VIRTUAL EMBRYOS AS TOOLS FOR 3D GENE EXPRESSION ANALYSES

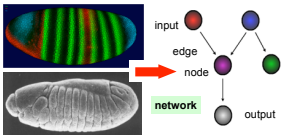
S.V.E. Keränen*, C.L. Luengo Hendriks*, C.C. Fowlkes^, L. Simirenko*, G.H. Weber``, O. Rübel~, M.-Y. Huang~
A.H. DePace*, C. Henriquez*, X.-Y. Li*, H.C. Chu*, D.W. Kaszuba*, A. Beaton*, S. Celniker*, B. Hamann~,
M.B. Eisen*, J. Malik^, D.W. Knowles*, M.DBiggin*.

*) Life Sciences and Genomics Division, Lawrence Berkeley National Laboratory, Berkeley, CA.
^) Computer Science Division, University of California, Berkeley, CA.
``) Computational Research Division, Lawrence Berkeley National Laboratory, Berkeley, CA.
~) Institute for Data Analysis and Visualization, University of California, Davis, CA.

BDTNP
Berkeley Drosophila Transcription Network Project

http://bdtnp.lbl.gov/

1 The Berkeley Drosophila Transcription Network Project (BDTNP) is a multidisciplinary collaboration studying the developmental regulatory network of *Drosophila* blastoderm embryos.



One component of this project maps the blastoderm expression patterns of 37 principal developmental regulatory genes and hundreds of their targets at cellular resolution, and uses these data to model potential regulatory interactions.

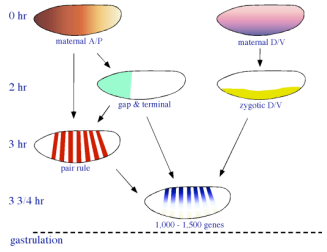
2 *Drosophila* embryo as a model for developmental regulatory networks

The basic bodyplan of *Drosophila melanogaster* embryo is determined during blastoderm stage by a cascade of regulatory interactions that read the maternal inputs into spatial information.

This converts genetically identical cells into different cell types.

Hence, to understand genome function, we need to record the development of local expression differences in a whole developing organism at a cellular resolution.

The BDTNP has developed a pipeline and methods for producing and analyzing 3D expression data from *Drosophila* stage 5 blastoderms at cellular resolution (Luengo et al. 2006 Genome Biol. 7:R123).



3 We have now generated such 3D data for 24 of the principal regulators and over 80 putative target genes, the latter selected using BDTNP ChIP-chip binding data and BDGP expression data.

Sample preparation >> Fluorescent *in situ* stains

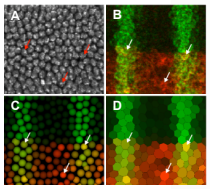
Imaging >> Confocal image stacks

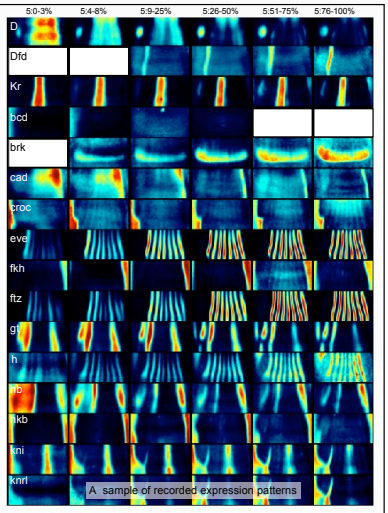
Nuclear segmentation >> PointCloud data

Registration >> Virtual PointCloud data

Visualization tools >> PointCloudXplore

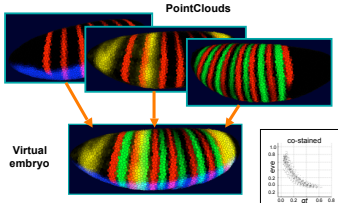
By recording nuclear DNA (A) and local levels of gene products (B), we can convert a 300 Mb confocal image stack into 1 Mb text file of the expression levels in each nucleus (C). This PointCloud data describes expression patterns in a computationally analyzable form (D).

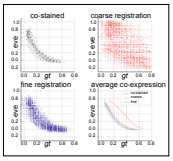




4 Because each imaged embryo contains expression information of only two genes, expression data from hundreds of embryos are mapped onto a virtual embryo to allow many genes' expression to be compared and modeled within each cohort.

The individual PointClouds each contain the information of only two gene products per cell for one embryo. Moreover, the equivalent cells in different embryos are in slightly different positions. To compare the spatial and temporal of many genes, we find the equivalent cells in multiple PointClouds by using the spatial information of one gene as a registration marker to align them all into a single *virtual embryo*. This significantly reduces the variability of the between embryo comparisons (Fowlkes et al. in preparation).

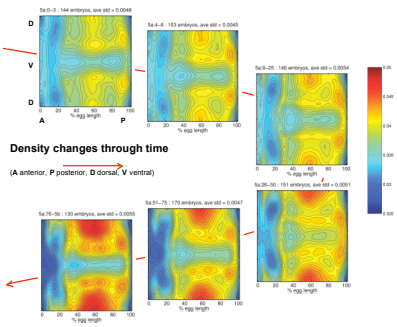




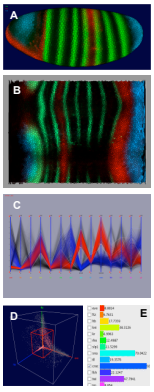
After fine registration, the expression profiles of *gt* and *eve* (stripe2) from two embryos closely resemble a co-stained embryo, while coarse registration produces wider scatter of expression (right) (Fowlkes et al. in preparation).

5 These virtual embryos contain nuclei placed to match the average density pattern and embryo shape for each cohort.

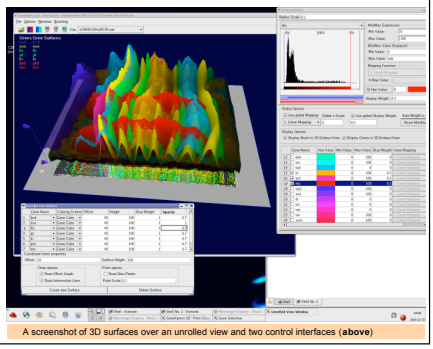
The nuclear positions shift during stage 5. Thus, any cellular resolution temporal analyses need nuclear resolution correspondence maps. We have generated a representative set of virtual embryos where the average number of nuclei at average local densities is traced through time (Keränen et al., 2006 Genome Biol. 7:R124).



6 Gene expression in such virtual embryos can be viewed with our tool called PointCloudXplore, which provides realistic interactive exploration of the 3D expression data as well as abstract views for analyzing the correlation between expression patterns within the N-dimensional gene expression space.



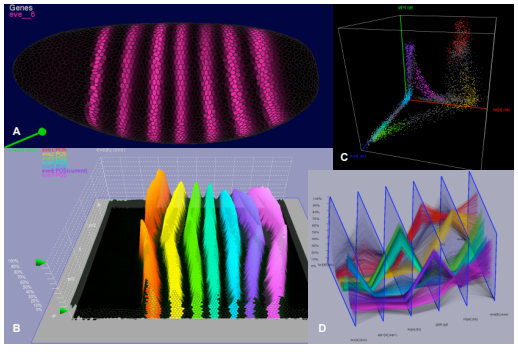
The expression data can be displayed in a realistic 3D embryo view (**A**) or in unrolled embryo view (**B**) that shows all cells simultaneously. The data can also be displayed in abstract expression space such as parallel co-ordinate view (**C**) or 3D scatter plot (**D**) which both show data for all cells, or in a cell magnifier (**E**), that shows the data from a single cell (Rübel et al., 2006 in *Data Visualization 2006*).



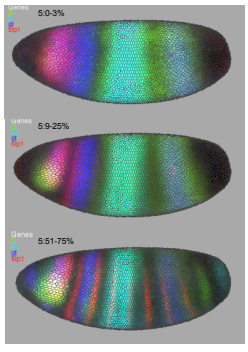
A screenshot of 3D surfaces over an unrolled view and two control interfaces (**above**)

7 The use of standardized virtual embryos allows temporal comparison within each nucleus between earlier expression of regulators in one cohort and the later expression of target gene patterns in another cohort, as well as better estimates of the developmental increase in complexity.

The expression of *eve* in cohort 6 (stage 5:76-100%) (**A**, **B**) compared to the expression patterns of gap genes in cohort 4 (stage 5:26-50%) (**C**, **D**). The expression levels of each *eve* stripe can be visualized and analyzed separately (**B**), and the relative expression levels of their putative regulators can be then studied in the corresponding nuclei either by analyzing their clustering in 3D scatter plots (**C**) or by observing the multigene expression profiles in parallel co-ordinate view (**D**) that also can contain 1D spatial information.

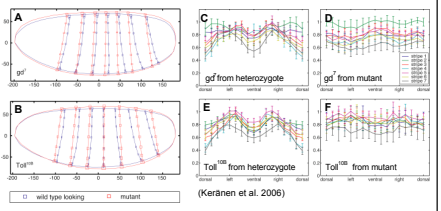


For example, the spatial expression of *D, Kr, gt* and *slp1* change during stage 5. In virtual embryos, such changes in multiple genes can be computationally analyzed in a standardized environment.



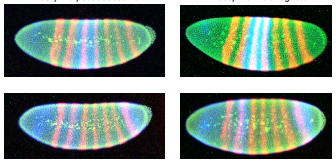
8 Gene expression data in regulatory factor mutant embryos and other *Drosophila* species is also being collected.

Dorsal-ventral signals affect the anterior-posterior pattern formation. In wild type *D. melanogaster*, the seven *ftz* stripes move closer together dorsally than ventrally. In dorsalized gd^7 mutants the ventral *ftz* stripes resemble wild type dorsal *ftz* stripes (**A**), whereas in ventralized $Toll^{10B}$ mutants the dorsal *ftz* stripe positions resemble the wild type ventral *ftz* stripes (**B**). The embryos from mutant mothers are compared to wild type looking embryos from heterozygous mothers of same stock. Also the intensity profiles of the stripes lose their dorsoventral polarity (**D**, **F**) that is seen in wild type looking embryos (**C**, **E**). Separate ChIP-chip data from BDTNP (X.-Y. Li et al. in preparation) indicates that dorsal-ventral factors dl, twi and sna often bind to regulatory regions of ftz and other anterior-posterior genes.



(Keränen et al. 2006)

Thumbnails of confocal image stacks for *D. melanogaster* and *D. pseudoobscura* stage 5 embryos (not in scale) triple stained for two mRNAs and nuclear DNA (right). The methods applicable to *D. melanogaster* blastoderm work with minor modifications on smaller *D. pseudoobscura* blastoderms (courtesy of AH DePace). *Kr* or *hb* are shown in blue, *eve* in red, and nuclei in green.



Unrolled view on comparison of *gt* mRNA expression in *D. melanogaster* (green) and *D. pseudoobscura* (red) aligned on *eve* mRNA expression (below). Note the relative positions of anterior *gt* stripes.