

# **Generation of a Centralized and Integrated Resource for Exposure Data**

**North Carolina State University – Raleigh**

PI: Carolyn J. Mattingly

Grant Number:3R01ES019604

Our objective is to provide a centralized, publicly available resource with comprehensive, well-annotated data and analysis tools that informs design and interpretation of environmental health studies and promotes novel insights into the etiologies of environmentally influenced diseases. Most human diseases involve interactions between genetic and environmental factors; however, the basis of these complex interactions are not well understood and limit improvements in toxicity prediction, risk assessment, research prioritization and therapeutic interventions. We developed the Comparative Toxicogenomics Database (CTD; [://ctd.mdibl.org](http://ctd.mdibl.org)) to enhance understanding about environment-disease connections by providing manually curated data describing chemical-gene/protein interactions and chemical- and gene/protein-disease relationships from the peer-reviewed literature and integrating these data with select external data sets (e.g., pathways and biological process data) and novel data analysis tools. In this application, we propose to leverage our expertise and CTD infrastructure to: 1) enhance the capacity to identify environment-disease connections by curating and integrating exposure data into CTD; and 2) expand the capacity for prediction, analysis and interpretation of environment-disease networks by developing novel analysis and visualization tools that include exposure data. This proposal responds to the needs expressed by the NIEHS and partner agencies for inclusion of exposure data when prioritizing research and performing toxicity testing, it addresses the need for centralization of exposure data in a broader biological context and it will provide "real-world" exposure context for existing data in CTD. The resulting resource will enable new opportunities for understanding and prioritizing human health effects from exposure and their underlying etiologies and coordinate data key to enhancing the capacity for toxicity prediction and risk assessment. PUBLIC HEALTH RELEVANCE: Most human diseases involve interactions between genetic and environmental factors; however, the basis of these complex interactions is not well understood. Despite the importance of real-world exposure information for understanding connections between the environment and human health and prioritizing environmental health research, to date there has been no effort to centralize these data within a broader biological framework and their full value remains unexploited. This proposal will leverage our established database development expertise and infrastructure for the publicly available Comparative Toxicogenomics Database (CTD), which aims to promote understanding about environment-disease relationships, to curate and integrate exposure data into CTD and enhance the capacity for prediction, analysis and interpretation of environment-disease networks by developing novel analysis and visualization tools that include exposure data.