**High-throughput literature mining to support read-across predictions of toxicity**

Nancy C. Baker[1], Thomas Knudsen[2], Kevin M. Crofton[2], Grace Patlewicz[2]

[1]Lockheed Martin, RTP, NC; [2]National Center for Computational Toxicology, Office of Research and Development, U.S. Environmental Protection Agency, RTP, NC

Baker.Nancy@epa.gov

Building scientific confidence in the development and evaluation of read-across remains an ongoing challenge. Approaches include establishing systematic frameworks to identify sources of uncertainty and ways to address them. One source of uncertainty is related to characterizing biological similarity. Many research efforts are underway such as structuring mechanistic data in adverse outcome pathways and investigating the utility of high throughput (HT)/high content (HC) screening data. A largely untapped resource for read-across to date is the biomedical literature. This information has the potential to support read-across by facilitating the identification of valid source analogues with similar biological and toxicological profiles as well as providing the mechanistic understanding for any prediction made. A key challenge in using biomedical literature is to convert and translate its unstructured form into a computable format that can be linked to chemical structure. We developed a novel text-mining strategy to represent literature information for read across. Keywords were used to organize literature into toxicity signatures at the chemical level. These signatures were integrated with HT *in vitro* data and curated chemical structures. A rule-based algorithm assessed the strength of the literature relationship, providing a mechanism to rank and visualize the signature as literature ToxPIs (LitToxPIs). LitToxPIs were developed for over 6,000 chemicals for a variety of toxicity signatures such as developmental toxicity, genotoxicity, reproductive toxicity, and thyroid toxicity. We then developed a user interface that facilitates exploration of the literature evidence behind the records of biological activity. This tool allows researchers to substantiate structure based read-across predictions with literature reports of *in vitro* and *in vivo* toxicity and thereby achieve a higher level of confidence in those predictions. *This abstract does not necessarily represent U.S. EPA policy*.