Industry-academic collaboration streamlines decades of data

By Raj Gosavi

Knowing the toxic effects of pharmaceutical drugs is important in drug development as well as in environmental science, but the relevant data are typically scattered throughout the scientific literature and a host of independent databases. Increasingly, scientists are recognizing the compelling need for finding and integrating this information, systematically, using common search terminology to make the data useful and accessible for academic and industrial researchers, as well as the general public.

As demonstrated by a new study, *(http://www.ncbi.nlm.nih.gov/pubmed/24288140)* funded in part by NIEHS, scientists at the pharmaceutical corporation Pfizer Inc. and academic researchers affiliated with the Comparative Toxicogenomics Database (CTD) *(http://ctdbase.org/)* have made major progress toward that goal, by integrating information about the toxicity of more than 1,200 pharmaceutical drugs.

Molecular toxicologist Carolyn Mattingly, Ph.D., *(http://harvest.cals.ncsu.edu/biology/index.cfm?pageID=4824&ID=2358)* associate professor at North Carolina State University (NCSU) and lead researcher on the study, has been directing development of CTD with NIEHS support since 2001 (see story). "We felt that the information that was lacking were the mechanisms that connect exposure to environmental chemicals with diseases," she said.

CTD is a publicly available research resource developed initially at the Mount Desert Island Biological Laboratory, where Mattingly was affiliated prior to moving to NCSU in January 2012. "This is the only database out there that connects mechanisms of chemical action to potential impacts on human health," Mattingly said. The collaboration with Pfizer, which took CTD biocurators - professional scientists who curate, collect, annotate, and validate information - a year to complete, significantly increased information about pharmaceutical and environmental chemicals.

Shown, above, at a data integration meeting in July 2013 (see story), Mattingly, like many of her colleagues working in big data applications, feels that networking data about chemical-gene and chemical-protein interactions, as well as chemical-disease and gene-disease relationships, can help scientists develop novel hypotheses about the origins of environmentally influenced diseases. (Photo courtesy of Steve McCaw)

Mapping hundreds of thousands of chemical-disease interactions

The database provides chemical-gene-disease information and associated functional and pathway data. The Pfizer collaboration specifically added data for chemicals that may be involved in cardiovascular, neurological, kidney, and liver disorders.

Current screening methods for drugs, although advanced, need to be complemented with newer methods for assessing compound toxicity. With the ability to predict toxic side effects of drugs, the database can potentially improve drug development, by allowing generation of new early-stage mechanistic screening of the drugs.

Data collection and integration took advantage of Pfizer's efficient text-mining procedures, focusing on select chemicals, combined with CTD's manual curation of more than 88,000 research articles. The information from the scientific literature was then integrated with the existing database to expand its network of chemical-gene-disease interactions.

Highlighting the common goal of the collaboration with Pfizer, Mattingly said, "It was a win-win situation. What pharmaceutical companies are trying to do is essentially what environmental scientists are trying to do - understand and prevent toxicity."

"The type of partnership demonstrated by the success of this joint Pfizer-CTD effort is essential to translating scientific data and knowledge into meaningful information that can be utilized to advance our understanding of environmental agents and their impacts and ultimately to improving public health," said Allen Dearry, Ph.D., director of the NIEHS Office of Scientific Information Management.

Data from CTD is increasingly cited and integrated with other biological databases, thereby enhancing access to the information. As she looked forward to further developments in CTD, Mattingly observed, "We continue to work closely with collaborators to develop additional tools and data modules to improve the utility of the database."

Citation: Davis AP, Wiegers TC, Roberts PM, King BL, Lay JM, Lennon-Hopkins K, Sciaky D, Johnson R, Keating H, Greene N, Hernandez R, McConnell KJ, Enayetallah AE, Mattingly CJ.
(http://www.ncbi.nlm.nih.gov/pubmed/24288140)
2013. A CTD-Pfizer collaboration: manual curation of 88,000 scientific articles text mined for drug-disease and drug-phenotype interactions. Database 2013; doi:10.1093/database/bat080.

(Raj Gosavi, Ph.D., is a research fellow in the NIEHS Structure and Function Research Group.)

The Environmental Factor is produced monthly by the National Institute of Environmental Health Sciences (NIEHS)
(http://www.niehs.nih.gov/)
, Office of Communications and Public Liaison. We welcome your comments and suggestions.
(bruskec@niehs.nih.gov)

This page URL: NIEHS website: http://www.niehs.nih.gov/
Email the Web Manager at webmanager@niehs.nih.gov