Minireview

Deciphering the genetic code of morphogenesis using functional genomics

Athea Vichas and Jennifer A Zallen

Address: Developmental Biology Program, Sloan-Kettering Institute, 1275 York Avenue, New York, NY 10065, USA.

Correspondence: Jennifer Zallen. Email: zallenj@mskcc.org

Abstract

A fundamental goal of developmental biology is to define the molecular mechanisms that control cell behavior during morphogenesis. A recent study in *BMC Biology* combines gene expression profiling, bioinformatics and functional analysis to identify genes that shape the *Drosophila* embryo.

See related research article http://biomedcentral.com/1741-7007/7/61

Morphogenesis of the *Drosophila* embryo requires the three-dimensional organization of populations of cells with characteristic shapes and behaviors that give rise to the adult structure of the animal. Genetic screens have been instrumental in identifying genes that specify cell fate and cell behavior. Many of these include transcription factors that are expressed in precise spatial patterns and are required for distinct morphological behaviors. However, the downstream pathways that translate cell fate decisions into cell shape and tissue structure are not fully understood. Current advances in genome-wide expression profiling, *in vivo* transcription factor binding site analysis and high-throughput studies of gene-expression patterns have made it possible to identify novel genes that are expressed during morphogenetically active periods of development.

Since the sequencing of the *Drosophila* genome, large-scale expression studies using DNA microarrays have been employed in diverse ways to reveal the genetic networks that drive development. Time-course experiments using DNA microarrays have identified subsets of genes expressed at specific developmental stages throughout *Drosophila* embryogenesis [1-5]. The first few hours of embryogenesis are marked by active changes in cell fate, cell shape and cell rearrangements that are accompanied by dynamic changes in gene expression. Recent studies have determined that more than 2,000 genes, or approximately 15% of the genome, display increased expression at these early stages [3-6]. These systematic genome-wide approaches provide a global perspective on the dynamics of gene expression during embryonic development.

A major challenge now posed by these studies is how to use the overwhelming amount of information from whole-genome studies to identify specific genes with key roles in embryonic morphogenesis. The ability to integrate data from stage-specific gene-expression studies, high-throughput spatial analysis of gene-expression patterns and bioinformatics tools for protein sequence analysis is critical to the process of candidate gene selection (Figure 1). The paper published in *BMC Biology* by Cambiazo and colleagues (Zúñiga *et al.* [7]) describes an integrated genome-wide approach to finding genes that are specifically upregulated during early *Drosophila* development. The authors use differential gene-expression profiling, bioinformatics, *in situ* analysis and functional characterization to identify a novel gene that is required for embryonic morphogenesis, as well as a number of interesting candidate genes for future functional analysis. Here, we discuss how these and other genomic datasets have been coupled with targeted functional analysis to gain insight into the molecular mechanisms underlying morphogenesis.

An integrative approach to identify genes required for morphogenesis

Gastrulation in the early *Drosophila* embryo is characterized by a wholesale reorganization of the embryo driven by region-specific changes in cell shape and cell movement. For these morphogenetic events to be properly executed cells must communicate with each other to coordinate their behavior with other cells and tissues in the embryo. To elucidate the molecular mechanisms underlying these processes, an essential step is to identify the extracellular signals that mediate communication between cells and the transmembrane receptors that detect and interpret these signals and translate them into cell shape and behavior.

Using suppression subtractive hybridization (SSH) and microarray analysis, Zúñiga *et al.* [7] isolated a set of transcripts that are differentially expressed between gastrulating and blastoderm *Drosophila* embryos. The SSH technique is able to circumvent the inability of most high-throughput gene-expression profiling techniques to isolate transcripts of low abundance [8]. Using this method, the authors
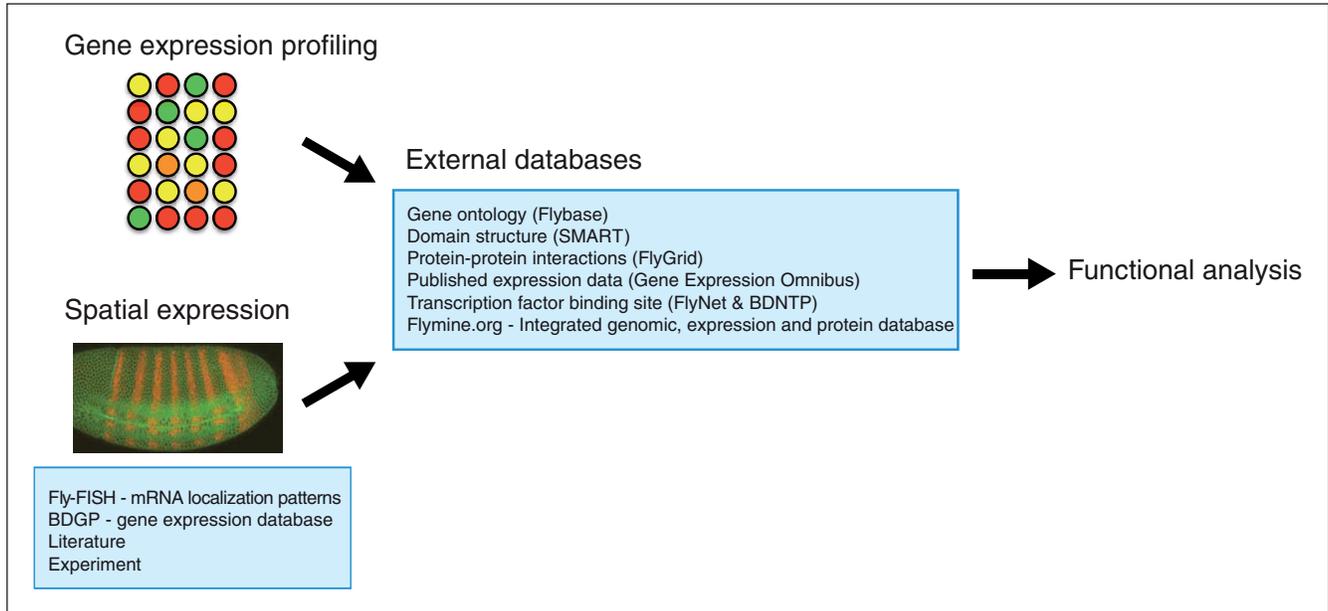
Figure 1

Strategies for selecting candidate genes from expression studies for functional analysis. This schematic depicts various approaches to selecting candidate genes for functional characterization. First, the temporal and spatial expression patterns of genes can be determined by integrating information from experimental and available datasets. Second, genes can be prioritized on the basis of predicted or known protein function. Finally, a subset of candidate genes can be functionally analyzed using genetic analysis, which can be labor intensive, or high-throughput methods such as RNA interference.

identified 114 genes and 4 noncoding RNAs that are more highly expressed in gastrulating embryos compared to the earlier blastoderm stages. To validate their experimental approach, the authors compared their gene list with the approximately 2,000 genes found to be expressed in the early embryo in other genome-wide expression studies [1-6,9]. More than half of the genes identified were previously reported to be expressed during gastrulation. In addition, their list contains 55 genes that have not been functionally characterized. Two of these genes are not represented on commercially available gene-expression chips, indicating that this technique is able to isolate transcripts that were not previously predicted. Of note, 19 of the novel genes are predicted to encode secreted or transmembrane proteins that may play a direct role in cell-cell communication during morphogenesis. Using *in situ* analysis the authors were able to determine that 12 out of 15 of these genes are expressed in a spatially restricted pattern in regions undergoing distinct morphogenetic processes.

This series of studies led to the identification of a novel secreted protein with a putative transmembrane domain that is specifically expressed in the dorsal region of the embryo. Using an RNA interference (RNAi) knock-down strategy, the authors provide evidence for a requirement for this gene during germband retraction, a process that requires regulation of cell shape changes and cell death. Consistent with these findings, this gene is a direct target of Medea, a transcription factor required for dorsal specification in the embryo [10,11]. It will be of interest to determine whether other genes identified in this screen are direct targets of the transcription factors that control morphogenesis during development.

Genome-wide analysis to identify genes downstream of cell fate

Given the extensive information from genetic studies about the transcriptional regulators that direct the early morphogenetic events of embryogenesis, it is now possible to use whole-genome comparative analysis to look for changes in gene expression in response to altering the levels of particular transcription factors. To identify direct targets of these transcription factors, chromatin immunoprecipitation (ChIP) followed by microarray hybridization on whole-genome tiling arrays can be used to determine *in vivo* protein-DNA interactions. Work from the Berkeley *Drosophila* Transcription Network Project (BDTNP) has characterized the *in vivo* DNA binding sites of 21 transcription factors in the *Drosophila* blastoderm embryo [11]. These studies find that each transcription factor is bound to more than 1,000 different sites, suggesting a complex transcriptional program downstream of these regulators. Combining this approach with microarray studies has been successful at

identifying direct targets of transcription factors required for embryonic segmentation [12].

While microarray analysis and ChIP:chip data can provide insight into the temporal expression and transcriptional regulation of genes during development, these approaches do not offer clues as to the spatial distribution of gene expression. Combining *in situ* expression patterns with genomic data on expression levels is one way to identify sets of genes involved in related developmental processes. In *Drosophila,* two groups have determined the spatial and temporal expression patterns of around 25% of the genes expressed in the embryo [2,6]. Lecuyer *et al.* [6], used fluorescent *in situ* hybridization (FISH) to generate high-resolution images for around 2,500 mRNAs throughout early *Drosophila* embryogenesis. This approach, which allows for single-cell resolution, led to the surprising discovery that over 70% of detected mRNAs are subcellularly localized, suggesting that uncharacterized transcripts can be classified on the basis of their localization as well as their expression. Both groups have generated web-based databases that offer a range of search options, including stage, tissue, gene name and predicted function. Analysis of these rich datasets may identify genes involved in morphogenesis that have failed to be identified in genetic screens.

The ultimate goal of developmental biology is to define how each gene contributes to cell fate, cell shape and cell behavior during morphogenesis. However, functional studies are often labor-intensive and are not readily adapted to high-throughput analysis, creating a bottleneck in going from expression to function. RNAi screening is emerging as a powerful technique for functional analysis *in vivo* [13]. Expression profiling has been coupled with double-stranded RNA injections in early embryos to identify genes required for cellularization and embryonic viability [3]. One limitation of traditional genetic screens is the inability to identify genes with subtle or redundant phenotypes, as well as components involved in multiple processes throughout development. The genome sequence creates the potential to identify gene families, which can be tested for redundant functions using combinatorial RNAi [14].

Future directions: the shape of things to come

In the past decade, genome-wide studies have provided insight into the dynamics of gene expression during development [1,4,5]; revealed the spatial localization of mRNA transcripts at the cell and tissue levels [2,6]; and mapped the DNA binding sites of essential transcription factors [10,11]. With this rapid proliferation of data, the current challenge is to develop the tools to translate the information contained in these data sets into meaningful biological insight.

The study by Zúñiga *et al.* [7] highlights the value of integrating information from temporal and spatial expression studies with protein sequence analysis to generate a short list of candidate genes for functional analysis. Moreover, the experimental approach undertaken by Zúñiga *et al.* has provided a dataset enriched for genes with potential regulatory roles in *Drosophila* embryogenesis. Future studies of this kind will help to determine how the expression of cell-fate determinants ultimately leads to the cell-shape changes and cell movements that shape the embryo.

References
1. Arbeitman MN, Furlong EE, Imam F, Johnson E, Null BH, Baker BS, Krasnow MA, Scott MP, Davis RW, White KP: **Gene expression during the life cycle of *Drosophila melanogaster*.** *Science* 2002, **297:**2270-2275.
2. Tomancak P, Beaton A, Weiszmann R, Kwan E, Shu SQ, Lewis SE, Richards S, Ashburner M, Hartenstein V, Celniker SE, Rubin GM: **Systematic determination of patterns of gene expression during *Drosophila* embryogenesis.** *Genome Biol* 2002, **3:**research0088.1-0088.14.
3. Pilot F, Philippe JM, Lemmers C, Chauvin JP, Lecuit T: **Developmental control of nuclear morphogenesis and anchoring by charleston, identified in a functional genomic screen of *Drosophila* cellularisation.** *Development* 2006, **133:**711-723.
4. De Renzis S, Elemento O, Tavazoie S, Wieschaus EF: **Unmasking activation of the zygotic genome using chromosomal deletions in the *Drosophila* embryo.** *PLoS Biol* 2007, **5:**1036-1051.
5. Hooper SD, Boue S, Krause R, Jensen LJ, Mason CE, Ghanim M, White KP, Furlong EE, Bork P: **Identification of tightly regulated groups of genes during *Drosophila melanogaster* embryogenesis.** *Mol Syst Biol* 2007, **3:**72.
6. Lecuyer E, Yoshida H, Parthasarathy N, Alm C, Babak T, Cerovina T, Hughes TR, Tomancak P, Krause HM: **Global analysis of mRNA localization reveals a prominent role in organizing cellular architecture and function.** *Cell* 2007, **131:**174-187.
7. Zúñiga A, Hödar C, Hanna P, Ibáñez F, Moreno P, Pulgar R, Pastenes L, González M, Cambiazo V: **Genes encoding novel secreted and transmembrane proteins are temporally and spatially regulated during *D. melanogaster* embryogenesis.** *BMC Biol* 2009, **7:**61.
8. Cao W, Epstein C, Liu H, DeLoughery C, Ge N, Lin J, Diao R, Cao H, Long F, Zhang X, Chen Y, Wright PS, Busch S, Wenck M, Wong K, Saltzman AG, Tang Z, Liu L, Zilberstein A: **Comparing gene discovery from Affymetrix GeneChip microarrays and Clontech PCR-select cDNA subtraction: a case study.** *BMC Genomics* 2004, **5:**26.
9. Furlong EE, Andersen EC, Null B, White KP, Scott MP: **Patterns of gene expression during *Drosophila* mesoderm development.** *Science* 2001, **293:**1629-1633.
10. Li XY, MacArthur S, Bourgon R, Nix D, Pollard DA, Iyer VN, Hechmer A, Simirenko L, Stapleton M, Luengo Hendriks CL, Chu HC, Ogawa N, Inwood W, Sementchenko V, Beaton A, Weiszmann R, Celniker SE, Knowles DW, Gingeras T, Speed TP, Eisen MB, Biggin MD: **Transcription factors bind thousands of active and inactive regions in the *Drosophila* blastoderm.** *PLoS Biol* 2008, **6:**e27.
11. MacArthur S, Li XY, Li J, Brown JB, Chu HC, Zeng L, Grondona BP, Hechmer A, Simirenko L, Keränen SV, Knowles DW, Stapleton M, Bickel P, Biggin MD, Eisen MB: **Developmental roles of 21 *Drosophila* transcription factors are determined by quantitative differences in binding to an overlapping set**

of thousands of genomic regions. *Genome Biol* 2009, **10:**R80.

12. Liu J, Ghanim M, Xue L, Brown CD, Iossifov I, Angeletti C, Hua S, Nègre N, Ludwig M, Stricker T, Al-Ahmadie HA, Tretiakova M, Camp RL, Perera-Alberto M, Rimm DL, Xu T, Rzhetsky A, White KP: **Analysis of *Drosophila* segmentation network identifies a JNK pathway factor overexpressed in kidney cancer.** *Science* 2009, **323:**1218-1222.

13. Dietzl G, Chen D, Schnorrer F, Su KC, Barinova Y, Fellner M, Gasser B, Kinsey K, Oppel S, Scheiblauer S, Couto A, Marra V, Keleman K, Dickson BJ: **A genome-wide transgenic RNAi library for conditional gene inactivation in *Drosophila*.** *Nature* 2007, **448:**151-156.

14. Schmid A, Schindelholz B, Zinn K: **Combinatorial RNAi: a method for evaluating the functions of gene families in *Drosophila*.** *Trends Neurosci* 2002, **25:**71-74.