

Visualizing and Refining Connectivity Map Query Results

October 17, 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 17, 2013
Anticipated Date of Graduation: October 17, 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 (Justin Lamb, 2006).

1.2 Graphical Models

1.3 Graphical Depictions of Biological Phenomena

Although they originated in the field of computer science, graphs have frequently been used as tools to model biological phenomena. Graphical models of protein interaction networks, gene expression networks, and other similar phenomena are commonplace. Because of its widespread use and adoption, I believe the graphical model is an appropriate and effective means to depict connectivity in the CMap sense.

1.4 Approach for This Work

In this work, a graph will be used to depict the existence and strength of connections between gene expression signatures in the CMap database. Signatures will be represented as nodes and their pairwise connections by the edges. The final app might look something like this.

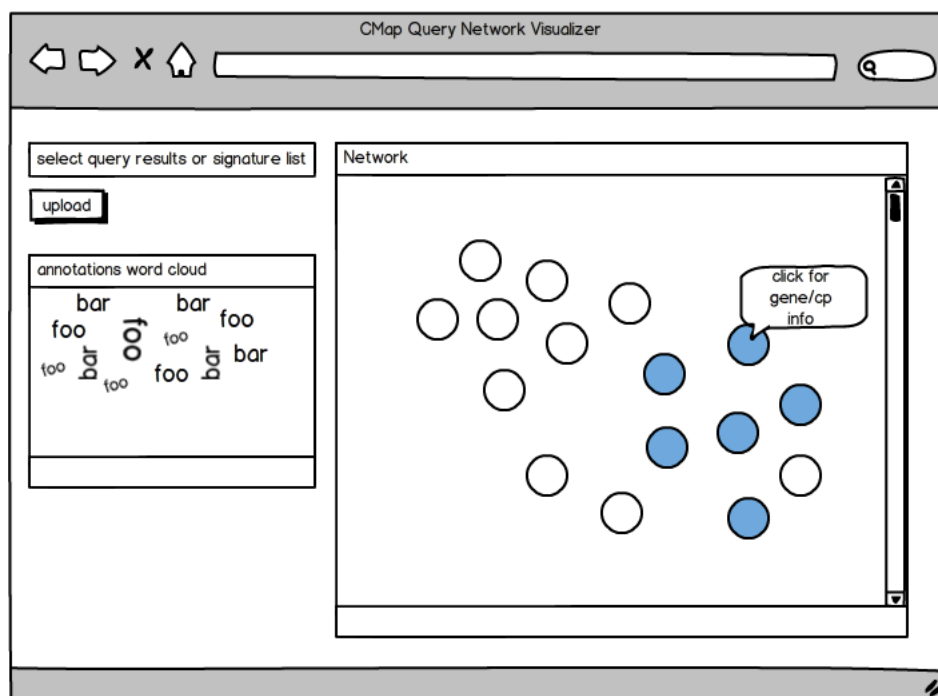


Figure 1: The App

2 Methods

2.1 Software Components

2.1.1 Front End: HTML & D3.js

2.1.2 Back End: Node.js & MongoDB

3 Potential Challenges

4 Preliminary Timeline

References

Justin Lamb, e. a. (2006). The connectivity map: Using gene expression signatures to connect small molecules, genes, and disease. *Science*, 313.