Expressing Genes

Felice Frankel

One of my colleagues, Angela DePace, a postdoctoral scientist at Berkeley, is part of a multidisciplinary group focused on analyzing and visualizing gene expression during development. I decided that Sightings *readers might be interested in what I believe is a fascinating effort to use graphics as an interactive tool to visually explore information. The following is an edited conversation with Angela (below, A. H. D.) and two other members of the group, Gunther H. Weber (G. H. W.) and Charless C. Fowlkes (C. C. F.). The group also includes Soile V. E. Keränen, Oliver Rübel, Min-Yu Huang, Cris Luengo Hendriks, Lisa Simirenko, Damir Sudar, Hans Hagen, David W. Knowles, Jitendra Malik, Mark D. Biggin and Bernd Hamann. In-depth information on the project can be found on their Web site,* http://bdtnp.lbl.gov.*.*

F. F. Can you give me a general overview of the project?

A. H. D. We're trying to understand how genes turn on and off in space and time during early development. How do you build a whole animal, with all kinds of different tissues, from a single cell? Part of the answer is that you create cell types by selectively expressing different sets of genes in different cells. We've been working to create a three-dimensional model of the early fruit fly embryo at cellular resolution, so that we can ask specific quantitative questions about how gene expression is controlled.

F. F. Can you briefly explain how you imagine a researcher would explore data in your model?

C. C. F. Part of what makes understanding animal development difficult is that it involves many different genes that all interact to correctly turn a homogenous set of cells into the specific tissues that make a functioning animal. Our

Visualization tools being developed by the Berkeley *Drosophila* Transcription Network Project (BDTNP) use multiple views to show aspects of gene expression during animal development—specifically in the early *Drosophila* embryo. The Unrolled View maps the entire *Drosophila* embryo continuously to a plane using cylindrical projection. This is possible because nearly all the cells in the *Drosophila* blastoderm lie in an ellipsoidal monolayer surface prior to gastrulation (the early migration of cells in the embryo that begins implementation of the animal's body plan). This surface can be "unrolled" by the following process: The embryo *(top)* is surrounded with a cylinder that is aligned with the embryo's anterior-posterior axis. All detected cells on the blastoderm surface are then projected onto this cylinder, and the resulting surface unrolled by cutting the cylinder along a line corresponding to the dorsal midline of the embryo. This process yields a continuous mapping of the embryo surface to a two-dimensional plane and allows users of the BDTNP visualization tool to trace expression patterns over the entire embryo.
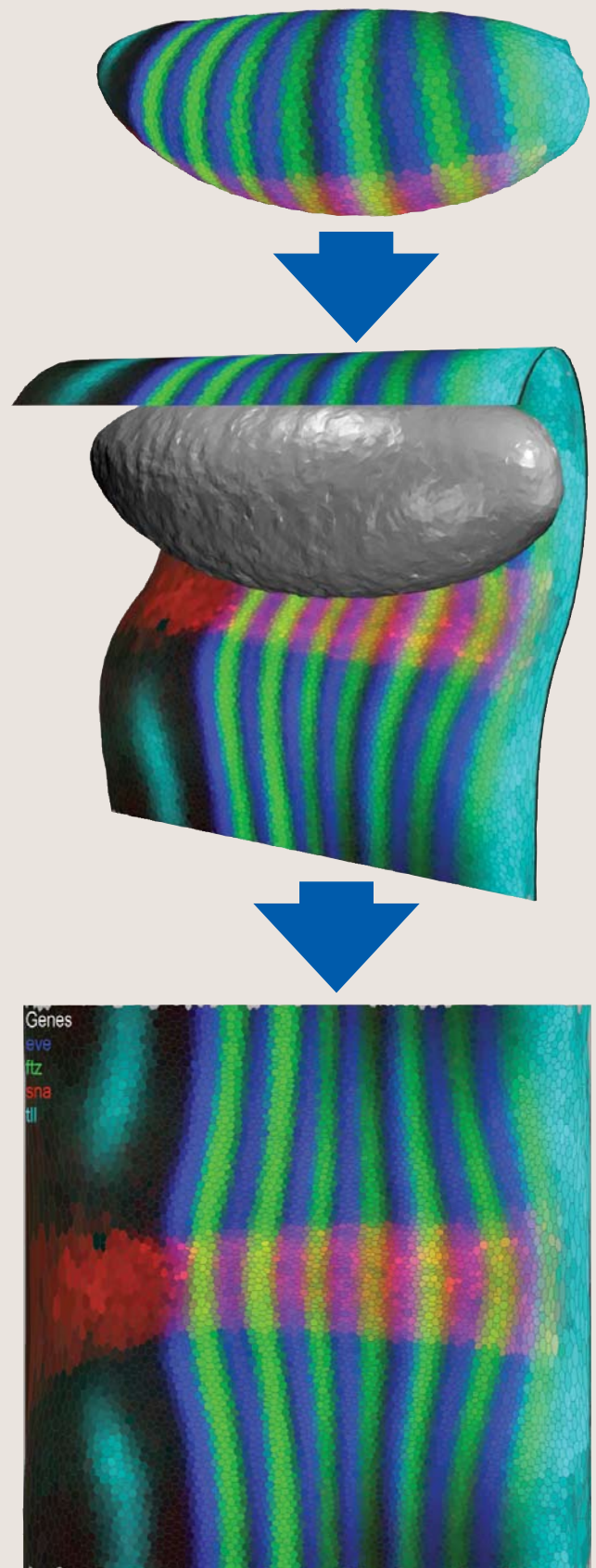
virtual *Drosophila* embryos contain composite data captured on several dozen genes from hundreds of different experiments. The virtual embryo gives a compact description of the "average" expression levels for these genes, which are ultimately responsible for determining how the cells develop into legs, wings, eyes etc. to make a functioning fruit fly.

Combining data for a large number of genes in one place overcomes a limitation of our experiments, where we can only stain for two genes at once. We have developed visualization tools that allow the user to interactively move around the virtual embryo in three dimensions and examine the expression patterns for many different genes as a biologist might see them through the microscope. We have also found it useful to display this 3-D data in a 2-D projection, much as one uses a Mercator projection to display a map of the whole planet at once. This "unrolling" (shown in the large image on the previous page) allows one to take in a whole atlas of gene expression at a glance.

In addition to visually exploring the qualitative spatial patterns of gene expression (as has long been done with the microscope), our data allow more sophisticated

quantitative analysis. For each cell in the embryo we have estimates of the concentration of different gene products. These concentrations can be thought of as specifying the coordinates of each cell in "expression space." The scatterplot below shows the locations of all the embryo cells in a 3-D expression space given by the genes hunchback, even-skipped and snail.

The genes we are measuring act as transcription factors that regulate the expression of other genes. For example, gene A may only be expressed when protein transcribed from gene B and gene C are both present. It is the network of interactions between genes that is responsible for the intricate patterns seen in the figures. Recording quantitative expression data will ultimately allow us to build computational simulations of these interactions and understand how the fruit fly is "built" from the ground up by its genes.

A. H. D. Developing visualization tools specific to this data set has allowed biologists to explore the data without needing to learn advanced computational methods. By having a visual environment where you can see

the same set of cells in different representations, a researcher can consider a variety of relationships rapidly, which can lead to new hypotheses to be tested. For example, the light blue cells in the images on these pages are all the same, but represented in different ways. The "striped" image is an unrolled view, displaying their spatial relationship to other cells. The scatterplot shows the quantitative relations among cells in terms of the amounts of three different genes expressed in those cells. Finally, the bar graph shows the levels of all genes in our data set in a particular cell.

F. F. Are there any things you found particularly difficult to represent? If so, why was it difficult?

C. C. F. There is a clear difficulty in simultaneously representing the expression levels for tens of different genes in cells distributed throughout the embryo. We can display up to three different genes simultaneously by making them each a different color (say red, green and blue) and then displaying each unique combination of these three

genes as a blend of these three colors. However, if we try to use this trick for more than three genes we run into trouble, because there will be distinct combinations of four colors that result in the same resulting blended color. Human color perception is inherently three-dimensional.

G. H. W. Here I agree only partially. While it is only possible to show three genes when each color is supposed to represent a unique gene-expression combination, it can still be useful to show more genes, depending in part on their spatial distribution. For example, the "Unrolled View" example on page 69 shows four genes: *eve* (blue), *ftz* (green), *sna* (red) and *tll* (light blue). Using this combination it is no longer possible to tell whether a cell is light blue because of high *tll* expression or because *eve* and *ftz* are both expressed at high levels. However, if one chooses the colors carefully one can still look at more than three patterns. For example, in the above example we know that *ftz* and *eve* are not highly expressed in the same cells and thus that the light blue results from *tll*. Another drawback of this lack of color uniqueness is that cells with high *tll* expression are always light blue, independent of whether *ftz* or *eve*—or both—are also expressed.

C. C. F. Ultimately it is not just visualizing the expression patterns but understanding the relations between expression levels for different genes that is important. We are now working on tools that focus specifically on revealing these interrelations in an intuitive way.

G. H. W. Some of these tools should, we hope, allow us to come up with alternate coloring schemes that allow display of more than three genes by automatically determining some relationships between genes and choosing colors in a different way.

All in all there is a need for automated methods. But visualization allows humans to "playfully" explore a data set and discover new and perhaps unexpected behavior. In the future it should also allow one to steer automated analyses that could then confirm subjective determinations made by a human.
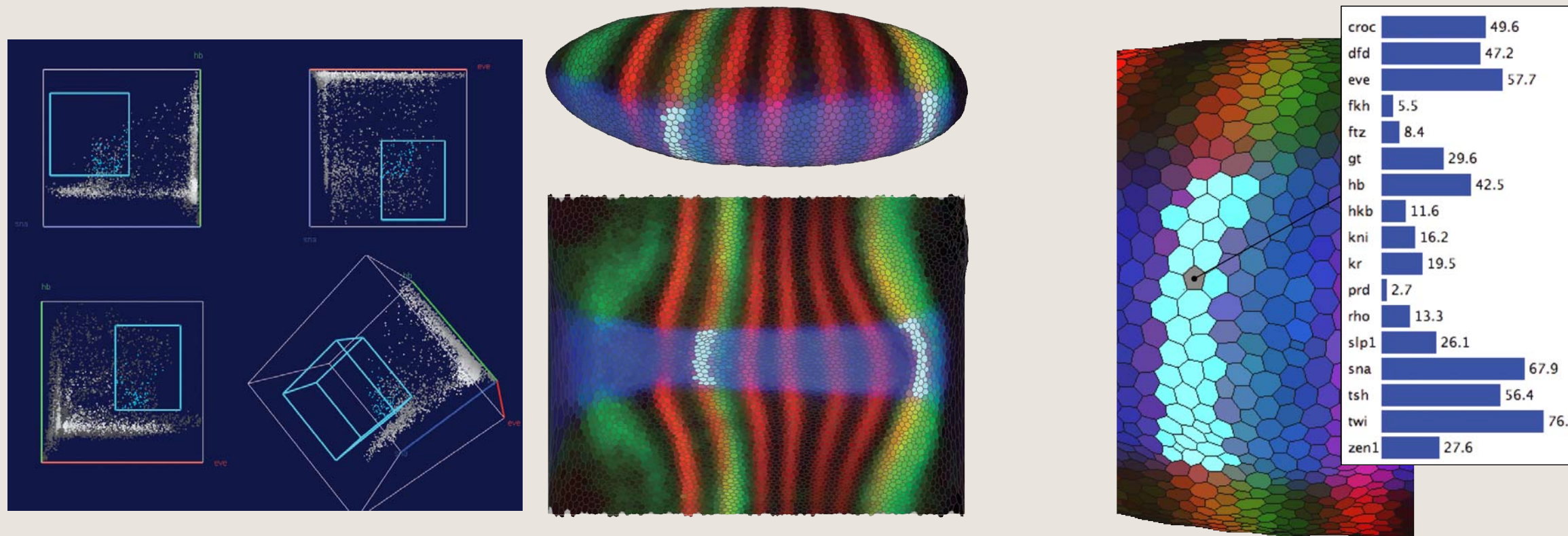
Three-dimensional scatterplots can show expression patterns of up to three genes. The plots define an axis for each gene using the colors that map gene expresion on the physical model of the embryo (*right*)—green for hunchback (*hb*), red for even-skipped (*eve*) and blue for snail (*sna*). In the 3-D plot at left, each cell is represented by a point whose location is based on the relative expression levels of the three selected genes in that cell. A subset of cells can be selected and highlighted to display the "expression space" of those cells. The space bounded in blue in the plots here (and containing light-blue "cells") corresponds to the blue-highlighted cells in the physical view at lower right.

Each cell in the physical embryo view has a "pseudo" stain. A viewer interested in the expression levels within a specific cell can select the cell and display the gene-expression levels of that cell in detail as a bar graph.

Felice Frankel is a senior research fellow in the Faculty of Arts and Sciences at Harvard University, where she heads the Envisioning Science project at Harvard's Initiative in Innovative Computing (IIC). She holds a concurrent appointment as a research scientist at the Massachusetts Institute of Technology. Address: c/o American Scientist, P.O. Box 13975, Research Triangle Park, NC 27709-3975.

www.americanscientist.org

MIT OpenCourseWare

Resource: Making Science and Engineering Pictures: A Practical Guide to Presenting Your Work
Felice Frankel

The following may not correspond to a particular course on MIT OpenCourseWare, but has been
provided by the author as an individual learning resource.