

## DIFFERENTIAL GENE EXPRESSION AND CONCENTRATION-RESPONSE MODELING WORKFLOW FOR HIGH-THROUGHPUT TRANSCRIPTOMIC (HTTr) DATA: RESULTS FROM MCF7 CELLS

Joshua Harrill<sup>1</sup>, Imran Shah<sup>1</sup>, Woodrow Setzer<sup>1</sup>, Matthew Martin<sup>1</sup>, Derik Haggard<sup>1,2</sup>, Russell Thomas<sup>1</sup>

<sup>1</sup>USEPA NCCT, RTP, NC

<sup>2</sup>ORISE,RTP,NC

Increasing efficiency and declining cost of generating whole transcriptome profiles has made high-throughput transcriptomics a practical option for chemical bioactivity screening. The resulting data output provides information on the expression of thousands of genes and is amenable to a variety of downstream applications. However, HTTr chemical screening presents challenges that are not inherent in traditional cell-based screening assays which produce univariate outputs. We present a microfluidics-based laboratory workflow for HTTr screening of MCF7 cells in 384-well format and a HTTr analysis pipeline for data quality control, differential gene expression and concentration-response modeling. MCF7 cells were plated in either DMEM+10% HI-FBS or phenol red-free DMEM+10% CS-HI-FBS and allowed a 24 h recovery period prior to exposure. A total of 44 chemicals were screened in 8-point concentration-response (0.03-100  $\mu$ M) in each media at 3 exposure durations (6, 12, 24 h) in 3 independent cultures (n=1 /treatment/culture). Chemicals were applied using an acoustic dispenser. Each test plate was uniquely randomized with respect to treatment positioning. Cell lysates were analyzed using a Tempo-Seq human whole transcriptome assay to a target read depth of 3M. Cell viability and apoptosis assays were run in parallel to exclude conditions causing cytotoxicity. For count data, total and percent mapped reads were used to exclude samples with poor performance. Data were subset by chemical x media x time with matching controls. Probes were filtered using a median raw read count  $\geq$  5. Count data was scaled, transformed and differentially expressed genes (DEGs) determined using DESeq2. Reproducibility of read counts in technical and biological replicates was high with pairwise correlations  $>$  0.85 and median CVs of between 20-40%. The correlation of  $\log_2$ FC values for DEGs among biological replicates was also high (median  $>$  0.75) within each media x time condition. ANOVA results demonstrated a broad range (10s to 1000s) in the number of concentration-responsive genes across chemicals. Concentration-response modeling demonstrated a broad range of probe and pathway level BMDs across chemicals for each media x time combination, and facilitated identification of no observable transcriptional response levels. *This abstract does not necessarily reflect USEPA policy.*