Project 8: Ontological Discovery for Ethanol Research

Elissa Chesler, Ph.D., Principal Investigator
Oak Ridge National Lab/University of Tennessee, Knoxville, TN

A. Specific Aims

The purpose of this project is to address the data integration challenge across multiple labs, species and research approaches and to create a data archive and analysis tool for knowledge discovery of broad utility to the alcohol abuse and addiction research community and beyond. The approach is to define relationships among diverse phenotypes through the use of common biological substrata, represented as related gene sets. What makes this approach unique is that we will be using empirically derived gene sets from a wide variety of sources to connect genes to phenotypes, and developing query tools that allow users to pose higher order questions about the relations among these phenotypes. The majority of these gene sets will come from INIA-STRESS projects and related ongoing work of the consortium members. We will also work with the Bioinformatics Tools core to develop reciprocal connections with Gene Network, WebGestalt, GeneKeyDB and GeneNetViz. Further, the existing tool will be useful to the Bioinformtatics Array Core in the annotation and integration of microarray data sets. This is a project, rather than a part of the Bioinformatics Tools core because it is an experimental development project, organized around a research questions in ontological work, "what are the essential gene categories networks and pathways that subserve common and unique biological functions in stress, alcohol and related behavioral phenotypes?"

The project has three objectives: 1) to develop gene sets and analytic tools for integration and hypothesis generation within our consortium, 2) to develop and publicize an interface and public repository for use in the alcohol research community, 3) to place the work in a theoretical framework of empirical derivation of ontological categories as a solution to the subjectivity of phenotypic description and its indirect mapping onto intrinsic biological substrates.

We have assembled a unique team to achieve our highly interdisciplinary goals, which apply cutting edge developments in computer science, systems biology, bioinformatics and genetics to address challenges posed by neurobehavioral scientists for decades. Dr. Chesler has expertise in behavioral neuroscience, quantitative genetics and statistical programming. Dr. Michael Langston is and expert in algorithmics and computability and Dr. Erich Baker is a database expert with medical education. Members of this team have collaborated previously on the Tennessee Mouse Genome Consortium's Mutrack system for mouse tracking, data storage and analysis (Baker, Galloway et al. 2004) the genetic analysis of gene expression (Baldwin, Chesler et al. 2005; Chesler, Lu et al. 2005) and the biological applications of high-memory computing (Zhang, Abu-Khazam et al. 2005). By combining our technical expertise, we propose to tackle the challenge of data integration posed by the Integrative Neuroscience Initiative

on Alcoholism, a user group with extra-ordinarily diverse research methods, levels of analysis and research questions, all centered on a single biological theme, the relation between stress and alcoholism.

The project will proceed with three specific aims:

Aim 1. To work with investigators to develop a collection of ethanol-related gene sets from existing microarray projects, bioinformatics tools, and literature analysis. The resource generated in this aim will be an integrated collection of relevant gene sets, pathways and networks derived from the experimental work of the INIA. This alone will be a remarkable resource to users in the field, who often leap from intracellular interaction networks to larger neural-systems networks in their analysis of molecular substrates of behavior.

Aim 2. To implement an analytic approach and database of the gene sets build on a graph-based computational engine to allow dynamic data integration. The integration and operation on theses gene sets, networks and pathways is a more complex computational challenge than it appears. We will harness graph-based approached and cutting edge data modeling to the challenges posed by this project.

Aim 3. To develop human and machine interfaces and displays to enable a broad community of users to use the resource. A unique feature of the INIA consortium is that it brings together researchers with highly varied expertise in behavior, genetics, genomics, molecular biology and bioinformatics. Many of the consortium members, collaborators and users are interested in high-level questions about psychological and physiological processes. The tools that we develop will be targeted to members of this diverse group, not just to those who explicitly study genomics and genetics. Rather, we will harness the achievements of genome biology and make them accessible for researchers of higher order systems.