

Graphical Representation of Cell/Tissue Type Relationships for the Web

Raymond Wan¹, Hiroshi Mamitsuka¹, Larisa Kiseleva², and Paul Horton²

¹Bioinformatics Center, Institute for Chemical Research, Kyoto University, Gokasho, Uji, 611-0011, Japan

²Computational Biology Research Center, AIST, Tokyo, 135-0064, Japan

Microarrays allow thousands of genes to be analyzed simultaneously under various experimental conditions. Applying unsupervised machine learning techniques on these data sets enable researchers to construct dendrograms and clusters depicting the similarity between genes or experiments. In these dendrograms, edges are drawn between the most similar objects (whether they be genes or experiments).

Kiseleva et al. [2005] proposes a similar type of analysis specific to experiments. To be precise, the experiments show the expression levels of genes in cells and tissues from a multi-cellular organism. An example of such a data set is GDS596 from the NCBI Gene Expression Omnibus¹ (GEO) [Su et al., 2004]. It consists of expression levels of genes from 79 samples from the human body, including muscles and nervous tissue. Using this data, Kiseleva et al. first calculated the similarity between each cell or tissue using Spearman's rank correlation coefficient. Then, the relationships between tissues were depicted graphically by applying Kruskal's minimum spanning tree algorithm across the 79 nodes. The end result is a tree that shows tissue similarity in a way quite different from traditional clustering.

In this poster, we describe a web server that generalizes this method for other data sets by allowing users to upload GEO-formatted data. Since an online system imposes execution time and latency constraints, we show how the web server is made feasible through the use of gene (feature) selection criteria, such as variance and random selection. We also propose the use of Gene Ontology (GO) terms as a means to select genes. This reduces the overall execution time and improves the interpretability of the resulting tree. We believe this tool will become increasingly useful as more microarray data is made available, giving a larger picture of gene expression in multi-cellular organisms such as humans.

References

- L. Kiseleva, P. Horton, and W. Fujibuchi. Cell-distance networks: A clue for understanding of cell diversity. In *Proc. 16th International Conference on Genome Informatics, Posters and Software Demonstrations*, pages P029-1 to P029-2, December 2005.
- A. I. Su et al. A gene atlas of the mouse and human protein-encoding transcriptomes. *Proc. National Academy of Sciences USA*, 101(16):6062-6067, April 2004. Data: http://www.ncbi.nlm.nih.gov/geo/gds/gds_browse.cgi?gds=596.

¹<http://www.ncbi.nlm.nih.gov/geo/>