

Building a Quantitative Spatio-temporal Atlas of Gene Expression in the *Drosophila* Blastoderm

Charless C. Fowlkes^{1,2}, Cris L. Luengo Hendriks^{1,3}, Soile V. E. Keränen^{1,4}, Angela H. DePace^{1,4}, Gunther H. Weber^{1,5}, Oliver Rübél^{1,5}, Min-Yu Huang^{1,5}, Sohail Chatoor^{1,3}, Lisa Simirenko^{1,4}, Clara Henriquez^{1,4}, Amy Beaton⁴, Richard Weiszmann⁴, Susan Celniker^{1,3}, Bernd Hamann^{1,5}, Michael B. Eisen^{1,4}, David W. Knowles^{1,3}, Mark D. Biggin^{1,4}, Jitendra Malik^{1,2}

¹ Berkeley *Drosophila* Transcription Network Project, <http://bdtnp.lbl.gov>

² Department of Electrical Engineering and Computer Science, UC Berkeley

³ Life Sciences Division, Lawrence Berkeley National Laboratory

⁴ Genomics Division, Lawrence Berkeley National Laboratory

⁵ Inst. for Data Analysis and Visualization (IDAV) and Dept. of Computer Science, UC Davis

* email: fowlkes@cs.berkeley.edu

The output of animal gene transcription networks are dynamic, three-dimensional patterns of expression. To build comprehensive models of these networks, it is essential to accurately measure the spatial and temporal distributions of transcription factors and target gene expression. Such measurements are now being systematically collected over a range of spatial and temporal resolutions in several animal models. However, current strategies for quantitating spatially resolved gene expression do not allow labeling of more than a few gene products in a given animal or tissue. This poses a serious limitation, as even simple portions of animal transcription networks can comprise tens of regulators and hundreds of target genes. We present a computational approach that overcomes this limitation by compositing expression measurements from hundreds of embryos into a common spatio-temporal atlas where the average expression patterns of many gene products can be studied simultaneously.

Our technique involves two key components, outlined in Figure 1. The first is a *spatial registration* algorithm that uses a marker gene expression pattern common to all labeled embryos to help identify equivalent *corresponding* cells or nuclei across multiple images of embryos at the same stage of development. The second component is a dynamical *morphological template* that specifies the average measured locations of nuclei over time, providing temporal correspondences between nuclei in embryos imaged at different developmental time points. This morphological template consists of an average number of nuclei whose three-dimensional motions track the average blastoderm shape and density of nuclear packing. Once correspondences have been established among embryos within and between cohorts, expression measurements are combined into a single composite model *VirtualEmbryo* that describes the average patterns of expression for many genes at multiple time points. Figure 2 shows example patterns from an initial atlas of spatio-temporal expression data during the 50 minutes prior to gastrulation based on 1822 embryos stained for 95 different genes.

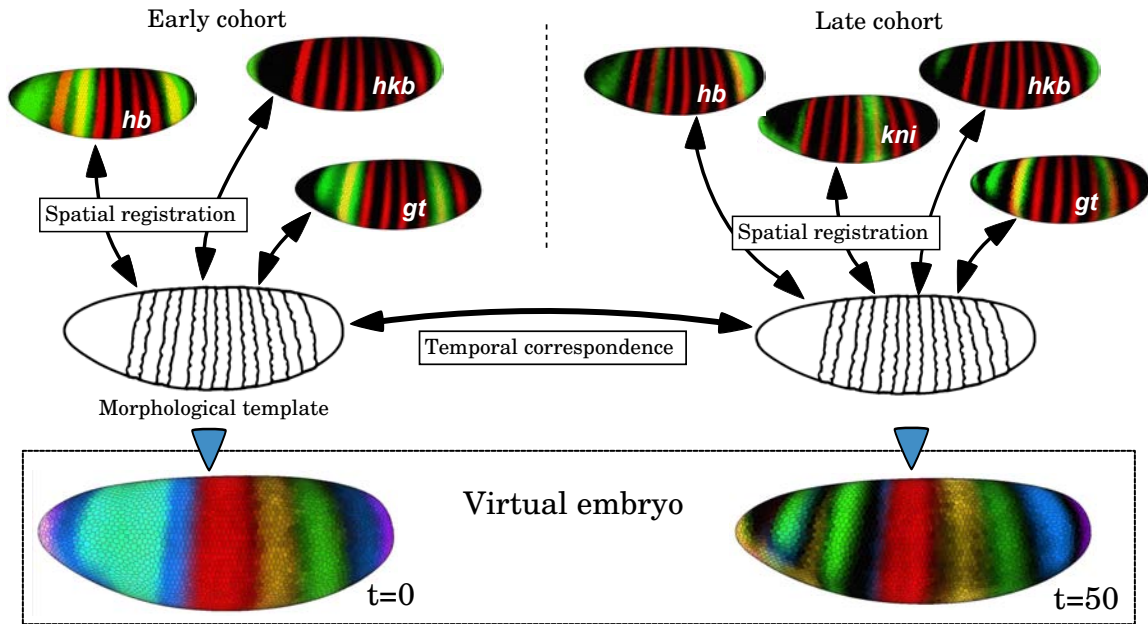


Figure 1. Data from hundreds of individual embryos is registered and composited into VirtualEmbryo in which coexpression of many genes can be analyzed.

Our approach takes into account significant *geometric variability* between individual embryos in their size, shape, number and positions of cells and in the locations of gene expression patterns. By measuring these variations and additional non-geometric variability in the relative concentrations of regulator and target gene products, we show that it is possible to establish upper-bounds on the degree of geometric and regulatory variability within a population. This provides a natural baseline for evaluating the accuracy of registration methods in the presence of biological variation. Using this baseline, we demonstrate that the VirtualEmbryo accurately describes average patterns of gene expression present in individual embryos, and thus provides a suitable framework for building quantitative models of gene regulation. Factoring out geometric variability will also be valuable in isolating differences in regulatory mechanisms between closely related species.

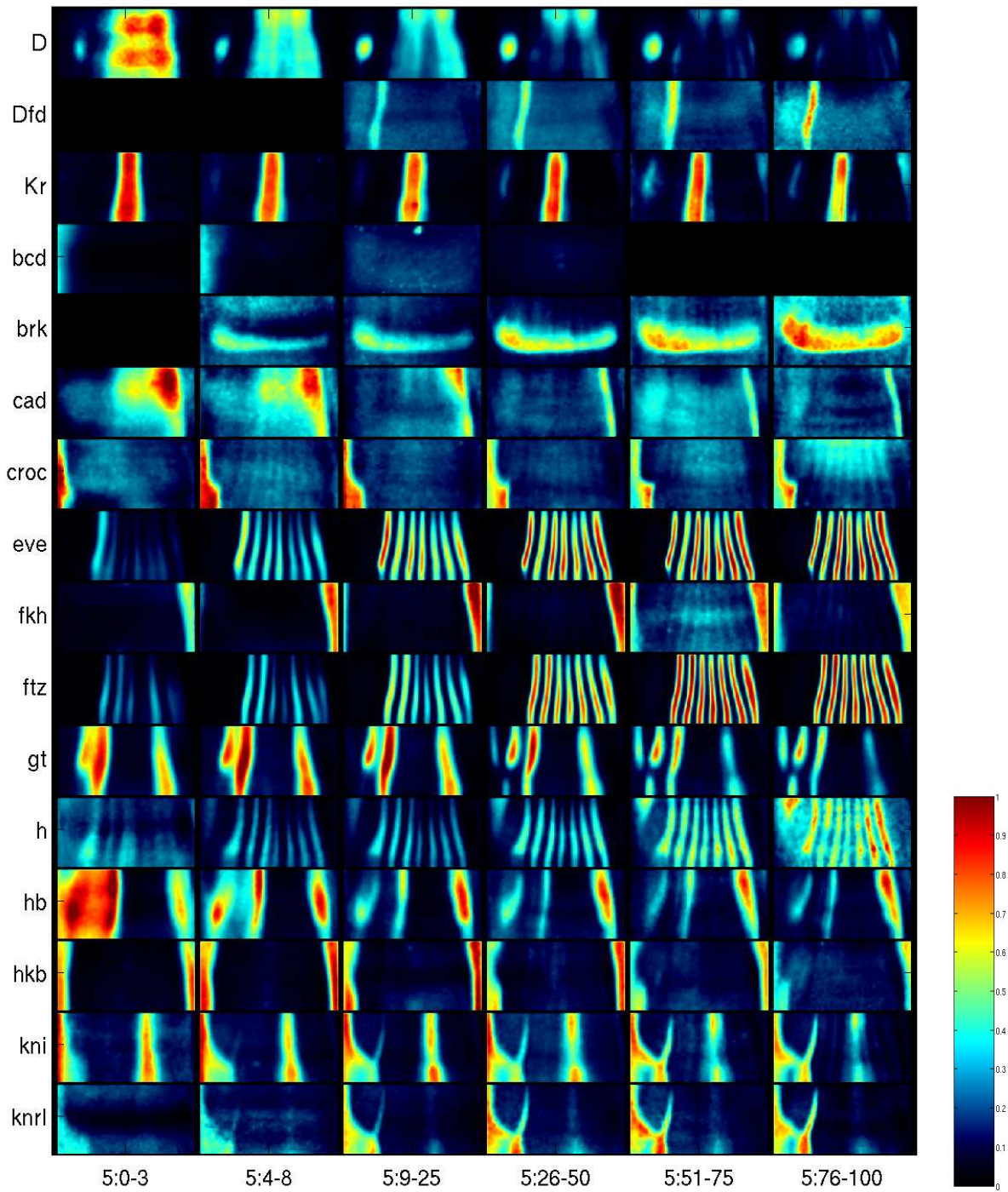


Figure 2. Average pattern dynamics of mRNA expression recorded in the VirtualEmbryo for several genes (rows) and six temporal cohorts (columns). Each rectangle shows a lateral view of the blastoderm in cylindrical projection with dorsal at top, anterior to the left.