High-Throughput Chemical Safety Screening Using Targeted RNA-seq

Logan J. Everett, Joshua A. Harrill, Derik Haggard, Joseph Bundy, Beena Vallanat, Imran Shah, Richard Judson

U.S. Environmental Protection Agency, Office of Research and Development, Center for Computational Toxicology & Exposure
Research Triangle Park, North Carolina, U.S.A.

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

Overview

- U.S. EPA has proposed a tiered testing strategy using New Approach Methodologies (NAMs) to identify hazards from chemical exposure and characterize their dose-response relationships. The first tier of testing will utilize NAMs that are high-throughput and provide broad biological coverage [1].
- Targeted RNA-seq of cultured human cells provides a platform for high-throughput transcriptomics (HTTr) screening that covers >20,000 genes and a wide array of biological responses and pathways [2].
- We successfully piloted the targeted RNA-seq approach for HTTr to predict biological pathway altering concentrations (BPACs) [3] and have now scaled this approach to screen over 1,200 chemicals in three distinct cell types.
- We have developed rigorous quality control procedures to remove aberrant samples that are scalable to tens of thousands of sequencing libraries per study.
- Differential expression models [4] reveal dose-responsive accumulation of transcriptional changes across cell types and chemicals tested, and signature-based analysis is used to derive potency and mechanism of action information for each chemical.

Screen Design & Analysis

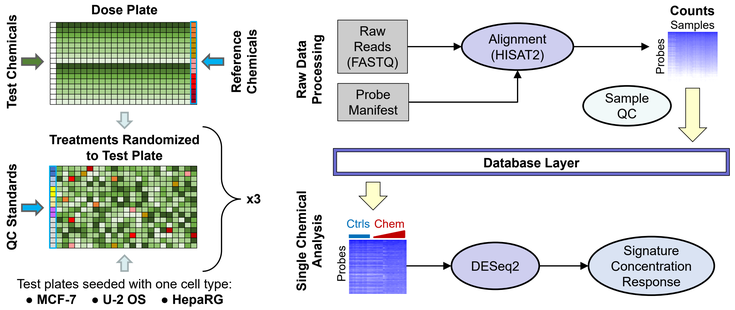

Figure 1. Design and analysis of large-scale high-throughput transcriptomic screens. Left: Dose plates are prepared with ~40 test chemical samples at 8 concentrations (half $\log_{10}$ spacing, single replicates) and a standard set of reference chemicals for each cell type. Cells are grown on the test plate, then treated with chemical samples from the dose plate. Positions of each chemical treatment are randomized on every plate by an automated liquid handling system. Each test plate is generated in triplicate using the same test chemicals, but with separate cell culture batches. QC reference standards are manually added to each test plate before transcriptomic profiling. Right: Raw data from targeted RNA-seq samples is rapidly aligned to known probe sequences producing counts of uniquely aligned reads for each probe in each sample. Probe counts are used to derive QC metrics and are stored in a database layer. Subsequent analysis is performed independently for each chemical. Count data for all concentrations, replicates, and plate-matched vehicle controls are extracted. DESeq2 [4] is used to compute moderated fold-changes, which are then input to a novel method for modeling concentration-responsive activity for a catalog of known gene signatures.

Scalable Quality Control

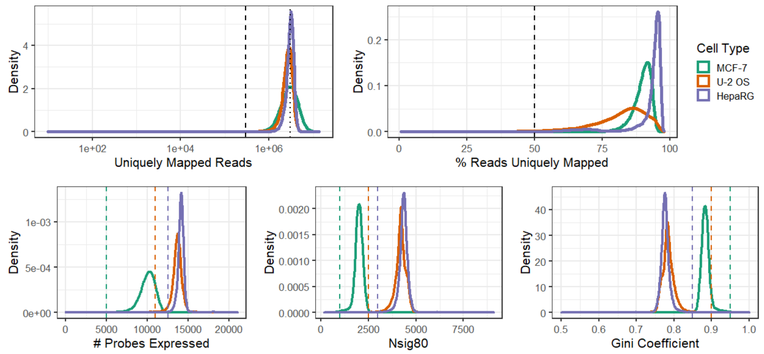

Figure 2. Quality control metrics by cell type. The QC process described in [4] was applied to >100,000 total samples from HTTr screening of >1,200 chemicals in three cell types. Target read depth of 3M reads was achieved in all cell types (upper left), and the majority of reads mapped uniquely to a known probe (upper right). Additional QC metrics differed by cell type, and a distinct threshold was set for each case using Tukey's outer fence of the overall distribution (dashed lines). Nsig80 is defined as the minimum number of probes capturing ≥80% of the total reads. Gini coefficient is a metric of inequality, where 1 = all reads aligned to a single probe and 0.5 = all probes having an equal number of aligned reads.

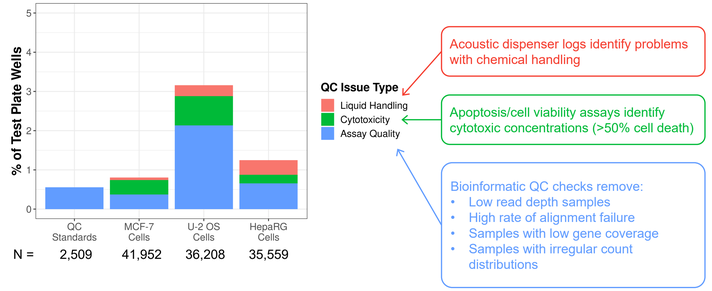

Figure 3. Quality control failure rates by sample and cell type. Individual HTTr profiles were excluded from further analysis due to errors in acoustic dispensing of chemicals (red), >50% cell death indicated by cell viability assay (green), or abnormal parameters from a battery of bioinformatic checks applied to sequencing data results (blue, see Figure 2). Greater than 98% of samples pass all QC filters, demonstrating the ability to reliably scale this workflow to studies encompassing thousands of chemicals and samples.

Differential Gene Expression Analysis

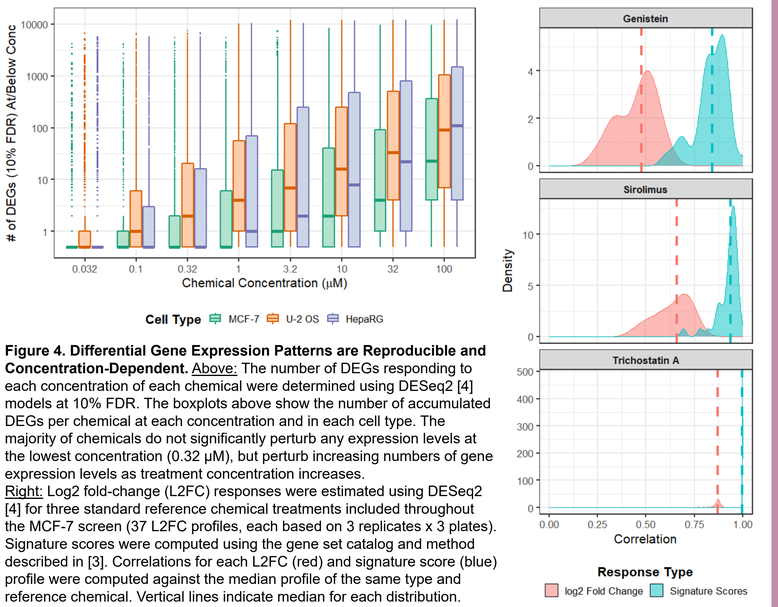

Figure 4. Differential Gene Expression Patterns are Reproducible and Concentration-Dependent. Above: The number of DEGs responding to each concentration of each chemical were determined using DESeq2 [4] models at 10% FDR. The boxplots above show the number of accumulated DEGs per chemical at each concentration and in each cell type. The majority of chemicals do not significantly perturb any expression levels at the lowest concentration (0.32 µM), but perturb increasing numbers of gene expression levels as treatment concentration increases. Right: Log2 fold-change (L2FC) responses were estimated using DESeq2 [4] for three standard reference chemical treatments included throughout the MCF-7 screen (37 L2FC profiles, each based on 3 replicates x 3 plates). Signature scores were computed using the gene set catalog and method described in [3]. Correlations for each L2FC (red) and signature score (blue) profile were computed against the median profile of the same type and reference chemical. Vertical lines indicate median for each distribution.

Summary & References

- **>98% of samples pass rigorous QC tests, even when generating >10,000 targeted RNA-seq samples.**
- **Reference chemicals included throughout the large screening studies demonstrate the reproducibility of differential expression profiles for bioactive treatments.**
- **Differential expression patterns across chemicals and concentrations can be mined to infer potency and mechanism of action for each chemical.**

1. Thomas, et al. *The Next Generation Blueprint of Computational Toxicology at the U.S. Environmental Protection Agency.* Toxicological Sciences 2019, 169(2):317-332
2. Yeakley, et al. *A Trichostatin A Expression Signature Identified by TempO-Seq targeted Whole Transcriptome Profiling.* PLoS One 2017, 12(5)e0178302
3. Harrill, et al. *High-Throughput Transcriptomics Platform for Screening Environmental Chemicals.* Toxicological Sciences 2021, doi:10.1093/toxsci/kfab009
4. Love, et al. *Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2.* Genome Biology 2014, 15(12):550

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