Topic Introduction

Three-Dimensional Morphology and Gene Expression Mapping for the *Drosophila* Blastoderm

David W. Knowles

To properly understand the transcriptional network of animals, we must have full quantitative comprehension of the spatial and temporal expression patterns of transcription factors and their targets. Visual inspection of embryos stained to reveal the patterns of genes shows levels of expression that change from cell to cell in a complex manner. With our current wealth of knowledge regarding the basic biology of animal genomes and the components of their transcriptional regulatory networks, combined with current technologies in optical microscopy, computing, and image and vision analysis, we should be able to capture quantitative, three-dimensional (3D) information about the transcriptional network (all factors and targets) for an entire animal at cellular resolution. It should also be possible to assemble these data into a single computationally analyzable database—an atlas—that could be the basis for uncovering new biology governing regulatory gene networks. This article describes progress toward realizing these goals, with the focus on *Drosophila melanogaster*. It describes a suite of high-throughput methods that have been used to create the first quantitative 3D description of gene expression and morphology at cellular resolution in a whole animal, and it presents some of the new biology that has been revealed by this quantitative atlas.

Cold Spring Harbor Protocols
www.cshprotocols.org

INTRODUCTION

This article describes the efforts of the Berkeley *Drosophila* Transcription Network Project (BDTNP) (http://bdtnp.lbl.gov) to produce a computationally analyzable atlas of morphology and gene expression in a whole blastoderm embryo at cellular resolution. This work has been a multidisciplinary collaboration involving expertise in biology, imaging, image analysis, mathematics, computer science, vision, visualization, and database construction, and so it is not feasible to create a practical manual of all these areas in one article. However, the methods and references used in the project are described in detail in the BDTNP publications. The first part of this article describes the thinking required to develop the strategy for this project, and the second part briefly outlines the techniques that enabled the atlas to be produced and some of the consequent biological discoveries. The aim is to promote understanding and to inspire readers to think of the many endeavors that become possible with comprehensive morphology and gene expression data of an entire animal.

THINKING ABOUT A STRATEGY

Looking at Gene Expression in *Drosophila* Embryos

If you walk down Bungtown Road to the sandbar at Cold Spring Harbor, you will be fascinated by the prehistoric-looking horseshoe crabs found there on the beach. But, for now, imagine you are in the

Adapted from *Imaging in Developmental Biology* (ed. Sharpe and Wong). CSHL Press, Cold Spring Harbor, NY, USA, 2011.

Cite this article as *Cold Spring Harbor Protoc*; 2012; doi:10.1101/pdb.top067843

laboratory peering down a microscope at an early-stage embryo of another arthropod, *Drosophila melanogaster*, one of the most studied model organisms. Your goal is to better understand how the genome orchestrates transcriptional control during morphogenesis to produce such incredible biological structures. This has been a long-standing problem that will continue to baffle and delight us. The transcriptional regulatory network in each animal is composed of thousands of different genetically controlled components, many of which interact directly and/or indirectly with one another with varying binding strengths. Such networks give rise to intricate patterns of expression that vary in space and time and direct morphogenesis.

Drosophila embryos are ~500-µm long, which is an ideal size for high-resolution fluorescence imaging. Moreover, this animal comes with >100 years of research data, including extensive work on the genetic basis of morphogenesis (for which the Nobel Prize was awarded in 1995) and, more recently, the entire genetic sequence (Adams et al. 2000). In our imaginary scene, the embryo, which is termed a blastoderm before gastrulation, has been fluorescently stained for one or two of the many genes that are involved in *Drosophila* development. The blastoderm, which exists for the first 3 h of embryogenesis, is a multinucleated syncytium that comprises a central yolk surrounded by a single layer of some 6000 nuclei (Campos-Ortega et al. 1997). The single layer of cells means that the gene expression patterning (which controls the body plan) is on the embryo surface, and this is why the patterning is so evident as you look under the microscope. The transcriptional regulatory network is also relatively simple, comprising only ~50 transcription factors, which control ~1000 target genes (Lawrence 1992), most of which have been well characterized.

Switching to phase-contrast microscopy, you see that the embryo is in the process of cellularizing. Cell membranes are invaginating from the surface of the embryo, on the apical sides of the nuclei, and growing toward the basal side through the cytoplasm that separates the newly forming cells. Closer observation of several embryos at different orientations reveals that cellularization is more complete on the ventral side than on the dorsal side. Cellularization occurs through stage 5 and does not quite complete before the embryo starts to gastrulate. You also notice that nuclei are more spherical at the beginning of cellularization and elongate during stage 5 with their long axis aligning with the normal vector of the embryo surface.

Perhaps the pattern being viewed under fluorescence is one of the maternally expressed genes such as *bicoid*, or maybe it is one of the later, more intricate patterns expressed by the developing embryo along either its anteroposterior (A–P) or dorsoventral (D–V) axis. What is striking is how clever the underlying transcriptional regulatory information encoded in animal genomes must be for such exquisitely varying patterns to be created. Because the patterning is on the surface, a simple adjustment of the focus is all that is needed to clearly see the rise and fall of expression levels along the A–P axis. It becomes clear that this complexity can only occur if expression levels vary in precisely determined, quantitative ways from one cell to another. Furthermore, although these cell-to-cell expression variations are clearly observed along the A–P axis, they are difficult to appreciate around the D–V axis. This is caused by a combination of the geometry of the embryo, the limited depth of focus of high-numerical-aperture lenses, and the difficulty of perceiving patterns in the third or optical *z* dimension. Nevertheless, there is no reason to think that expression level variations around the D–V axis are any less complex.

You then lower the magnification by switching objectives, or you view the embryo under a binocular dissecting fluorescence microscope. The longer depth of focus created by the lower numerical aperture of the lens (i.e., a simple geometric pinhole effect) allows the 3D nature of the pattern, which envelops the embryo, to be revealed. And although the numerical aperture and resolving power at this magnification are lower, you can still appreciate the cellular level of the patterning. To overcome the limited perception around the D–V axis, you imagine inventing a rotisserie system for the embryo or a potato peeler, whereby the embryo could be skinned and the skin laid flat for better observation. The embryo could perhaps be squashed and analyzed on one side, but that would perturb the delicate system being studied and make it difficult to determine the D–V orientation. You wish for a better way than either skinning or squashing to quantitatively examine this beautifully patterned developmental canvas of life.

Pattern Dynamics: Looking at Fixed versus Live Embryos

Gene expression pattern formation is dynamic. Broadly expressed maternal proteins initiate the patterning of increasingly complex target gene expression. The earliest patterns in the system can be divided into those having dominantly A–P or D–V dependence. These early A–P and D–V patterns act largely independently at first, but at some stage, their interaction is needed to create later patterns of greater complexity that have both A–P and D–V dependences. How do these regulators interact, and what is the shape of the network?

Pattern dynamics result from individual nuclei that change their expression level with time. Are these nuclei uniformly distributed on the embryo surface, or are they also dynamic like the pattern? Nuclear spatial dynamics would certainly impact the underlying mechanism of pattern formation. It was long assumed that nuclei were uniformly distributed at the periphery of the embryo during stage 5, although they underwent great dynamics during earlier stages when they migrated to the embryo surface. Also, as soon as the embryo gastrulates, the nuclei are neither stationary nor distributed uniformly. So there is really no evidence to suggest that nuclei should be uniformly distributed during stage 5. To understand embryonic pattern formation, it is likely important to map both gene expression levels and nuclear position through time. In fact, it would not be surprising if there were significant rearrangements of nuclear positions that correlated spatially with imminent gastrulation events.

But why are we looking at fixed embryos when, for studying dynamics, live animals would certainly be better? A few calculations quickly reveal that live animals bring other complexities. First, the yolk is opaque because it is granulated and optically similar to a bag of glass marbles in air. Fixed embryos are carefully optically cleared by a mounting medium that makes the central yolk transparent, in much the same way as those marbles would become clear by mounting them in molten glass. The mountant also has a refractive index that matches that of the cover glass, and this removes the possibility of spherical aberration (caused by refractive index mismatch at the optical interface of the cover glass and biology), which increasingly obscures image formation with acquisition depth. Having a single refractive index throughout the object allows you to see clearly all the way through the embryo. Second, stage 5 only lasts for a little <1 h. To properly capture the pattern dynamics, each image must be acquired within 1 min or so. This is a problem because the acquisition time for a high-resolution 3D image of an entire blastoderm takes almost an order of magnitude longer than this with a laser-scanning microscope. Finally, it is far easier to create probes that can be hybridized to any product of the 1000 genes of interest than it is to create live animals that express fluorescent proteins for 1000 genes. In light of these points, fixed embryos provide an excellent starting point for developing an image-based strategy to capture the morphology and the gene expression patterns of the developing embryo.

In summary, assessment of the imaginary scenario described above allows you to make the following notes and observations regarding the development of a strategy for mapping 3D morphology and gene expression in *Drosophila* embryos.

1. The size of a *Drosophila* embryo makes it well suited for microscopy because it fits into a single field of view.
2. The blastoderm is a syncytium in which the expression patterning that controls the body plan is completely revealed on the embryo surface.
3. Only ~50 transcription factors and 1000 genes are thought to be involved in the transcriptional network of the blastoderm.
4. Cellularization begins during stage 5, the period before the onset of gastrulation. Cell membranes grow from the embryo surface inward, and nuclei become elongated.
5. Complex patterning requires expression level variations between individual cells. These variations are precisely controlled over the entire animal.
6. Variations of expression are clearly seen along the A–P axis. It is difficult to see around the D–V axis, but patterning in this direction is probably just as complex.

Cite this article as *Cold Spring Harbor Protoc*; 2012; doi:10.1101/pdb.top067843

7. To track expression pattern dynamics, a precise temporal staging system is required.
8. How do the early A–P and D–V expression patterns interact?
9. What is the role of nuclear position on the regulatory canvas? Are the nuclei distributed uniformly through stage 5? Do the nuclei reorganize in preparation for gastrulation?
10. Live animals would be ideal, but they complicate the imaging. Initially, let us use fixed embryos.

High-Throughput Imaging and Analysis to Create a Computationally Analyzable Gene Expression Atlas

High-throughput biology has become commonplace in many areas, including proteomics, genomics, and drug discovery, in which the research involves screening and analysis of repetitious biological experiments. Fluorescence microscopy is a well-established tool in biology, but increasingly sophisticated applications are replacing point-and-click experiments in which a single representative image is used to elucidate a biological question. Current technologies in microscopy and computing allow new kinds of experiments to be envisioned and to be performed, and these require automated image acquisition, large repositories for data management and storage, and fast automated image-analysis and quantitative visualization techniques.

For the *Drosophila* blastoderm atlas, the goal is to produce a computationally analyzable record of the pattern of protein expression of ~50 transcription factors and the mRNA expression of 1000 genes, which together encompass the primary patterning regulators and their targets. Although it is relatively straightforward to produce probes that are specific for any given gene, current staining and imaging technologies do not allow 1000 different genes to be stained and to be imaged in a single embryo. Instead, several thousand embryo images will be required if we are to capture all of the patterns of interest in a single atlas. To enable such a large project, automation must be developed for acquiring and analyzing the images, and a database will be needed to track the associated metadata. Embryos should be imaged at as high a numerical aperture as possible, but the lens must have sufficient working distance to allow the whole embryo to be imaged. Each image will have to be analyzed to remove the inherent image acquisition artifacts and to quantify and to tabulate the gene expression and morphology at cellular resolution to produce computationally analyzable data sets. This analysis will condense the information from a large 3D multichannel image into a much smaller point cloud—a table of the nuclear coordinates and the relative expression per nucleus for the genes imaged (Luengo Hendriks et al. 2006).

To create the atlas, the point clouds of multiple embryos will need to be registered together onto an average morphological framework. This framework will allow new patterns to be added, and the atlas can be built in a sequential manner. Image registration is a field of research in computer science and vision. As microprocessor power increases, new computational techniques are being developed to automatically find objects of interest in an image and to track them over multiple images. One of the goals of such work is the identification of unique fiducial points in two images that correspond to the same object. Here, we will apply image registration techniques to the point cloud data rather than try to register the large raw image files. Embryo point cloud registration is complicated by the simplicity of the blastoderm morphology, which limits the availability of fiducial marks needed to coordinate reciprocal positions between different embryos. Point clouds can be coarsely registered by aligning the embryo's A–P and D–V axes, as well as by registering their surfaces, but this coarse registration is only the first step and does not provide the final accuracy needed. Different embryos at the same stage differ in size, shape, and in the relative location of expression patterns (Luengo Hendriks et al. 2006; Keränen et al. 2006). At late stage 5, for example, the maxima of pair-rule stripes are only two or three cells wide, and the relative positions of these stripes vary slightly between different embryos. Coarsely registered embryos would average these biological variations by incorrectly blurring the pair-rule stripe patterns over too many nuclei and thereby would produce representations of patterns that are incorrect. To circumvent this, expression patterns need to be registered on a nucleus-by-nucleus basis. Even so, it would be difficult to uniquely identify nuclei based on

morphology alone. However, if each embryo is stained for a common gene that has a complex pattern of expression, then many nuclei could be uniquely identified based on their relative position to the A–P and D–V axes and their relative expression level. These fiducial marks could then be used to establish one-to-one correspondences between all the nuclei of two embryos to produce pairwise averages that do not blur the expression patterns (Fowlkes et al. 2008).

The imaging and registration strategy now becomes clear. To allow nuclear segmentation, each embryo must be stained and imaged for fluorescent markers of total DNA. One common marker gene is needed to aid registration, and other expression patterns that are imaged can be registered to the atlas. Embryo point clouds are first coarsely registered to the framework based on the A–P and D–V axes and then are finely registered based on a set of fiducial marks defined by the coordinates of individual nuclei. For practical reasons, based on the number of acquisition channels of the microscope and the complexity of hybridizing probes to an embryo, three-channel imaging provides a simple solution. This means that one fluorescent channel will be used to capture new expression patterns, which will be added to the atlas. Finally, the atlas will be complex and 3D, and we will need a sophisticated visualization tool to allow the data to be quantitatively explored. This strategy then defines an imaging and acquisition pipeline that involves the following steps.

1. Staining, mounting, and imaging. The procedures involved include construction of a cDNA probe, in situ hybridization, tyramide signal amplification, immunohistochemistry, and other techniques for fluorescently labeling mRNA, protein, and nuclear DNA in whole embryos. Imaging is accomplished with a two-photon laser-scanning microscope, and a managed data repository is established for the image and its associated metadata.
2. Temporal staging.
3. Conversion of images to point clouds. Image analysis is achieved by 3D nuclear and cellular segmentation, and extraction analysis algorithms are used to turn each embryo image into a computationally analyzable data point cloud. Computers with sufficient capability are required.
4. Registration of point clouds into a virtual embryo. Three-dimensional image registration methods are used to register multiple point clouds at cellular resolution and thereby create a virtual embryo atlas of morphology and gene expression.
5. PointCloudXplore. This is a visualization tool for viewing and analyzing the atlas.

Details for many of the methods described here are provided in Luengo Hendriks et al. 2006. All the software tools that were developed to produce the BDTNP atlas are freely available.

THE PIPELINE

Staining, Mounting, and Imaging

Some of the crucial steps in acquiring images are proper staining and mounting. Poor images result from an inadequate interface between the disciplines of biology and imaging; for example, the staining has not worked well, the autofluorescence is overwhelming, and/or the mounting is not optimal. Often, many experiments are needed to determine the proper staining and mounting conditions for a particular application. For high-resolution optical imaging, attention to detail is required down to the scale of 0.5 μm. This dimension is that of the fluorescence emission wavelength, which determines the diffraction-limited resolution of the optical system.

In the BDTNP pipeline, RNA probes for specific genes can be visualized with tyramide signal amplification reactions using many fluorescent dyes (e.g., Cy3 and coumarin), and total DNA can be stained with SYTOX Green (Luengo Hendriks et al. 2006). These dyes have well-separated emission spectra and are all excited by two-photon absorption of a single wavelength, which allows three-channel images to be acquired simultaneously. The stained embryos can then be mounted in DePex, a xylene-based plastic (Electron Microscopy Sciences, Hatfield, PA). This mountant has the

 Cite this article as *Cold Spring Harbor Protoc;* 2012; doi:10.1101/pdb.top067843

advantage of creating permanent solid slides that protect the fluorophores from oxygen, thereby making the samples resistant to photobleaching. The mountant also optically clears the embryo yolk and has a refractive index similar to that of the cover glass. Accurate determination of the refractive index of the mountant is critical because it sets the length scale along the optical *z* direction. The refractive index can be determined from geometric measurements made from the images of embryo morphology, which are assumed to be independent of the orientation of the embryo when it is imaged.

Three-channel images can then be acquired using two-photon excitation at 750 nm (Fig. 1A–C) (Luengo Hendriks et al. 2006). Simultaneous multichannel acquisition of conventional dyes results in channel cross talk owing to the long-wavelength emission tails of the dyes. For quantitative analysis, it is critical to unmix the channels so relative fluorescence levels can be measured correctly. For a large number of images, an automated channel unmixing algorithm is critical (Luengo Hendriks et al. 2007). One bottleneck for imaging is the wait time while the microscope is acquiring the image. Most manufacturers allow the user to develop macros to automate microscope control. However, hopefully, the manufacturers will soon realize that high-throughput biology has arrived, and they will develop and provide these tools. In the meantime, to automate image acquisition, a macro should be written that lets the user record the state of the microscope for each embryo to be imaged. The macro can then sequentially reset the microscope to the appropriate configuration for each embryo, and the embryo images can be acquired and saved together with the microscope configuration data. This approach minimizes the necessary interactions between user and microscope,

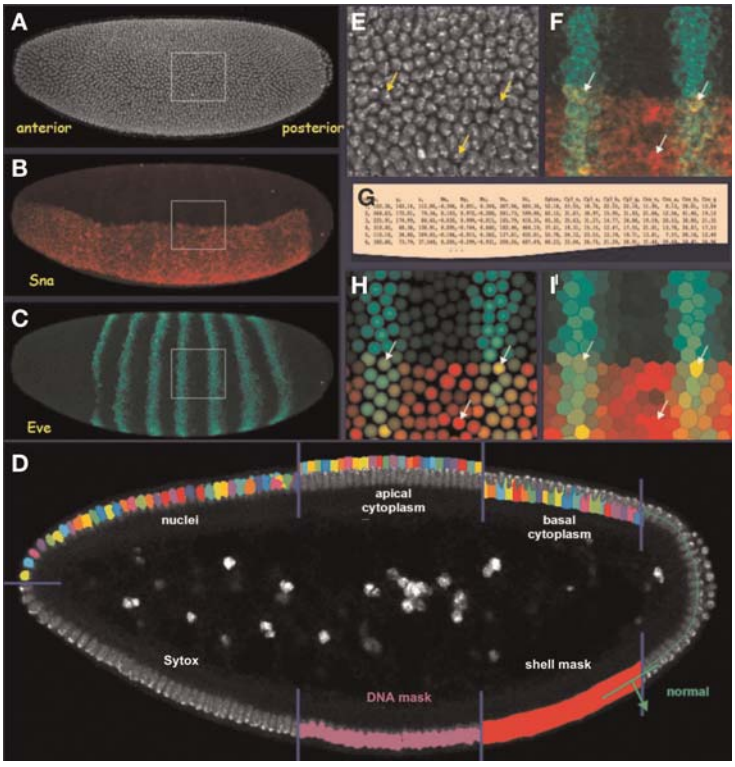

FIGURE 1. Segmentation analysis: Creating a point cloud. (*A–C*) The three channels of a 3D embryo image: (*A*) a nuclear stain (white); (*B*) a *snail* (*sna*) mRNA stain (red); and (*C*) an *even-skipped* (*eve*) mRNA stain (green). The position and the extent of the nuclei on the surface of the blastoderm embryo are defined by a model-based 3D segmentation analysis (*D*). The analysis defines the embryo surface. It locates individual nuclei based on a DNA stain, using a gray-weighted watershed algorithm to grow nuclear seeds along the embryo surface normals. Apical and basal cytoplasmic regions are estimated around each nucleus by Voronoi tessellation. Magnifications of the raw image are shown for the nuclear stain (*E*) and for the mRNA stains for *eve* and *sna* (*F*). The resulting segmentation mask is used to quantify expression levels, and these data are simply tabulated to create a PointCloud file (*G*). Two different renderings of the point cloud are shown for this region of interest, both derived using the visualization tool PointCloudXplore (*H, I*). The arrows indicate the locations of the same three cells in these panels.

and allows the microscope to sequentially image multiple embryos without user intervention. Because all the acquisition parameters are recorded, any slide can be put back on the microscope, and any or all of the embryos that have been previously imaged can be automatically revisited and/or reimaged.

Temporal Staging

To study pattern dynamics, a precise temporal staging system is required. Embryos can be staged by visual inspection under phase-contrast microscopy and can be broadly classified via the stages previously described (Campos-Ortega et al. 1997). However, a finer temporal staging system is needed to fully capture the rapid changes in morphology and gene expression. During stage 5, the percentage of membrane invagination can be used as a temporal marker. Membrane invagination is clearly seen under phase-contrast microscopy but is only accurately measured in the *x*–*y* plane. Further, because embryos are mounted in random D–V orientations and membrane invagination is more advanced on the ventral side than on the dorsal side, the final membrane invagination percentage must be extrapolated from the value measured on the ventralmost surface and the D–V orientation of the embryo, which can be accurately determined after the point cloud from that image is made.

Conversion of Images to Point Clouds

To quantify the relative expression levels from images, the image positions of all the nuclei are required. The total-DNA SYTOX image (Fig. 1A) can be used to determine the position and the extent of all embryonic nuclei to produce a segmentation mask. Many basic algorithms have been developed in the field of image analysis to enable image segmentation, but segmentation remains a problem that is not trivial. Model-based approaches are usually required in which the segmentation algorithm relies on a priori knowledge of the objects to be segmented and high image quality is essential.

For speed, the segmentation analysis can be restricted to the embryo surface by creating a 3D binary mask, the shell mask (Fig. 1D). The mask can be produced by taking an adaptive threshold of the SYTOX image, which isolates the bright DNA fluorescence and adapts for signal attenuation with image depth. The shell mask is then used to direct spectral unmixing and to allow attenuation correction of the SYTOX channel required for the segmentation. The shell mask can be refined to produce a DNA mask (Fig. 1D) that identifies regions in the image belonging to the nuclei. The SYTOX image within this mask is then convolved with a narrow Gaussian, and the positions of individual nuclei, termed seeds, are defined by local maxima in this image. To eliminate multiple seeds per nucleus, the embryo surface normal for each seed is computed, and neighboring seeds that lie along this normal are removed. Each seed is then grown to fill the nuclei, using a region-growing technique. To use both geometrical and gray-value information of the image, the region growing should combine a watershed algorithm and a gray-weighted distance transform (Luengo Hendriks et al. 2006). The combination of these two algorithms creates nuclear boundaries that match actual boundaries when visible, yet divides distances between seeds equally when boundaries are not distinguishable.

To capture the labeled mRNA expression levels, the cellular extent surrounding each nucleus can be estimated by growing the nuclear segmentation mask along the apical and basal directions into the cytoplasm by tessellation (Fig. 1D). The distances grown were established by determining the cytoplasmic extent, which can be estimated by its autofluorescence from the raw image. The resulting cytoplasmic mask is then used in combination with the nuclear mask to divide each cell into three regions: apical, nuclear, and basal (Fig. 1D). Expression levels are estimated on a per cell basis as the average value within each of these regions for all the channels (Fig. 1E,F). These values, together with the center of mass of the nuclei, the volumes of the various cellular regions, and the neighborhood relationships between cells, are then written to a text PointCloud file (Fig. 1G). The PointCloud file captures the information from the image and allows the data to be visualized (Fig. 1H,I). For subsequent analysis, expression values from the point clouds are corrected for signal attenuation by dividing the results from the Cy3 and coumarin channels by those from the SYTOX channel. This approach assumes that the average SYTOX intensity is constant from nucleus to nucleus and that it is representative of the attenuation of the other channels.

 Cite this article as *Cold Spring Harbor Protoc;* 2012; doi:10.1101/pdb.top067843

Registration of Point Clouds into a Virtual Embryo

The atlas is constructed by registering individual point clouds onto a common framework, created as an average over an ensemble of point clouds. The registration takes into account the significant variation between embryo sizes, number of cells, and relative locations of gene expression patterns. The key is identifying corresponding cells, or fiducial marks, between images of multiple embryos and using these correspondences to combine measurements of expression and morphology (Fowlkes et al. 2008) (Fig. 2A). Fiducial marks can be established by using the pair-rule patterns of *eve* and *fushi tarazu* (*ftz*) as common expression markers. These patterns can be used, along with the A–P and D–V coordinates, to define a set of cells along pattern boundaries that act as the correspondence points between pairs of embryos. Once these are established, it is straightforward to find the correspondence for the remaining cells in the embryo so that an average virtual embryo can be built. Multiple frameworks are required because both the embryo morphology and the relative position of nuclei change during stage 5. Thus, to properly understand the dynamics, a temporal correspondence is also required between different frameworks for the temporal cohorts (Fig. 2A) (Fowlkes et al. 2008).

The accuracy of the resulting registration can be tested by comparing average expression levels recorded in the virtual embryo with those measured directly in embryos costained for the expression of two genes. For this experiment, a set of cells along the anterior edge of *eve* stripe 2 are selected, in which the transcription factor *gt* is known to repress *eve* expression (Fig. 2B). *gt* versus *eve* expression is plotted for these cells from a set of embryos costained for *eve* and *gt* (Fig. 2C) and from a set of *eve-ftz* and *gt-ftz* embryos, registered using *ftz* expression (Fig. 2D). As Figure 2D shows, the virtual measurement of *eve* and *gt* should replicate the relation derived from the costained embryos (Fig. 2C). Inferences about regulatory relationships made using the virtual embryo are, thus, similar to those made from individual embryos. As a comparison, the coexpression of *eve* and *gt* can be plotted for a set of coarsely registered embryos (Fig. 2E). This experiment shows the superiority of the fine registration technique and shows that building cellular resolution expression atlases from multiple embryos is not only possible but is biologically accurate and also provides useful data for understanding transcriptional regulation.

PointCloudXplore: A Visualization Tool

To date, the BDTNP has registered PointCloud data from ~3000 embryos containing around 100 different mRNA and protein expression patterns to form a VirtualEmbryo representation of the blastoderm. One of the challenges of such a large multidimensional data set is making it accessible for visual observation and quantitative exploration. The BDTNP has developed a suite of visualization tools (PointCloudXplore), which show the data in multiple views (Weber et al. 2009; Rübel et al. 2010). Using PointCloudXplore, expression patterns such as those of *eve* and *sna* can be displayed on an average embryo morphology for multiple temporal cohorts (Fig. 3A), or these patterns can be unrolled and displayed as a surface projection (Fig. 3B). Other relationships can be seen for multiple genes (Fig. 3C), and height maps of the expression patterns can be added (Fig. 3D). More complex analyses such as scatter plots and parallel coordinates are also possible using PointCloudXplore.

NEW BIOLOGY REVEALED

The PointCloud and VirtualEmbryo data can be used to reveal new features of the blastoderm system. Calculations of the local spatial density of nuclei have revealed an intricate morphology of nuclear density patterns that changes during stage 5 (Fig. 4A–F) (Keränen et al. 2006). An overlay of these density maps onto the expression patterns of *eve* and *sna* (e.g., Fig. 4G) shows that the pattern of lower nuclear densities corresponds to the embryo locations that are about to undergo gastrulation and to form the cephalic and the ventral furrows. The location of pair-rule expression stripes has also been shown to move relative to the field of nuclei (Keränen et al. 2006). These two results have a critical impact on the choice of method for temporal registration of the PointCloud data.

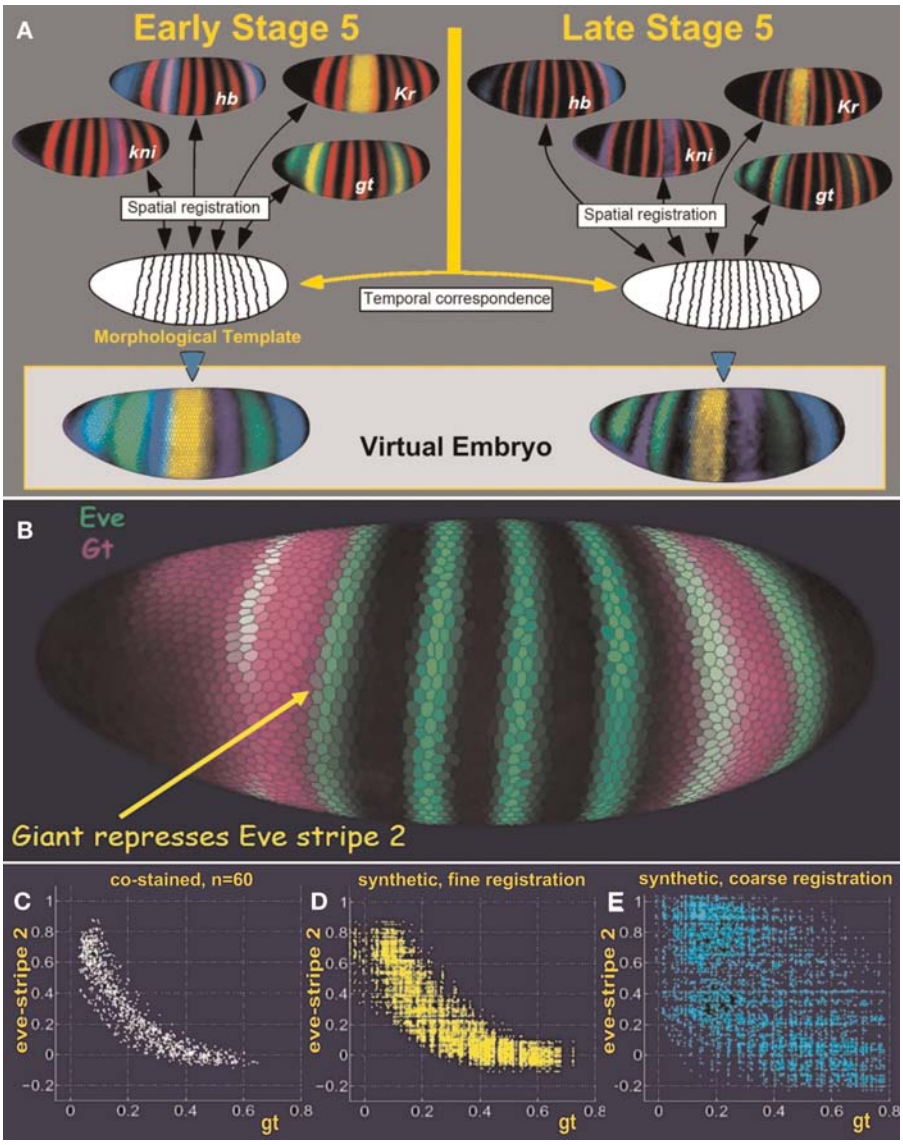

FIGURE 2. Construction of a virtual embryo. Images from hundreds of embryos are grouped into temporal cohorts, and each cohort is registered based on a common pair-rule marker gene, which is shown in red. Expression patterns of many other genes stained in different embryos are, thus, brought into a common frame. Temporal correspondence between registered cohorts produces the virtual embryo atlas of multiple genes over multiple times (*A*). The accuracy of the virtual embryo was shown by plotting the coexpression of *giant* (*gt*) and *eve* for a set of cells along the anterior border of eve stripe 2 that are known to be repressed by *gt* (*B*). A plot of the coexpression of *gt* and *eve* in embryos costained for *eve* and *gt* shows a clear anticorrelation (*C*). A plot of the coexpression of *gt* and *eve* for these cells in embryos costained for *eve-ftz* and *gt-ftz* and registered using *ftz* shows the same anticorrelation (*D*). However, a plot of coexpression of *gt* and *eve* in embryos costained for *eve-ftz* and *gt-ftz* that were only coarsely registered is much less accurate and does not reveal the anticorrelation (*E*).

The latter result indicates that a reference gene pattern cannot be used to register data between temporal cohorts, and the former result shows that temporal correspondences cannot be obtained by using common spatial coordinates. Thus, to register average frameworks of different temporal cohorts, a model of the nuclear position dynamics is required (Fowlkes et al. 2008).

The atlas makes it possible to describe expression patterns quantitatively along both the A–P and the D–V axes. Expression pattern gradients can be used to define the pattern boundaries, and their dynamics and relative expression levels can be plotted (Keränen et al. 2006). For example, it is now

Cite this article as *Cold Spring Harbor Protoc*; 2012; doi:10.1101/pdb.top067843

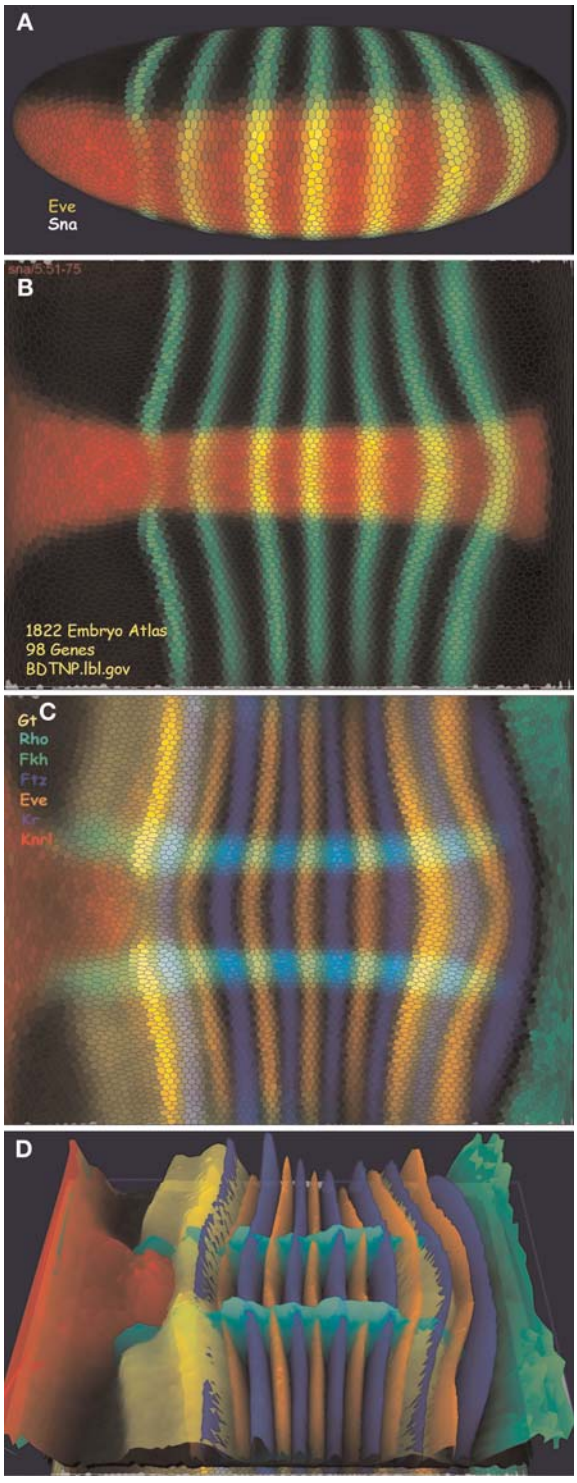

FIGURE 3. PointCloudXplore. The visualization tools of the blastoderm project provide multiple physical views. This figure shows expression patterns of *eve* and *sna* (*A,B*) and *gt*, *rho*, *fkh*, *ftz*, *kr*, and *knrl* (*C,D*) from the blastoderm atlas. The patterns can be visualized in 3D on a virtual embryo, which can be rotated to any angle (*A*), or the embryo surface can be unrolled to reveal the entire pattern in a 2D view (*B,C*). This view can also be rotated, and height maps of the relative expression for each gene can be added (*D*).

easy to plot the relative expression of pair-rule stripes around the D–V axis (Fig. 4H,I). These plots show that there are unsuspected quantitative changes along the D–V axis of pair-rule stripes (Luengo Hendriks et al. 2006). Interestingly, the D–V dependencies of pair-rule expression stripes are almost lost in dorsalizing mutants, suggesting that D–V acting factors regulate A–P expression patterns (Keränen et al. 2006). Although the atlas is an average over thousands of embryos, the biological variability is not lost and can be studied from individual embryo point clouds. For example, plotting

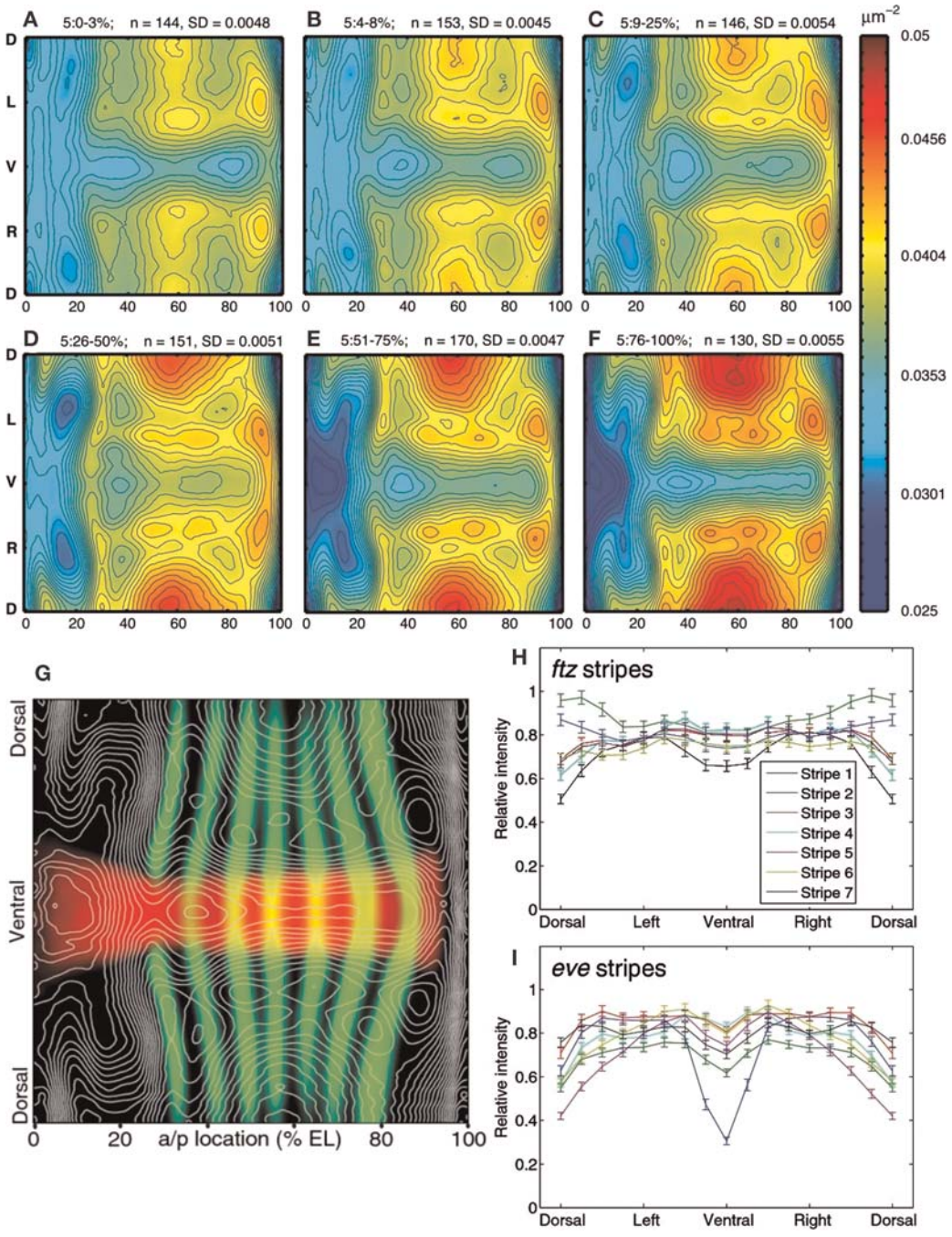

FIGURE 4. Complex motions of nuclear density and patterns. Stage 5 blastoderm embryos show a complex dynamic pattern of nuclear packing densities. Spatial nuclear densities were calculated for several temporal cohorts of embryos through stage 5 as the number of nuclear centers per μm^2. The resulting density maps were computationally unrolled into a 2D projection (*A–F*). Nuclear densities are represented as a color heat map with lowest densities (0.025 nuclei/μm^2) in blue and the highest densities (0.05 nuclei/μm^2) in red. Isodensity contours were overlayed with expression patterns of *sna* and *eve* (*G*). Expression of pair-rule stripes of *ftz* and *eve* vary significantly around the D–V axis (*H, I*). The stripes show marked differences in expression profiles and have unique modes of expression variation along the D–V axis. The error bars give the 95% confidence intervals for the means. The data were derived from 155 embryos. EL, embryo length; SD, standard deviation.

Cite this article as *Cold Spring Harbor Protoc*; 2012; doi:10.1101/pdb.top067843

the number of nuclei per embryo versus A–P egg length (Fowlkes et al. 2008) shows that embryo size correlates with number of nuclei, suggesting that egg size depends on mitosis rate.

The chief motivation for developing the blastoderm expression atlas was to understand the regulation of target genes by transcription factors. By using the atlas, it has been shown that the patterns of many transcription factors correlate or anticorrelate with the patterns of other genes (Fowlkes et al. 2008) in the same way that the expression of *gt* and *eve* stripe 2 anticorrelate (Fig. 2B–E). From these correlations, it is possible to infer potential regulatory relationships within the network, a number of which meet the expectations of previous molecular genetic data (Fowlkes et al. 2008). Although such inferences cannot be used to imply that a transcription factor directly binds and regulates a target gene, in combination with other classes of data, such as genome-wide in vivo binding data and in vitro DNA specificity data (Li et al. 2008), they will provide a significant constraint on possible models for the structure of the regulatory network.

CONCLUSION

Current technologies in optical microscopy and computing allow high-throughput, image-based investigations of multicellular biological systems. The data obtained in these studies have facilitated the construction of multidimensional computationally analyzable maps, which provide unprecedented opportunities for revealing the biology of a system at the macromolecular level and on a per cell basis. The BDTNP has produced the first quantitative, 3D atlas of gene expression and morphology of early embryonic development in *Drosophila*. These technologies are profound and have a broad and exciting future.

ACKNOWLEDGMENTS

I am privileged in collaborating with an outstanding group of scientists from multiple disciplines who, as part of the BDTNP, brought their expertise to bear on this problem. Special thanks are due to Mark D. Biggin. Many biologists are overwhelmed by today's image-based technologies and what they can provide. Mark quickly understood the possibilities and provided leadership and drive, without which this project would not have been realized. Thanks to Soile V.E. Keränen and Mark for helpful suggestions with this article.

REFERENCES

Adams MD, Celniker SE, Holt RA, Evans CA, Gocayne JD, Amanatides PG, Scherer SE, Li PW, Hoskins RA, Galle RF, et al. 2000. The genome sequence of *Drosophila melanogaster*. *Science* **287:** 2185–2195.

Campos-Ortega JA, Hartenstein V. 1997. *The embryonic development of Drosophila melanogaster*, 2nd ed. Springer, Berlin.

Fowlkes CC, Luengo Hendriks CL, Keränen SVE, Weber GH, Rübel O, Huang MY, Chatoor S, DePace AH, Simirenko L, Henriquez C, et al. 2008. A quantitative spatiotemporal atlas of gene expression in the *Drosophila* blastoderm. *Cell* **133:** 364–374.

Keränen SVE, Fowlkes CC, Luengo Hendriks CL, Sudar D, Knowles DW, Malik J, Biggin MD. 2006. Three-dimensional morphology and gene expression in the *Drosophila* blastoderm at cellular resolution II: Dynamics. *Genome Biol* **7:** R124. doi: 10.1186/gb-2006-7-12-r124.

Lawrence PA. 1992. *The making of a fly: The genetics of animal design* Blackwell-Scientific, Oxford.

Li X, MacArthur S, Bourgon R, Nix D, Pollard DA, Iyer VN, Hechmer A, Simirenko L, Stapleton M, Luengo Hendriks CL, et al. 2008. Transcription factors bind thousands of active and inactive regions in the *Drosophila* blastoderm. *PLoS Biol* **6:** e27. doi: 10.1371/journal.pbio.0060027.

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

Luengo Hendriks CL, Keränen SVE, Biggin MD, Knowles DW. 2007. Automatic channel unmixing for high-throughput quantitative analysis of fluorescence images. *Opt Express* **15:** 12306–12317.

Rübel O, Weber GH, Huang M, Bethel EW, Biggin MD, Fowlkes CC, Luengo Hendriks CL, Keränen SVE, Eisen MB, Knowles DW, et al. 2010. Integrating data clustering and visualization for the analysis of 3D gene expression data. *Trans Comput Biol Bioinform* **7:** 64–79.

Weber GH, Rübel O, Huang M, DePace AH, Fowlkes CC, Keränen SVE, Luengo Hendriks CL, Hagen H, Knowles DW, Malik J, et al. 2009. Visual exploration of three-dimensional gene expression using physical views and linked abstract views. *Trans Comput Biol Bioinform* **6:** 296–309.