# Building predictive gene signatures through simultaneous assessment of transcription factor activation and gene expression

**Natalia M. VanDuyn, Jill A. Franzosa, Keith Houck, William Ward, Brian N. Chorley, J. Christopher Corton**

Exposure to many drugs and environmentally-relevant chemicals can cause adverse outcomes. These adverse outcomes, such as cancer, have been linked to molecular initiating events (MIE) and downstream key events to define adverse outcome pathways (AOP). Identification of gene sets (signatures) that are predictive of either MIEs (e.g., transcription factor (TF) activation) or key events (e.g., cell proliferation) would be useful in predicting AOP modulation after chemical exposure. The goal of this project is to identify signature genes for TF activation via simultaneous assessment of TF activity and global gene expression in the same cell system. RNA isolated from HepG2 cells exposed in concentration-response to ~800 ToxCast or reference chemicals was used to assess TF activation at over 50 cis elements with Attagene FACTORIAL assays and expression of more than 47,000 RNA targets was measured for 100 of these samples using the Illumina HumanHT-12 v4 Expression BeadChip. To evaluate the effects of chemical exposure on gene expression and TF activation, these data were fit to a four-parameter Hill function with the ToxCast program's data processing pipeline. Pearson's correlation coefficient and associated p-value were used to identify genes with a significant correlation between the change in expression and activation of one or more TF/cis-elements. These results were used to derive gene expression signatures for the aryl hydrocarbon receptor (AhR), thyroid hormone receptor alpha (TRα), and peroxisome proliferator-activated receptor gamma (PPARγ) activation. Utilizing the fold-change rank-based Running Fisher's algorithm, the signatures were compared to ~1500 biosets in an annotated human primary hepatocyte gene expression database to identify those that exhibited a significant positive correlation to the signature. Each signature identified an independent set of chemicals known to activate the TF. For example, the AhR signature identified biosets associated with known AhR-activating chemicals including TCDD, benzo[a]pyrene and quercetin (p-value < $1x10^{-33}$), thus validating the method. Future work will focus on expanding the analysis to other TFs, allowing a comprehensive, simultaneous assessment of the modulation of multiple human TFs by chemicals in large, publically available, genomic datasets. (This abstract does not represent EPA policy).