Available online at www.sciencedirect.com
ScienceDirect
ELSEVIER
Current Opinion in Genetics & Development

Mapping the gene expression universe

Eric Lécuyer[2] and Pavel Tomancak[1]

Methods to globally survey gene expression provide valuable insights into gene function during development. In particular, comprehensive *in situ* hybridization studies have demonstrated that gene expression patterns are extraordinarily diverse and new imaging techniques have been introduced to capture these patterns with higher resolution at the tissue, cellular, and subcellular levels. The analysis of massive image databases can be greatly facilitated by computer vision techniques once annotated image sets reach the crucial mass sufficient to train the computer in pattern recognition. Ultimately, genome-wide atlases of gene expression during development will record gene activity in living animals with at least cellular resolution and in the context of morphogenetic events. These emerging datasets will lead to great advances in the field of comparative genomics and revolutionize our ability to decipher and model developmental processes for a variety of organisms.

Addresses
[1] Max Planck Institute of Molecular Cell Biology and Genetics, Pfotenhauerstr. 108, Dresden D-01307, Germany
[2] Donnelly Centre for Cellular & Biomolecular Research, University of Toronto, 160 College Street, Toronto, Ontario, Canada M5S 3E1

Corresponding author: Lécuyer, Eric (eric.lecuyer@utoronto.ca) and Tomancak, Pavel (tomancak@mpi-cbg.de)

Current Opinion in Genetics & Development 2008, **18**:506–512

This review comes from a themed issue on
Genomes and evolution
Edited by Sarah Teichmann and Nipam Patel

Available online 20th September 2008

0959-437X/$ – see front matter
© 2008 Elsevier Ltd. All rights reserved.

DOI 10.1016/j.gde.2008.08.003

Introduction

Determining the spatial and temporal expression pattern of a gene is often the first step toward understanding its function during development. This simple notion was perhaps best demonstrated by classical experiments in *Drosophila*, where remarkable embryonic phenotypes, such as the absence of alternating body segments, were linked to genes expressed in corresponding stripes of cells along the anterior–posterior axis of the early embryo [1]. RNA *in situ* hybridization (ISH) is the method of choice for visualizing gene expression patterns in organisms, simply because the technique requires sequence information alone to generate gene-specific probes and is universally applicable to a broad range of model organisms and developmentally interesting tissues [2]. While generally less quantitative than microarrays, the fine spatial resolution of ISH techniques is essential for studying gene-regulatory processes in the proper developmental and cellular context.

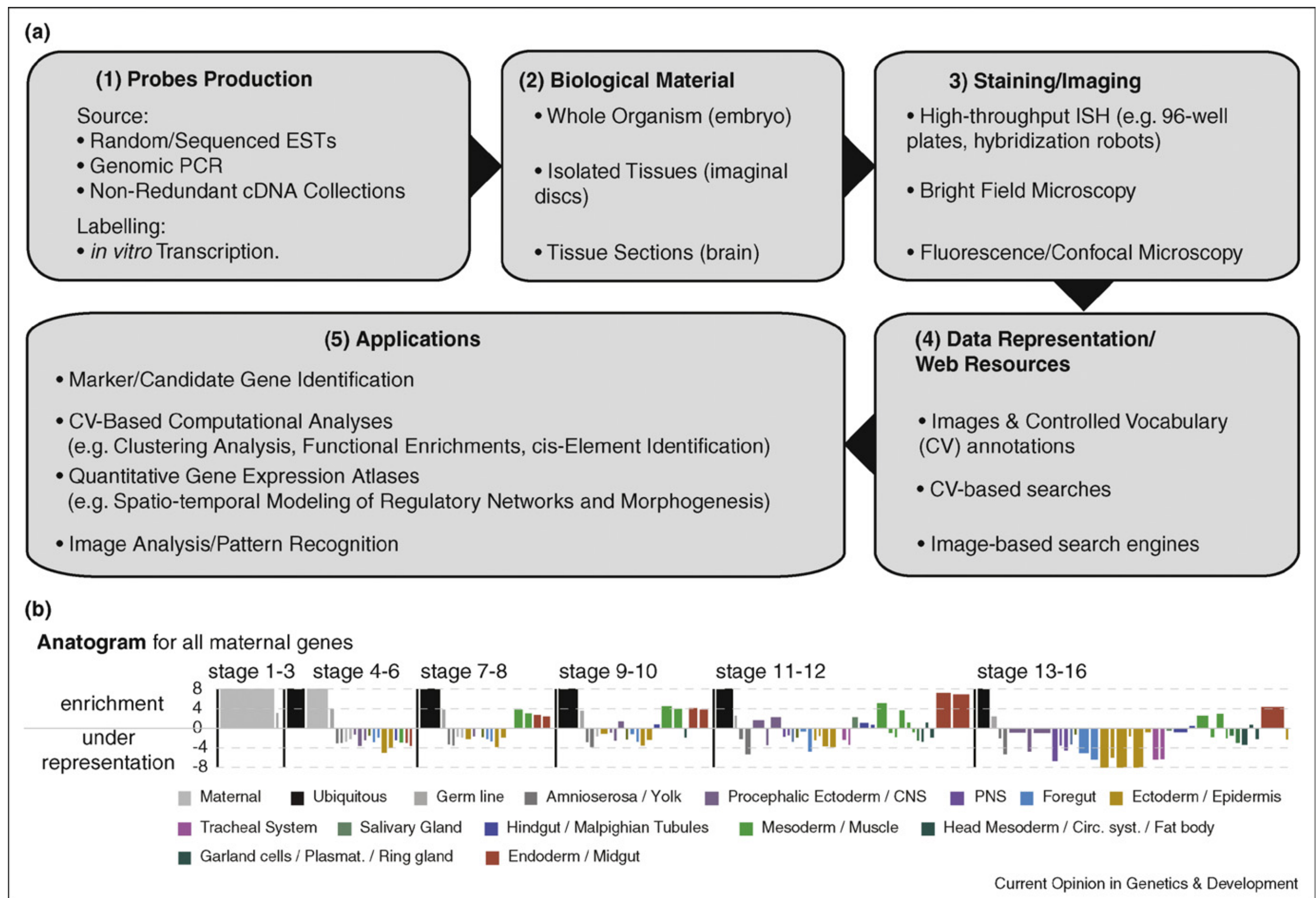
Systematic RNA-ISH screens are currently being performed in all major developmental model systems using gene-specific probes of expressed sequence tags (ESTs) [3•,4–8,9•,10•]. These projects use high-throughput RNA-ISH protocols [11,12] to collect digital microscopy images of stained specimens accompanied with expert-entered annotations, all of which are organized in relational databases (Figure 1a). Project-specific resources are linked to general-purpose model organism databases and, in some cases, also to each other for seamless 'evo–devo' comparisons of the expression patterns of orthologous genes [13].

In this report, we review bioinformatics approaches to the analysis of spatio-temporal gene expression data and contrast them with the emerging, interdisciplinary field of computer vision-assisted expression analysis. We then describe the current trends in high-throughput ISH gene profiling toward more quantitative analysis of expression patterns at the highest attainable level of spatial resolution. We further propose that expression patterns forming in the context of dynamic developmental processes need to be resolved with high-temporal resolution and discuss the 'state-of-the-art' genetic strategies that will enable recording of gene expression patterns in living biological systems.

Organizing gene expression data by annotations and image analysis

Comprehensive ISH screens have many potential applications. In addition to supplying an extensive catalog of candidate genes for use as cellular markers or prioritizing focused studies of gene function, they also offer an unbiased view of how the genome is deployed during development, which can lead to unexpected insights into gene-regulatory mechanisms. In general, two types of data are generated from high-throughput ISH studies: firstly, digital images of stained specimens at specific stages of development and secondly, associated controlled vocabulary (CV) annotations devised by the curators, which provide a standardized description of gene expression characteristics throughout the dataset (Figure 1a). When employed diligently, a CV enables the application of sophisticated clustering algorithms to ISH datasets to group genes together with similar expression patterns and to search for statistically significant enrichments within these groups, such as correlations

Figure 1



(a) Flowchart depicting the steps involved in ISH screening projects, the types of data generated, and the useful applications for the data. **(b)** The 'Anatogram' provides a compact representation of the CV annotations for a group of genes, in this case for genes that encode maternally deposited transcripts in *Drosophila* [3•]. A vertical black line delimits developmental stages, and each colored bar represents an individual CV term. The width of each bar is proportional to the number of times a term was used in the entire dataset, and the height represents the relative enrichment of the given term within the particular gene set (in this case, all 3334 maternally expressed genes versus the entire dataset of 4759 genes expressed in the embryo). Enrichment is given in units of standard deviation above or below the expected sample count based on the background frequencies (*z*-score). Terms with bars below the zero line are under-represented in the sample. The color code is based on organ systems shared by metazoan organisms and can be adapted to spatio-temporal gene expression data from other animals, providing an organism-independent way to present spatial gene expression data.

in gene function or shared regulatory elements [3•,14,15•,16]. For example, a combined analysis of CV annotations and quantitative microarray data was recently utilized to characterize gene expression during *Drosophila* embryogenesis [3•]. Using this approach, the authors were able to distinguish dozens of gene clusters exhibiting either broad or highly restricted expression patterns, most of which showed enrichments for specific gene functions. By utilizing powerful statistical tools they organized the combined gene expression and gene function space into a network representing the relationship between tissue-specific expression regulation and gene function in development. The authors further devised several useful computational tools to visualize the expression properties of groups of genes, such as the 'anatogram' depicted in Figure 1b [3•].

One drawback of CV annotations is that they are established by human annotators, which may incorporate biases into a dataset due to errors in interpretation. Computer-assisted image analysis represents a promising alternative approach for organizing gene expression data independent of CV annotation [17]. At first glance, it may seem that the application of image analysis techniques without thorough reference atlases would be doomed to fail. Yet, machine-learning techniques combined with segmentation and feature extraction algorithms can yield surprisingly consistent results in expression pattern classification from a large collection of expert-annotated images. The atlas of patterns of gene expression during *Drosophila* embryogenesis [3•,14] has been used by several groups to test supervised and unsupervised image analysis approaches to classify *Drosophila* embryo ISH

images. For example, the task-oriented FlyExpress search engine combines manual segmentation and BLAST-like comparison to retrieve similar looking images and represents an annotation-independent gateway into the dataset [18,19]. Others have applied state-of-the-art computer vision approaches to decompose the patterns, group similar patterns together, and annotate based on learned examples [20•,21•,22•]. These sophisticated approaches are yet to be incorporated into publicly available tools for the benefit of biologists. In another study, quantitative microarray time-course data were used in combination with CV annotations [3•] and image analysis to identify syn-expressed genes [23]. Other promising approaches use statistical significance testing against randomized datasets to compare images of patterns and identify similarities (Uwe Ohler and Erwin Frise, unpublished data). Ultimately, the combination of expert annotation and computer vision approaches should provide a useful framework for exploiting rich ISH screen datasets in order to extract novel biological insights.

Image analysis is not limited to fly expression data. In the plant field, confocal images of green fluorescent protein (GFP) reporter constructs were processed to assign gene expression to tissues in flexible *Arabidopsis* roots [24]. Similar nonrigid structures are commonplace in biological applications and the image analysis field is well equipped for their registration as demonstrated by the straightening of *C. elegans* worms to align them to a reference atlas [25]. The fixed, known lineage of *C. elegans* makes it particularly suitable for the application of image analysis techniques and several impressive approaches have been developed to trace cells during early worm development [26–28]. It will be interesting to see these approaches were applied to model systems with less rigid development.

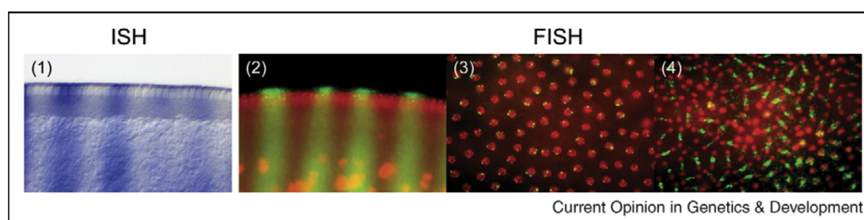
High-resolution approaches

While most ISH screens conducted to date have proven valuable for documenting developmental gene expression at the tissue level, the methods employed offer limited resolution at the cellular and subcellular levels. Achieving single-cell resolution is an important consideration, as many of the processes that guide morphogenesis are manifested at the cellular level through changes in cell fate, morphology, migration, proliferation, and survival [29]. Moreover, gene activity is crucially influenced by subcellular compartmentalization of gene products and techniques with subcellular resolution are required to reveal this dimension of gene regulation. The implementation of higher resolution staining and imaging techniques, with the capacity to unambiguously reveal specimen features at the cellular and subcellular dimensions, thus represents an essential step in the further development of comprehensive gene expression studies.

The potential of enhanced imaging approaches is illustrated by a recent study in which a high-resolution fluorescent ISH (FISH) procedure was used to probe the subcellular localization properties of over 3000 mRNA species during *Drosophila* embryogenesis [15•]. Using a detection method that provides strikingly enhanced subcellular resolution compared to standard ISH detection (Figure 2), the study revealed that the majority of expressed mRNAs are subcellularly localized in fly embryos. The authors further uncovered numerous correlations between transcript distribution patterns and protein localization and function, suggesting that mRNA localization plays a much broader role in organizing cellular protein networks than previously suspected. The accompanying FLY-FISH database provides a useful resource for forming hypotheses about functions of individual genes based on localization phenotypes and for characterizing localization signals by informatics analysis of groups of genes displaying similar localization patterns. ISH screens have been applied to study the subcellular distribution of mRNAs and expression dynamics of other categories of regulatory RNAs in various model systems [30–34], and the use of high-resolution approaches should prove extremely valuable for illuminating novel facets of post-transcriptional regulatory pathways.

Another complexity regarding most ISH procedures is the limitation in the number of genes that can be

Figure 2



Comparison of staining features revealed by standard ISH versus high-resolution FISH. This example shows the patterns observed following hybridization of a probe for *Runt* mRNA in *Drosophila* embryos, detected by ISH (panel 1; mRNA in blue) or by FISH (panels 2–4; mRNA in green, DNA in red). While some subcellular features, such enriched localization in the apical cytoplasm (panels 1–2), are revealed with both detection methods, the high-resolution FISH method provides additional details, including fine spatio-temporal mapping of zygotic mRNA expression in nuclear foci (panel 3) and localization of transcripts at the spindle midzone during anaphase (panel 4).

simultaneously monitored in a given experiment, despite advances in multiplexing methodologies [35,36]. Since gene-regulatory networks generally comprise large numbers of regulators and target genes [37], building predictive models to decipher these networks will require strategies that allow the simultaneous quantitative measurement of all network components. In *Drosophila*, several groups have developed quantitative imaging approaches aimed at understanding the transcriptional networks that drive early embryonic development [38•,39–42]. Recently, the members of the Berkeley *Drosophila* Transcription Network Project developed a powerful imaging and data registration pipeline that provides a quantitative description of gene expression and morphology at cellular resolution in whole embryos [38•,43•,44•]. Their staining methodology (i.e. simultaneous costaining for reference and test genes and nuclei) and data registration in the form of point clouds enables the averaging of data from individually imaged embryos into a composite VirtualEmbryo, in which the expression patterns of many genes can be simultaneously analyzed while taking into account changes in nuclear density [38•]. Ultimately, these strategies will allow for more accurate modeling of spatio-temporal gene expression networks and should lead to great insights into the molecular mechanisms that drive development.

Gene expression patterns in living specimen

Tissue-specific gene expression drives differential cellular behavior during development and conversely morphogenetic events serve as triggers for gene expression regulation [45]. The techniques discussed so far operate on fixed specimens and insights into the temporal interplay between gene regulation and morphogenesis is limited by sparse sampling of developmental stages. Ideally one would like to examine the patterns of gene expression in living specimens with high-temporal resolution. The utility of such live reporters was spectacularly demonstrated in recent studies of Bicoid (Bcd) morphogen gradient formation by time lapse imaging of Bcd-GFP fusions [46•,47•].

In *Drosophila*, transposon-mediated gene trapping is a standard genetic method to randomly generate live gene expression reporters [48,49]. Mobilization of a transposon containing 'exonized' fluorescent reporter genes can be used to create protein fusions suitable for subcellular localization studies [50]. Recently, the efficiency of the technique was increased significantly by combining transposon remobilization with fluorescent-activated embryo sorting [51,52•]. Unfortunately, the strong selection imposed by the sorting step leads to repeated isolation of strongly expressing genes around transposon hot-spots, which limits the spectrum of genes that can be isolated. Still, large-scale gene trap screens led to the identification of many useful markers, such as atypical mononucleated muscle cells around ovarioles in *Drosophila* female gonads [53].

Reverse genetic approaches to mark tissues were pioneered in *C. elegans* by hooking fluorescent reporters to 5′-flanking regions of genes [54•]. An obvious extension of this approach is to insert GFP in the context of large genomic clones that contain all regulatory sequences necessary for wild-type expression. Recently developed high-throughput liquid culture recombineering pipelines enable systematic modification of genomic clones [55•], paving the way for genome-wide gene trapping in nematodes. The high-throughput recombineering approach is applicable to genomic clones derived from any organism [56•]; however, transgenesis of large constructs may prove to be an obstacle in some developmental systems. The current situation is especially favorable in *Drosophila*, where the recent advances in transgenesis technologies allow the insertion of large genomic clones in a site-specific manner, largely reducing complications arising from positional effects [57•]. Together with the extensive atlases documenting gene expression at the tissue, cellular, and subcellular levels, *Drosophila* is ideally positioned to yet again serve as a pioneering model system for genome-wide studies of gene regulation at the highest spatial and temporal resolution.

Conclusions

The common denominator of all ISH projects published or initiated to date is that they are not finished. An argument can be made that they will in fact never be, since the currently used approaches are surpassed by higher resolution techniques. Yet, it is important to complete genome-wide surveys using a single standard approach. Perhaps it should be considered to establish a publishing venue for such wrapping up efforts so that the gene expression atlases may reach the same level of maturity as genome sequences.

Computer vision is a vibrant research field that represents an uncharted territory for most biologists. Advances in image analysis of gene expression data are usually first reported at specialized computer vision conferences and frequently these conceptual prototypes never mature into usable tools for solving biological problems. As the computer vision community has a lot to offer, it will be important to establish venues to foster collaborative interactions with developmental and systems biologists.

Probing gene expression patterns with high spatial and temporal resolution offers a unique opportunity to study the evolutionary mechanisms that lead to the emergence of gene expression innovations and phenotypic variation between closely related species. Once again, *Drosophila* researchers are well equipped to tackle these problems, as the complete genome sequences of 12 related *Drosophila*

species were recently published [58,59] and toolkits for probing gene activity can be seamlessly applied to all *Drosophilids*. It will be exciting to swap entire gene regions between related fly species and assay the impact of *cis*-regulatory sequences and *trans*-acting factors on the expression patterns [60,61].

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Hafen E, Kuroiwa A, Gehring WJ: **Spatial distribution of transcripts from the segmentation gene fushi tarazu during *Drosophila* embryonic development**. *Cell* 1984, **37**:833-841.

2. Wilcox JN: **Fundamental principles of *in situ* hybridization**. *J Histochem Cytochem* 1993, **41**:1725-1733.

3. • Tomancak P, Berman BP, Beaton A, Weiszmann R, Kwan E, Hartenstein V, Celniker SE, Rubin GM: **Global analysis of patterns of gene expression during *Drosophila* embryogenesis**. *Genome Biol* 2007, **8**:R145.
The authors present a detailed analysis of *Drosophila* embryonic gene expression patterns for over 6000 genes based on ISH-derived CV annotations and microarray analysis. They also utilize innovative computational tools to highlight functional enrichments within clusters of coexpressed genes.

4. Pollet N, Muncke N, Verbeek B, Li Y, Fenger U, Delius H, Niehrs C: **An atlas of differential gene expression during early *Xenopus* embryogenesis**. *Mech Dev* 2005, **122**:365-439.

5. Bell GW, Yatskievych TA, Antin PB: **GEISHA, a whole-mount in situ hybridization gene expression screen in chicken embryos**. *Dev Dyn* 2004, **229**:677-687.

6. Quiring R, Wittbrodt B, Henrich T, Ramialison M, Burgtorf C, Lehrach H, Wittbrodt J: **Large-scale expression screening by automated whole-mount in situ hybridization**. *Mech Dev* 2004, **121**:971-976.

7. Thisse B, Heyer V, Lux A, Alunni V, Degrave A, Seiliez I, Kirchner J, Parkhill JP, Thisse C: **Spatial and temporal expression of the zebrafish genome by large-scale in situ hybridization screening**. *Methods Cell Biol* 2004, **77**:505-519.

8. Imai KS, Levine M, Satoh N, Satou Y: **Regulatory blueprint for a chordate embryo**. *Science* 2006, **312**:1183-1187.

9. • Lein ES, Hawrylycz MJ, Ao N, Ayres M, Bensinger A, Bernard A, Boe AF, Boguski MS, Brockway KS, Byrnes EJ *et al.*: **Genome-wide atlas of gene expression in the adult mouse brain**. *Nature* 2007, **445**:168-176.
This paper describes a massive undertaking from researchers at the Allen Institute for Brain Science, in which the entire transcriptome of the mouse brain was revealed by high-throughput ISH analysis of brain sections for all 21 500 mouse genes.

10. • Visel A, Carson J, Oldekamp J, Warnecke M, Jakubcakova V, Zhou X, Shaw CA, Alvarez-Bolado G, Eichele G: **Regulatory pathway analysis by high-throughput in situ hybridization**. *PLoS Genet* 2007, **3**:1867-1883.
The authors describe a robotic ISH platform used to probe gene expression patterns in sectioned mid-gestation mouse embryos. On the basis of elegant computational analysis of CV annotations, they highlight a novel regulatory cascade centered on the Pax6 transcription factor. This work provides an excellent example of the usefulness of annotated ISH data in illuminating novel regulatory pathways.

11. Thisse C, Thisse B: **High-resolution in situ hybridization to whole-mount zebrafish embryos**. *Nat Protoc* 2008, **3**:59-69.

12. Lecuyer E, Parthasarathy N, Krause HM: **Fluorescent in situ hybridization protocols in *Drosophila* embryos and tissues**. *Methods Mol Biol* 2008, **420**:289-302.

13. Haudry Y, Berube H, Letunic I, Weeber PD, Gagneur J, Girardot C, Kapushesky M, Arendt D, Bork P, Brazma A *et al.*: **4DXpress: a database for cross-species expression pattern comparisons**. *Nucleic Acids Res* 2008, **36**:D847-D853.

14. Tomancak P, Beaton A, Weiszmann R, Kwan E, Shu S, Lewis SE, Richards S, Ashburner M, Hartenstein V, Celniker SE *et al.*: **Systematic determination of patterns of gene expression during *Drosophila* embryogenesis**. *Genome Biol* 2002, **3**: RESEARCH0088.

15. • 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.
This manuscript presents a large-scale fluorescent ISH study of mRNA localization for over 3000 *Drosophila* mRNAs. The authors find that the majority of mRNAs are subcellularly localized in patterns that often closely correlate with the distribution of the encoded protein products. The study suggests that mRNA localization plays a broad role in organizing cellular protein modules, and underscores the power of imaging techniques with subcellular resolution.

16. Down TA, Bergman CM, Su J, Hubbard TJ: **Large-scale discovery of promoter motifs in *Drosophila melanogaster***. *PLoS Comput Biol* 2007, **3**:e7.

17. Peng H: **Bioimage informatics: a new area of engineering biology**. *Bioinformatics* 2008, **24**:1827-1836.

18. Gurunathan R, Van Emden B, Panchanathan S, Kumar S: **Identifying spatially similar gene expression patterns in early stage fruit fly embryo images: binary feature versus invariant moment digital representations**. *BMC Bioinform* 2004, **5**:202.

19. Kumar S, Jayaraman K, Panchanathan S, Gurunathan R, Marti-Subirana A, Newfeld SJ: **BEST: a novel computational approach for comparing gene expression patterns from early stages of *Drosophila melanogaster* development**. *Genetics* 2002, **162**:2037-2047.

20. • Peng H, Long F, Zhou J, Leung G, Eisen MB, Myers EW: **Automatic image analysis for gene expression patterns of fly embryos**. *BMC Cell Biol* 2007, **8(Suppl 1)**:S7.
The authors describe an integrative analysis of a subset of the BDGP *Drosophila* embryo ISH database using feature representation of the patterns to cluster genes with similar expression. Coregulated gene sets are used as the basis for transcription factor binding site motif prediction.

21. • Zhou J, Peng H: **Automatic recognition and annotation of gene expression patterns of fly embryos**. *Bioinformatics* 2007, **23**:589-596.
This paper extends work described in Ref. [20•] to automatically annotate fruit fly embryonic ISH images.

22. • Ji S, Sun L, Jin R, Kumar S, Ye J: **Automated annotation of *Drosophila* gene expression patterns using a controlled vocabulary**. *Bioinformatics* 2008, **24**:1881-1888.
This paper introduces an annotation framework relying on the integration of several invariant feature extraction algorithms that are widely used in the computer vision community. The approach takes into account the specific structure of the combined image and annotation dataset and operates within the framework of the FlyExpress resource [18,19], which should enhance its usability.

23. Costa IG, Krause R, Opitz L, Schliep A: **Semi-supervised learning for the identification of *syn*-expressed genes from fused microarray and in situ image data**. *BMC Bioinform* 2007, **8(Suppl 10)**:S3.

24. Mace DL, Lee JY, Twigg RW, Colinas J, Benfey PN, Ohler U: **Quantification of transcription factor expression from *Arabidopsis* images**. *Bioinformatics* 2006, **22**:e323-e331.

25. Peng H, Long F, Liu X, Kim SK, Myers EW: **Straightening *Caenorhabditis elegans* images**. *Bioinformatics* 2008, **24**:234-242.

26. Bao Z, Murray JI, Boyle T, Ooi SL, Sandel MJ, Waterston RH: **Automated cell lineage tracing in *Caenorhabditis elegans***. *Proc Natl Acad Sci U S A* 2006, **103**:2707-2712.

27. Murray JI, Bao Z, Boyle TJ, Boeck ME, Mericle BL, Nicholas TJ, Zhao Z, Sandel MJ, Waterston RH: **Automated analysis of embryonic gene expression with cellular resolution in *C. elegans***. *Nat Meth* 2008, **5**:703-709.

28. Long F, Peng H, Liu X, Kim S, Myers G: **Automatic recognition of cells (ARC) for 3D images of *C. elegans***. In *RECOMB 2008.* Edited by Wong MVaLa.. Berlin, Heidelberg: Springer-Verlag; 2008:128-139.

29. Megason SG, Fraser SE: **Imaging in systems biology**. *Cell* 2007, **130**:784-795.

30. Aboobaker AA, Tomancak P, Patel N, Rubin GM, Lai EC: ***Drosophila* microRNAs exhibit diverse spatial expression patterns during embryonic development**. *Proc Natl Acad Sci U S A* 2005, **102**:18017-18022.

31. Wienholds E, Kloosterman WP, Miska E, Alvarez-Saavedra E, Berezikov E, de Bruijn E, Horvitz HR, Kauppinen S, Plasterk RH: **MicroRNA expression in zebrafish embryonic development**. *Science* 2005, **309**:310-311.

32. Nishikata T, Yamada L, Mochizuki Y, Satou Y, Shin-i T, Kohara Y, Satoh N: **Profiles of maternally expressed genes in fertilized eggs of *Ciona intestinalis***. *Dev Biol* 2001, **238**:315-331.

33. Ason B, Darnell DK, Wittbrodt B, Berezikov E, Kloosterman WP, Wittbrodt J, Antin PB, Plasterk RH: **Differences in vertebrate microRNA expression**. *Proc Natl Acad Sci U S A* 2006, **103**:14385-14389.

34. Kingsley EP, Chan XY, Duan Y, Lambert JD: **Widespread RNA segregation in a spiralian embryo**. *Evol Dev* 2007, **9**:527-539.

35. Levsky JM, Shenoy SM, Pezo RC, Singer RH: **Single-cell gene expression profiling**. *Science* 2002, **297**:836-840.

36. Kosman D, Mizutani CM, Lemons D, Cox WG, McGinnis W, Bier E: **Multiplex detection of RNA expression in *Drosophila* embryos**. *Science* 2004, **305**:846.

37. Stathopoulos A, Levine M: **Genomic regulatory networks and animal development**. *Dev Cell* 2005, **9**:449-462.

38. • Fowlkes CC, Hendriks CL, Keranen SV, Weber GH, Rubel O, Huang MY, Chatoor S, DePace AH, Simirenko L, Henriquez C *et al.*: **A quantitative spatiotemporal atlas of gene expression in the *Drosophila* blastoderm**. *Cell* 2008, **133**:364-374.

This study utilizes the quantitative whole-embryo imaging strategies described in Refs. [43•,44•] to average expression data from many genes into a composite virtual embryo. This analysis pipeline provides a framework for the application of virtual multiplexing approaches that should prove extremely useful in predicting gene-regulatory interactions.

39. Surkova S, Kosman D, Kozlov K, Manu, Myasnikova E, Samsonova AA, Spirov A, Vanario-Alonso CE, Samsonova M, Reinitz J: **Characterization of the *Drosophila* segment determination morphome**. *Dev Biol* 2008, **313**:844-862.

40. Zinzen RP, Senger K, Levine M, Papatsenko D: **Computational models for neurogenic gene expression in the *Drosophila* embryo**. *Curr Biol* 2006, **16**:1358-1365.

41. Janssens H, Hou S, Jaeger J, Kim AR, Myasnikova E, Sharp D, Reinitz J: **Quantitative and predictive model of transcriptional control of the *Drosophila melanogaster* even skipped gene**. *Nat Genet* 2006, **38**:1159-1165.

42. Jaeger J, Surkova S, Blagov M, Janssens H, Kosman D, Kozlov KN, Manu, Myasnikova E, Vanario-Alonso CE, Samsonova M *et al.*: **Dynamic control of positional information in the early *Drosophila* embryo**. *Nature* 2004, **430**:368-371.

43. • Keranen SV, Fowlkes CC, Luengo Hendriks CL, Sudar D, Knowles DW, Malik J, Biggin MD: **Three-dimensional morphology and gene expression in the *Drosophila* blastoderm at cellular resolution II: dynamics**. *Genome Biol* 2006, **7**:R124.

See annotation to [38•].

44. • Luengo Hendriks CL, Keranen SV, Fowlkes CC, Simirenko L, Weber GH, DePace AH, Henriquez C, Kaszuba DW, Hamann B, Eisen MB *et al.*: **Three-dimensional morphology and gene expression in the *Drosophila* blastoderm at cellular resolution I: data acquisition pipeline**. *Genome Biol* 2006, **7**:R123.

See annotation to [38•].

45. Supatto W, Debarre D, Moulia B, Brouzes E, Martin JL, Farge E, Beaurepaire E: **In vivo modulation of morphogenetic movements in *Drosophila* embryos with femtosecond laser pulses**. *Proc Natl Acad Sci U S A* 2005, **102**:1047-1052.

46. • Gregor T, Tank DW, Wieschaus EF, Bialek W: **Probing the limits to positional information**. *Cell* 2007, **130**:153-164.

Along with Ref. [47•], this report uses live imaging of Bcd-GFP fusion proteins to study the fine-scale dynamics of morphogen gradient formation in *Drosophila* embryos, revealing a crucial contribution of nuclear accumulation dynamics in the establishment of precise and reproducible gradients.

47. • Gregor T, Wieschaus EF, McGregor AP, Bialek W, Tank DW: **Stability and nuclear dynamics of the bicoid morphogen gradient**. *Cell* 2007, **130**:141-152.

See annotation to [46•].

48. Morin X, Daneman R, Zavortink M, Chia W: **A protein trap strategy to detect GFP-tagged proteins expressed from their endogenous loci in *Drosophila***. *Proc Natl Acad Sci U S A* 2001, **98**:15050-15055.

49. Clyne PJ, Brotman JS, Sweeney ST, Davis G: **Green fluorescent protein tagging *Drosophila* proteins at their native genomic loci with small P elements**. *Genetics* 2003, **165**:1433-1441.

50. Kelso RJ, Buszczak M, Quinones AT, Castiblanco C, Mazzalupo S, Cooley L: **Flytrap, a database documenting a GFP protein-trap insertion screen in *Drosophila melanogaster***. *Nucleic Acids Res* 2004, **32**:D418-D420.

51. Quinones-Coello AT, Petrella LN, Ayers K, Melillo A, Mazzalupo S, Hudson AM, Wang S, Castiblanco C, Buszczak M, Hoskins RA *et al.*: **Exploring strategies for protein trapping in *Drosophila***. *Genetics* 2007, **175**:1089-1104.

52. • Buszczak M, Paterno S, Lighthouse D, Bachman J, Planck J, Owen S, Skora AD, Nystul TG, Ohlstein B, Allen A *et al.*: **The carnegie protein trap library: a versatile tool for *Drosophila* developmental studies**. *Genetics* 2007, **175**:1505-1531.

Through mobilization of GFP-marked transposons in *Drosophila*, the authors established a collection of protein and enhancer trap transgenic lines labeling a variety of cell lineages and cellular structures. The collection represents a valuable resource for studying the developmental dynamics of gene/protein expression in live animals.

53. Hudson AM, Petrella LN, Tanaka AJ, Cooley L: **Mononuclear muscle cells in *Drosophila* ovaries revealed by GFP protein traps**. *Dev Biol* 2008, **314**:329-340.

54. • Hunt-Newbury R, Viveiros R, Johnsen R, Mah A, Anastas D, Fang L, Halfnight E, Lee D, Lin J, Lorch A *et al.*: **High-throughput in vivo analysis of gene expression in *Caenorhabditis elegans***. *PLoS Biol* 2007, **5**:e237.

The authors used PCR stitching of putative 5′-regulatory regions to create GFP reporter transgenic lines for 1886 genes in *C. elegans*. The resulting expression data, encompassing images and pattern annotations, were organized in a public database (http://gfpweb.aecom.yu.edu/index).

55. • Sarov M, Schneider S, Pozniakovski A, Roguev A, Ernst S, Zhang Y, Hyman AA, Stewart AF: **A recombineering pipeline for functional genomics applied to *Caenorhabditis elegans***. *Nat Methods* 2006, **3**:839-844.

This paper describes a powerful recombineering pipeline allowing for systematic manipulation of *C. briggsae* genomic bacterial artificial chromosome (BAC) clones in order to generate tagged transgenes in high-throughput fashion. Variations of this method are now being applied to other model systems, as described in Ref. [56•].

56. • Poser I, Sarov M, Hutchins JR, Heriche JK, Toyoda Y, Pozniakovsky A, Weigl D, Nitzsche A, Hegemann B, Bird AW *et al.*: **BAC TransgeneOmics: a high-throughput method for exploration of protein function in mammals**. *Nat Methods* 2008, **5**:409-415.

See annotation to [55•].

57. • Venken KJ, He Y, Hoskins RA, Bellen HJ: **P[acman]: a BAC transgenic platform for targeted insertion of large DNA fragments in *D. melanogaster***. *Science* 2006, **314**:1747-1751.

A breakthrough study demonstrating an efficient site-specific transgenesis method for large genomic clones in *Drosophila* melanogaster. The authors introduce the *Drosophila* research community to powerful recombineering approaches for the manipulation of large transgenes.

58. Stark A, Lin MF, Kheradpour P, Pedersen JS, Parts L, Carlson JW, Crosby MA, Rasmussen MD, Roy S, Deoras AN *et al.*: **Discovery of functional elements in 12 *Drosophila* genomes using evolutionary signatures**. *Nature* 2007, **450**:219-232.

59. Clark AG, Eisen MB, Smith DR, Bergman CM, Oliver B, Markow TA, Kaufman TC, Kellis M, Gelbart W, Iyer VN *et al.*: **Evolution of genes and genomes on the *Drosophila* phylogeny**. *Nature* 2007, **450**:203-218.

60. Prud'homme B, Gompel N, Rokas A, Kassner VA, Williams TM, Yeh SD, True JR, Carroll SB: **Repeated morphological evolution through *cis*-regulatory changes in a pleiotropic gene**. *Nature* 2006, **440**:1050-1053.

61. Gompel N, Prud'homme B, Wittkopp PJ, Kassner VA, Carroll SB: **Chance caught on the wing: *cis*-regulatory evolution and the origin of pigment patterns in *Drosophila***. *Nature* 2005, **433**:481-487.