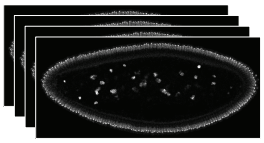
Building Composite Maps of Gene Expression Patterns and Morphology: Registering 3D Representations of Drosophila Embryos

BDTNP
Berkeley Drosophila Transcription Network Project

Charless C. Fowlkes[1], C.L. Luengo Hendriks[2], S.V.E. Keränen[2], M.D. Biggin[2], DW Knowles[2], D. Sudar[2], and J. Malik[1]
1) Computer Science Division, University of California, Berkeley, CA
2) Life Sciences and Genomics Divisions, Lawrence Berkeley National Laboratory

Introduction

The Berkeley Drosophila Transcription Network Project is developing a suite of methods to convert image stacks generated by confocal microscopy into numerical representations of gene expression at cellular resolution. One key difficulty is that fluorescence microscopy can only capture expression levels for a few (2-4) gene products in a given animal. Here we report on a method for registering 3D expression data from different Drosophila embryos stained for overlapping subsets of target genes in order to build a composite atlas, ultimately containing collocation information for 1000's of genes. Registration is also shown to be a useful tool in studying morphology. We observe complex, dynamic patterns of cell density during cellular blastoderm prior to gastrulation.



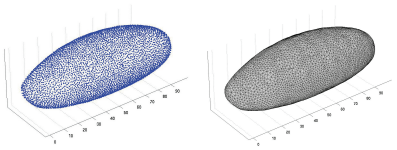
2-photon fluorescent confocal microscopy (20x lens) captures the entire embryo (~500µm by 200µm) at high enough resolution to distinguish individual nuclei (voxel size 0.45x0.45x1.5µm)

↓ Nuclear segmentation

```
id, x,      y,      z,      Vn,     Vc,     Sytox, Cy3_n,…
 1, 102.36, 142.14, 112.00, 207.96, 605.36, 52.18, 23.55,…
 2, 264.63, 172.01,  79.36, 281.73, 599.90, 82.12, 31.67,…
 3, 225.91, 174.99,  88.65, 185.79, 418.35, 85.32, 35.63,…
                      . . .
```

Each confocal image stack is segmented into cellular volumes with the help of a nuclear stain and the fluorescence level is quantified in each volume (see poster 358A by *C.L. Luengo Hendriks et al)*. The resulting pointcloud contains position, volumes and expression levels associated with each detected nucleus in the blastoderm.

↓ Geometric modeling & Feature extraction



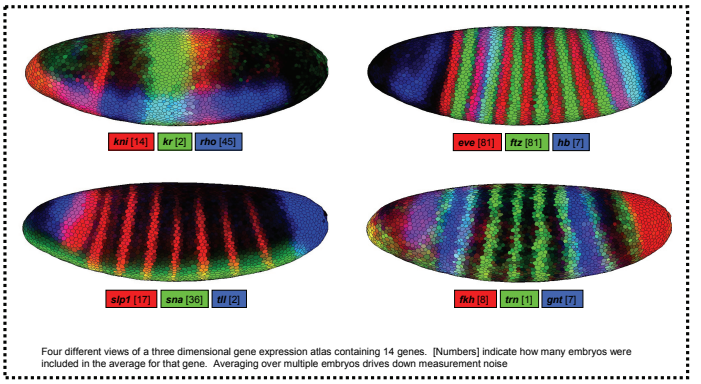
Geometric modeling recovers neighborhood relationships between cells, capturing the blastoderm topology. Distinctive local patterns of gene expression around each cell provide features that drive the correspondence process.

↓ Registration onto composite map

To combine expression data from two or more embryos, it is necessary to identify corresponding cells in each embryo. This is accomplished by registering each embryo onto a composite map where multiple expression measurements can be meaningfully averaged.

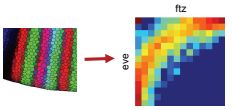
Application: Building a 3D Atlas of Gene Expression

In order to build models of transcription networks in developing embryos, we need to measure the spatial co-expression of active genes. If we treat each image separately, then for **N** genes it is necessary to perform $O(N^2)$ hybridizations, one for each pair. Registration allows us to construct a complete picture with only $O(N)$ hybridizations, one for each gene paired with a reference gene. This makes it feasible to aim for large scale models of 1000s of genes which would otherwise require millions of hybridizations. Here we show an initial composite atlas built from 150 images of 14 different genes.

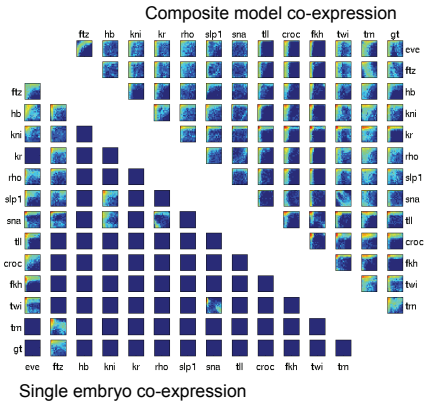


Four different views of a three dimensional gene expression atlas containing 14 genes. [Numbers] indicate how many embryos were included in the average for that gene. Averaging over multiple embryos drives down measurement noise

The figure at right demonstrates the utility of registering multiple embryos. Each square shows the joint distribution of the expression levels for a pair of genes. For example, **ftz** and **eve** are expressed in a disjoint set of cells so the distribution shows anti-correlation. In contrast, **sna** and **twi** show strong co-expression



The lower triangle shows the statistics measured for single embryo pointclouds. The upper triangle is generated by registering these same embryos onto the composite atlas, allowing us to fill in statistics for gene pairs which were not originally imaged.



Registering 3D Blastoderm Models

Registration Algorithm

1. Perform coarse registration
2. Extract shape-context descriptor
3. Find corresponding cells
4. Warp embryo model to align with corresponding cells on target model
5. Repeat 3-4 until convergence

Local Descriptors of Expression Patterns

For each cell we compute a descriptor which captures the local pattern of gene expression. The descriptor, called a *shape context*, integrates expression levels over bins arranged in a polar fashion around the given cell.

Bins far from the cell of interest integrate over a larger area and are less sensitive to variations in the shape of expression patterns.

Correspondence as Optimization

X_{ij} = ***1*** if cell **i** is matched to cell **j**
0 otherwise
C_{ij} = dissimilarity of local descriptors for cells **i** and **j**
D_{ij} = distance between cells **i** and **j**

$$\text{minimize} : \Sigma_{ij} (C_{ij} + \lambda D_{ij}) \cdot X_{ij}$$
$$\text{subject to} : \Sigma_i X_{ij} = 1$$
$$\Sigma_j X_{ij} = 1$$

λ sets the relative importance of distance versus descriptor similarity

A Non-linear Deformation Model

Once the initial correspondence has been computed we would like to find a global deformation that brings the matched cells into alignment. We fit a regularized *thin-plate spline* to model the deformation between embryos. The thin plate spline is the higher dimensional generalization of the cubic spline. The smoothness of the spline fit is governed by a regularization term. We start with a high degree of regularization, enforcing nearly rigid transformations and gradually relax this constraint over the course of several iterations.

Application: Pre-gastrulation Dynamics of Cell Density

Registration is also useful in the quantitative study of embryo morphology. Just as we can use reference gene correspondence to average fluorescence measurements, we can also average morphological measurements, in this case the density of cell packing. Here we find a complex dynamic pattern which wouldn't have been apparent in casual visual inspection of the raw data. Of interest is the steep density gradient on either side of the collection of cells that will soon collapse into the ventral furrow.

The figures at right show the local nuclear density in an cylindrical (mercator) projection. This is a convenient way to display the entire embryo at once. Densities were averaged inside a moving window whose radius was ~5 times the average inter-nuclear distance.

The title on each figure indicates the temporal staging and number of embryos in each cohort. Staging is quantified by the degree to which cellularization has progressed (e.g. 15-35% cell wall invagination). Embryos showing any signs of gastrulation were excluded from consideration.