

COMPARATIVE AND COLLABORATIVE BIOINFORMATICS REQUIRED TO STUDY GENE NETWORKS THAT CREATE PHENOTYPES

The genome sequences from several chordates are being completed; the bioinformatics largely exists in the research community to discover the protein-coding potential of those genomes. However, the bioinformatics to elucidate gene regulation encoded in genomes and gene regulatory networks is not so developed. New bioinformatics, new model organism resources, new experimental approaches, and new collaborations are needed if the community is to understand the gene networks that help create phenotypes of interest.

A research team at ORNL and the University of Tennessee are developing some needed bioinformatics. The overall projects include 1) supplying several web services and collaborative bioinformatics that supports large consortia of experimental researchers and 2) developing comparative bioinformatics and new data mining environments that can ultimately help understand the nature and evolution of gene regulatory networks. Collaborative projects under way at ORNL, the University of Tennessee, the Tennessee Mouse Genome Consortium (www.tnmouse.org) and the new Integrative Neuroscience Initiative on Alcoholism (www.iniastress.org) are good testbeds for determining and validating what are the new bioinformatics needs.

Progress on several comparative and collaborative bioinformatics projects will be demonstrated. One shared requirement in several projects is a need for data mining environments that can better process large sets of genes and gene products as the linked nodes in a network. We will present a prototype version of one data mining environment, called GeneKeyDB. Several applications of GeneKeyDB will be shown, including in the comparative analysis of large sets of orthologous genes and in candidate gene finding.

Research sponsored by the National Institutes of Health and, in part, by the US Department of Energy under contract DE-AC05-00OR22725 with UT-Battelle, LLC