



Advanced Scientific Computing Research
Computer Science
FY 2007 Accomplishment

Visualization and Analysis of 3D Gene Expression Data

E. Wes Bethel*, Oliver Rübél, Gunther H. Weber

Lawrence Berkeley National Laboratory

Bernd Hamann

University of California, Davis

Hans Hagen

University of Kaiserslautern, Germany

Summary

Recent methods for extracting precise measurements of spatial gene expression patterns from three-dimensional (3D) image data opens the way for new analysis of the complex gene regulatory networks controlling animal development. To support analysis of this novel and highly complex data we developed PointCloudXplore (PCX), an integrated visualization framework that supports dedicated multi-modal, physical and information visualization views along with algorithms to aid in analyzing the relationships between gene expression levels. Using PCX, we helped our science stakeholders to address many questions in 3D gene expression research, e.g., to objectively define spatial pattern boundaries and temporal profiles of genes and to analyze how mRNA patterns are controlled by their regulatory transcription factors.

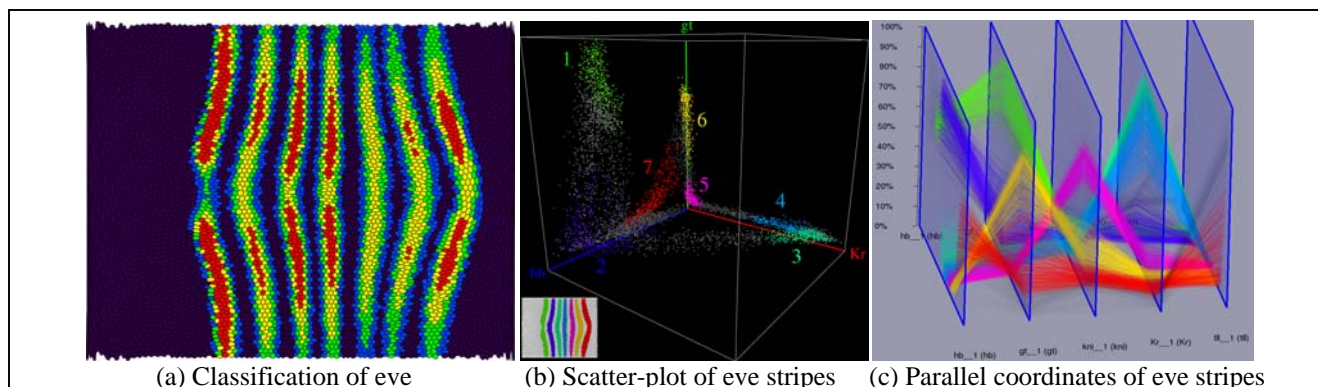


Figure 1. a) Classification of the expression pattern of *even skipped* (*eve*). Clustering detects the seven stripes of the *eve* pattern as well as intra-stripe variations. b) Each stripe of the *eve* pattern is marked by one cluster (bottom left). The clusters are then highlighted in a scatter-plot of the genes *Krüppel* (*Kr*), *hunchback* (*hb*), and *giant* (*Gt*), where large differences between the stripes are readily visible. The available data has implicated *gt*, *hb*, and *Kr* in controlling some stripes, but the scatter-plot suggests that these factors have the potential to each regulate all stripes by their unique combinations of expression levels. c) 3D parallel coordinates –with lines sorted according to physical cell positions— allow analysis of expression characteristics of the stripes in many gene dimensions.

Understanding the control of embryo development is a fundamental question in biology. Specific combinations of developmental regulatory factors, which form part of complex genetic regulatory networks and ultimately coordinate the expression of all genes, determine a cell's fate. As a result, the developing embryo exhibits an extraordinarily complex set of spatial

and temporal gene expression patterns. To gain insight into the complex genetic regulatory networks controlling gene expression, the *Berkeley Drosophila Transcription Network Project* (BDNTP) is generating 3D gene expression data that describes the output of the network at cellular resolution for about 100 genes at different stages of embryonic development. A

* (510) 486-7353, ewbethel@lbl.gov

large variety of questions can be addressed by performing analysis of these new data sets.

Analysis of such highly complex data is a challenging task. Of particular interest is the interplay between different types of genes, particularly determining the relationships between expression levels across a set of genes. For example, cells in a given region of the organism may exhibit high expression levels of one gene but a low expression level of another. Determining these relationships is a non-trivial effort.

In collaboration with the BDTNP life sciences researchers, we developed PCX, a flexible tool for integrated interactive visual exploration and analysis of the temporally varying 3D gene expression data. Multi-modal physical and information visualization views, such as scatter-plots and parallel coordinates, are linked via the concept of cell selection. A user selects cells of interest in any view, and then those selected cells are highlighted in all other views.

We integrated data clustering into PCX to aid in exploration and analysis of 3D gene expression data. We developed a dedicated cluster quality measure based on physical scattering of clustering results to improve cluster validation and to aid in determining the appropriate number of clusters. The interplay of data visualization and data clustering leads to improved visualization and enables more detailed analysis than previously possible.

Using PCX, many relevant questions in 3D gene expression research can be addressed. Clustering can, e.g., be used to objectively define spatial pattern boundaries (see Figure 1). Characteristic details within a pattern can be identified (Figure 1a) as well as detailed comparison of expression characteristics of different domains of an expression pattern becomes possible (Figure 1b, c). PCX has proven useful for temporal gene profile analysis. Using PCX, biologists were able to discover more complex changes in expression patterns than before. To analyze how expression patterns are controlled by their regulatory transcription factors, we used PCX to identify groups of cells that behave similar with respect to the expression of potential regulators (see Figure 2). In this case, the analysis suggested, e.g., that the gene *hb* may be able to mediate differential transcription along

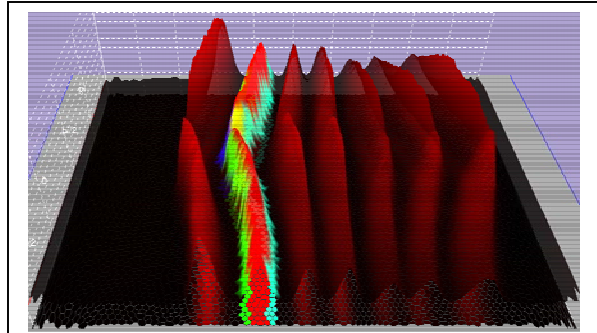


Figure 2. Surface plot of *eve*. By using the regulators *gt*, *hb*, and *Kr* of *eve* as input to data clustering different characteristic parts of the second stripe can be identified indicating that these parts are regulated differently.

both body axes of the embryo and not only on the anterior/posterior axis as typically thought.

The software, PCX, is distributed using an Open Source license. The software is available through the BDTNP website (bdtnp.lbl.gov) alongside the time-varying, 3D gene expression data.

Publications:

[1] O. Rübél, G. H. Weber, M.-Y. Huang, E.W. Bethel, M.D. Biggin, C.C. Fowlkes, C.L. Luengo Hendriks, Member, IEEE, S.V.E. Keränen, M.B. Eisen, D.W. Knowles, J. Malik, H. Hagen, and B. Hamann, "Integrating Data Clustering and Visualization for the Analysis of 3D Gene Expression Data," Submitted to IEEE Transactions on Computational Biology and Bioinformatics, 2007.

[2] G.H. Weber, O. Rübél, M.-Y. Huang, A.H. DePace, C.C. Fowlkes, S.V.E. Keränen, C. L. Luengo Hendriks, H. Hagen, D. W. Knowles, J. Malik, M. D. Biggin, and B. Hamann, "Visual exploration of threedimensional gene expression using physical views and linked abstract views," Accepted for Publication in IEEE Transactions on Computational Biology and Bioinformatics, 2007.

[3] O. Rübél, G.H. Weber, S.V.E. Keränen, C.C. Fowlkes, C.L. Luengo Hendriks, L. Simirenko, N.Y. Shah, M.B. Eisen, M.D. Biggin, H. Hagen, J.D. Sudar, J. Malik, D.W. Knowles, and B. Hamann, "PointCloudXplore: Visual analysis of 3d gene expression data using physical views and parallel coordinates," in Data Visualization 2006 (Proceedings of EuroVis 2006), B. S. Santos, T. Ertl, and K. Joy, Eds., Aire-la-Ville, Switzerland, 2006, pp. 203–210.

For further information on this subject contact:

Name: E. Wes Bethel.

Organization: Lawrence Berkeley National Laboratory.

Email: ewbethel@lbl.gov

Phone: (510) 486-7353

Posters and Interactive Demos:

[4] O. Rübél, G.H. Weber, M.-Y. Huang, E.W. Bethel, M.D. Biggin, C.C. Fowlkes, C.L. Luengo Hendriks, S.V.E. Keränen, M.B. Eisen, D.W. Knowles, J. Malik, H. Hagen, and B. Hamann, Applications of Visualization and Data Clustering to 3D Gene Expression Data, accepted for IEEE Visualization 2007, Sacramento, California, USA, October, 2007

[5] O. Rübél, G.H. Weber, M.-Y. Huang, C.C. Fowlkes, S.V.E. Keränen, C.L. Luengo Hendriks, M.D. Biggin, H. Hagen, D.W. Knowles, J. Malik, J.D. Sudar and B. Hamann, Interactive visualization of measured gene-expression patterns in three dimensions at cellular resolution, Exhibit at: Supercomputing 2005, Seattle, Washington, U.S.A., November, 2005

[6] G.H. Weber, O. Rübél, M.-Y. Huang, C. C. Fowlkes, S.V.E. Keränen, C. L. Luengo Hendriks, M.D. Biggin, H. Hagen, D.W. Knowles, J. Malik, J.D. Sudar and B. Hamann, Interactive visualization of measured gene-expression patterns in three dimensions at cellular resolution, Interactive demo at: IEEE Visualization 2005, Minneapolis, Minnesota, U.S.A., October, 2005.