Robust Workflows for the Analysis of High-Throughput Transcriptomics Data in Chemical Safety Screening

Logan J. Everett[1], Joshua A. Harrill[1], Derik Haggard[2], 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 HTTr Studies

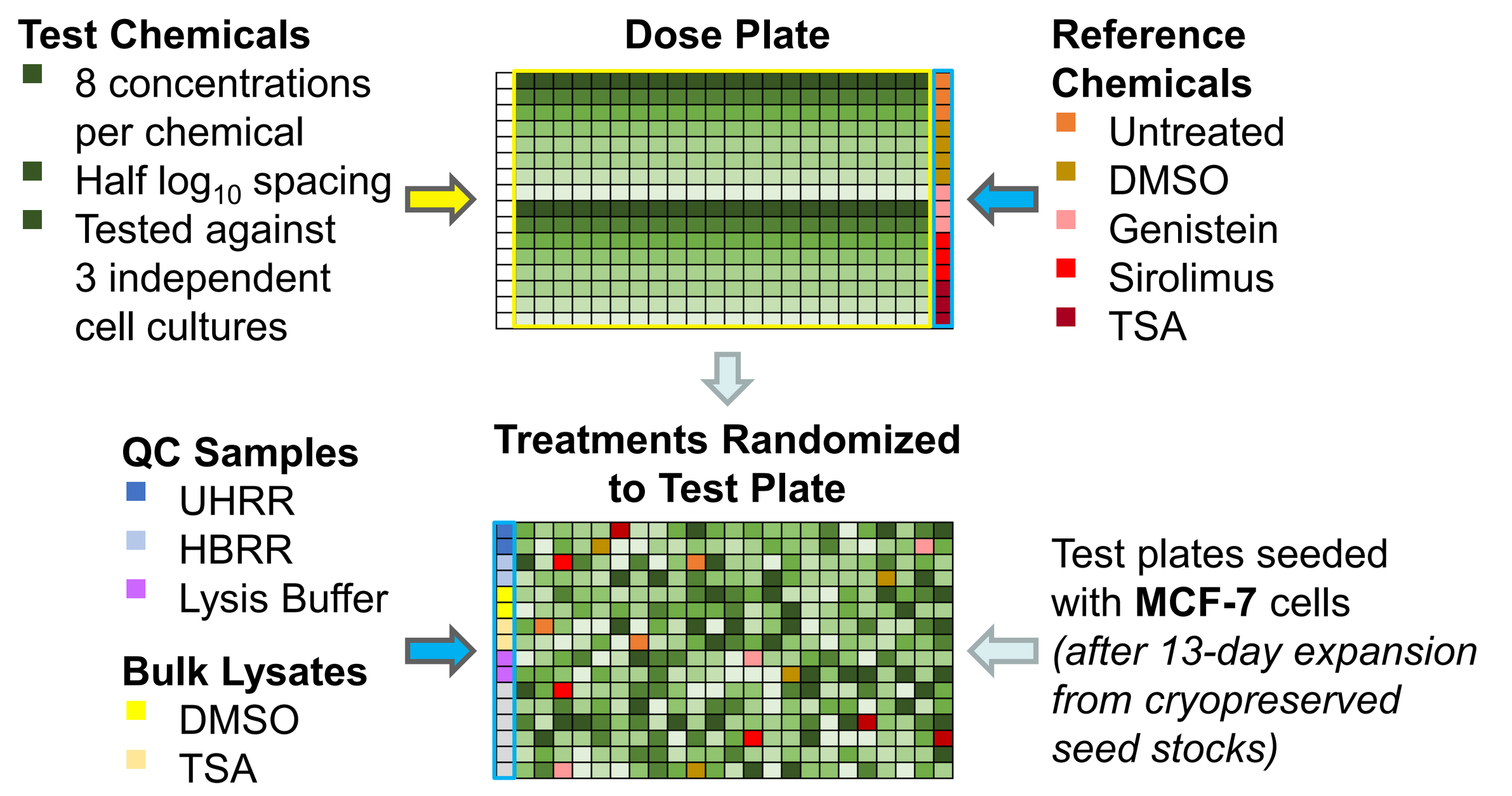
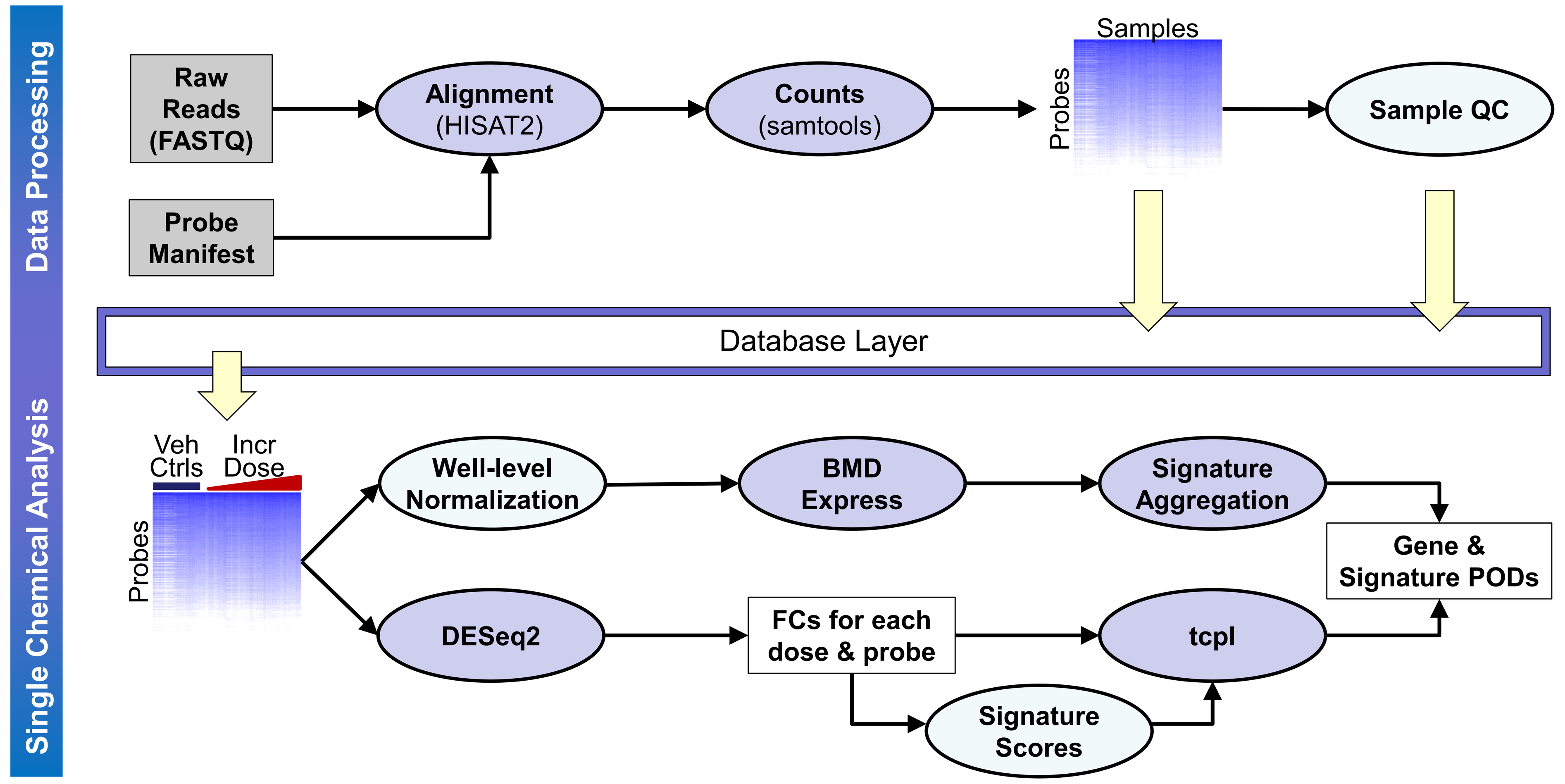
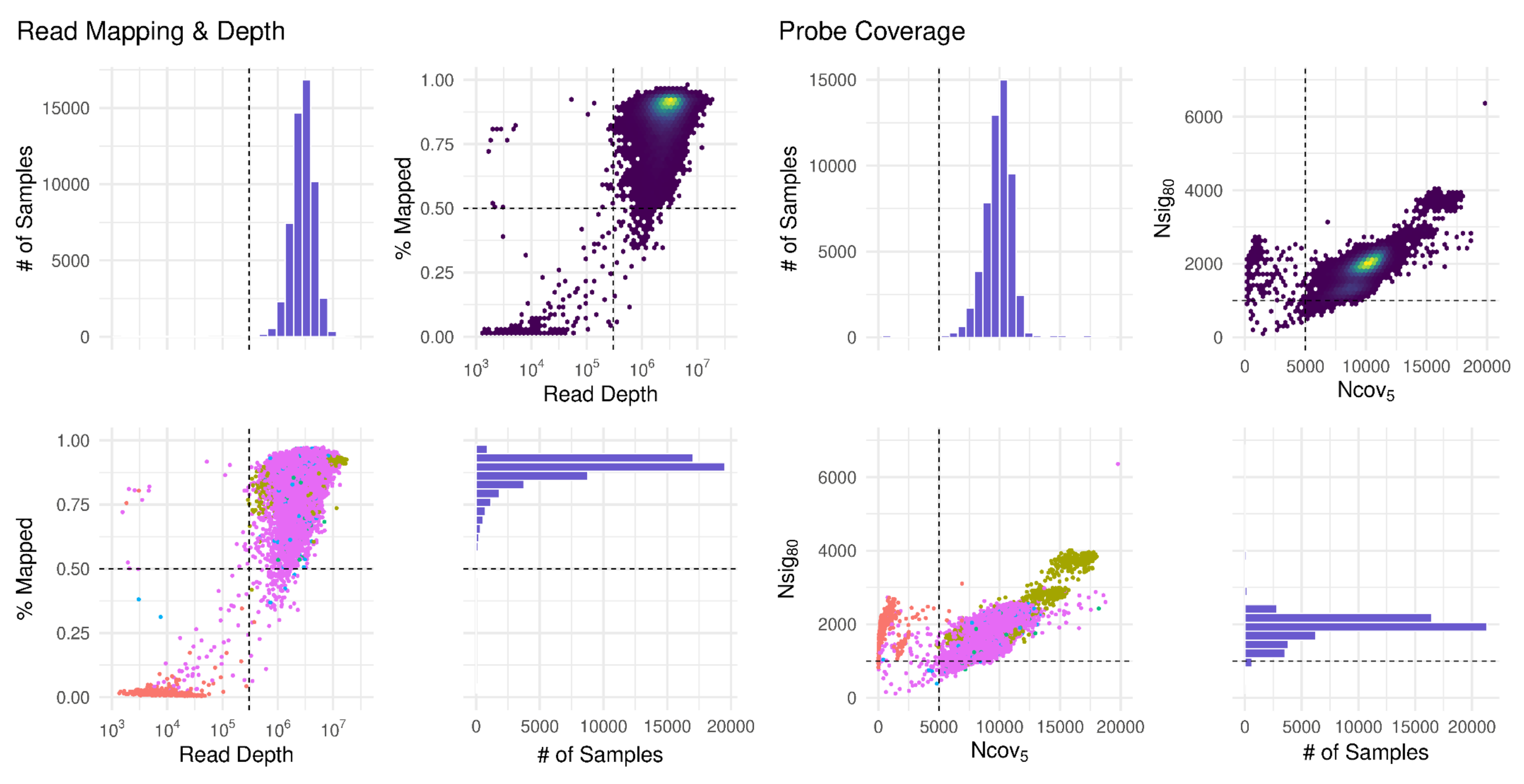


Figure 1 (Left) Study design for plating cells and chemicals in high-throughput screens. Dose plates are prepared with a set of test chemicals and standard reference chemicals. Test plates are first seeded with cells, then chemicals are transferred from dose plates to test plates in random arrangements 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.



Method Validation



Quality Control Metrics:

- Read depth = total # of reads uniquely aligned to probe manifest
- % Mapped = % of all sequenced reads uniquely aligned to probe manifest
- $Nsig_{80}$ = # of probes capturing top 80% of signal
- $Ncov_5$ = # of probes with minimum coverage of 5 reads

Figure 3 (Above) Distribution of QC metrics across ~55,000 samples from MCF-7 screening studies covering ~2,000 chemicals. QC metrics are used to identify a small proportion of individual samples with quality issues, such as low input material.

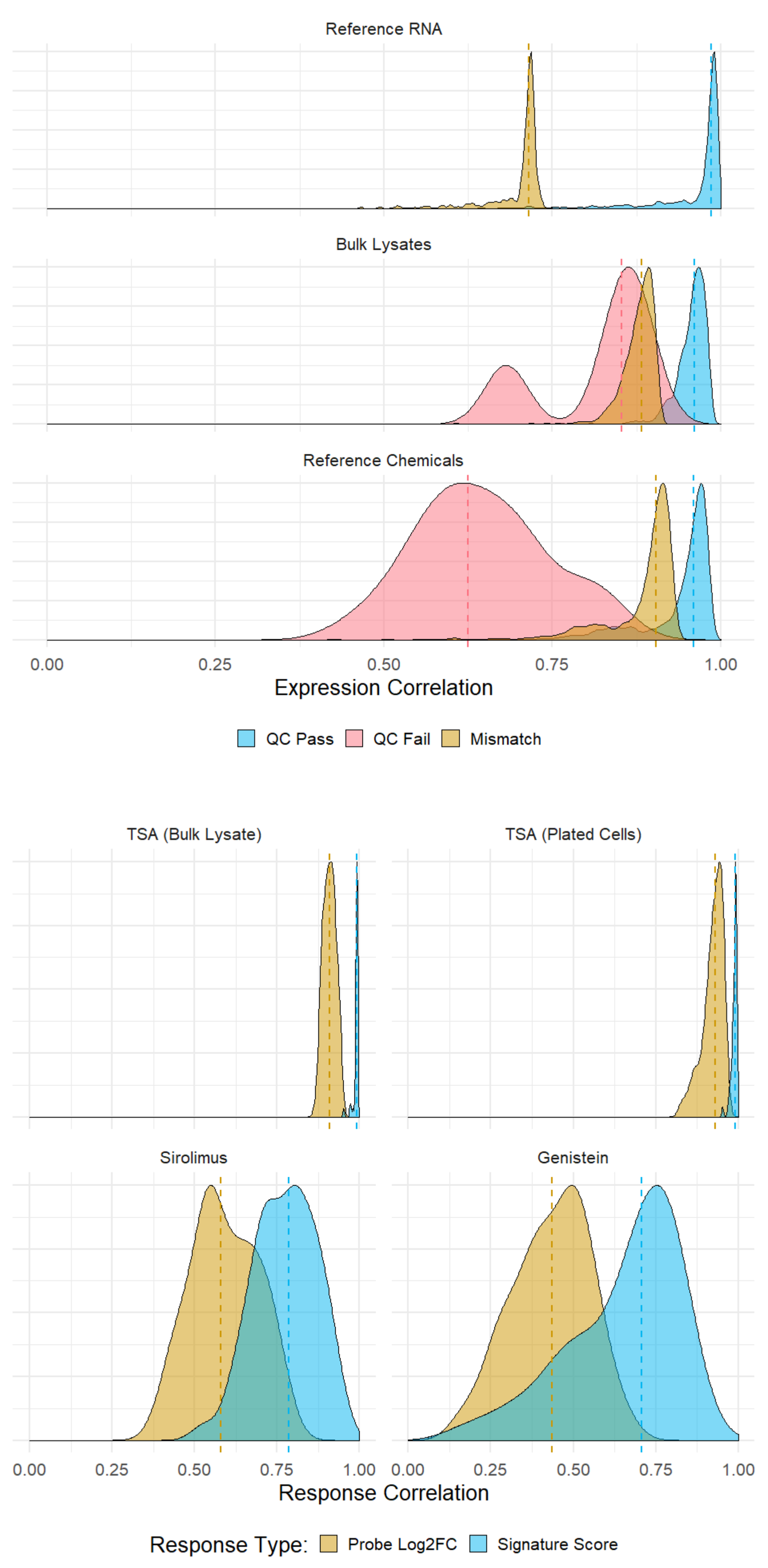
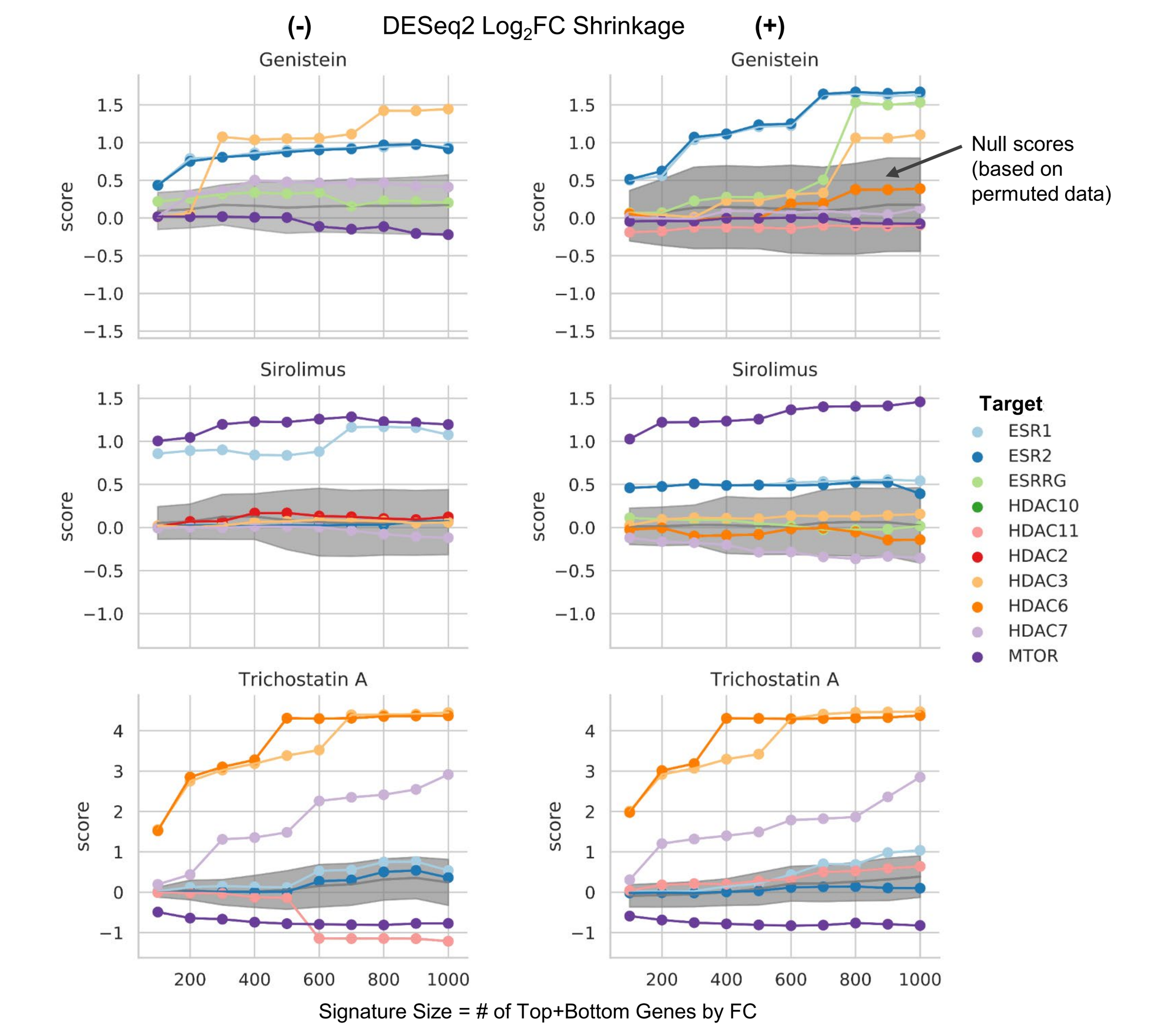


Figure 4 (Upper Left) Distribution of expression correlations for QC and reference samples across the screening study. Correlations are higher for replicates passing QC (blue) than for replicates failing QC (red) or for correlations between different sample types (yellow). (Lower Left) Distribution of correlations for response profiles of reference chemical treatments. Correlations are higher using aggregate signature scores (blue) than using Log2 fold-change of individual probes (yellow).

Figure 5 (Below) Optimization of DESeq2 and signature scoring method using pilot data. Scores are shown for each reference chemical against signatures for known targets. Scores are shown +/- DESeq2 fold-change shrinkage and for multiple signature sizes (x-axis). Shrunken fold-changes and signature size of 200 were chosen for primary analysis.



Pilot Study Results

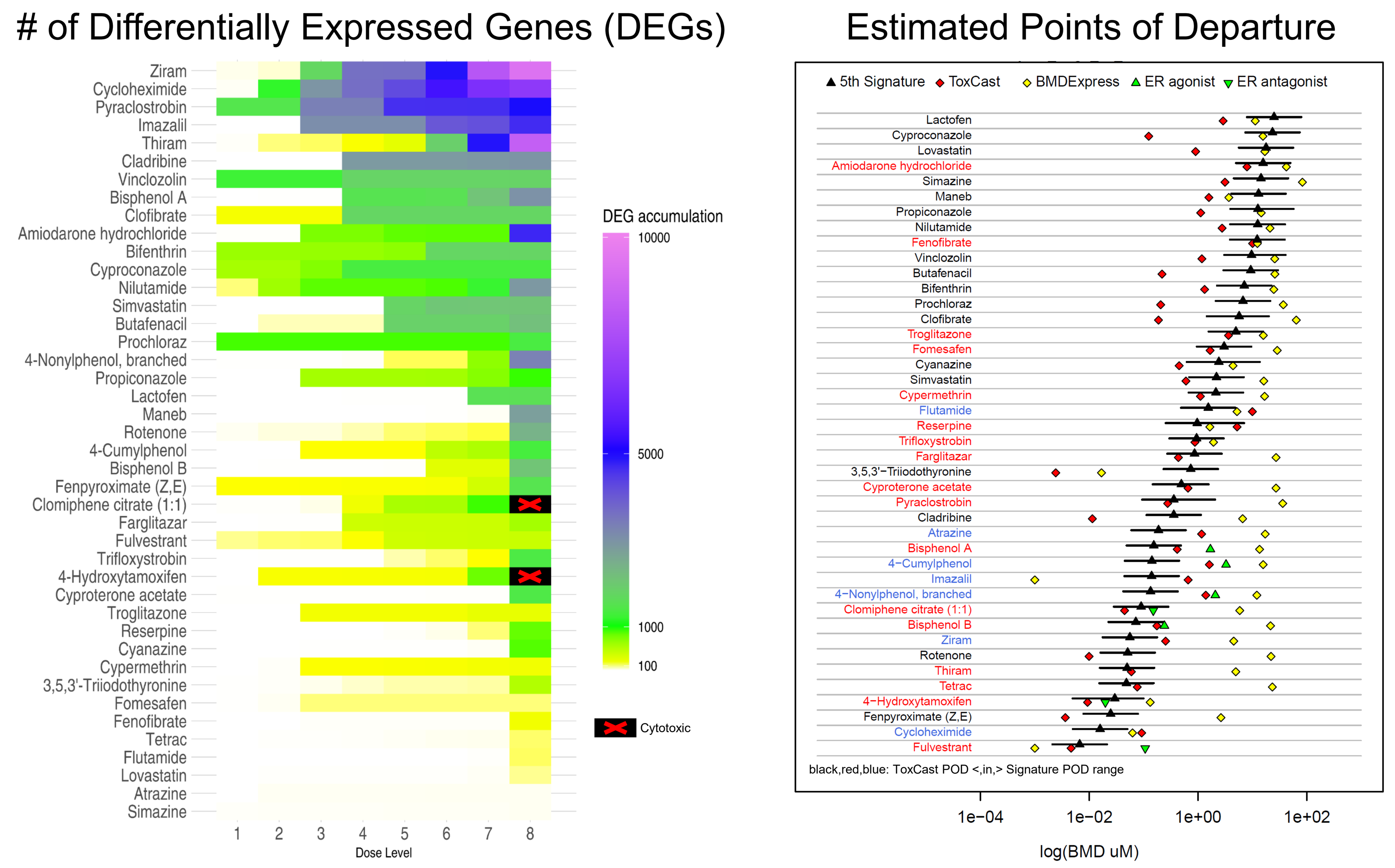


Figure 6. Results from a pilot study using MCF-7 cells. (Left) Accumulation of total differentially expressed genes (DEGs) at increasing dose levels (low to high) for 42 pilot chemicals. DEGs are based on DESeq2 analysis with 10% FDR (black = cytotoxic dose level). (Right) Overall transcriptional point of departure (POD) for each chemical, determined by two parallel methods. Black triangles are based on signature aggregation of DESeq2 fold-changes, followed by concentration-response modeling of the signature scores, with 5th lowest signature BMD used as overall POD. Yellow diamonds are based on concentration-response modeling of individual probes using BMDExpress 2.0 followed by signature-level aggregation of BMDs as proposed by the National Toxicology Program [4]. 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.

References & Acknowledgements

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 authors would like to acknowledge additional contributions from Joseph Bundy.

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