Gene Signature Concentration-Response Modeling of High-Throughput Transcriptomics (HTTr) Data for Mechanistic Prediction and Potency Estimation

Joshua Harrill[1], Logan Everett[1], Derik Haggard[1], Joseph Bundy[1], Laura Taylor[1], Bryant Chambers[1], Clinton Willis[1], Imran Shah[1], Richard Judson[1]

[1]USEPA Center for Computational Toxicology and Exposure (CCTE), RTP, NC,

Joshua Harrill | harrill.joshua@epa.gov | https://orcid.org/0000-0003-4317-6391

www.epa.gov

Background

- In an effort to accelerate the pace of chemical risk assessment, Thomas et al. (2019) (PMID: 30835285) have developed a tiered hazard evaluation strategy using New Approach Methods (NAMs).
- The first tier specifies the use of high-throughput profiling (HTP) assays to rapidly evaluate the biological activity of hundreds of chemicals in human-derived cell models.
- High-throughput transcriptomics (HTTr) using the TempO-Seq assay has been identified as an HTP assay that can be used to:
  1) identify molecular points of departure (mPODs) for perturbation of cellular biology by chemicals.
  2) generate gene expression profiles for mechanism of action prediction.
- Laboratory workflows and data analysis pipelines have been developed that facilitate large scale HTTr screens of hundreds of chemicals in 384-well culture format (Harrill et al. 2020) (PMID: 30835285).
- Here, the results from an HTTr screen of 1794 chemicals from the ToxCast collection in MCF7 cells are summarized.

Experimental Design

A

Cell Type:	MCF7 (ATCC® HTB-22)
Chemicals:	n = 1784 ToxCast Chemicals
Concentrations:	n = 8 (0.03 – 100 uM)
Time Point:	6 hours
Assays:	TempO-Seq (Whole Transcriptome) Cell Viability

B

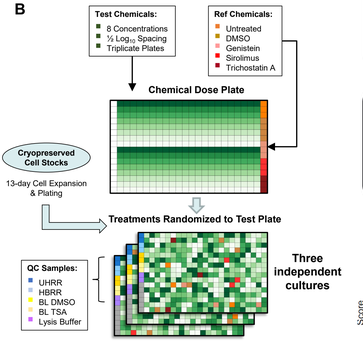


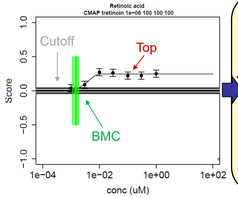
Figure 1. (A) Table listing experimental parameters for this study. (B) Eight-point dilution series of test chemicals were arranged on a dose plate (green) along with quadruplicate wells of DMSO (vehicle control, brown) and triplicate wells of three reference chemicals (pink, red, maroon). Treatments were randomized with respect to position on assay plates. Each assay plate in the experiment was uniquely randomized. Column 1 of each plate contains reference materials used to track assay performance. UHRR = Universal Human Reference RNA; HBRR = Human Brain Reference RNA; BL DMSO = Bulk Lysate DMSO; BL TSA = Bulk Lysate Trichostatin A.

Signature Modeling Approach

Signature Collection Curated From:
- Molecular Signatures Database (PMID: 21546393)
- Connectivity Map (PMID: 17008526)
- DisGeNET (PMID: 25877637)
- BioPlanet (PMID: 31133849)

Signature Scoring Approach:
- For each treatment, calculate a signature score using single sample gene set enrichment analysis (ssGSEA) (PMID: 19847166)

Signature Concentration-Response Modeling:
- Use tcplfit2 (PMID: 34791027) for high throughput concentration-response modeling of signature scores.
- Calculate benchmark concentration (BMC) for each signature x chemical combination.



Signed, Scaled Area Under the Curve (ssAUC)

$$[\,3 - \log_{10}(BMC)\,] \times |\,Top / Cutoff\,| \times Sign(Top)$$

Assay Quality Control (QC)

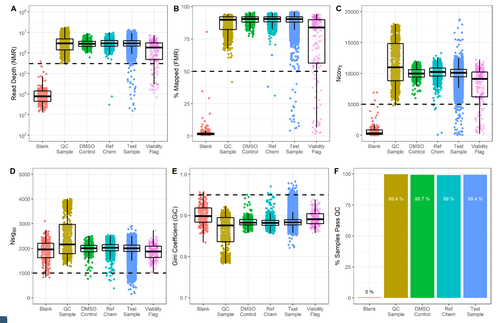


Figure 2. Quality Control. Quality control (QC) metrics described in Harrill et al. (2020) were calculated for all 33,886 samples. (A) Number of Mapped Reads (NMR). (B) Fraction of Mapped Reads (FMR). (C) Number of probes with at least 5 reads (N_{OxN}). (D) Number of probes capturing top 80% of signal ($NSig_{80}$). (E) Gini coefficient, a measure of the degree of inequality in a distribution of values. (F) Percent-age of samples passing QC criteria, stratified by sample type. Points below (A-D) or above (E) the horizontal dotted line in each panel are samples that did not pass QC and were discarded from the study.

- The high-throughput transcriptomics (HTTr) study contained 32,886 samples.
- > 99% of samples were of high quality based on QC criteria established in Harrill et al. (2020) (PMID: 33538836)

Signature Modeling Identifies Expected Biology for Reference Chemicals

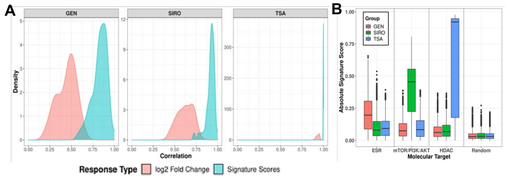


Figure 2. Reference Chemical Results. (A) Distribution of log2 fold-change (red) and signature scores (blue) correlations in like QC samples across all plates in this study. (B) Reference chemicals produce higher scores for signatures matching their known mechanism of action.

- Correlation of signature scores are higher than correlation of fold-change.
- Signature scores identify the expected biological responses for reference chemicals.

Consistent Results for Chemicals Screened In Duplicate

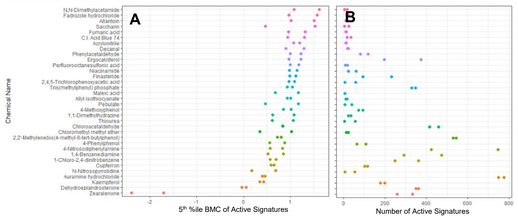


Figure 3. Duplicate Chemical Results. 34 chemicals were screened in duplicate in the HTTr study. (A) 5th percentile of signature-level BMCs. (B) The number of active signatures, defined as having a hitcall ≥ 0.9, top / cutoff ≥ 1 and BMC within the tested concentration range.

- BMCs varied by less than 1 order of magnitude for the 34 chemicals screened in duplicate.
- The number of active signatures was comparable for duplicates of the same chemical.

Signature Modeling Reveals Patterns in Biological Activity of Chemicals

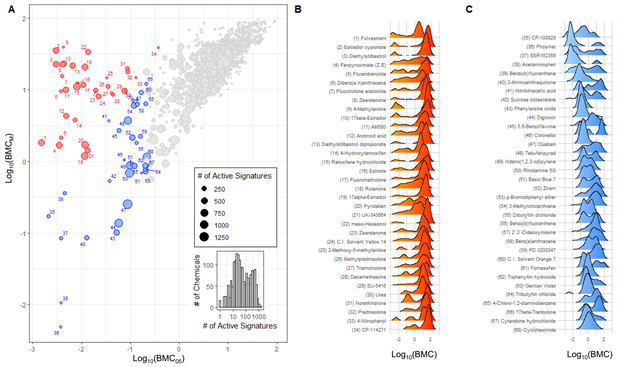


Figure 4. HTTr Screening Summary Result. (A) Scatterplot of the 5th percentile BMC (BMC_{05}) of all active signatures versus the median BMC (BMC_{50}) of all active signatures. Each bubble represents a chemical. Potent chemicals with ≥ 2 orders of magnitude difference between BMC_{05} and BMC_{50} are shaded red and numbered. Potent chemicals with ≤ 2 orders of magnitude difference between BMC_{05} and BMC_{50} are colored blue. (B-C) Ridgeline plots of the distribution of signature level BMCs for chemicals are shaded red (B) or blue (C) in panel A. Chemicals in red display a peak of biological activity at low concentrations, followed by a larger peak of biological activity at low concentrations.

- 1767 out of 1784 chemicals (99%) had at least one active signature.
- The number of active signatures / chemical ranged from one to several hundred.
- A peak of biological activity at low concentrations is observed for known modulators of estrogen receptor (ER), glucocorticoid receptor (GR) and retinoic acid receptor (RAR).
- Other highly potent chemicals included metals, dyes, polyaromatic hydrocarbons (PAHs), sulfhydryl reactive chemicals and ATPase inhibitors. These did not display the "early" peak of biological activity observed with ER, GR and RAR modulators.

Chemicals of the Same Type Elicit Similar Transcriptomic Responses

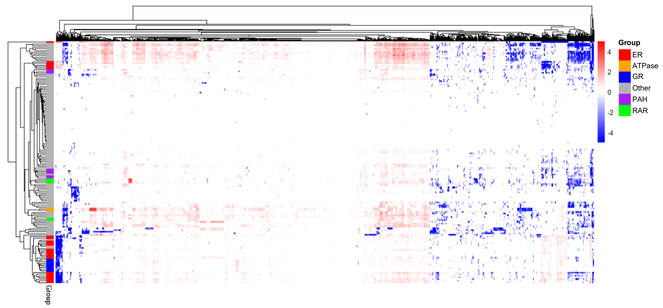
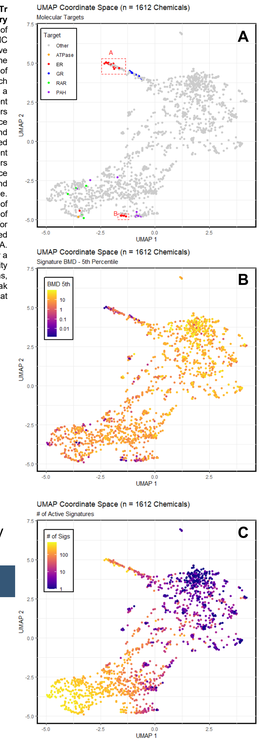


Figure 5. Clustering of Chemicals Based on ssAUC. Heatmap of ssAUC values for all 68 chemicals annotated in Figure 4 (rows) and all signatures that were active in ≥ 10% of those chemicals. Chemicals and signatures were clustered using option == "complete" in the *pheatmap* R-package. Color coding on the lefthand side denotes chemicals where the primary molecular mechanism of action is known.

- ER, GR, RAR and ATPase modulators, as well as PAHs, cluster together using the ssAUC metric.

Chemicals With Similar Activity Identified with UMAP



D Group A (ER Agonists) Chemical Name	ToxCast ER Pathway Model (AUC) Agonist	Antagonist
17alpha-Estradiol	1.06	0
17beta-Estradiol	0.935	< 0.01
17beta-Trenbolone	0.475	0
5alpha-Dihydrotestosterone	0.4	0
Daidzein	0.44	0
Dehydroepiandrosterone	0.365	< 0.01
Diethylstilbestrol	0.943	< 0.01
Diethylstilbestrol dipropionate	-	-
dl-Norgestrel	-	-
Estradiol cypionate	-	-
Estrone	0.807	< 0.01
Ethisterone	-	-
Genistein	0.538	0
Levonorgestrel	0.394	0
meso-Hexestrol	0.993	0
Norethindrone	0.524	0
Zearalenone	0.71	0
Celestolide	0	< 0.01
1,2,4-Trichlorobenzene	0	0
17alpha-Ethinylestradiol	1	0
17-Methyltestosterone	0.495	0
2-(Phenylmethylene)octanal	< 0.01	0
2,2',4,4'-Tetrahydroxybenzophenone	0.397	< 0.01
3,3'-Dimethylbisphenol A	-	-
4-[1,1,3,3-Tetramethylbutyl]phenol	0.393	0
4,4'-Sulfonyldiphenol	0.263	0
4,4'-Thiodianiline	-	-
4-Hydroxybenzophenone	-	-
4-Phenylphenol	0.219	< 0.01
beta-Sitosterol	-	-
Bisphenol B	0.491	< 0.01
Bromofos	-	-
Butylparaben	0.251	< 0.01
Carbophenothion	-	-
EPTC	0	0
FR900409	-	-
Kaempferol	0.252	0
Mestranol	0.742	0
Perfluorodecanoic acid	< 0.01	0
Pregnenolone	-	-
Sodium fluoroacetate	0	0
Tetrac	0	< 0.01
Undecane	0	0

"-" = info not available for ToxCast model.

Group B (ER Antagonists) Chemical Name	ToxCast ER Pathway Model (AUC) Agonist	Antagonist
4-Hydroxytamoxifen	< 0.01	0.686
Clomiphene citrate (1:1)	< 0.01	0.588
Fulvestrant	0	0.635
Raloxifene hydrochloride	< 0.01	0.671
Tamoxifen citrate	< 0.01	0.546

"-" = info not available for ToxCast model.

Small groups of chemicals with similar biological activity can be identified using UMAP.

Figure 6. Uniform Manifold Approximation and Projection (UMAP) of HTTr Results. A matrix of ssAUC values for n = 1612 chemicals having ≥ 10 active signatures (rows) and n = 620 signatures (columns) that were active in > 5% of those chemicals was constructed. A pairwise Euclidean distance matrix was then computed and used as input for the UMAP R-Package to generate a coordinate space. Each point in the coordinate space represents a chemical. Points closer to one another have a more similar HTTr profile than points that are relatively further away from one another. (A) UMAP coordinate space overlaid with "landmark" chemicals from Figure 4. (B) UMAP coordinate space overlaid with BMD_{05}. (C) UMAP coordinate space overlaid with number of active signatures. (D) Table of chemicals found in close proximity to estrogenic "landmark" chemicals using UMAP. Values are ToxCast ER Pathway Model results.

Conclusions

Concentration-response modeling of HTTr signature scores:
- Yields potency estimates for perturbation of cellular biology.
- Provides insight into chemical mechanism-of-action.
- Can be used to group / identify chemicals with similar biological response profiles.