Predicting ToxCast™ and Tox21 Bioactivity Using Toxprint Chemotypes

Jeremy Fitzpatrick and Grace Patlewicz

The EPA ToxCast™ and Tox21 programs have generated bioactivity data for nearly 9076 chemicals across ~1192 assay endpoints; however, for over 70% of the chemical-assay endpoint pairs there is no data. To help fill the gaps, we constructed random forest models for each assay endpoint, using chemotypes which are fragment based descriptors. For each model, the assay endpoint data was split into a training set containing 80% of the active chemicals and an equal number of in actives, with the remainder used as the test set. Many assay endpoints still lacked enough data to build effective models. The 277 models with at least 200 compounds in their training sets were effective at predicting bioactivity,with 250 of 277 generating predictions with greater than 60% balanced accuracy. Our models predict ToxCast™ and Tox21 bioactivity values in the absence of experimental data and may be useful for building predictive toxicity models. This abstract may not reflect U.S. EPA views or policy.