Optimizing Transcriptomics for High-Throughput Bioactivity Screening of Environmental Chemicals

www.epa.gov

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

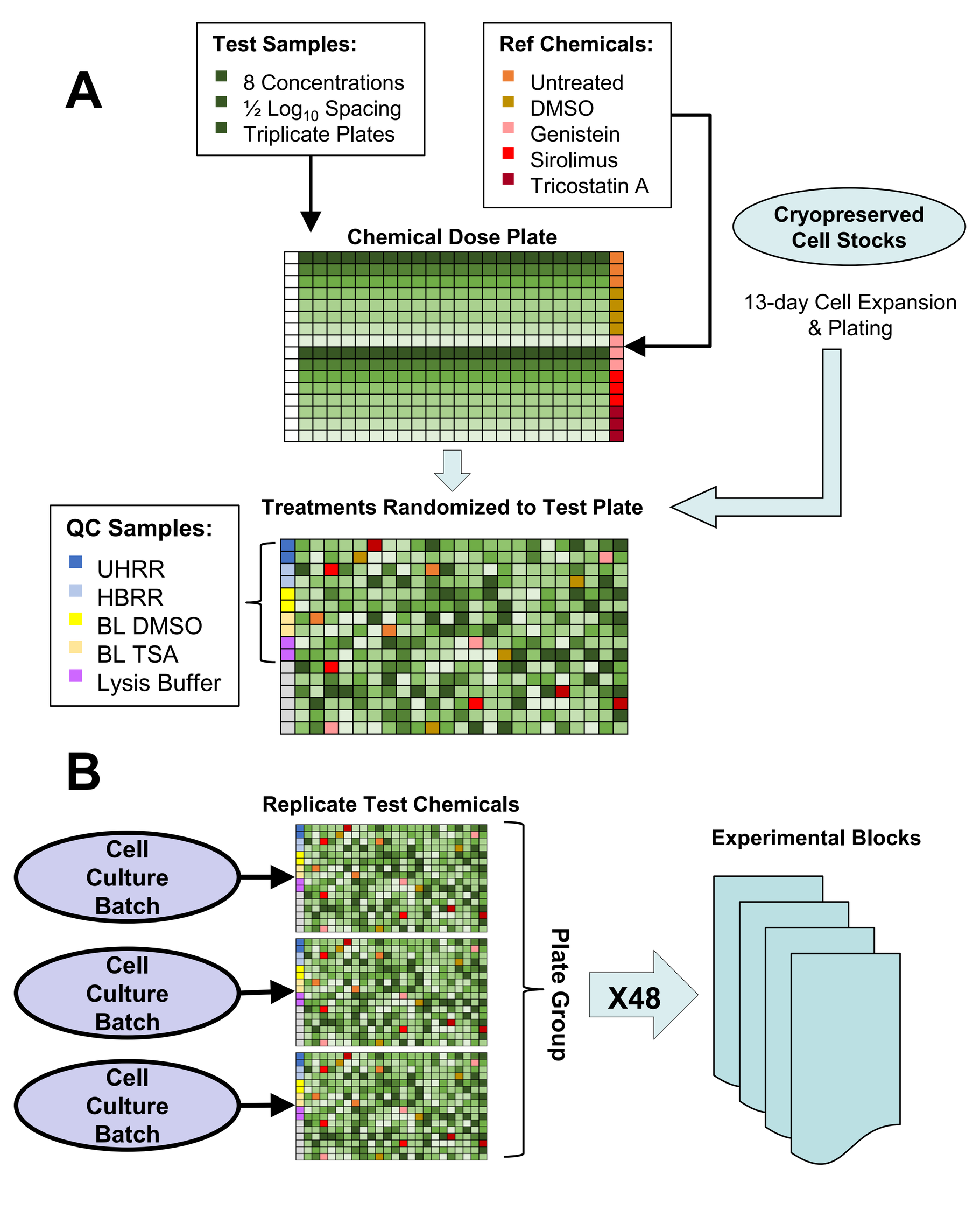
[1]ORISE Postdoctoral Research Participant [2]Center for Computational Toxicology and Exposure, Office of Research and Development, U.S. Environmental Protection Agency

Derik E. Haggard I haggard.derik@epa.gov I 919-541-0745

Background

- There is a need to develop New Approach Methodologies (NAMs) with the goal to replace low-throughput *in vivo* studies for chemical risk assessment [1].
- Implementing NAMs will likely require a tiered screening strategy for chemical bioactivity characterization and assessment [2].
- High-throughput transcriptomics (HTTr) shows promise as a first tier screening approach due to its ability to detect numerous mechanisms of bioactivity.
- The US EPA has developed a HTTr screening assay using targeted RNA-seq and have screened 2012 chemicals with multiple reference standards in MCF7 breast cancer cells [3].
- Evaluation of the reference standards is needed to determine the stability and reproducibility of the HTTr assay.

HTTr MCF7 Assay Experimental Design



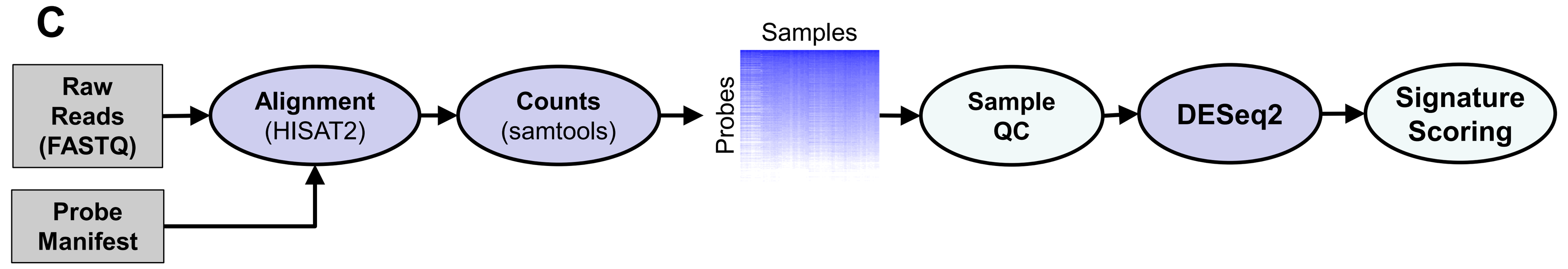


Figure 1. (A) Strategy for plating chemicals in high-throughput screens. Dose plates are prepared with a set of test chemicals, reference chemicals, and reference standards. Test plates are first seeded with cells, then chemicals are transferred from dose plates to test plates in random arrangements by LabCyte Echo® 550 Liquid Handler. Additional QC standards are added to each plate. Parallel test plates are prepared and an apoptosis/cell viability assay is performed to determine cytotoxicity of chemicals or concentrations.

(B) All subsets of test chemicals are plated in triplicate (one replicate per test plate) using independent cell culture batches (~10 test plates per culture batch) and are considered a plate group. A total of 48 plate groups were screened across four independent experimental locks.

(C) Sequence alignment of FASTQ files were performed using HISAT2 and uniquely aligned reads were quantified using samtools. Multiple quality control metrics were calculated for all samples and samples with sufficient quality were further analyzed using DESeq2 and gene set enrichment analysis using single-sample GSEA (ssGSEA) [4].

Overall Sample Quality of the HTTr MCF7 Screen

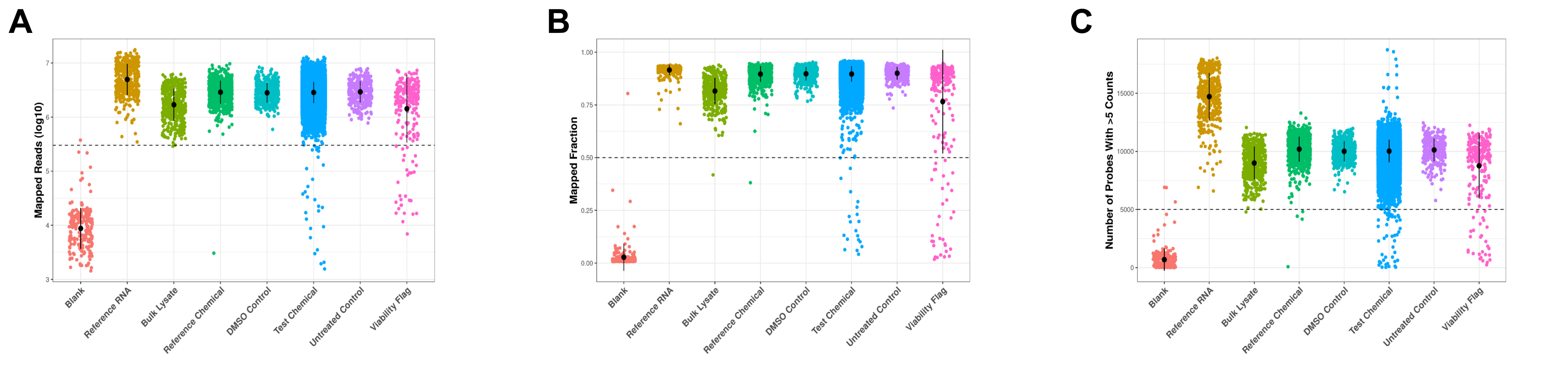


Figure 2. Sample quality of the HTTr MCF7 screen measured by several standard quality metrics. Distributions of total mapped reads **(A)**, mapped fraction **(B)**, and total number of probes with >5 total counts **(C)** by treatment type across all samples. Dashed line indicates threshold to flag low quality samples. The "Viability Flag" category (pink points) represent samples where cell viability in a parallel apoptosis/cell viability assay was ≤50%**.** A total of 53313 of 54432 (~98%) of all samples passed these quality metrics.

Reproducibility of the HTTr MCF7 Reference Standards

Reference RNA and Bulk Lysates:

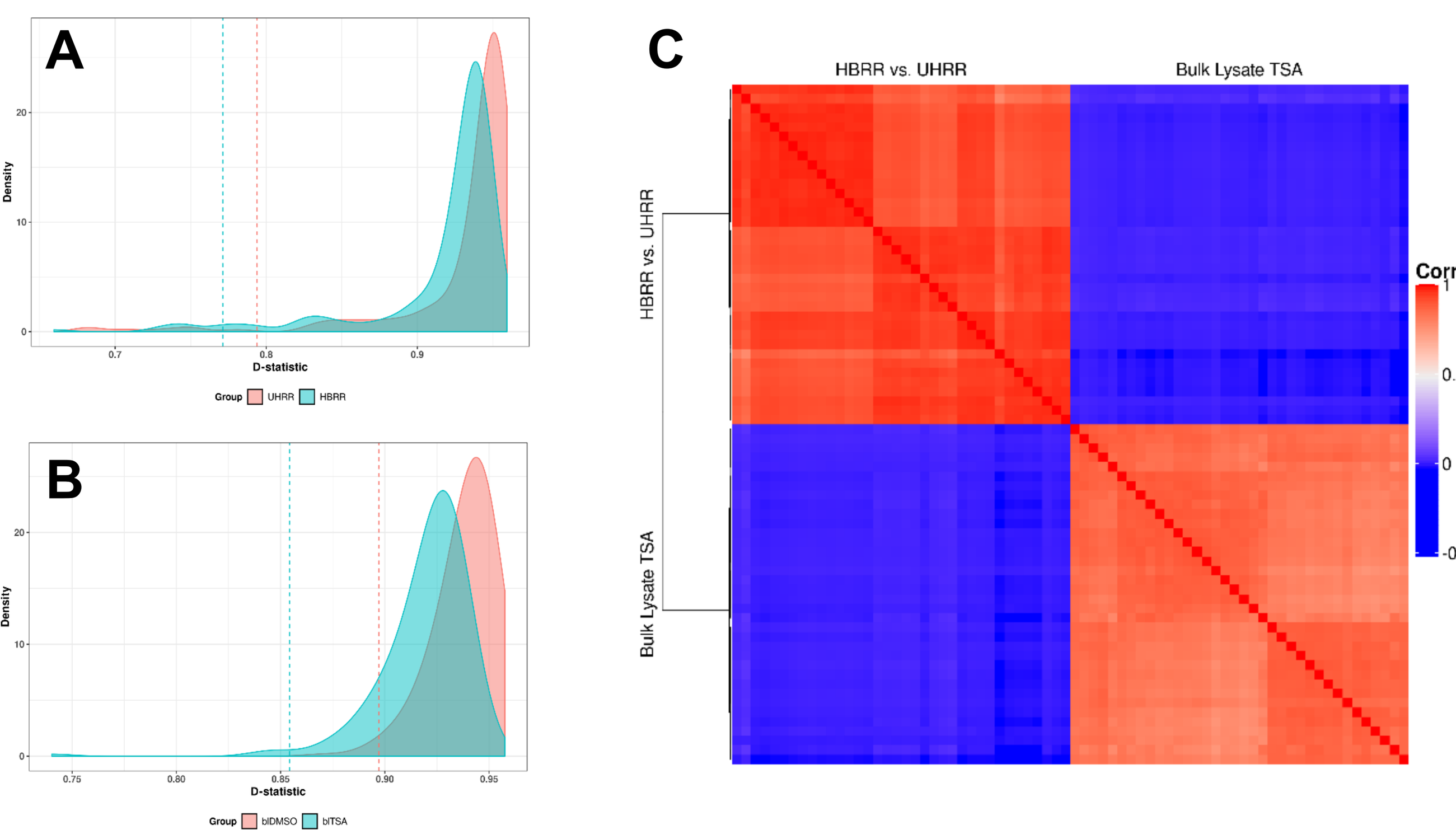
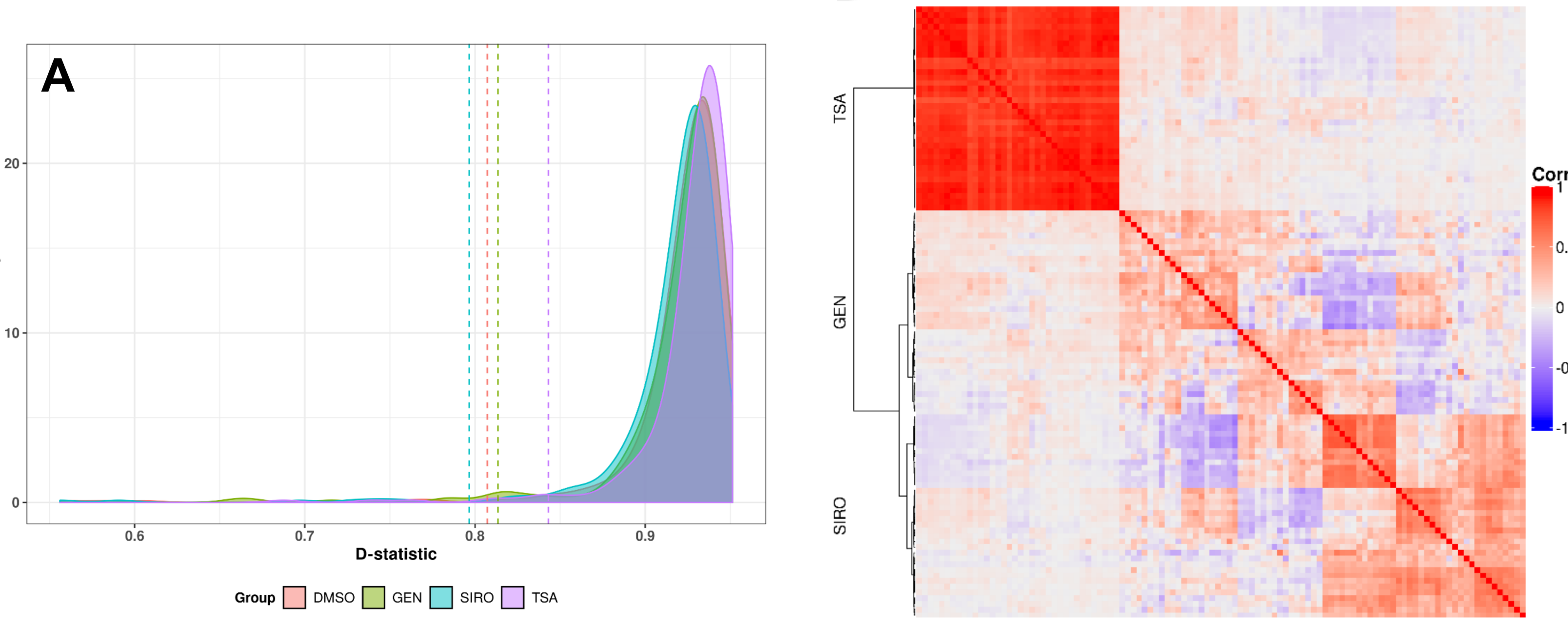


Figure 3. Reproducibility and performance of reference RNA, bulk lysate samples, and reference chemical treatments. The D Statistic (mean correlation of log2 counts per million for each treatment group to all other like treatment groups) was calculated for **(A)** reference RNAs and **(B)** bulk lysate samples [5]. Dashed vertical lines denote the threshold for flagging samples with poor D Statistic (three standard deviations below the mean). **(C)** K-means (K=2) clustered Heatmap of DESeq2 shrunken log2 fold change values for reference RNA (HBRR vs. UHRR) and bulk lysate (bulk lysate TSA vs. bulk lysate DMSO) treatment groups across the HTTr MCF7 screen.

Reference Chemical Treatments:

Figure 4 Reproducibility and performance of chemical treatments. **(A)** The D Statistic for the reference chemical treatments [5]. Dashed vertical lines denote the threshold for flagging samples with poor D Statistic (three standard deviations below the mean). **(B)** K-means (K=3) clustered Heatmap of DESeq2 shrunken log2 fold change values for reference chemical treatment groups (TSA, SIRO, GEN; DMSO as control for all comparisons).



Reference Chemical Treatments are Enriched in Predicted Target Mechanism of Bioactivity

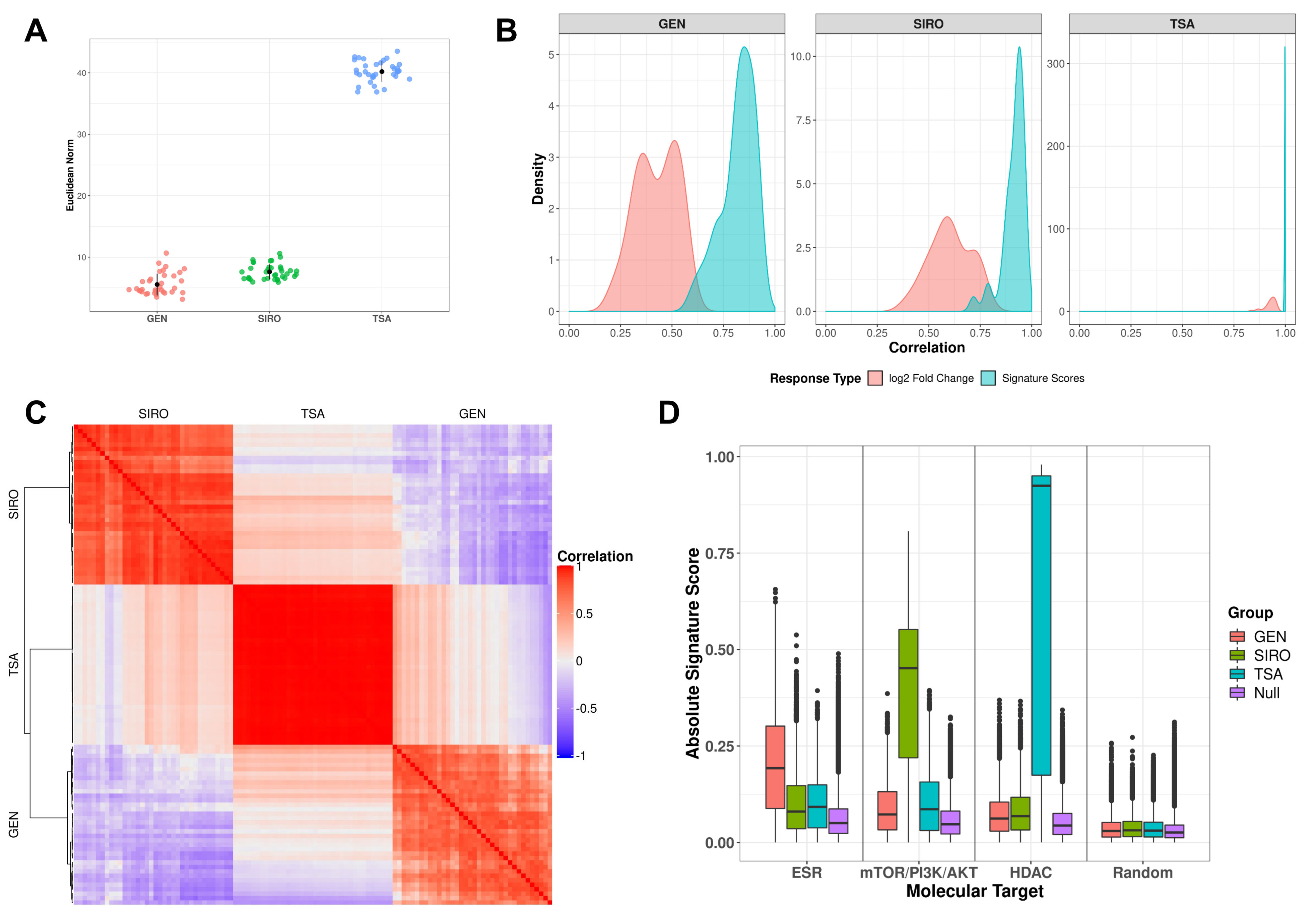


Figure 5. (A) Transcriptional signal strength of the three reference chemical treatments across treatment groups using the Euclidean norm of the top and bottom 100 probes based on DESeq2 log2 fold changes. **(B)** Correlation of DESeq2 log2 fold change (pink) or ssGSEA scores (blue) for each reference chemical. Signature scores were calculated using the ssGSEA algorithm with a signature set comprising signatures with the molecular targets associated with each reference chemical and a set of randomly derived signatures [4]. **(C)** Clustered correlation matrix of ssGSEA scores for the three reference chemicals across plate groups. **(D)** Distribution of ssGSEA signature scores for each reference chemical or empirically-derived null data. Signatures were separated based on known molecular target specific to each reference chemical (ESR for GEN, mTOR/PI3K/AKT for SIRO, and HDAC for TSA) or random signatures (Random).

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. Hanzelmann, et al. *GSVA: gene set variation analysis for microarray and RNA-Seq data.* BMC Bioinformatics 2013, 14(7):1-15
5. House et al. *A Pipeline for High-Throughput Concentration Response Modeling of Gene Expression for Toxicogenomics*. Froniers in Genetics 2017, 8(168):1-11

This poster does not reflect US EPA policy. D.E.H. was supported by appointment to the Research Participation Program of the U.S. Environmental Protection Agency, Office of Research and Development, administered by the Oak Ridge Institute for Science and Education through an interagency agreement between the U.S. Department of Energy and the U.S. EPA.