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

Joshua Harrill[1], Imran Shah[1], Woodrow Setzer[1], Derik Haggard[1,2], Matthew Martin[1], Richard Judson[1], Sophie Malcomber[3], Andrew White[3], Russell Thomas[1]

[1]USEPA National Center for Computational Toxicology, RTP, NC [2]Oak Ridge Institute for Science and Education, Oak Ridge, TN [3]Safety and Environmental Assurance Centre, Unilever, Colworth, UK

www.epa.gov

This research is part of the US EPA and Unilever collaborative research effort

Joshua Harrill | harrill.joshua@epa.gov | ORCID: 0000-0003-4317-6391

Background and Objectives

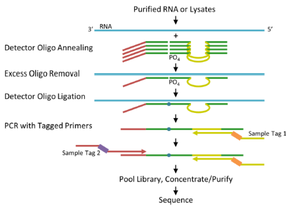
Background

- Current initiatives in toxicological research include integrating toxicity testing data from *in vitro* and alternative methods into risk assessment practices.
- Previously, *in vitro* point-of-departure (PODs) estimates suitable for use in screening-level risk assessments have been based on panels of *in vitro* assays that report a limited number of endpoints and that may not provide comprehensive coverage of all biological pathways.
- Increasing efficiency and declining cost of generating whole transcriptome profiles has made high-throughput transcriptomics (HTTr) a practical option for determining *in vitro* bioactivity thresholds and increasing coverage of biological space in chemical screening.

Technology

- The TempO-Seq human whole transcriptome assay measures the expression of greater than 20,000 transcripts.
- Requires only picogram amounts of total RNA per sample.
- Compatible with purified RNA samples or cell lysates.
- Transcripts in cell lysates generated in 384-well format are barcoded according to well position and combined in a single library for sequencing using industry standard instrumentation.
- Per sample fastq files are generated and aligned to BioSpyder sequence manifest to generated integer count tables.

TempO-Seq Assay Illustration



Objectives

- Develop a laboratory workflow for concentration-response screening of chemicals using a targeted RNA-Seq human whole transcriptome assay (TempO-Seq).
- Perform concentration-response screening of a small number of chemicals (n=44) with well annotated mechanisms of action in a representative cell type (MCF7) at three time points (6, 12, 24 hours post-exposure.
- Establish a computational pipeline for data normalization, differentially expressed gene (DEG) identification, concentration-response modeling and mechanism-of-action (MoA) prediction.
- Select a time x media combination for large-scale screening of chemicals from the USEPA Toxicity Forecaster (ToxCast) compound library.

Experimental Design

Parameter	Multiplier	Notes
Cell Type(s)	1	MCF7
Culture Condition	2	DMEM + 10% HI-FBS PRF-DMEM + 10% CS-HI-FBS
Chemicals	44	see subsequent slides
Time Points:	3	6, 12, 24 hours
Assay Formats:	3	TempO-Seq HCI Cell Viability & Apoptosis
Concentrations:	8	3.5 $\log_{10}$ units; ½ $\log_{10}$ spacing
Biological Replicates:	4	3 TempO-Seq; 1 Reserve

Chemical Test Set

Chemical Name	MIE Family	Chemical Name	MIE Family	Chemical Name	MIE Family	Chemical Name	MIE Family
Bisphenol A	ESTROGEN	Flutamide	ANTIANDROGEN	Rotenone	MITOCHONDRIA (COMPLEX I)	Imazalil	STEROIDOGENESIS
Bisphenol B		Nilutamide		Fenpyroximate (Z,E)		Prochloraz	
4-Nonylphenol, branched		Cyproterone acetate		Trifloxystrobin	MITOCHONDRIA (COMPLEX II)	Cyproconazole	
4-Cumylphenol		Vinclozolin		Pyraclostrobin		Propiconazole	
4-Hydroxytamoxifen	ANTIESTROGEN	Maneb	SH REACTIVE	Atrazine	cAMP INDUCER / PDE INHIBITOR	Fomesafen	PPO INHIBITOR
Clompiphene Citrate (1:1)		Ziram		Cyanazine		Butafenacil	
Fulvestrant		Thiram		Simazine	PS2 INHIBITOR	Lactofen	PPO INHIBITOR / PPAR
Reserpine	VMAT	Cadribrine	CYTOTOXICANT	Tetrac	THR	PFOS	PPAR
Amiodarone hydrochloride		Cycloheximide		3,5,3'-Triiodothyronine		PFOA	
Lovastatin	HMGCR	Bifenthrin	NA+ CHANNEL	Clofibrate	FIBRATES	Troglitazone	
Simvastatin		Cypermethrin		Fenofibrate		Farglitazar	

MIE = Molecular Initiating Event

- Chemical set covers broad range of mechanistic diversity with redundancy within mechanistic class.

Laboratory Workflow

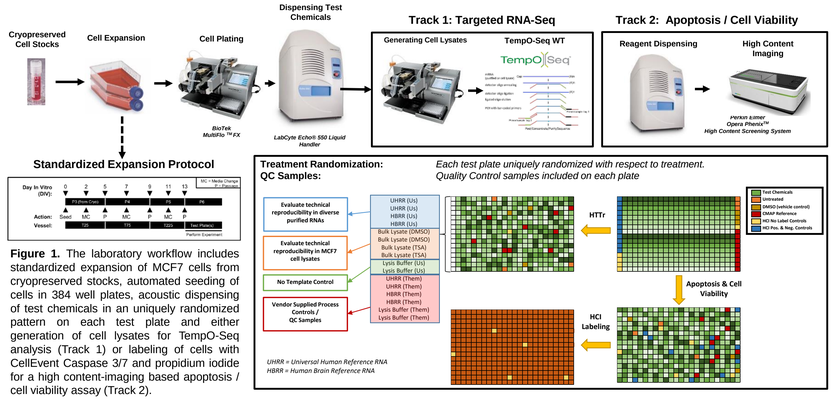


Figure 1. The laboratory workflow includes standardized expansion of MCF7 cells from cryopreserved stocks, automated seeding of cells in 384 well plates, acoustic dispensing of test chemicals in an uniquely randomized pattern on each test plate and either generation of cell lysates for TempO-Seq analysis (Track 1) or labeling of cells with CellEvent Caspase 3/7 and propidium iodide for a high content-imaging based apoptosis / cell viability assay (Track 2).

HTTr Data Analysis Pipeline

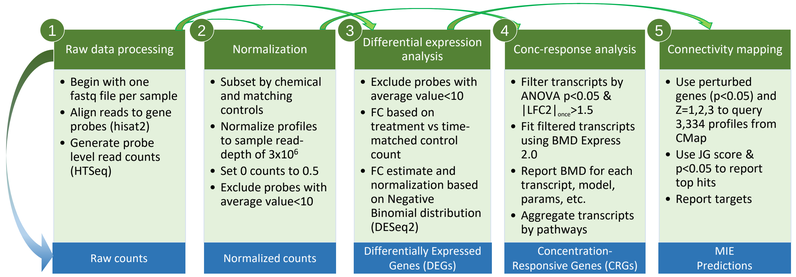


Figure 2. The HTTr data analysis pipeline begins with (1) alignment of FASTQ files to the BioSpyder human whole transcriptome probe manifest. (2) Data are subset by chemical with matching controls and counts are normalized, scaled to target per sample read depth ($3x10^6$) and floored. (3) Calculation of FC estimates and identification of DEGs is performed using DESeq2. (4) Concentration response analysis, calculation of probe level BMDs and aggregation of pathways is performed using BMDExpress2.0, (5) Connectivity mapping to the CMAP v2 Affymetrix database and within the BSP dataset is used for MoA prediction.

Benchmark Dose Modeling

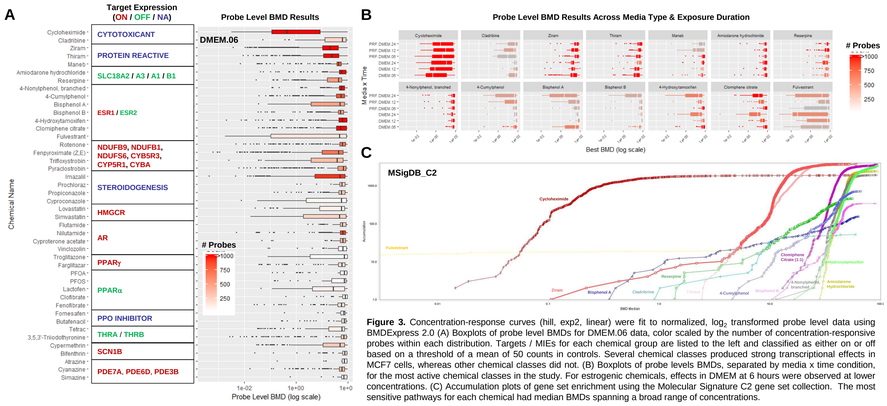
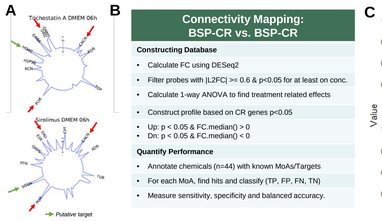


Figure 3. Concentration-response curves (hill, exp2, linear) were fit to normalized, $\log_2$ transformed probe level data using BMDExpress 2.0 (A) Boxplots of probe level BMDs for DMEM.06 data, color scaled by the number of concentration-responsive probes within each distribution. Targets / MIEs for each chemical group are listed to the left and classified as either on or off based on a threshold of a mean of 50 counts in controls. Several chemical classes produced strong transcriptional effects in MCF7 cells, whereas other chemical classes did not. (B) Boxplots of probe levels BMDs, separated by media x time condition, for the most active chemical classes in the study. For estrogenic chemicals, effects in DMEM at 6 hours were observed at lower concentrations. (C) Accumulation plots of gene set enrichment using the Molecular Signature C2 gene set collection. The most sensitive pathways for each chemical had median BMDs spanning a broad range of concentrations.

MIE Prediction



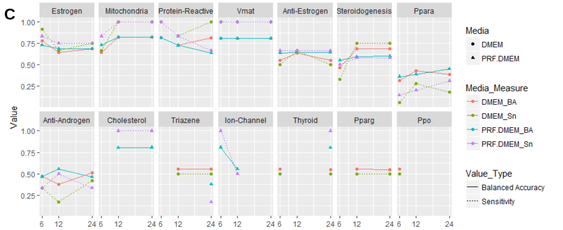


Figure 4. (A) Radial plots of putative molecular targets identified using Connectivity Mapping of single point concentration BioSpyder (BSP) data from reference chemicals (Trichostatin A, Sirolimus) compared to the CMAP Affymetrix database. This approach was successful in identifying known molecular targets, but a large degree of promiscuity in predicting targets was observed. (B) A novel approach for Connectivity Mapping was developed that incorporates information on concentration-responsive probes. A "leave-one-out" cross validation comparison within the BSP dataset was performed. Results are shown in panel C. (C) High sensitivity (Sn) and balanced accuracy (BA) for identifying similarly annotated chemicals was observed in the cross-validation analysis of the BSP dataset. There was not a consistent media x time combination with superior performance across all chemical groups.

Summary & Conclusions

- A laboratory workflow for HTTr concentration-response screening was established using microfluidics and randomization of treatments within test plates.
- A computational pipeline for processing TempO-Seq human whole transcriptome data was established using publically available tools and resources.
- BMD analysis with respect to chemical grouping demonstrated that some chemical groups produced marked transcriptional responses in MCF7 cells (e.g. Estrogens, antiestrogens, VMAT, protein reactive, mitochondrial toxicants) while other chemical groups had low biological activity (e.g. anti-androgens, PPARγ, PPARα, THRA/B, Triazenes).
- Leave-one-out cross-validation within the TempO-Seq dataset demonstrated high sensitivity and balanced accuracy for identifying like chemicals that produce marked transcriptional effects.
- Known MOAs for well annotated chemicals could be identified using Connectivity Mapping to the CMAP Affymetrix database, however high promiscuity in putative target identification was observed using this method.

U.S. Environmental Protection Agency
Office of Research and Development