.



**Fig. 3.** Annotation terms associated with major local clusters. (A) Event map, such as that in the Fig. 1B, obtained for ‘peak on’ input pattern. (B) Event map constructed only for genes associated with annotation term ‘maternal’. (C) Area on the time/duration parameter space enriched for the term ‘maternal’. One can see that maternal genes may have peaks of expression at various time points of embryogenesis, however, only the first peak of expression is significantly enriched for the term ‘maternal’. (D and E) Genes associated with ‘brain’ term may have local peak of expression nearly at any time of embryogenesis (D), but significant enrichment of the term ‘brain’ (D) is observed only for a single cluster (cluster #7, nerve system on panel A). (F and G) Cuticle genes have a single peak of expression (F), which coincides with the area of enrichment for the term ‘cuticle’ (G).

### 3.6 Re-engagement of genes and gene networks in embryogenesis

Some genes have multiple peaks of expression during embryogenesis. Such re-engagement of gene batteries has been previously reported (Goltsev *et al.*, 2009; White *et al.*, 1999). For instance, many ‘maternal’ genes are shut down at ~3 h (stage 6), but are reactivated by 4.5 h (after GBE, stage 10). Some of these ‘maternal’ genes are shut down again at 8 h and activated again at 11 h (after GBR, stage 14). Every time, such ‘maternal’ expression coincides with other waves of gene expression, associated with different developmental events (compare Fig. 3A and B). However, the annotation term ‘maternal’ is significantly enriched only in the earliest peak of expression (Fig. 3C). This suggests that the fraction of ‘maternal’ genes in the later peaks of expression is significantly smaller. Multiple engagements have been detected in the case of genes associated with the term ‘brain’ (Fig. 3D), while the ‘brain’ genes were significantly enriched only in one temporal cluster (cluster #7, Fig. 3E). Altogether, (see more examples in Supplementary Fig. S9) these observations suggest that similar, but not identical gene networks participate in distinct events. Known association of the same CV annotation terms with many genes tested by *in situ* hybridization supports this view (Tomancak *et al.*, 2002, 2007).

There are also examples of gene clusters with a single engagement during embryogenesis. For example, most genes associated with the GO term ‘cuticle’ had only one peak of expression (Fig. 3F) and this peak occurs late in embryogenesis, just before the secretion of the larval cuticle (Fig. 3G). Interestingly, in the mosquito, *A. gambiae*,

these genes are engaged twice in development due to the secretion of a serosal cuticle at earlier stages of development; this cuticle is absent in *Drosophila* (Goltsev *et al.*, 2009).

Examples of genes with multiple peaks of expression were retrieved as profiles shared by clusters #7, #16 (86 genes) and the profiles shared by clusters #6, #9, #16 (30 genes). The profiles shared by the maternal (#16) and the neural (#7) clusters contained genes associated with DNA replication (Supplementary Fig. S10A). Early expression of these genes may correspond to pre-blastoderm mitotic cycles 1–9 and blastoderm cycles 10–13. The pause in expression seen at ~3 h coincides with cellularization. The second peak in expression occurs during the time when mitotic cycles are reinstated during gastrulation. Genes shared by three local clusters (#6, #9, #16, three engagements) were associated with multiple GO/CV terms and had multiple functions (pleiotropic genes). For instance, *Daughterless* (*da*) (Supplementary Fig. S10B) is a class I HLH protein that forms heterodimers with class II HLH proteins to activate transcription (Massari and Murre, 2000). *Daughterless* has both maternal and zygotic functions, including sex determination, neurogenesis, and the growth of imaginal discs (Cline, 1989). Pellino (*pli*) family proteins are regulators of kinase activity, components of the Toll and Toll-like receptor signaling pathways (Schauvliege *et al.*, 2007). In *Drosophila*, Toll-receptor signaling is involved in dorsal-ventral patterning (Rusch and Levine, 1994) and the innate immune response (Aggarwal and Silverman, 2008). The gene *pelota* (*pelo*) is required for the progression of the cell cycle (Adham *et al.*, 2003), and for maintenance and self-renewal of germline stem cells (Xi *et al.*, 2005). The gene *shark* (Supplementary Fig. S10B) is a tyrosine kinase that regulates the c-Jun N-terminal kinase (JNK) pathway (Fernandez *et al.*, 2000). This pathway influences a number of developmental processes, including DC (Igaki, 2009).

### 3.7 Gene networks involved into cell motility and morphogenesis

While the most extensive waves of coherent gene expression were observed in the time intervals separating major morphogenetic events, two waves, containing relatively few genes, were observed at the time of GBE (cluster #3) and DC (cluster #10). Genes expressed during these morphogenetic movements might be expected to play a role in the reorganization of the cytoskeleton and regulation of actin polymerization. Indeed, *cdc42* exhibits peak expression at the time of both GBE and DC (Supplementary Fig. S10C). It encodes a GTPase involved in the regulation of actin filament polymerization. Major morphogenetic events, such as GBE and DC involve large-scale reshaping of cells and cell movements.

Neurotactin (Nrt) and Ecdysone-induced protein 28/29kD (Eip71CD) were among other genes expressed at the time of GBE and DC (Supplementary Fig. S10C). Both genes are expressed in invaginating mesodermal cells (Supplementary Fig. S10E, G and H), while Eip71C is expressed in the dorsal epidermis during DC (Supplementary Fig. S10F). Presumably, Eip71C (Ecdysone-induced protein 28/29kD) is downstream of EcR, a receptor for Ecdysone that regulates many morphogenetic events, including DC, cuticle synthesis, and molting (Chavoshi *et al.*, 2010; Thummel, 1997). Another component of ecdysone signaling, *ecdysoneless* (*ecd*) (Gaziova *et al.*, 2004; Kozlova and Thummel, 2003) also peaks at the time of GBE and DC.

## 4 CONCLUSIONS

The application of time-duration analysis to *Drosophila* embryogenesis identified a link between coherent waves of gene expression and specific morphogenetic processes such as GBE and DC. We also presented evidence that the re-engagement of pleiotropic genes is associated with common features of these different processes, such as reorganization of the cytoskeleton. We suggest that local clustering analysis can be used to supplement conventional clustering methods, such as hierarchical or *k*-means clustering.

## ACKNOWLEDGEMENTS

The authors thank Yury Goltsev for his help with fly and mosquito data sets, Jacques Bothma for critical remarks and help with preparation of the article and members of the Levine lab for stimulating discussions. Time-duration event maps, gene annotations and software are available at [http://flydev.berkeley.edu/cgi-bin/GTEM/emap\\_dm/index\\_dmap.htm](http://flydev.berkeley.edu/cgi-bin/GTEM/emap_dm/index_dmap.htm). The source code can be obtained from DP upon request.

**Funding:** National Institutes of Health (grant GM34431).

**Conflict of Interest:** none declared.

## REFERENCES

- Adams,M.D. et.al. (2000) The genome sequence of *Drosophila melanogaster*. *Science*, **287**, 2185–2195.
- Adham,I.M. et al. (2003) Disruption of the pelota gene causes early embryonic lethality and defects in cell cycle progression. *Mol. Cell Biol.*, **23**, 1470–1476.
- Aggarwal,K. and Silverman,N. (2008) Positive and negative regulation of the *Drosophila* immune response. *BMB Rep.*, **41**, 267–277.
- Arbeitman,M.N. et al. (2002) Gene expression during the life cycle of *Drosophila melanogaster*. *Science*, **297**, 2270–2275.
- Ashburner,M. et al. (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.*, **25**, 25–29.
- Balasubramanian,R. et al. (2005) Clustering of gene expression data using a local shape-based similarity measure. *Bioinformatics*, **21**, 1069–1077.
- Biemar,F. et al. (2005) Spatial regulation of microRNA gene expression in the *Drosophila* embryo. *Proc. Natl Acad. Sci. USA*, **102**, 15907–15911.
- Bolstad,B.M. et al. (2003) A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics*, **19**, 185–193.
- Chavoshi,T.M. et al. (2010) Tissue-autonomous EcR functions are required for concurrent organ morphogenesis in the *Drosophila* embryo. *Mech. Dev.*, **2010**, 20.
- Cline,T.W. (1989) The affairs of daughterless and the promiscuity of developmental regulators. *Cell*, **59**, 231–234.
- de Hoon,M.J. et al. (2004) Open source clustering software. *Bioinformatics*, **20**, 1453–1454.
- Dhollander,T. et al. (2007) Query-driven module discovery in microarray data. *Bioinformatics*, **23**, 2573–2580.
- Eisen,M.B. et al. (1998) Cluster analysis and display of genome-wide expression patterns. *Proc. Natl Acad. Sci. USA*, **95**, 14863–14868.
- Fernandez,R. et al. (2000) The *Drosophila* shark tyrosine kinase is required for embryonic dorsal closure. *Genes Dev.*, **14**, 604–614.
- Gazivola,I. et al. (2004) Cell-autonomous roles of the ecdysoneless gene in *Drosophila* development and oogenesis. *Development*, **131**, 2715–2725.
- Goltsev,Y. et al. (2009) Developmental and evolutionary basis for drought tolerance of the *Anopheles gambiae* embryo. *Dev. Biol.*, **330**, 462–470.
- Hong,J.W. et al. (2008) How the Dorsal gradient works: insights from postgenome technologies. *Proc. Natl Acad. Sci. USA*, **105**, 20072–20076.
- Hooper,S.D. et al. (2007) Identification of tightly regulated groups of genes during *Drosophila melanogaster* embryogenesis. *Mol. Syst. Biol.*, **3**, 72.
- Igaki,T. (2009) Correcting developmental errors by apoptosis: lessons from *Drosophila* JNK signaling. *Apoptosis*, **14**, 1021–1028.
- Katagiri,F. and Glazebrook,J. (2009) Pattern discovery in expression profiling data. *Curr. Protoc. Mol. Biol.*, 22.5.1–22.5.15.
- Kozlova,T. and Thummel,C.S. (2003) Essential roles for ecdysone signaling during *Drosophila* mid-embryonic development. *Science*, **301**, 1911–1914.
- Massari,M.E. and Murre,C. (2000) Helix-loop-helix proteins: regulators of transcription in eucaryotic organisms. *Mol. Cell Biol.*, **20**, 429–440.
- Rusch,J. and Levine,M. (1994) Regulation of the dorsal morphogen by the Toll and torso signaling pathways: a receptor tyrosine kinase selectively masks transcriptional repression. *Genes Dev.*, **8**, 1247–1257.
- Samsonova,A.A. et al. (2007) Prediction of gene expression in embryonic structures of *Drosophila melanogaster*. *PLoS Comput. Biol.*, **3**, e144.
- Sasik,R. et al. (2002) Extracting transcriptional events from temporal gene expression patterns during *Dictyostelium* development. *Bioinformatics*, **18**, 61–66.
- Schauvliege,R. et al. (2007) Pellino proteins: novel players in TLR and IL-1R signalling. *J. Cell Mol. Med.*, **11**, 453–461.
- Stolc,V. et al. (2004) A gene expression map for the euchromatic genome of *Drosophila melanogaster*. *Science*, **306**, 655–660.
- Thummel,C.S. (1997) Dueling orphans–interacting nuclear receptors coordinate *Drosophila* metamorphosis. *Bioessays*, **19**, 669–672.
- Tomanac,P. et al. (2002) Systematic determination of patterns of gene expression during *Drosophila* embryogenesis. *Genome Biol.*, **3**, R0088.
- Tomanac,P. et al. (2007) Global analysis of patterns of gene expression during *Drosophila* embryogenesis. *Genome Biol.*, **8**, R145.
- van den Berg,B.H. et al. (2009) Comparing gene annotation enrichment tools for functional modeling of agricultural microarray data. *BMC Bioinform.*, **10**, S9.
- White,K.P. et al. (1999) Microarray analysis of *Drosophila* development during metamorphosis. *Science*, **286**, 2179–2184.
- Xi,R. et al. (2005) Pelota controls self-renewal of germline stem cells by repressing a Bam-independent differentiation pathway. *Development*, **132**, 5365–5374.