

Modeling and Predicting Cancer from ToxCast Phase I Data

Authors: Richard Judson, David Dix, Robert Kavlock, Imran Shah, Keith Houck, Thomas Knudsen, David Reif, Matt Martin, National Center for Computational Toxicology, USEPA, RTP, NC

The ToxCast™ program is generating a diverse collection of *in vitro* cell free and cell based HTS data to be used for predictive modeling of *in vivo* toxicity. We are using this *in vitro* data, plus corresponding *in vivo* data from ToxRefDB, to develop models for prediction and prioritization. This poster will focus on a set of machine learning based models that produce toxicity signatures, which are algorithms that yield a toxicity class prediction based on an association between *in vitro* assay data and an *in vivo* endpoint, derived from training examples. We demonstrate this approach with a signature for rat liver proliferative lesions using data from the chemicals with rat chronic/cancer data in ToxRefDB, 61 of which are positive for this endpoint. We also demonstrate the use of derived gene and pathway perturbation scores which are more aggregated predictors that can be used in machine learning approaches. Qualitative uses of these perturbation scores are demonstrated with relation to other *in vivo* endpoints in rodents and humans, including their use in predicting whether a chemical will be a probable human carcinogen.

This work was reviewed by EPA and approved for publication but does not necessarily reflect official Agency policy.