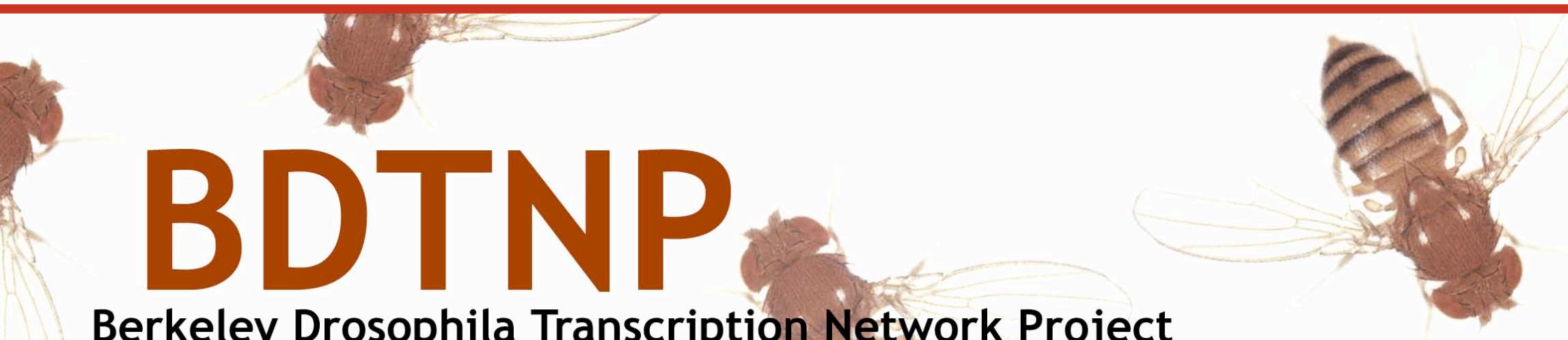
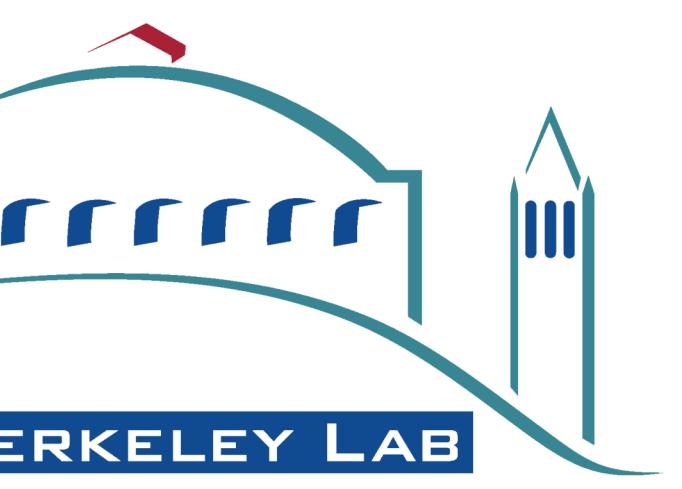


On computational analysis of quantitative, 3D spatial expression in *Drosophila* blastoderm

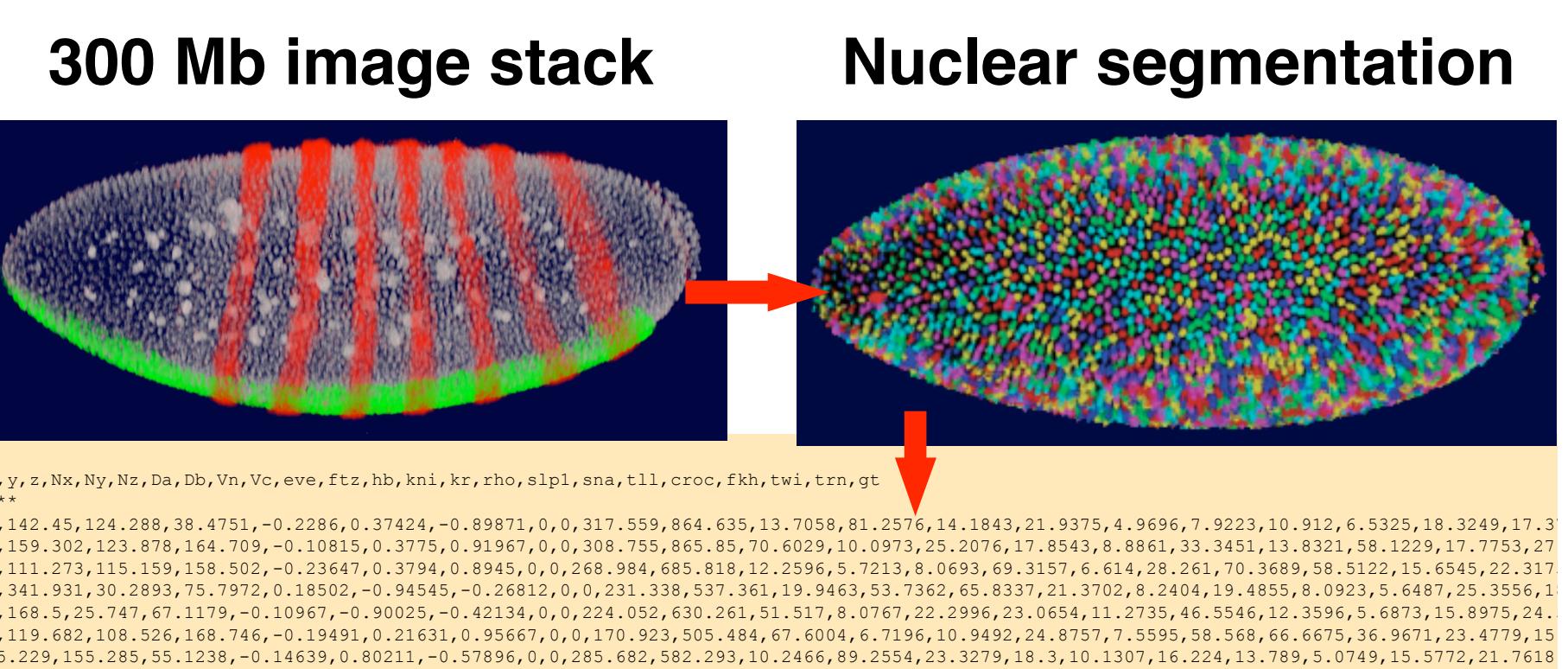


Soile V. E. Keränen¹, Angela DePace², Ann Hammonds¹, Bill Fisher¹, Oliver Rübel¹, Gunther Weber¹, Clara Henriquez¹, Charless Fowlkes³, Cris L. Luengo Hendriks⁴, Lisa Simirenko⁵, E. Wes Bethel¹, Hans Hagen⁶, Bernd Hamann⁷, Jitendra Malik⁸, Susan E. Celniker¹, David W. Knowles¹, Michael B. Eisen⁵, Mark D. Biggin¹

¹ Lawrence Berkeley National Laboratory, Berkeley, CA, USA; ² Harvard Medical School, Boston, MA, USA; ³ UC Riverside, Riverside, CA, USA; ⁴ Uppsala University, Uppsala, Sweden; ⁵ UC Berkeley, Berkeley, CA, USA; ⁶ University of Kaiserslautern, Kaiserslautern, Germany; ⁷ UC Davis, Davis, CA, USA

3D cellular resolution quantitation of spatial gene expression

The development of species specific morphologies results from complex, quantitative action of gene expression networks. The analysis of such networks requires computationally analyzable, cellular resolution datasets of spatial gene expression. The methods to construct them should be as robust and automated as possible to increase the speed of analysis and to decrease the error.

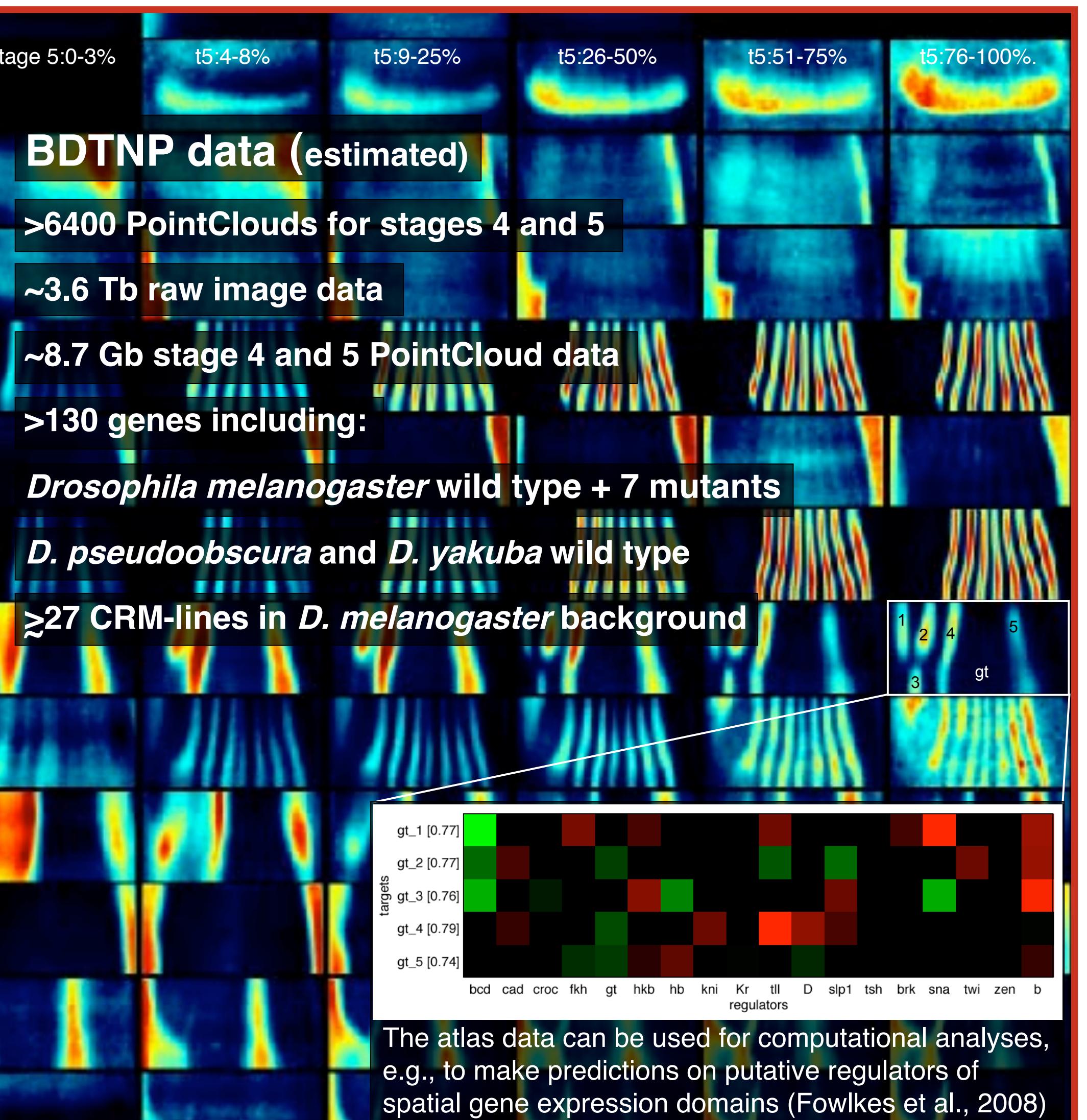
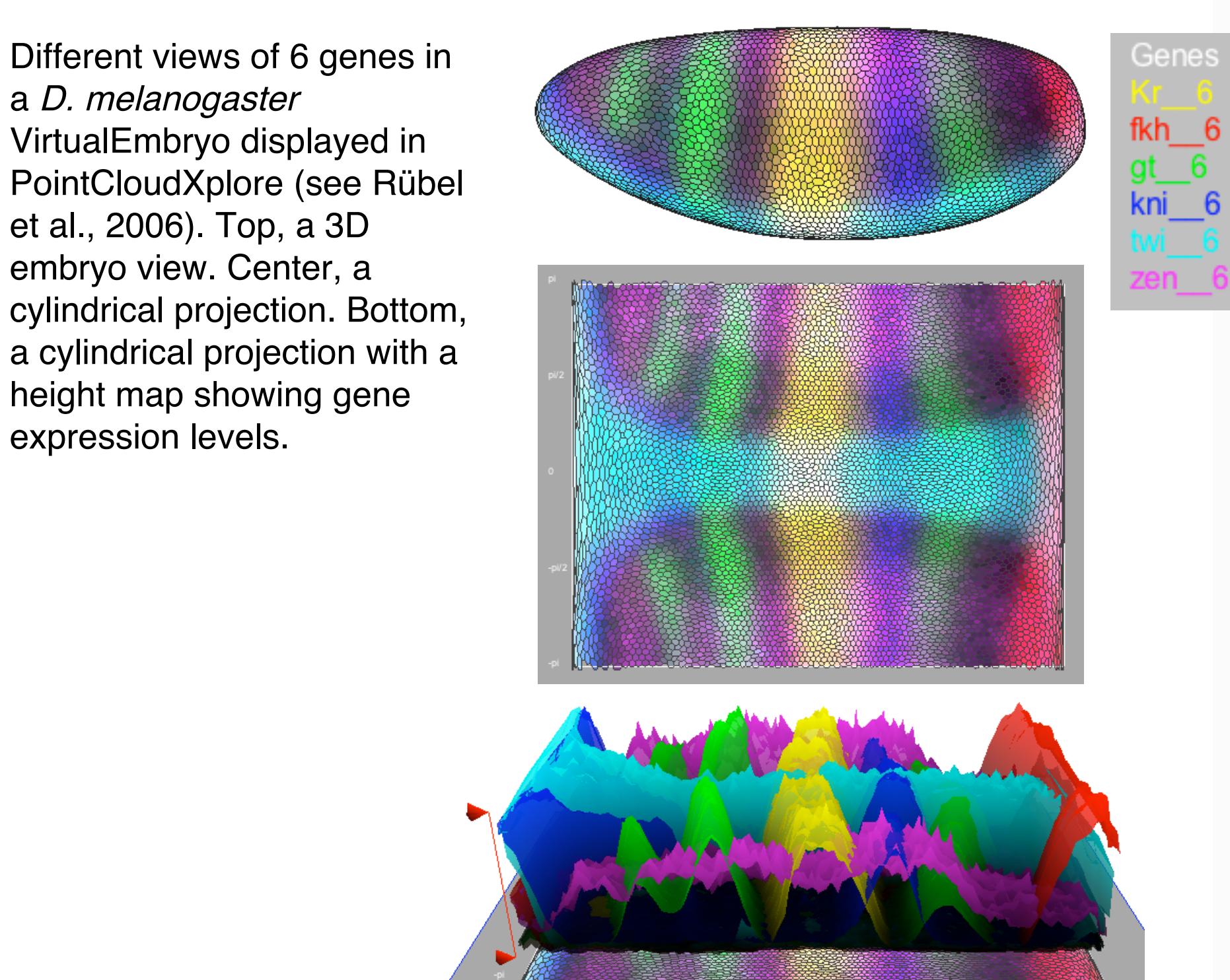


1 Mb text file for computational analyses

The Berkeley *Drosophila* Transcription Network Project (BDTNP) is developing cellular resolution, quantitative expression maps of *Drosophila* embryos in a computationally analyzable format. Files representing data from one embryo are called **PointClouds**. Multiple PointClouds are then aligned to a common framework, termed a **VirtualEmbryo**, to allow modeling and simulation of multiple genes regulation in a 3D environment.

For *Drosophila* blastoderm stage embryos, we now have data for the protein and mRNA expression of over 100 genes, 7 patterning mutant strains, ≥ 27 transgenic promoter constructs, and 3 *Drosophila* species.

(<http://bdtnp.lbl.gov/>)

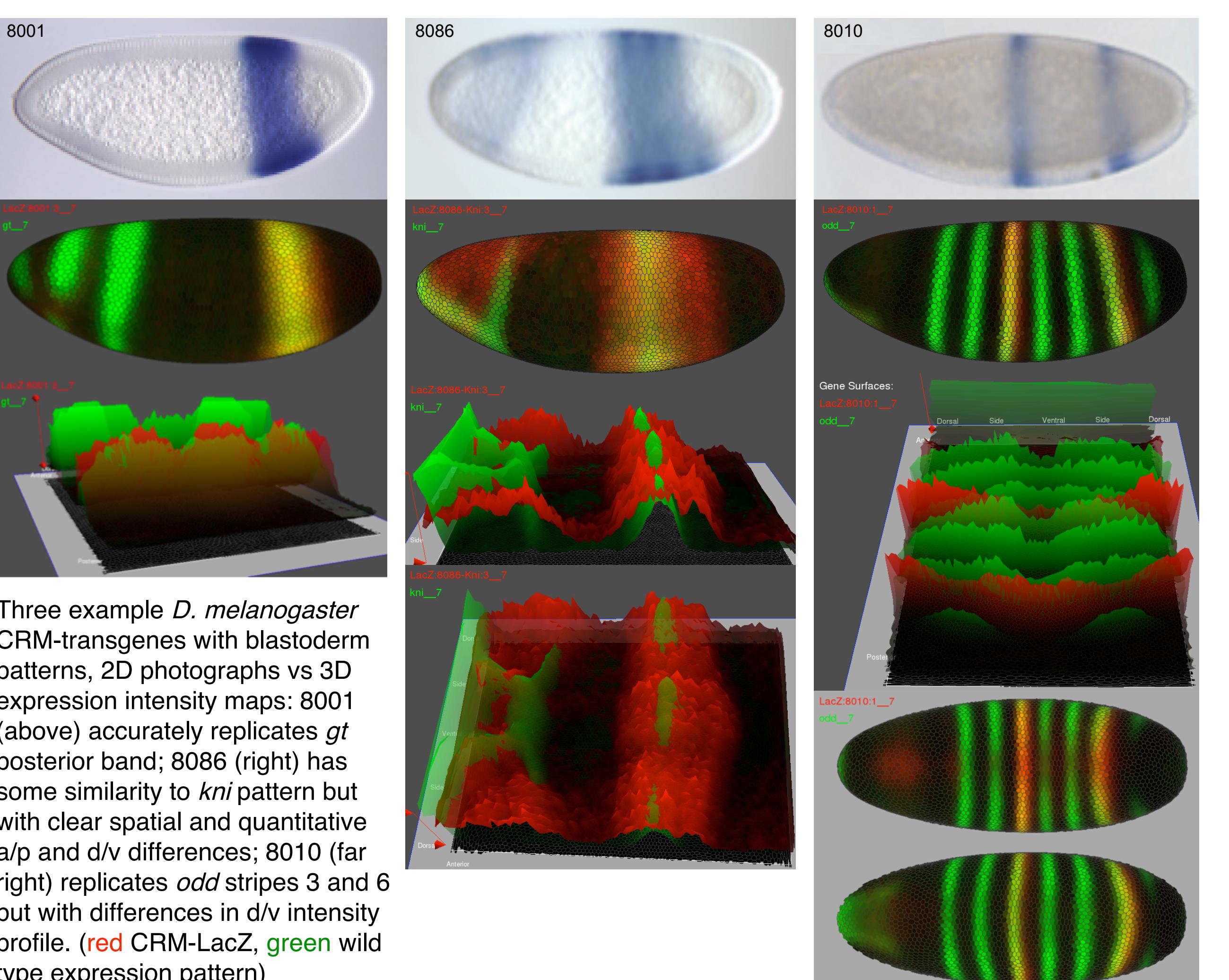


3D analyses of cis-regulatory output show subtle quantitative differences to the wild type pattern

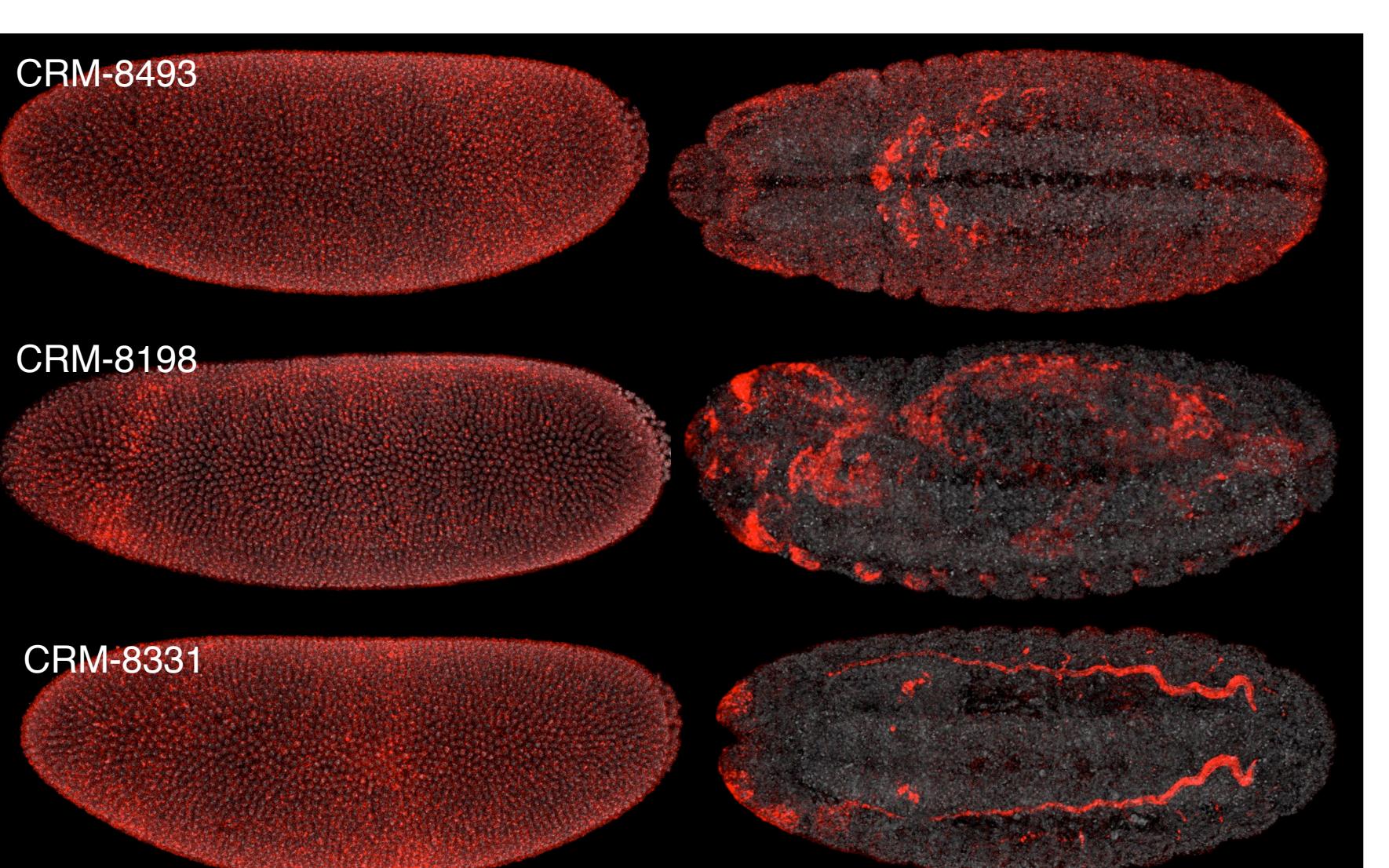
To analyze how sequence is read into expression pattern, the BDTNP are testing with expression constructs suspected or known *cis*-regulatory modules (CRMs), identified with sequence analyses. Visual examination of a 2D photographs shows that many of these putative CRMs partially recapitulate the wild type blastoderm patterns.

To make cellular resolution comparisons between the expression output of a CRM-transgene and the wild type expression, we have measured the 3D expression patterns of a set of CRM-transgenes and aligned them to a VirtualEmbryo containing wild type gene expression patterns.

We found that most of the studied CRM-transgenes have subtle or not so subtle quantitative differences in expression compared to intact genes. While discrepancies between CRMs and intact gene patterns have been noted in a few case before (e.g., Schroeder et al., 2004) more often the similarities have been emphasized. In comparison, our results suggest that quantitative and qualitative differences are so common that the current gene regulatory models based on CRM sequences need to be calibrated against actual experimental data.



CRMs showing weak to moderate blastoderm expression are expressed at much higher levels in late embryos, often at multiple stages and cell types.



Three example *D. melanogaster* CRMs with weak early pattern and strong late pattern in *D. melanogaster* embryos.

We believe that many CRMs that drive early gene expression are multifunctional. This is likely to impose extra evolutionary constraints in the sequence and structure of the CRM-promoter combinations. It is also possible that the early expression patterns are largely non-functional and simply tolerated because of the later gene function. Such patterns would be free to acquire new, essential functions, thus imposing further constraints on the CRM.

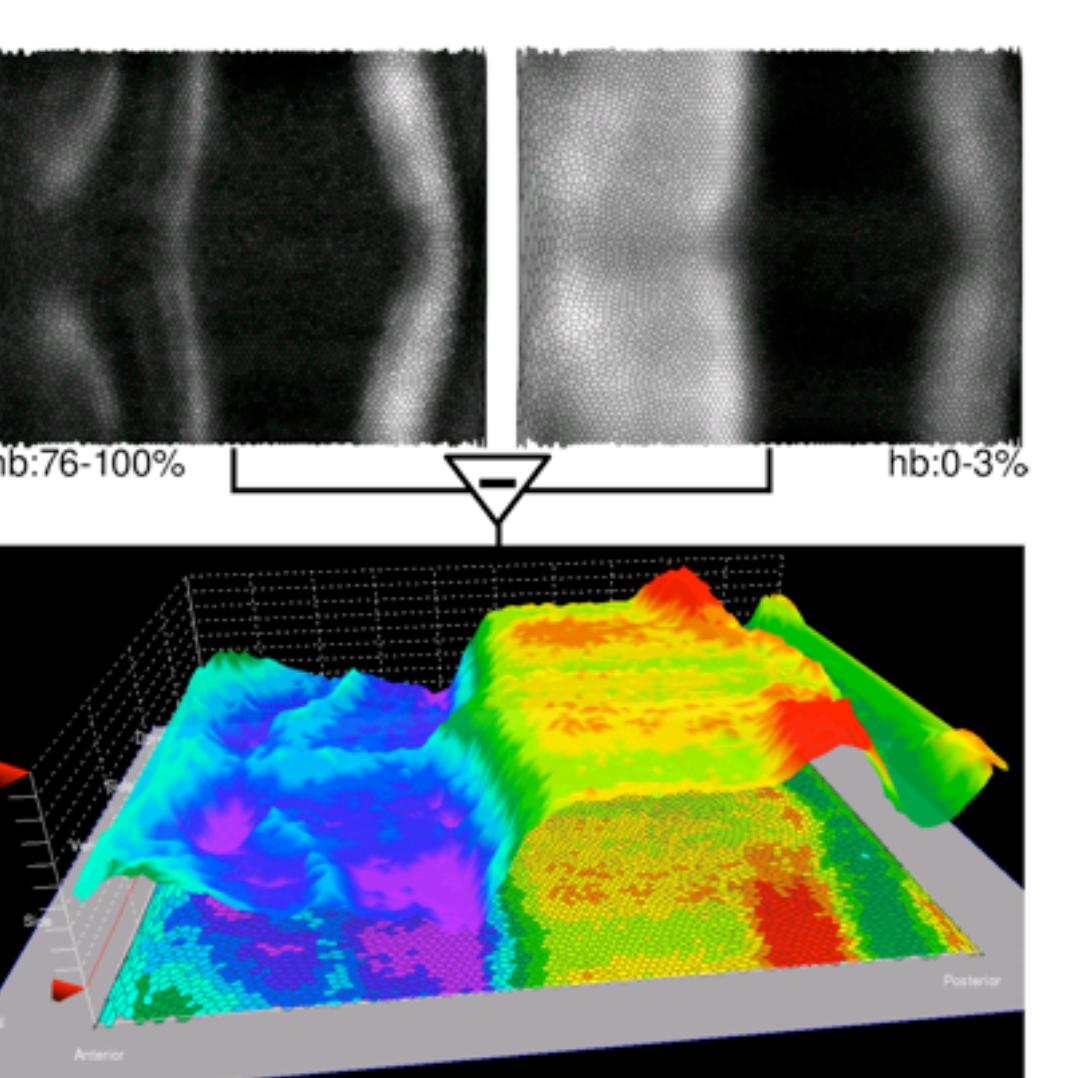
Interactive displays in computational modeling and mining the PointCloud data: PointCloudXplore/Matlab - interface

While the spatial patterns of regulator availability (network input) may have multiple implications on the function and evolution of regulatory sequences, a VirtualEmbryo represents the spatial expression profile of a blastoderm in large text file (tens of Mb), which can be nonintuitive to an average biologist. To facilitate the use of these quantitative cellular resolution embryo atlases we have provided a visualization and analysis tool, PointCloudXplore, which can display the expression data both as an embryo and as an abstract view.

To increase the flexibility of the PointCloudXplore, have now added to it a Matlab interface which can be used for exporting the PointCloud data for Matlab analyses, running the analysis scripts and/or importing the results back to the PointCloudXplore (Rübel et al., submitted). This enables versatile use of various Matlab functions for rapid development of new analysis and data plotting scripts without having to alter the PointCloudXplore itself.

As an example for a newly added data mining capability, we chose the early and late *hb* expression patterns (above right) with PointCloudXplore/Matlab interface. An associated short Matlab script then computed their difference and returned this output to the PointCloudXplore to be viewed as an artificial expression channel (below right).

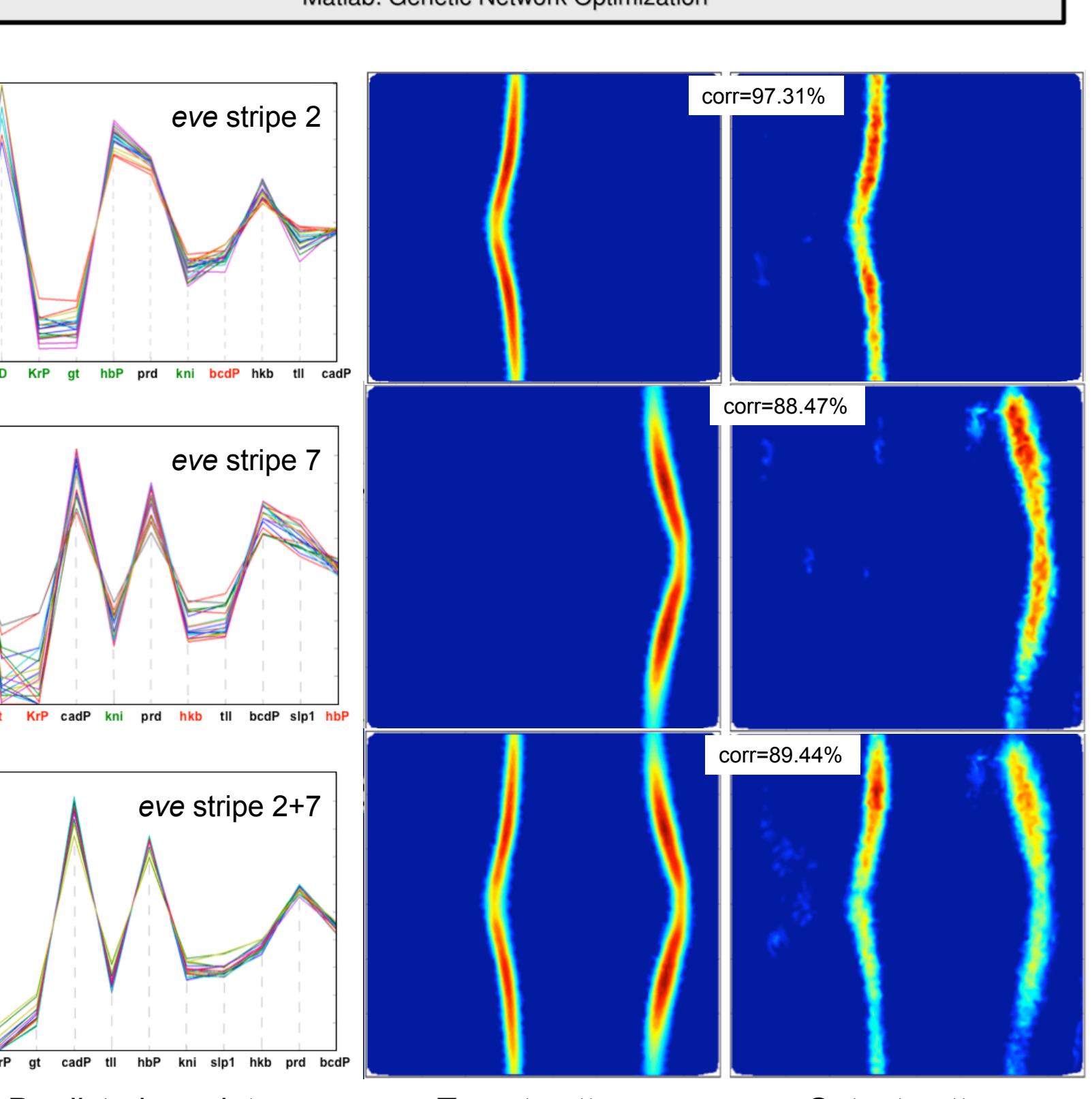
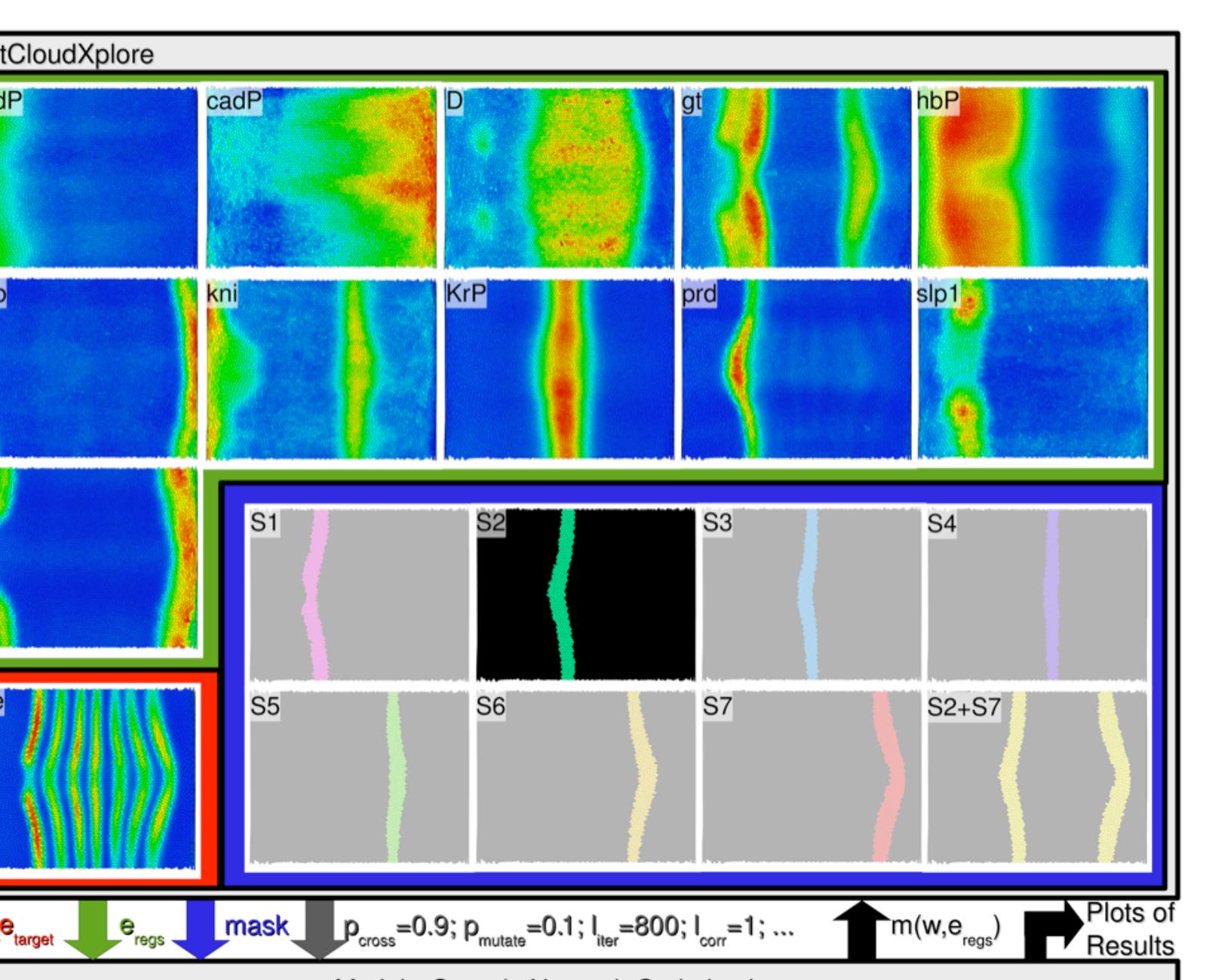
Sharing Matlab scripts for PointCloud data manipulation, for example, via curated databank would be of further advantage to the user community.



The PointCloudXplore/Matlab-interface can be used also for more complicated simulation and analysis programs. We have used the PointCloudXplore/Matlab-interface to run a genetic algorithm to see if we can identify potential regulators of individual *eve* stripes. We found that while the simulations were quite successful in identifying the known regulators of *eve* stripe 2, the results from *eve* stripe 7 were less promising. We assume that this is due to insufficient gene regulatory model for our optimization algorithm. We also found that *eve* stripes 2 and 7 could theoretically be co-regulated, as also suggested by some earlier authors (Hare et al. 2008, Janssens et al. 2006).

Since differences between computational modeling and biological experiments can be useful both for revealing weaknesses in theoretical models and for discovering new phenomena, we believe that cellular resolution computer atlases of embryos and organs and *in silico* experimentation with these will become an important tool in biology, parallel to *in vitro* and *in vivo* experiments.

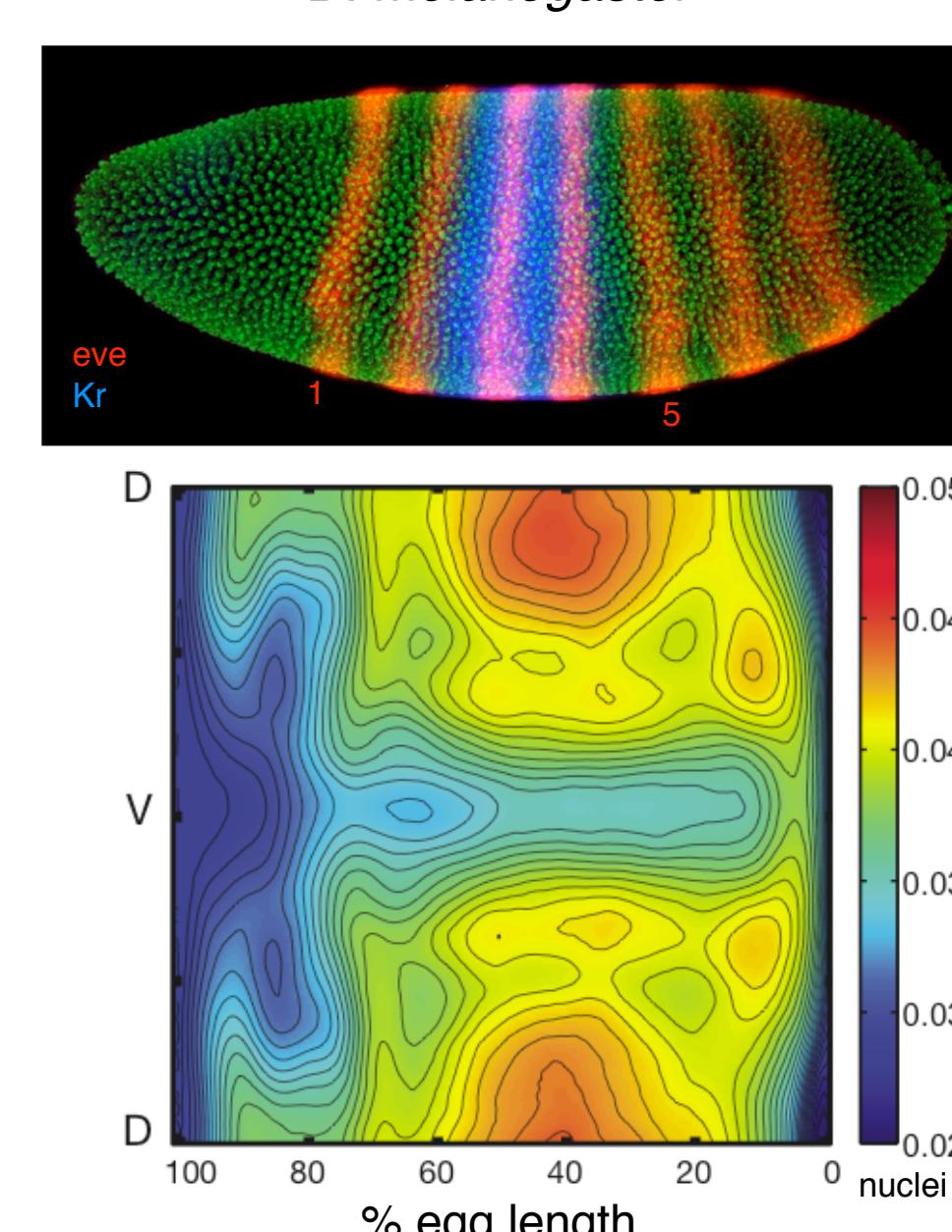
The input data (above right) used for attempt to computationally find of potential regulators of individual *eve* stripes, and the results (right). The predictions that agree with experimental data are shown in green, the predictions that disagree are shown in red and the results that neither agree nor disagree with experimental data are shown in black, as are the stripe 2+7 results. The stripe 7 results show that the output of a regulatory network can be mimicked by a set of regulators that is different from the experimentally verified one.



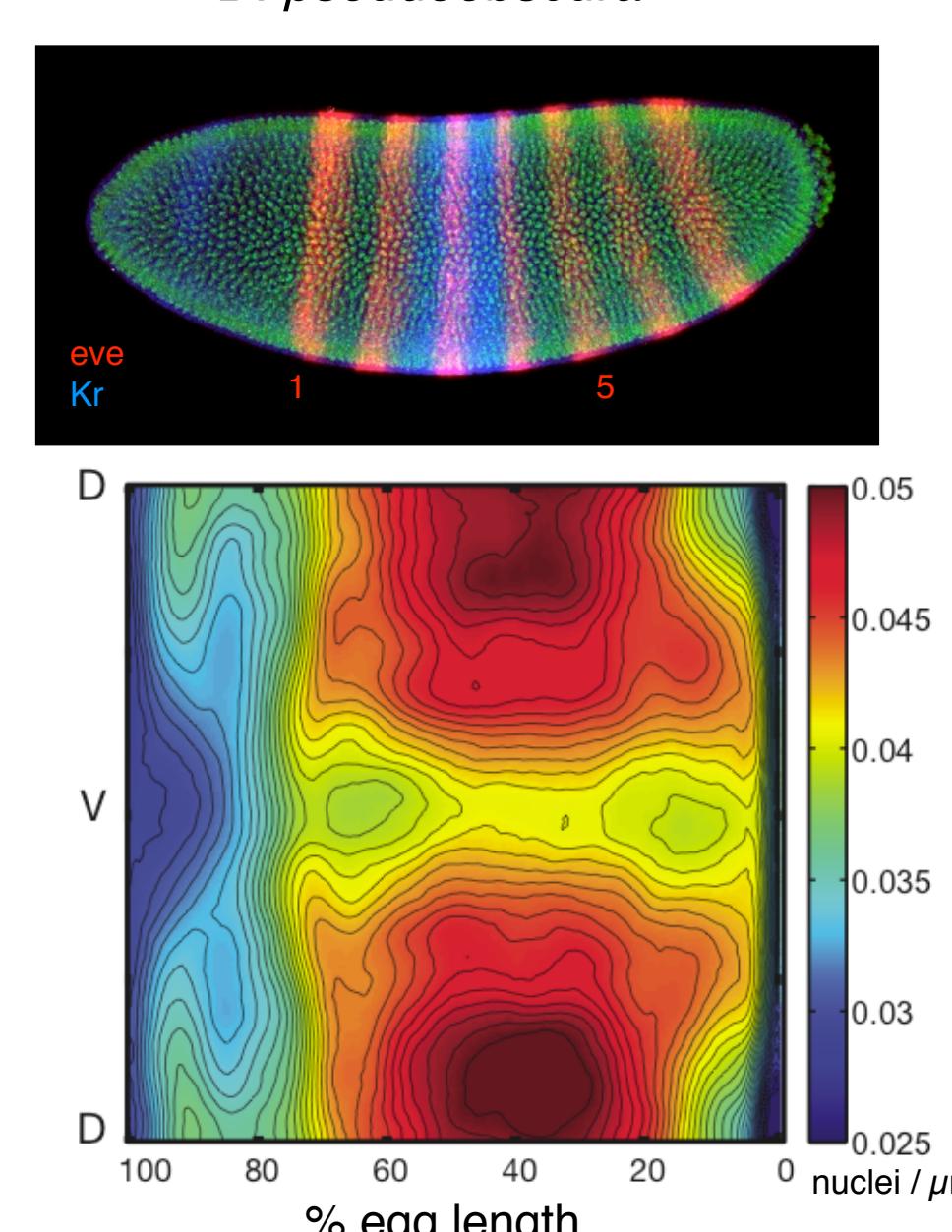
Computational embryology can reveal subtle differences between species

With cellular resolution data, new morphological features and morphogenetic phenomena can be discovered, which will also lead to new insights in comparative embryology.

D. melanogaster



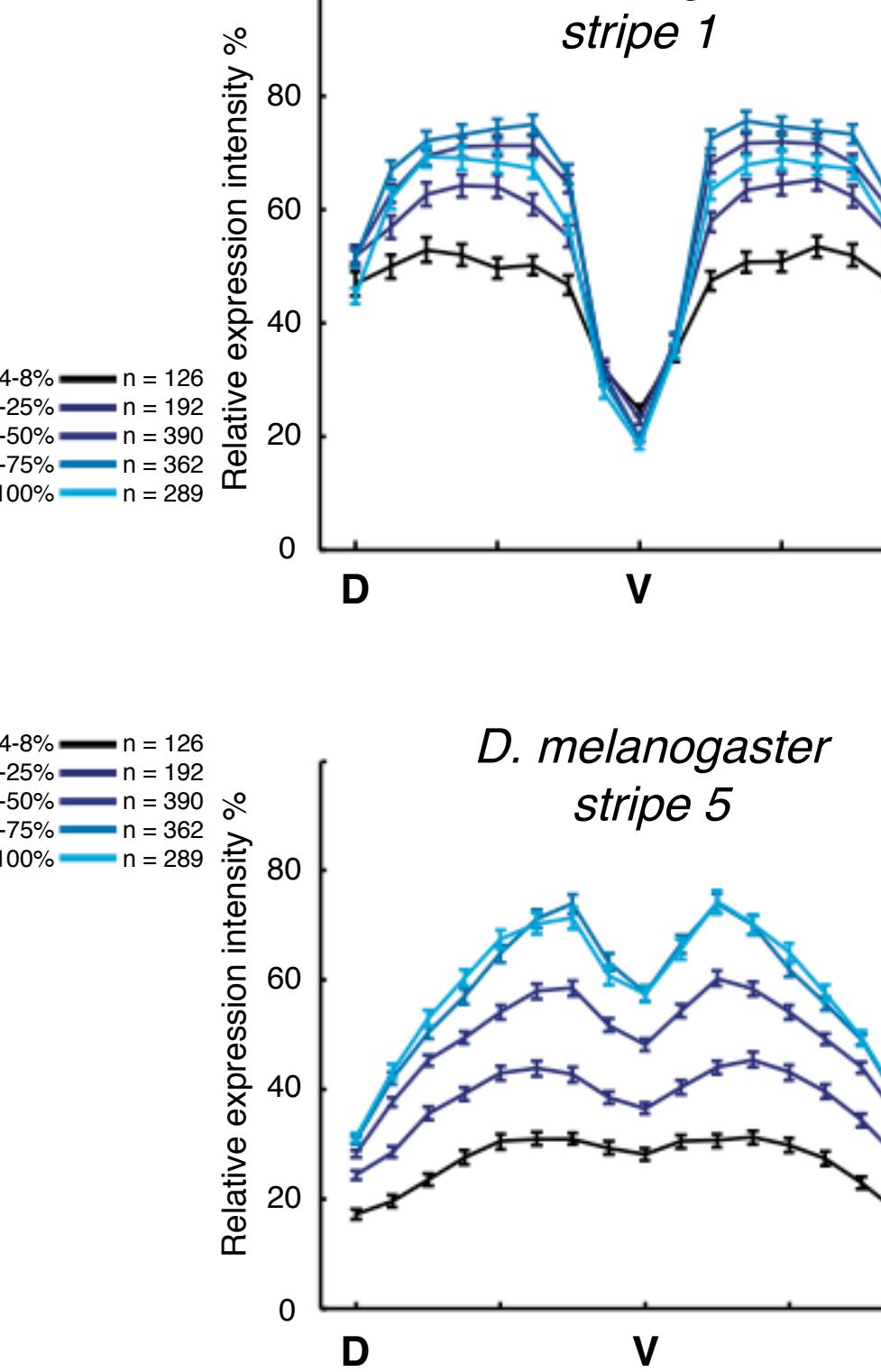
D. pseudoobscura



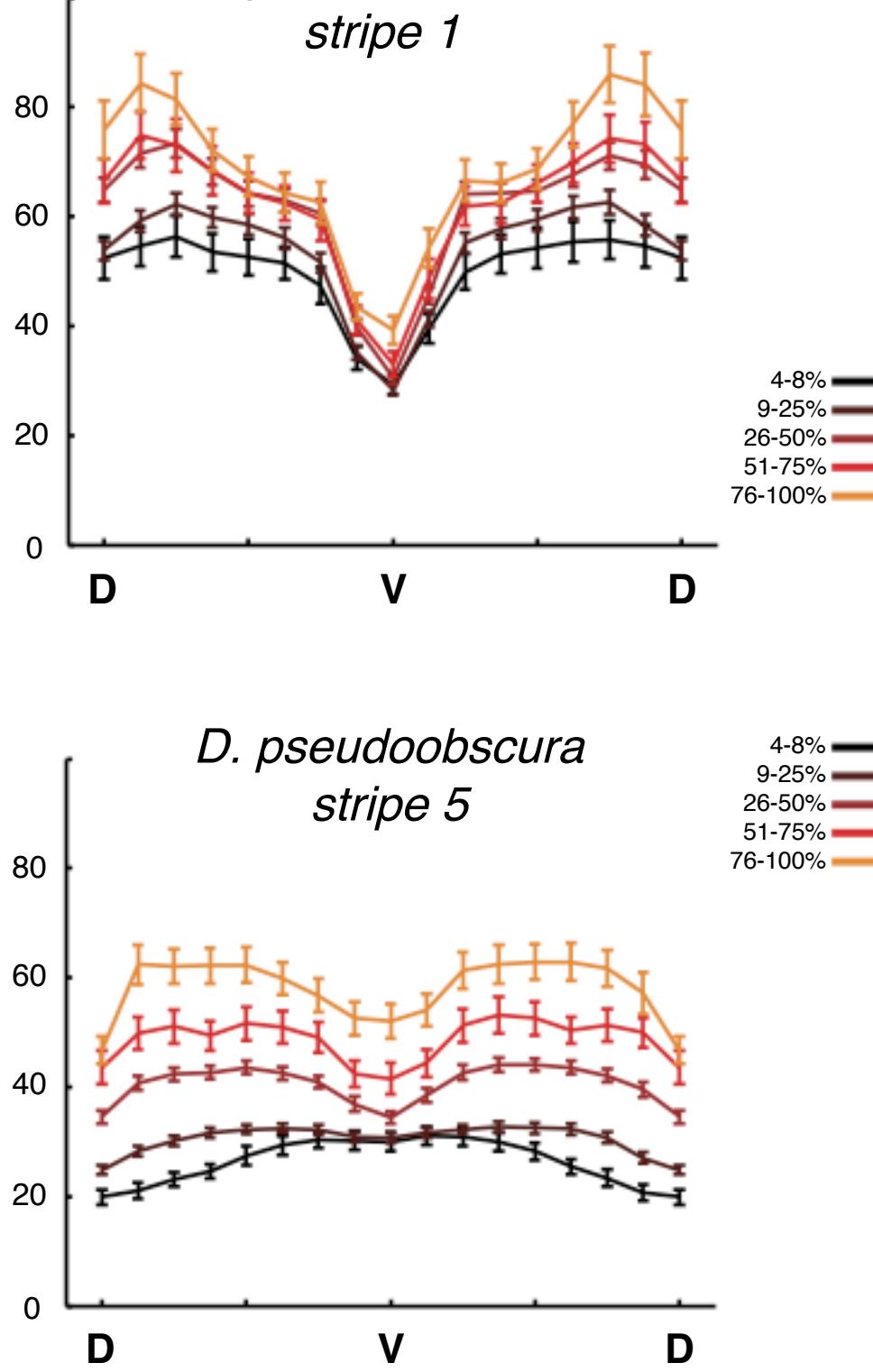
Analysis of *Drosophila melanogaster* blastoderm (left) shows significant anterior-posterior (AP) and dorsal-ventral (DV) nuclear density differences. Similar analyses on *D. pseudoobscura* blastoderm (right) show analogous but different density patterns.

The density differences between the species are likely to depend on subtle quantitative differences in the expression patterns of developmental regulators and in the cellular responsiveness to developmental signaling between the species.

D. melanogaster



D. pseudoobscura



Computational analyses of gene expression show the qualitatively similar *eve* stripe 1 and *eve* stripe 5 have relative intensity differences between *D. melanogaster* and *D. pseudoobscura*. The error bars show 95% confidence intervals.

We think that small quantitative expression differences are very common between even closely related species and that they play a significant role in evolution of species specific morphologies and in speciation in general.

Cellular resolution data can also be used for rapidly measuring features, that are theoretically analyzable by hand counts, such as the total number of nuclei and egg sizes, in large numbers of embryos (left). With a sufficiently large sample size, we can show that nuclear numbers at stage 5 scale to the egg size in both *Drosophila* species, and that the distributions of nuclear numbers are different, though individual embryos may fall within the parameters of another species.

Further basic measurements using large data sets are likely to reveal further developmental rules, or constraints to morphological evolution.

Future prospects

Uses of cellular resolution embryo atlases:

Digital reference maps of gene expression and fine level anatomical detail from anatomy to systems genomics ("computational histology")

Novel form of computational biology resources for data mining and rapid multidimensional analyses of morphogenesis, gene expression and potential modes of microevolution and speciation

Computational platforms for modeling and simulating gene regulatory network function and pattern formation in a realistic environment ("virtual embryos as *in silico* experimental model organisms")

We are currently:

Analyzing the transcription factor binding in putative CRMs to combine ChIP-chip and ChIP-seq data to the spatial pattern data

Expanding the expression pattern data set for wild type target gene mRNAs, wild type regulator proteins and CRM-LacZ reporter strains

Expanding the methods for generating cellular resolution embryo atlases in later stage *Drosophila* embryos

Developing approaches for computational analyses of spatial PointCloud data

Future needs:

PointCloud maps of multiple late embryonic stages and more species

Dissemination of new computational tools and 3D spatial datasets to *Drosophila* community and developmental biologists in general

Development of interactive ontology map for *Drosophila* embryo describing each cell's organ and tissue identity, expression profile, and lineage

More sophisticated algorithms and computational strategies for handling and analysis of spatial expression data and for integrating it with other forms of genomic data (sequence, microarray, protein-interaction, etc.)

References: Rübel et al., 2006 in Proc. Eurographics/IEEE-VGTC Symp. On Visualization
Schroeder et al., 2004 PLoS Biol. 2(9):E271.
Hare et al. 2008 PLoS Genet. 4(6):e1000106.
Janssens et al. 2006 Nature Genet. 38(10):1159-65.