Preliminary Data and Work Progress

Project 1 has initiated the analysis of the Phase I ToxCast data that has started to become available as of February of 2009. Initial work has focused on establishing computational approaches for establishing links between chemical activity and biological function, with a primary objective being to establish chemical activity within the context of functional biological networks and pathways. In addition, we expect that such contextual information well help establish new predictor variables that could potentially be used in modeling efforts in all projects. Similarly, the breadth of the assays represented in the ToxCast data presents an additional challenge and we have begun analysis across ToxCast assays, again with a focus on attempting to discern similarities and/or differences in assays with regard to biological context. We are initially attempting the use of graph-based methodologies for the integration of these different data types and the finding of frequent co-occurrence functional modules. In combination with additional external data sets describing known functional networks, such approaches can potentially reduce the dimension of the biological variables to a set that is both statistically predictive as well as biologically relevant. The application of a recently developed Bayesian data integration approach is also being undertaken, with the goal of linking assays into an integrated representation of biological relationships and associated perturbations.

Results to Date

We have submitted two abstracts to the upcoming ToxCast Data Analysis Summit in May 2009. These largely focus on establishing a biological context for ToxCast data and the preliminary analysis of the relative strengths/weaknesses of the individual assays with regard to functional information, as well as the similarities/differences regarding the assays' relevance to specific biological processes. In addition, we have submitted two papers for publication (both currently in review at BMC Bioinformatics). The first paper, *Comparison of phylogenetic trees via evolutionary structure alignment*, describes the development of a novel method for the comparison of trees, as well as other network-like structures, so as to determine regions of similarity and/or dissimilarity. This method has a wide range of applications, from helping to infer protein interactions to the comparison of pathway/network structures. We expect to be able to utilize this approach in the analysis of ToxCast data and the building of associated biological networks. Our second paper, *Enhancing metabolomic data analysis through NMR spectra alignment*, is a novel method for the alignment of NMR spectra that are typical of systems-level metabolomic efforts. This approach has potential use in the improved analysis of metabolomic data sets that we expect to be relevant in the analysis of future toxicity-relevant data sets.

With the recent release of the ToxCast Phase I data sets, we expect to increasingly focus on the modeling and prediction of how pathway architecture and dynamics is altered through chemical perturbation.

Activities for Subsequent Reporting Period

In depth analysis of ToxCast Phase I data will be the primary activity of the subsequent reporting period. In support of these efforts, a postdoctoral fellow (Ke Xu) will be joining the project starting August 1st, 2009. Dr. Xu's expertise is broadly in the field of applied mathematics and her work in the project will focus on the development of methodologies for linking high-level statistical models to lower-level mechanistic representations. In particular, understanding how perturbations at the network level lead to changes in system dynamics and behavior is of particular interest.

Significant emphasis will be placed on the development of methods for the integration of different data types. In addition, the challenge of understanding results across species needs to be addressed and we plan on initiating investigations of this using the ToxCast data. These efforts will initially be based on our earlier work on network structure inference using Bayesian approaches and other machine learning methodologies. Such approaches will generally provide

a network-based representation of the (potential) relationships between genes, proteins, and/or chemicals. Given such representations, substantial effort will be placed on the use of such network information to find “hidden” components of network structure that can be inferred from measurements of network properties under different perturbation conditions. Finally, Project 1 investigators have previously developed a mechanistic model of metabolism, focused on glycolysis/gluconeogenesis that includes the liver, muscle, fat and blood as communicating compartments. We plan on investigating the applicability of this model as a tool for the prediction of the possible effects of chemical perturbation of metabolic pathways.

We will enhance our interaction with Project 2, with the goal of integrating the eQTL analyses/approaches with the network-focused methodologies being developed in Project 1. Similarly, we plan on beginning more significant interaction with Project 3, with particular emphasis on trying to aid in establishing the network context in which a chemical or class of chemicals are likely to demonstrate an effect. We will also establish greater interactions and collaboration with EPA investigators at NCCT.

Publications Arising From this Project

Papers

1. Choi K, and Gomez SM. (2009) Comparison of phylogenetic trees via evolutionary structure alignment. BMC Bioinformatics (*In Revision)*.
2. Staab J, O’Connell TM, Gomez SM. (2009) Enhancing metabolomic data analysis through NMR spectra alignment. BMC Bioinformatics (*Submitted)*.

Posters/Abstracts

1. Choi K, Staab J, Elston TC and Gomez SM (2009). Dimension reduction and functional relevance in ToxCast chemical toxicity data. EPA ToxCast Data Analysis Summit, Durham, USA.
2. Staab J, Choi K, Elston TC and Gomez SM (2009). Establishing a biological context for ToxCast chemical toxicity data. EPA ToxCast Data Analysis Summit, Durham, USA.