PointCloudXplore: Visual Analysis of 3D Gene Expression Data Using Physical Views and Parallel Coordinates

Oliver Rübel¹, Gunther H. Weber², Soile V.E. Keränen³, Charless C. Fowlkes⁴, Cris L. Luengo Hendriks³, Lisa Simirenko³, Nameeta Y. Shah², Michael B. Eisen³, Mark D. Biggin³, Hans Hagen¹, Damir Sudar³, Jitendra Malik⁴, David W. Knowles³ and Bernd Hamann²

¹ International Research Training Group “Visualization of Large and Unstructured Data Sets,” University of Kaiserslautern, Germany

² Institute for Data Analysis and Visualization (IDAV), University of California, Davis, CA, USA

³ Life Sciences and Genomics Divisions, Lawrence Berkeley National Laboratory (LBNL), CA, USA

⁴ Computer Science Division, University of California, Berkeley, CA, USA
1. Genetic Fundamentals

2. Data- and Visualization Pipeline

3. Basic Principles

4. Visualization of Gene Expression in Physical Space
   4.1 Defining Embryo Orientation and Position
   4.2 Orthographic Views
   4.3 Unrolled View
   4.4 Expression Surfaces

5. Visualization in Gene Expression Space Using Parallel Coordinates
   5.1 Varying Line Properties
   5.2 Line Trace Highlighting and Animation
   5.3 Extending Parallel Coordinate View to 3D
   5.4 Brushing in Parallel Coordinate Views

6. Conclusions and Future Work

---

Oliver Rübel
Gene Expression is the amount of protein produced inside a cell using the genetic information of one gene.

Proteins are involved in practically every function performed by a cell e.g. as enzymes, structural proteins or as regulatory proteins.

Regulatory proteins are responsible for regulation of gene expression.

Development of any organism is guided by complex genetic regulatory networks.
2. Dataset- and Visualization Pipeline

- Dataset Acquisition
- Image Acquisition / IA
- Image Segmentation / IS
- Embryo Registration / ER
- PointCloudXplore / PCX

Oliver Rübel
EuroVis 2006 - PointCloudXplore: Visual Analysis of 3D Gene Expression Data Using Physical Views and Parallel Coordinates
2.1 Image Acquisition

- Embryo is *mounted* and *stained* for nuclear DNA and two genes.
- Image the embryo using 2-photon fluorescent confocal microscopy.
- One image contains the whole embryo with enough resolution to segment individual nuclei.
- One image stack is about 0.5 GB.
2.2 Image Segmentation

Image Acquisition ➔ Image Segmentation ➔ Embryo Registration

id, x, y, z, Nx, Ny, Nz, Vn, Vc, Sytox, Cy3n, Cy3a, Cy3b, Cy3g, Coun, Coua, Cub, Cou_g
1, 102.36, 142.14, 112.00, -0.396, 0.851, 0.644, 207.96, 605.36, 52.18, 23.55, 27.76, 22.55, 22.10, 11.95, 8.13, 28.01, 12.04
2, 264.63, 172.01, 79.36, 0.103, 0.972, -0.208, 281.73, 599.90, 82.12, 31.67, 34.97, 15.95, 51.93, 21.06, 12.56, 41.40, 19.12
3, 225.91, 174.99, 88.65, -0.030, 0.999, -0.015, 185.79, 418.35, 85.32, 35.63, 31.27, 14.77, 34.00, 19.59, 20.53, 38.80, 21.35
4, 318.42, 48.34, 138.91, 0.095, -0.744, 0.660, 182.46, 464.19, 37.61, 19.31, 15.15, 12.47, 17.55, 21.01, 13.76, 26.87, 17.53
5, 110.18, 34.40, 109.65, -0.186, -0.913, 0.362, 127.81, 432.01, 55.78, 24.12, 23.53, 12.39, 19.71, 13.81, 7.57, 28.16, 12.40
6, 340.48, 73.79, 37.54, 0.205, -0.299, -0.931, 208.26, 607.49, 80.23, 35.04, 25.75, 21.24, 28.91, 31.48, 20.65, 50.45, 26.96
2.3 Embryos Registration

Image Acquisition → Image Segmentation → Embryo Registration

- *eve Source Embryo*
- *snail Source Embryo*
- *Target Embryo*
3. Basic Principles

Multiple Views

- Different views emphasize different data properties.
- Viewing data from different perspectives supports detailed data analysis.

Brushing and Linking

- **Brushing**: Allow selection of data and emphasize selected data parts.
- **Linking**: Indicate visually which parts of one data display correspond to that shown in another display.

Fundamental idea

Mark interesting data parts according to data properties visualized in one view and look for other properties of selected data in other views.

→ One can use different ways to select data parts using different type of brushes, but all brushes can be displayed in all views.


4. Visualization of Gene Expression in Physical Space

Mapping gene expression to color:

\[
\begin{align*}
\text{color}_{ij} &= \text{HSV}(\text{geneColor}_j, 1.0, \text{value}_{ij}) \\
\text{value}_{ik} &= \begin{cases} 
0 & : \text{exprLevel}_{ij} < \text{minDisp}_j \\
\frac{\text{exprLevel}_{ij} - \text{minDisp}_j}{\text{maxDisp}_j - \text{minDisp}_j} & : \text{minDisp}_j \leq \text{exprLevel}_{ij} \leq \text{maxDisp}_j \\
1.0 & : \text{exprLevel}_{ij} > \text{maxDisp}_j
\end{cases}
\end{align*}
\]

Color of cell with index i:

\[
\text{color}_i = \sum_{j=1}^{n} (\text{color}_{ij} \ast \text{displayWeight}_j) + \sum_{k=1}^{m} (\text{color}_{ik} \ast \text{displayWeight}_k)
\]

4.1 Defining Embryo Orientation and Position

- Translate embryo into the point of origin of the world coordinate system by using the mean point of the PointCloud as reference point.
- Use principal component analysis to define the orientation of the PointCloud.
4.2 Orthographic Views

- Ventral
- Posterior
- Dorsal
- Anterior
- Left
- Right
4.3 Unrolled View

**Interacting with Embryo Views:**
- Interactive zooming, panning, and rotation
- Selection of cells made possible via drawing on the embryo surface

**Advantages**
- Qualitative analysis of gene expression values
- Analysis and comparison of spatial gene expression pattern
- Information about cell size and density
- Selection of cells according to cell positions
4.4 Expression Surfaces

Advantages

- Quantitative analysis of gene expression values
- Detailed view of data without losing overall context
- Spatial information preserved
5 Visualization in Gene Expression Space Using Parallel Coordinates

**Advantages**

- Overview of expression values of many genes, all cells
- Identification of possible correlations between genes
- Visual identification of line clusters
- Data selection in many data dimensions possible

---


5.1 Varying Line Properties

**Line transparency**
- Transparent data lines support detection of dense clusters in the data.

**Embryos Coloring**
- Color helps in tracing the topology of data lines and highlights line clusters.

**AP/DV Coloring**

**Distance to the first average**
- Transparent data lines support detection of dense clusters in the data.
- Color helps in tracing the topology of data lines and highlights line clusters.

5.2 Line Trace Highlighting and Animation

**Line Trace Highlighting**

- Possible to follow the topology of single data lines
- Moving the cursor along a parallel coordinate axis reveals information about the data line density and line distribution along that axis.

**Line Trace Animation**

- Subset of lines to be animated is defined by the user
- User can stop and pause the animation and also define the current time step of the animation.
5.3 Extending Parallel Coordinate View to 3D

**Advantage**

- Spatial dimensions are clearly separated from gene expression dimensions.
- Character of gene expression patterns are preserved in one dimension.
- Additional information about gene expression patterns and their relationships are revealed.
5.4 Brushing in Parallel Coordinate Views

Defining a Brush

Brush Bands

3D View

3D Dimensional Zoom
Future Work

- Integration of additional information visualization techniques, e.g., scatter plots
- Integration of automatic data analysis methods, e.g.,
  - unsupervised clustering
  - singular value decomposition (SVD)
Acknowledgements

http://graphics.idav.ucdavis.edu/research/vistools3dgeneexpression

- Visualization Group at the Institute for Data Analysis and Visualization, UC Davis
- Berkeley Drosophila Transcription Network Project at the Lawrence Berkeley National Laboratory
- Life Sciences and Genomics Divisions, Lawrence Berkeley National Laboratory
- Computer Science Division, UC Berkeley
- Visualization Group at the Lawrence Berkeley National Laboratory
- International Research Training Group “Visualization of Large and Unstructured Data Sets,” University of Kaiserslautern, Germany

- National Institutes of Health
- National Science Foundation
- Lawrence Berkeley National Laboratory
- Deutsche Forschungsgemeinschaft (German Research Foundation)