Chemical-Gene Interactions from ToxCast Bioactivity Data Expands Universe of Literature Network-Based Associations

Sean Watford, Imran Shah, Richard Judson, Matt Martin

UNC-CH, USEPA NCCT

Characterizing the effects of chemicals in biological systems is often summarized by chemical-gene interactions, which have sparse coverage in the literature. The ToxCast chemical screening program has produced bioactivity data for nearly 2000 chemicals and over 450 gene targets. To evaluate the information gained from the ToxCast project, a ToxCast bioactivity network was created comprising ToxCast chemical-gene interactions based on assay data and compared to a chemical-gene association network from literature. The literature network was compiled from PubMed articles, excluding ToxCast publications, mapped to genes and chemicals. Genes were identified by curated associations available from NCBI while chemicals were identified by PubChem submissions. The frequencies of chemical-gene associations from the literature network were log-scaled and then compared to the ToxCast bioactivity network. In total, 140 times more chemical-gene associations were present in the ToxCast network in comparison to the literature-derived network highlighting the vast increase in chemical-gene interactions putatively elucidated by the ToxCast research program. There were 165 associations found in the literature network that were reproduced by ToxCast bioactivity data, and 336 associations in the literature network were not reproduced by the ToxCast bioactivity network. The literature network relies on the assumption that chemical-gene associations represent a true chemical-gene interaction. This bias along with other confounders and known issues (e.g., term mapping) that will be better accounted for in future versions of the literature network with additional manual curation or natural language processing methods applied. Meanwhile, the comparison of the ToxCast and literature networks demonstrated the increase in chemical-gene associations being discovered through screening programs and the need to improve chemical and gene annotations in the literature. This approach provides a framework for estimating information gain from ToxCast and other chemical screening programs as well as highlighting areas of follow-up to confirm chemical-gene interactions previously unreported in the literature. *This abstract does not necessarily reflect U.S. EPA policy.*