ToxPlorer™: A Comprehensive Knowledgebase of Toxicity Pathways Using Ontology-driven Information Extraction

I Shah[1] , A Singh[2], C Haugh[1], J Jack[1], R Judson[1], T Knudsen[1], M Martin[1], J Wambaugh[1]
[1]National Center for Computational Toxicology (NCCT), US EPA, RTP, NC, USA.
[2]Lockheed Martin, RTP, NC, USA

Realizing the potential of pathway-based toxicity testing requires a fresh look at how we describe phenomena leading to adverse effects *in vivo*, how we assess them *in vitro* and how we extrapolate them *in silico* across chemicals, doses and species. We developed the ToxPlorer™ framework to extract experimental evidence from the literature about the effects of chemicals in living systems and to coherently synthesize prior knowledge into semantic networks. This was accomplished in four main steps. First, we developed an ontology to formally describe functional relationships in toxicology, which include molecules and their interactions, but also cell types, cellular processes and behaviors, phenotypes and histological effects. Second, we systematically analyzed the text of 655,271 PubMed abstracts about the mammalian liver using 363,472 diverse entities in our ontology. Third, we used this information to focus on a subset of 23,244 abstracts about nuclear receptor-mediated hepatocarcinogenesis. Out of 241,944 sentences a subset of 3,712 sentences from 2,199 abstracts produced more than 100,000 putative relationships of relevance. Fourth, we used ontology-driven information extraction tools to find less than 5,000 semantically-valid assertions about chemical-induced hepatotoxicity. By manually curating this information we found evidence relating 501 chemicals, 671 genes, 121 cell-events and 38 histological lesions (to date). Our findings recapitulate many of the events involved in nuclear receptor-mediated direct and indirect hyperplasia, formation of preneoplastic foci and development of neoplastic lesions in rodents. ToxPlorer™ is publicly available along with tools for analyzing and interactively reconstructing putative toxicity pathways.