SAMNetWeb: identifying condition-specific networks linking signaling and transcription

The MIT Faculty has made this article openly available. Please share how this access benefits you. Your story matters.

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>As Published</td>
<td><a href="http://dx.doi.org/10.1093/bioinformatics/btu748">http://dx.doi.org/10.1093/bioinformatics/btu748</a></td>
</tr>
<tr>
<td>Publisher</td>
<td>Oxford University Press</td>
</tr>
<tr>
<td>Version</td>
<td>Author's final manuscript</td>
</tr>
<tr>
<td>Accessed</td>
<td>Thu Dec 27 14:42:43 EST 2018</td>
</tr>
<tr>
<td>Citable Link</td>
<td><a href="http://hdl.handle.net/1721.1/108517">http://hdl.handle.net/1721.1/108517</a></td>
</tr>
<tr>
<td>Terms of Use</td>
<td>Creative Commons Attribution-Noncommercial-Share Alike</td>
</tr>
<tr>
<td>Detailed Terms</td>
<td><a href="http://creativecommons.org/licenses/by-nc-sa/4.0/">http://creativecommons.org/licenses/by-nc-sa/4.0/</a></td>
</tr>
</tbody>
</table>
Systems Biology

SAMNetWeb: identifying condition-specific networks linking signaling and transcription

Sara JC Gosline†, Coyin Oh† and Ernest Fraenkel†*

†Department of Biological Engineering, Massachusetts Institute of Technology, 77 Massachusetts Ave, Cambridge, MA 02139.
*Corresponding author

Received on XXXXX; revised on XXXXX; accepted on XXXXX

Associate Editor: XXXXXXX

ABSTRACT

Motivation: High throughput datasets such as genetic screens, mRNA expression assays, and global phospho-proteomic experiments are often difficult to interpret due to inherent noise in each experimental system. Computational tools have improved interpretation of these datasets by enabling the identification of biological processes and pathways that are most likely to explain the measured results. These tools are primarily designed to analyze data from a single experiment (e.g. drug treatment vs. control), creating a need for computational algorithms that can handle heterogeneous datasets across multiple experimental conditions at once.

Summary: We introduce SAMNetWeb, a web-based tool that enables functional enrichment analysis and visualization of high throughput datasets. SAMNetWeb can analyze two distinct data types (e.g. mRNA expression and global proteomics) simultaneously across multiple experimental systems to identify pathways activated in these experiments and then visualize the pathways in a single interaction network. Through the use of a multi-commodity flow based algorithm that requires each experiment ‘share’ underlying protein interactions, SAMNetWeb can identify distinct and common pathways across experiments.


Contact: fraenkel-admin@mit.edu

1 INTRODUCTION

Due to large consortium-based data collection efforts such as ENCODE (Birney et al., 2007) and TCGA/TCGA Network, 2012), the quantity of biological signaling data is growing at an astounding rate. These data repositories release mRNA expression data, chromatin accessibility data, DNA sequencing data and proteomics data for hundreds of tissues and thousands of patients. As a result, there is a growing need for algorithms to meaningfully interpret diverse types of data across experimental conditions such as different diseases and tissues.

There are numerous network-based analysis tools (Tuncbag, Braunstein, et al., 2012; Yeger-Lotem et al., 2009), including their web-server counterparts (Lan et al., 2011; Tuncbag, McCallum, Huang, & Fraenkel, 2012) that perform visualization and functional enrichment of high throughput datasets within a single experiment. These algorithms improve the functional enrichment of basic statistical approaches (Eden, Navon, Steinfeld, Lipson, & Yakhini, 2009; Jiao et al., 2012) by identifying “hidden” nodes that are likely involved in the experiment but not measured as changing. However, these tools only analyze one experimental condition at a time, making visualization of multiple networks cumbersome. Furthermore, when applied to different perturbations in the same experimental system, these algorithms select similar networks (Gosline, Spencer, Ursu, & Fraenkel, 2012).

SAMNetWeb is able to analyze multiple experiments by mapping both signaling and transcriptomic datasets to the same underlying protein-protein interaction network. Representing each experiment as a ‘commodity’, the algorithm identifies common and distinct paths between the two datasets for each experiment using the SAMNet multi-commodity flow-based algorithm (Gosline et al., 2012). This approach makes it easier to visualize multiple experiments in the same network as well as enhances the ability to perform pathway enrichment to identify biological processes perturbed in each pathway. SAMNetWeb is implemented in an easy-to-use web server interface that enables the experimentalist to analyze multiple high-throughput experiments with both transcriptional and signaling data from each experiment. The end result is the identification of pathways shared between experiments and those unique to each experiment.

2 DESCRIPTION

SAMNetWeb analyzes signaling and transcriptomic data from multiple experiments by mapping the data to a weighted graph and then running a constrained optimization algorithm that selects the combination of nodes and edges that best explains the observed data. Protein-level signaling changes are mapped to a protein interaction network which connects to mRNA expression changes via a transcriptional regulatory network. Each edge in the network is weighted by confidence, enabling the SAMNet algorithm to create a reduced network that best explains the connections among the inputs. Usage is described in Figure 1.

2.1 Required input

SAMNetWeb requires two sets of data, representing protein-level
Figure 1: The five required steps to SAMNetWeb data analysis (left) with the respective parts of the SAMNetWeb submission page (right).

2.2 SAMNetWeb results

Once the files are properly formatted they can be submitted (Step 5) for analysis. Results are queued and processed within ~15 minutes and displayed on a static web page. Users can download relevant Cytoscape (Shannon et al., 2003) files, resulting DAVID enrichment (Jiao et al., 2012) analysis and view the network using the Cytoscape plug-in (Lopes et al., 2010).

We demonstrate the results of SAMNetWeb at http://fraenkel.mit.edu/samnetweb/emt. In this analysis, we showcase the ability to study multiple conditions at once by combining phosphoproteomic measurements and mRNA expression measurements across four experimental models of epithelial-mesenchymal transition (EMT). The resulting network and DAVID enrichment demonstrate the individual interactions predicted in each EMT model as well as the biological processes that are enriched for each dataset.

3 DISCUSSION

With the publication of more high-throughput datasets there is a growing need for easy-to-use, integrative analysis tools. SAMNetWeb is able to integrate signaling and transcriptomic data across multiple experimental systems to facilitate hypothesis generation following initial high-throughput sequencing experiments. The intuitive web-server interface and pre-formatted interaction networks make the algorithm useable without prior programming knowledge. Furthermore, SAMNetWeb is very flexible in its inputs, allowing the user to map any type of data to the interactome. This flexibility will enable SAMNetWeb to remain useful as new technologies are developed to analyze genome-wide data.

ACKNOWLEDGEMENTS

The authors acknowledge Dr. Nurcan Tuncbag for her technical guidance and Scott Moskrin for technical support.

Funding: This work is supported by NIH grants U54CA112967 and R01GM089903 and used computing resources funded by the National Science Foundation under Award No. DB1-0821391.

REFERENCES


