Neuronal Stem Cell Differentiation

Gene Yeo & Alexander Rakhlin

Artificial Intelligence Laboratory and
The Center for Biological and Computational Learning
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139

http://www.ai.mit.edu

The Problem: To elucidate the important genes involved in neuronal stem cell (NSC) differentiation using Affymetrix time series gene expression values obtained from rat mRNA. To formulate algorithms which will work well on this type of problems.

Motivation: Stem cells are important in the treatment of many diseases. Particularly, neuronal stem cells are important in the treatment of degenerative brain diseases such as Parkinson and Alzheimer. Understanding the mechanisms and pathways of how NSC differentiate into astrocytes, mature neurons, and possibly other cell types is extremely important. With the current advent of GeneChip technology, we are able to analyze massive amounts of gene expression data. Moreover, we can look at the time-course of genes that regulate differentiation. We are interested in formulating computational tools to perform such analysis.

Previous Work: Eisen et al [1] devised a system of cluster analysis for genome-wide expression data from DNA microarray hybridization to group genes involved in similar cellular processes by some measure of similarity in their pattern of expression. They used a form of correlation coefficient as a similarity metric and hierarchical clustering to group genes of "similar" function.

Approach: We first used various normalization techniques, such as Maximum-Likelihood Method [2], in order to remove noise and scale the data appropriately. We explored the validity of keeping or removing Affymetrix APM calls. We modulated the fold-change criterion to get a reasonable number of genes as a pre-analysis step. These genes were clustered according to their behavior over the four time points. These genes were later analyzed from a biological perspective.

Impact: This work opens a door for further biological validation of selected important genes. We realized the inherent complexity of the problem due to the sparsity of the data and uncertainty attributed to the technology.

Future Work: Introducing chemicals into the cells that halt protein synthesis means that transcription factors will not be around to up-regulate gene transcription/expression. Adding specific growth factors will up-regulate only the genes specific to a pathway. As new data from such experiments become available, we are interested in exploring genetic networks and pathways using computational methods [3].

Research Support: Research at CBCL is sponsored by grants from: Office of Naval Research (DARPA) under contract No. N00014-00-1-0907, National Science Foundation (ITR) under contract No. IIS-0085836, National Science Foundation (KDI) under contract No. DMS-9872936, and National Science Foundation under contract No. IIS-9800032 Additional support was provided by: Central Research Institute of Electric Power Industry, Center for e-Business (MIT), Eastman Kodak Company, DaimlerChrysler AG, Compaq, Honda R&D Co., Ltd., Komatsu Ltd., Merrill-Lynch, NEC Fund, Nippon Telegraph & Telephone, Siemens Corporate Research, Inc., Toyota Motor Corporation and The Whitaker Foundation.

References:

[1] Michael Eisen, Paul T. Spellman, Patrick O. Brown, and David Botstein. Cluster analysis and display of genome-wide expression patterns. *Proceedings of National Academy of Science*, 95:14863–14868, December 1998.

[2] A. Hartemink, D. Gifford, T. Jaakkola, and R. Young. Maximum likelihood estimation of optimal scaling factors for expression array normalization. in microarrays. *Proceedings of SPIE*, 4266, 2001.

[3] Dana Pe'er, Aviv Regev, Gal Elidan, and Nir Friedman. Inferring subnetworks from perturbed expression profiles. *Bioinformatics*, 17:S215–S224, 2001.