2nd International Biocuration Meeting
San Jose, California October 25 - 28, 2007

Curating the Comparative Toxicogenomics Database: a knowledge and discovery environment for chemical-gene-disease associations

Allan Peter Davis[1], Cynthia Murphy[1], Michael C. Rosenstein[1], Thomas Wiegers[1], John N. Forrest[1,2], James L. Boyer[1,2], and Carolyn J. Mattingly[1]

[1]Department of Bioinformatics, The Mount Desert Island Biological Laboratory, Salisbury Cove, Maine 04672; and [2]Department of Medicine, Yale University School of Medicine, New Haven, Connecticut 06520.

Chronic diseases may arise from interactions between chemicals and genes that modulate important physiological processes. Understanding chemical-gene interactions will help resolve questions about disease predisposition, therapeutic drug interactions, and exposure risks. Towards this end, we are curating the Comparative Toxicogenomics Database (CTD; http://ctd.mdibl.org) as a resource to integrate diverse information for the cross-species analysis of chemical, gene, and disease relationships. Text mining and manual curation of the literature are used to generate specific chemical-gene interactions and chemical/gene-disease relationships. Our paradigm uses community-established terms (from MeSH, NCBI, and OMIM) combined with our own controlled vocabularies to produce detailed chemical-gene interactions in a structured format. This approach standardizes the way data is captured by curators, facilitates database querying, and allows data to be displayed from the perspective of a chemical, gene, or disease. Currently, CTD provides integrated data for 59,000 chemicals, 1.2 million sequences (with associated GO annotations), 83,000 taxonomical terms, and 6,000 human diseases. To date, more than 40,000 interactions have been manually curated for 2,900 chemicals and 5,500 genes in 220 species. Over 800 genes are associated with both a human disease and a chemical, enabling the exploration of putative chemical-gene-disease networks and the development of testable hypotheses.