Visualizing and Refining Connectivity Map Query Results

October 7, 2013

Proposal for a
Thesis in the Field of
Biotechnology

In Partial Fulfillment of the Requirements
for a Master of Liberal Arts Degree

Harvard University
Extension School

Theodore Natoli
51 Lewis Avenue, Apartment 3
Arlington, MA 02472
857-498-1946
tnatoli@fas.harvard.edu

Proposed Start Date: October 7, 2013
Anticipated Date of Graduation: October 7, 2013
Thesis Director: Aravind Subramanian

Abstract

The Connectivity Map (CMap) is a database of gene expression signatures obtained from experiments in which cultured human cells are treated with pharmacologic and genomic perturbagens. A typical use case of this database is for a researcher to query with a signature of a cell state of interest and use the matching perturbagens to develop a functional hypothesis for follow-up.

Current pattern matching algorithms that perform CMap queries suffer from a universal weakness – the enormous size and richness of signatures in CMap means that a query typically generates hundreds of strong correlations which are hard to distinguish thereby making prioritization on the basis of a distance metric difficult.

We hypothesize that one mode of prioritization is to highlight query results that are highly interconnected amongst themselves over singletons. The goal of this work is to provide a web-based tool for implementing an interconnectivity-based method of query result refinement and of visualizing CMap query results in a graph layout.

1 Background

1.1 The Connectivity Map

The connectivity map is a database containing the gene expression profiles resulting from treating cultured cells with various chemical and genomic perturbations lamb2006.

1.2 Graphical Depictions of Biological Phenomena

1.3 Approach for This Work

2 Methods

2.1 Software Components

2.1.1 Front End: HTML & D3.js

2.1.2 Back End: MongoDB

3 Potential Challenges

4 Preliminary Timeline

References

[1] Justin Lamb et al, *The Connectivity Map: Using Gene Expression Signatures to Connect Small Molecules, Genes, and Disease.* Science, 313, 1994.