Chemical Safety Screening Using High-Throughput Transcriptomics

Logan J. Everett[1], Joshua A. Harrill[1], Derik Haggard[2], Joseph Bundy[1], Imran Shah[1], Richard Judson[1], Thomas Sheffield[2], R. Woodrow Setzer[1]

[1]U.S. Environmental Protection Agency, Office of Research and Development, Research Triangle Park, NC

[2]Oak Ridge Institute for Science and Education, Research Triangle Park, NC

Logan J. Everett | everett.logan@epa.gov | 919-541-2495

Background

- New Approach Methodologies (NAMs) aim to replace vertebrate animal testing for chemical safety screening and assessment [1] .
- U.S. EPA has proposed a tiered testing strategy using NAMs to broadly identify hazards from chemical exposure and characterize their dose-response relationships [2].
- There is a need for NAMs that are both high-throughput and broad coverage for the first tier of testing.
- Targeted RNA-seq of cultured human cells provides a platform for high-throughput transcriptomics (HTTr) that covers >20,000 genes and a wide array of biological responses and pathways [3].
- HTTr is intended to predict the overall benchmark dose (BMD) for preliminary risk assessment, as well as specific hazards and molecular initiating events (MIEs) to aid in selection of orthogonal testing at later tiers.

Design & Analysis of Screening Studies

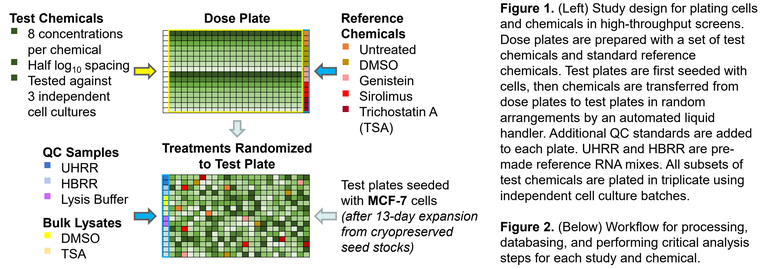
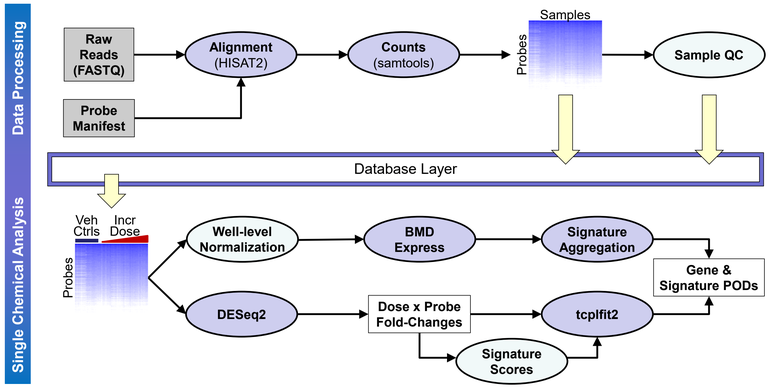


Figure 1. (Left) Study design for plating cells and chemicals in high-throughput screens. Dose plates are prepared with a set of test chemicals and standard reference chemicals. Test plates are first seeded with cells, then chemicals are transferred from dose plates to test plates in random arrangements by an automated liquid handler. Additional QC standards are added to each plate. UHRR and HBRR are pre-made reference RNA mixes. All subsets of test chemicals are plated in triplicate using independent cell culture batches.

Figure 2. (Below) Workflow for processing, databasing, and performing critical analysis steps for each study and chemical.



Pilot Study Method Validation

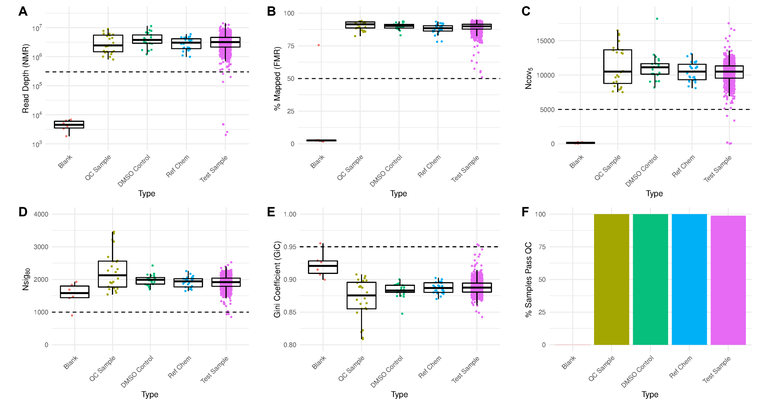
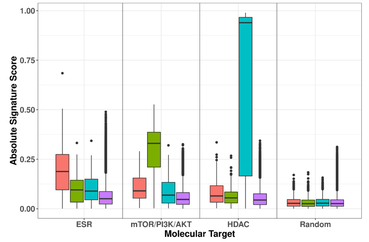


Figure 3. (Above) Distributions of QC metrics across 1,134 targeted RNA-seq profiles from MCF-7 pilot study of high-throughput screening methods, with corresponding thresholds for sample rejection (dashed lines). (A) Read depth = total # of reads uniquely aligned to probe manifest. (B) % Mapped = % of all sequenced reads uniquely aligned to probe manifest. (C) $Ncov_5$ = # of probes with minimum coverage of 5 reads. (D) $Nsig_{80}$ = # of probes capturing top 80% of signal. (E) Gini Coefficient is a generalizable metric of overall inequality in any distribution, applied here to read counts across all probes in a sample (lower values indicate more even distribution of reads across probes). (F) % of samples passing all QC thresholds by category. 13 samples corresponding to test chemical treatment conditions were flagged and removed.



	Signature	Absolute Score	Direction
Genistein	RYAN_ESTROGEN_RECEPTOR_ALPHA	0.6844617	+
	CMAP equilin 1.5e-05 100 5542 100	0.5054424	+
	CMAP fulvestrant 1e-08 100 1417 100	0.4857079	-
	CMAP fulvestrant 1e-06 100 6811 100	0.4519683	-
	BHAT_ESR1_TARGETS_NOT_VIA_AKT1	0.4270213	+
Sirolimus	CMAP sirolimus 1e-07 100 5437 100	0.5265263	+
	CMAP sirolimus 1e-07 100 6409 100	0.5166319	+
	CMAP sirolimus 1e-07 100 8359 100	0.5001286	+
	CMAP wortmannin 1e-06 100 577 100	0.4990501	+
	CMAP wortmannin 1e-08 100 1423 100	0.4971645	+
Trichostatin A	CMAP vorinostat 1e-05 100 817 100	0.9899307	+
	CMAP trichostatin A 1e-07 100 8791 100	0.9869679	+
	CMAP trichostatin A 1e-07 100 6547 100	0.9854099	+
	CMAP trichostatin A 1e-06 100 8056 100	0.9852016	+
	CMAP trichostatin A 1e-07 100 7972 100	0.9849593	+

Figure 4. Biological reproducibility and accuracy of reference chemical treatments based on signature scoring method. Three distinct reference chemicals were run at a single bioactive concentration in triplicate on each test plate: GEN = Genistein (red), which targets estrogen receptor (ESR); SIRO = Sirolimus (green), which targets mTOR/PI3K/AKT pathway; TSA = Trichostatin A (blue), which inhibits Histone Deacetylase (HDAC). DESeq2 fold-changes for each chemical treatment vs DMSO were scored for overall signature activity against a broad gene set collection using ssGSEA method, and gene sets were further aggregated based on relevant molecular targets. (Left) Distributions of absolute signature scores for relevant molecular target signatures across each reference chemical treatment group; Null (purple) = 1,000 simulated chemical treatments derived from re-sampling real Log_2 fold-change data; Random = Randomly generated gene sets with similar size and gene frequency as real gene sets. (Right) Table of the top 5 highest ranked signatures by absolute signature score for each reference chemical in the pilot study.

Pilot Study Test Chemicals

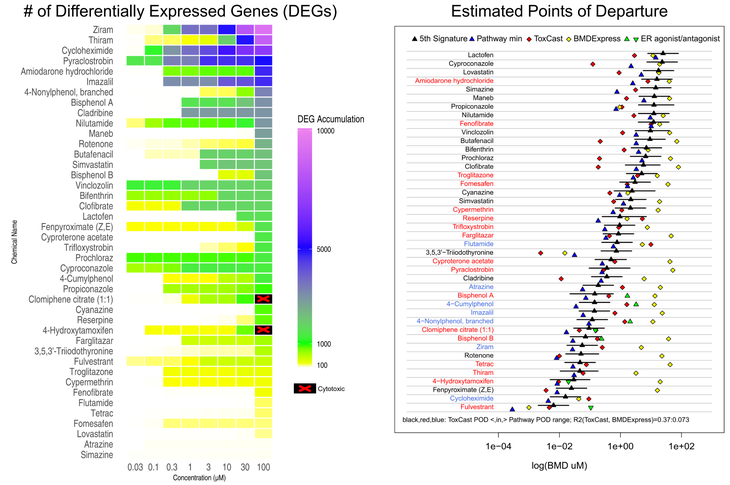


Figure 5. Transcriptomic responses of MCF-7 cells to test chemicals. (Left) Accumulation of total differentially expressed genes (DEGs) at increasing concentrations (low to high) for 42 test chemicals. DEGs are based on DESeq2 analysis with 10% FDR (black = cytotoxic dose level). (Right) Overall transcriptional point of departure (POD) for each chemical, determined by two parallel methods: i) signature aggregation of DESeq2 fold-changes, followed by concentration-response modeling of the signature scores, with minimum (blue triangles) or 5th lowest (black triangles) signature BMD used as overall POD; ii) probe-level concentration-response modeling followed by signature-level aggregation of BMDs as proposed by the National Toxicology Program [4] (yellow diamonds). PODs determined from the 5th percentile of all ToxCast High-Throughput Screening (red diamonds) and results from an integrative model for ER agonists/antagonists [5] (green triangles) are shown for comparison. Chemical names are color-coded based on the comparison between 5th lowest signature POD (black triangle) and ToxCast POD (red diamonds): Names in black indicate ToxCast POD < transcriptional POD; red indicates ToxCast POD within the confidence interval for transcriptional POD; blue indicates transcriptional POD < ToxCast POD.

References

1. 15 U.S.C. §2603(h): Reduction of Testing on Vertebrates
2. Thomas, et al. *The Next Generation Blueprint of Computational Toxicology at the U.S. Environmental Protection Agency.* Toxicological Sciences 2019, 169(2):317-332
3. Yeakley, et al. *A Trichostatin A Expression Signature Identified by TempO-Seq targeted Whole Transcriptome Profiling.* PLoS One 2017, 12(5)e0178302
4. National Toxicology Program. *NTP research report on National Toxicology Program approach to genomic dose-response modeling.* NTP RR 5, 2018.
5. Judson, et al. *Integrated Model of Chemical Perturbations of a Biological Pathway Using 18 In Vitro High-Throughput Screening Assays for the Estrogen Receptor.* Toxicological Sciences, 2015, 148(1)137-154

The views expressed are those of the presenter and do not necessarily reflect the views or policies of the U.S. Environmental Protection Agency.