Gene Interaction Validation Against BioGRID

Robert Güth[1], Zhengyu Ouyang[2], Yang Zhang[2], Frederick Mackmiller[2], Thomas J. Ha[3], Dan Goldowitz[3], and Joe Song[2]

***Short Abstract* — Biological systems can be represented with mathematical models. However, in order to ensure that these *in silico* representations are accurate, detected gene interactions between genes and their products have to be experimentally validated. The purpose of this work was to establish a programming environment that can be used to determine whether detected gene interactions have been previously reported experimentally in the BioGRID database. Initial results however indicate that lacking gene annotation is a major potential obstacle toward investigating this question.**

***Keywords* — Model validation, BioGRID, genetic interaction.**

I. Background

Computational models of biomolecular systems represent genetic interactions that occur in cells. In order to assess the appropriateness of such *in silico* models, detected interactions first need to be justified with evidence from the literature. One way to do so is by cross-reference to available databases for documented interactions. One of the most comprehensive data bases for genetic interactions is BioGRID [1], containing more than 160,000 genetic interactions derived from experimental data in various model organisms.

It was thus the purpose of this project to develop programs in the R language to search BioGRID for detected genetic interactions and evaluate the biological relevance of reconstructed network models.

II. The Approach

After candidate interactions were identified, the BioGRID database was searched for direct or indirect links between pairs of interactors. As BioGRID provides only direct interactions, a programming algorithm was developed to check for indirect interactions based on information in the BioGRID data set. The workflow is as follows:

1. *The latest BioGRID release (2.0.63 as of April 2010) was downloaded.*
2. *The RBGL package from Bioconductor [2] was used to represent BioGRID interactions graphically with genes as nodes and interactions as edges.*
3. *The shortest path between interactors from candidate interactions were determined by utilizing the Dijkstra's algorithm [3], implemented as dijkstra.sp() in RBGL.*
4. *Path lengths shorter than 3 are considered plausible and the paths are returned by the program. In contrast, paths with lengths greater than 3 are considered implausible unless experimental exists to suggest otherwise.*

[1]Biology Department, New Mexico State University, Las Cruces, NM, 88003. E-mail: rgueth@nmsu.edu

[2]Department of Computer Science, New Mexico State University, Las Cruces, NM, 88003. E-mail: joemsong@cs.nmsu.edu

3 Dept. of Medical Genetics, University of British Columbia, Canada.

III. Summary of Results

Transcriptomic data were obtained during embryonic and early post-natal cerebellar development in mice over 12 time points (embryonic days 12-19, post-natal days 0, 3, 6, 9). First, expression data from 1,235 known transcription factors (TFs) (W.W. Wasserman, personal communication, 2010) were selected and clustered into groups exhibiting similar dynamics in expression. Then comparative dynamical system modeling [4] was applied to detect conserved and differential interactions between cluster representatives under embryonic and postnatal development conditions. Upon comparison between all names in the gene clusters and all gene names (including aliases) from BioGRID, it was determined that 68% of all transcription factors in the gene clusters are covered by BioGRID.

IV. Conclusions

Many gene interactions were verified by BioGRID. These include differential or conserved interactions during various stages in embryonic and postnatal development.

References

[1] Stark C, Breitkreutz BJ, Reguly T, Boucher L, Breitkreutz A, Tyers M. **BioGRID: A General Repository for Interaction Datasets.** *Nucleic Acids Res.* **34** (2006):D535-539.

[2] Gentleman, R., Carey, V. Huber, W., Irizarry, R., Dudoit, S. **Bioinformatics and Computational Biology Solutions Using R and Bioconductor**. 2005 Springer.

[3] Dijkstra, E. W**. A note on two problems in connexion with graphs.** *Numerische Mathematik* **1** (1959):269–271.

[4] Ouyang, Z., Song M. **Comparative identification of differential interactions from trajectories of dynamic biological networks**. Proceedings of German Conference on Bioinformatics, Halle, Germany. 2009.