Structure-based Chemical Taxonomy to Focus Chemical Queries

Tripp LaPratt[1*], Victoria Hull[1], Emily Reinke[1], Nicole Kleinstreuer[2]

[1]Inotiv, Research Triangle Park, NC, USA; [2]NIH/NIEHS/DTT/NICEATM, Research Triangle Park, NC, USA

**Presenting author*

Navigating and characterizing chemical space, especially for large chemical queries, is a challenge. Tools that distinguish chemical structure can focus and inform queries by helping to identify substructures and chemical classes related to chemical activity. One such tool, ClassyFire (https://jcheminf.biomedcentral.com/articles/10.1186/s13321-016-0174-y), will be integrated into the Integrated Chemical Environment (ICE;) to aid with the exploration of toxicologically relevant data. ClassyFire is an automated, structure-based chemical taxonomy tool created by the Wishart Research Group with 4,285 unique classifications across 11 hierarchical levels. We used classifications for chemicals tested in the U.S. Environmental Protection Agency's Tox21 program to focus queries of bioactivity and chemical use data in ICE. Tox21 chemicals with ClassyFire classifications included 8,583 chemicals spanning eight levels of hierarchy (8,477 organic; 106 inorganic). "Benzenoids" and "organoheterocyclic compounds" were the most abundant Superclasses with 2,744 and 1,959 chemicals, respectively. Chemicals within these two Superclasses were most bioactive for gene expression and cell viability assays. Among benzenoids, 354 had use data available, and 186 organoheterocyclic compounds had use data. For both chemical superclasses, "consumer goods" were the most abundant use followed by "construction/maintenance" and "cleaning." By linking taxonomy to bioactivity and use, we can explore chemical structures and use cases that potentially contribute or relate to these endpoints. This can help focus chemical selection, aid in selection of alternative chemicals, and potentially identify chemical substructures that are correlated with chemical activity. This project was funded by the NIEHS under Contract No. HHSN273201500010C.