Building gene expression signatures indicative of transcription factor activation to predict AOP modulation

VanDuyn, Natalia[1]; Franzosa, Jill[2]; Houck, Keith[2]; Ward, William[2]; Chorley, Brian[2]; Corton, Chris[2]

1. ORISE, RTP, NC, USA.
2. NHEERL, U.S. EPA, RTP, NC, USA.

Adverse outcome pathways (AOPs) are a framework for predicting quantitative relationships between molecular initiating events (MIE) and downstream key events that lead to adverse outcomes. Defining gene sets (signatures) that predict chemical-induced MIEs (e.g., transcription factor (TF) activation) or key events (e.g., cell proliferation) would be useful in building models of AOP modulation that predict adverse outcomes. Current methods for identifying TF activation signatures have limitations. TF binding can be identified by experimental methods such as ChIP-Seq, but this is costly and time consuming, or by indirect measures such as in silico identification of TF binding sites, which may or may not be biologically relevant. Our approach is to identify the genes associated with TF activation via simultaneous assessment of TF activity and global gene expression in the same cell system (human hepatoblastoma cell line, HepG2). We have taken advantage of the Attagene FACTORIAL data (from ToxCast) and subjected the same RNA samples to microarray profiling of more than 47,000 RNA targets using the Illumina Human Expression BeadChip. Pearson's correlation and associated p-value were used to identify genes with a significant correlation between the change in expression and activation of AhR, TRalpha and PPARgamma across multiple chemical exposure experiments. The AhR signature was compared to a large human database of microarray experiments using the Running Fishers algorithm. These methods identified biosets associated with TCDD, benzo(a)pyrene and quercetin (p-value < $1x10^{-33}$), confirming the utility of the method. This analysis can be extended to other TFs, allowing comprehensive assessment of chemical-induced modulation of human TFs in large, publically available, genomic datasets. (This abstract does not represent EPA policy.)