The Comparative Toxicogenomics Database: connecting chemicals, genes and diseases

Allan Peter Davis[1], Carolyn Mattingly[1], Michael Rosenstein[1], Thomas Wiegers[1], John Forrest[1, 2], James Boyer[1, 2]

[1]The Mount Desert Island Biological Laboratory, Salisbury Cove, ME 04672, [2]Yale University School of Medicine, New Haven, CT 06520

The etiology of most chronic diseases involves interactions between environmental factors and genes that modulate important physiological processes. One critical environmental factor is chemical exposure. Understanding interactions between chemicals and genes/proteins at the molecular level will help resolve questions concerning disease predisposition, therapeutic drug interactions, and health risks associated with differential susceptibility to exposures. Towards this end, we are developing the Comparative Toxicogenomics Database (CTD; http://ctd.mdibl.org) as a public resource to integrate and curate diverse data to promote the understanding of chemical interactions with genes and proteins.

CTD currently incorporates 59,000 chemicals, 1.2 million sequences (with Gene Ontology annotations), 83,000 taxonomic terms, 6,000 human diseases, and 78,000 references to produce a unique database for the cross-species analysis of chemical, gene/protein, and disease interactions. We also now manually curate the literature to capture specific details of every chemical–gene/protein interaction using searchable controlled vocabularies. To date, more than 22,000 interactions have been curated for 2,200 chemicals and 3,500 genes. Currently, over 650 genes have both a disease association and a curated chemical interaction, opening up the way to explore putative chemical-disease connections. User-friendly query pages and vocabulary browsers allow scientists to ask sophisticated questions and explore the data in detail from a chemical, gene, and disease perspective; all results can be downloaded in spreadsheets for further analysis.

CTD is a public information resource that supports hypothesis-driven research about chemical actions, enhances understanding about complex chemical–gene/protein networks and their impact on human health, and helps turn dispersed toxicology information into knowledge.