

Project 8: Ontological Discovery for Ethanol Research

Elissa Chesler, Ph.D., Principal Investigator

Oak Ridge National Lab/University of Tennessee, Knoxville, TN

A major challenge for collaborative research on the neurobiology of alcoholism is the need to bring together data across all levels of biological scale from multiple strains and species to address questions about relations among biological processes and behaviors. Deep analytic integration is possible for planned experiments, reference populations and reference protocols. However, in a large and diverse consortium and extending to the biological research community, such integration is typically unobtainable. Knowledge synthesis and qualitative shifts in approach to the neurobiology of alcoholism requires highly interdisciplinary individuals who have masters multiple literatures. This challenge can be addressed through application of recent developments in bioinformatics and advances in graph algorithms for global analysis of empirically defined gene-to-phenotype relations. We will create an archive of gene-to-phenotype relations, and use graph algorithms to discover the naturally occurring sets, subsets and supersets of phenotypes and genes. This scale, scope and complexity of this problem will be tackled by exploiting advances in high-performance supercomputing, database design and graph algorithms. This purely empirical approach to the development of gene and phenotype ontologies is designed to extract emergent new knowledge from the wealth of genome-phenome association data, in addition to describing what is known. The archive will be initially populated using our existing resources and from new project data collected by members of the Integrative Neuroscience Initiative on Alcoholism Consortium (INIA). Cross-species integration will be achieved through the use of gene homology. We will also develop an open-source, open-access Web based "Ontological Discovery System" for access to the data archive, analytic method and results by member of the INIA consortium and the alcohol research community at large. Using this tool, investigators will be able to start a query with sets of genes or phenotypes of interest and interactively explore their relations to other entities in the data archive to discover highly similar biological categories and to organize knowledge about the many facets of alcohol related phenotypes and their relations to the broader context of neural pathways and systems. This development work will occur in the context of a team of expert colleagues who will be capable of testing software, sharing data, performing biological interpretation and validating results of our analyses.