Using high-throughput literature mining to support read-across predictions of toxicity

Nancy C. Baker¹, Thomas Knudsen², Kevin M. Crofton², Grace Patlewicz²

¹Leidos, RTP, NC, USA; ²NCCT, ORD, US EPA, NC, USA

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 such information 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 as keyword features (toxicity signatures) at the chemical level. The elements of the toxicity signatures were weighted using a rule-based algorithm that assessed the strength of the literature relationship. This weight was used to rank and visualize the signature as literature ToxPIs (LitToxPIs) for ~6,000 chemicals described in the biomedical literature for a variety of toxicity types including genetic toxicity, developmental toxicity, reproductive toxicity and thyroid toxicity. We then developed a user interface (UI) that facilitates exploration of the literature evidence behind the signatures. As an example, the literature evidence extracted from the 2,745 articles about bisphenol A resulted in a toxicity signature showing reproductive toxicity as the most significant type for this chemical. When ranked with all chemicals showing evidence of reproductive toxicity, bisphenol A was third in the total list of 2,092 chemicals. To demonstrate how these methods can enrich read-across for chemical categories, we generated and ranked LitToxPIs for a set of 64 benzene derivatives. This UI provides a tool that 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.