EPA
United States
Environmental Protection
Agency

Determining Transcriptional Points of Departure Using a Whole Transcriptome Screening Assay

Derik E. Haggard, PhD

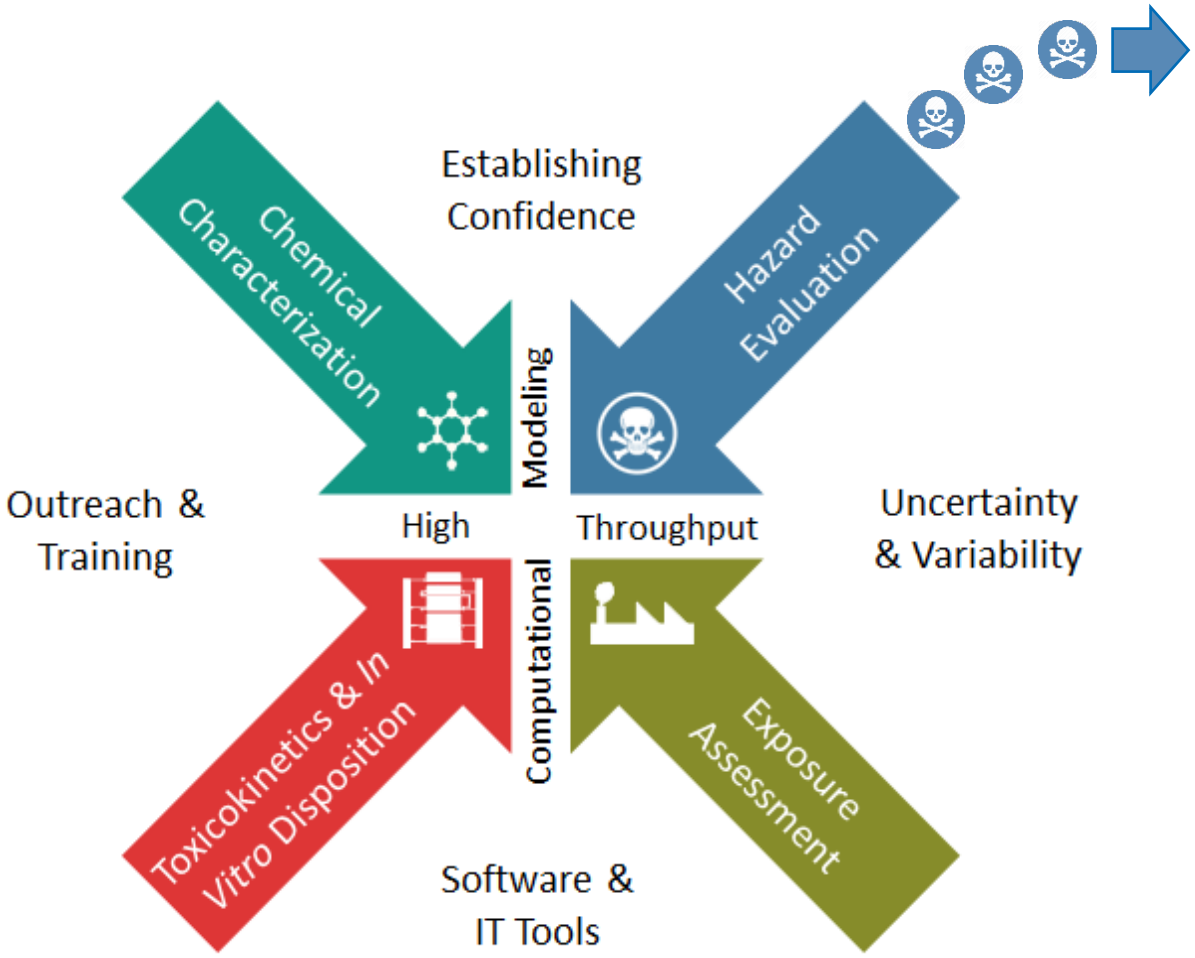
Scientific Analyst
USEPA Center for Computational Toxicology and Exposure (CCTE)

Disclaimer

The views expressed in this presentation are those of the author(s) and do not necessarily represent the views or policies of the U.S. Environmental Protection Agency, nor does mention of trade names or products represent endorsement for use.

Computational Toxicology Research Areas

CCTE research programs focus on developing the **tools, approaches and data** needed to accelerate the pace of chemical risk assessment and foster incorporation of non-traditional toxicity testing data into regulatory decision-making processes.



The NexGen Blueprint of CompTox as USEPA
Tox. Sci. 2019; 169(2):317-322

- **ToxCast:** Use of targeted high-throughput screening (HTS) assays to expose living cells or isolated proteins to chemicals and assess bioactivity and potential toxic effects.

	# of assays	# of chemicals	Types of chemicals
Phase 1 ***(2007 – 2009)***	500	300	Mostly pesticides
Phase 2 ***(2009 – 2013)***	700	2,000	Industrial, consumer product, food use, "green" chemicals

- Mostly targeted assays (*chemical X → target Y*)
- Incomplete coverage of biological space.

- **New Approach for Hazard Evaluation:** Employ broad-based (i.e. non-targeted) **profiling assays** that cast the broadest net possible for capturing the potential molecular and phenotypic responses of human cells in response to chemical exposures.

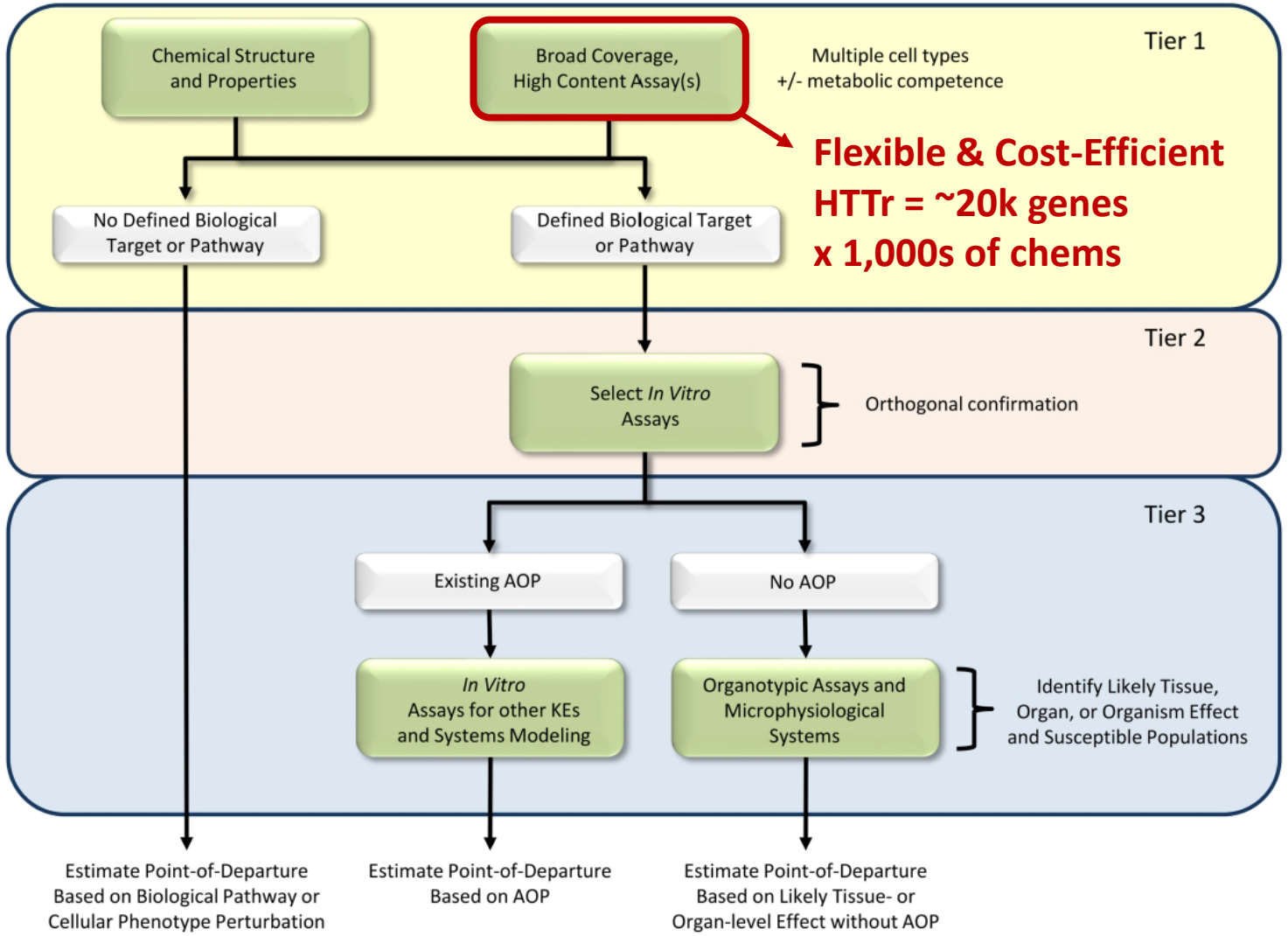
Tiered Chemical Safety Testing Strategy

Tier 1 Primary Goals:

- Prioritize chemicals by bioactivity & potency
- Predict biological targets for chemicals

HTTr Key Challenges:

- Curve-fitting on count-based data
- Summarization at pathway/chemical level



Thomas, et al. Toxicol Sci 2019

Automated *in vitro* Chemical Screening

Cryopreserved Cell Stocks

Cell Expansion

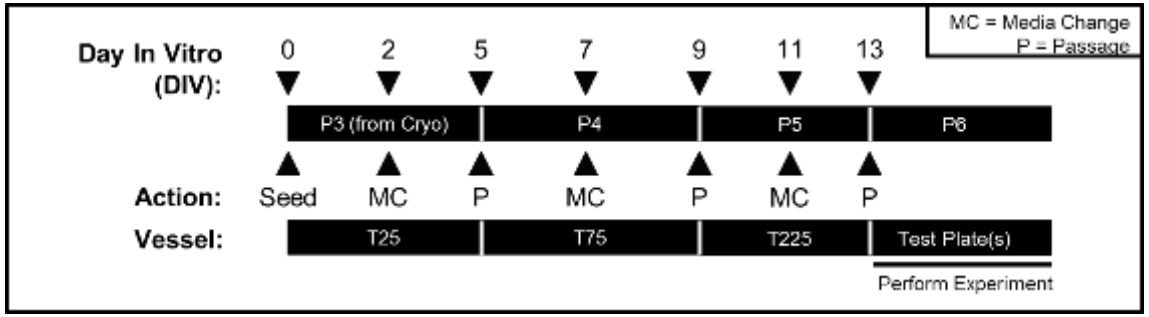
Cell Plating

BioTek
MultiFlo ™ FX

Dispensing Test Chemicals

LabCyte Echo® 550
Liquid Handler

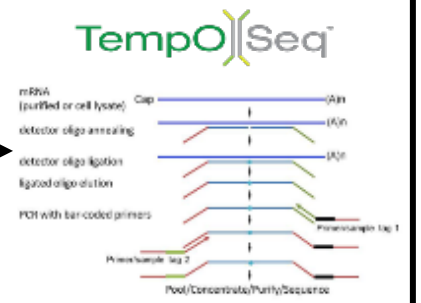
Standardized Expansion Protocol



Track 1: Targeted RNA-Seq

Generating Cell Lysates

TempO-Seq WT



Track 2: Apoptosis / Cell Viability

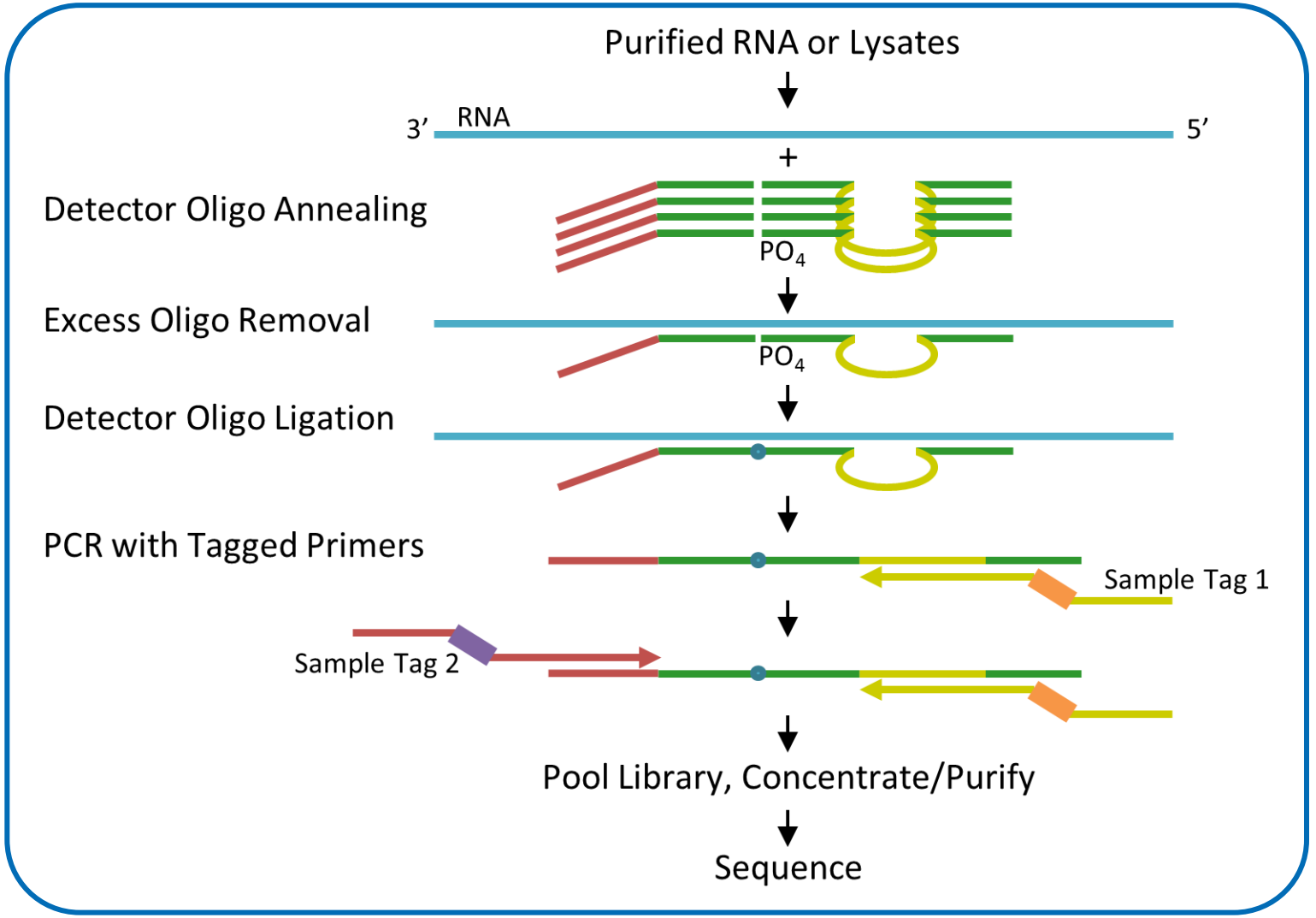
Reagent Dispensing

High Content Imaging

Perkin Elmer
Opera Phenix™
High Content Screening System

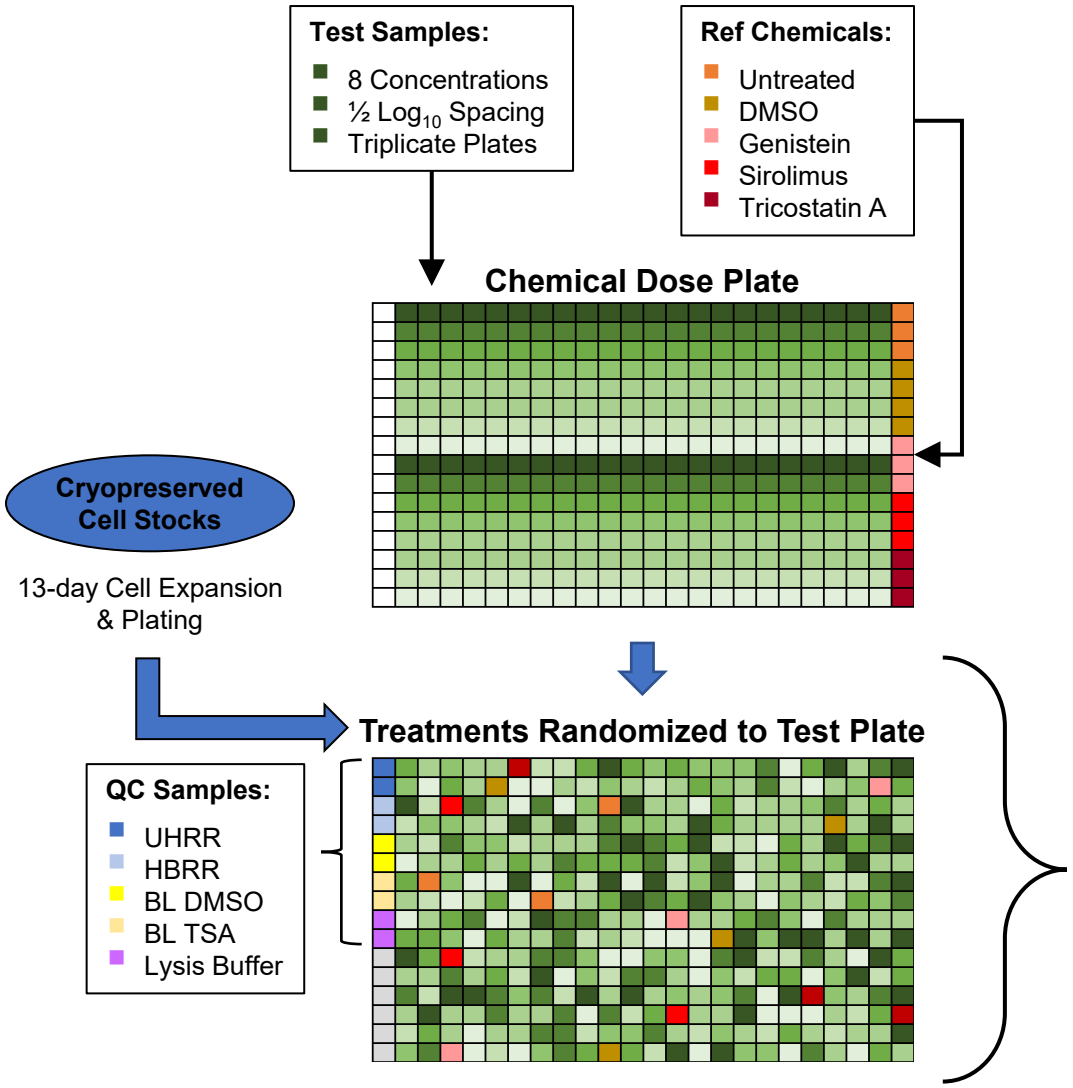
High-Throughput Transcriptomics (HTTr) Assay

- Targeted RNA-seq enables high-throughput profiling of **cell lysates** or purified RNA
- Probe set for whole human transcriptome targets ~21,000 human genes
- Captures majority of signal with much lower sequencing depth (~3M reads with attenuation)
- Barcoding and pooling allows multiplexing of hundreds of samples



Yeakley, et al. PLoS ONE 2017

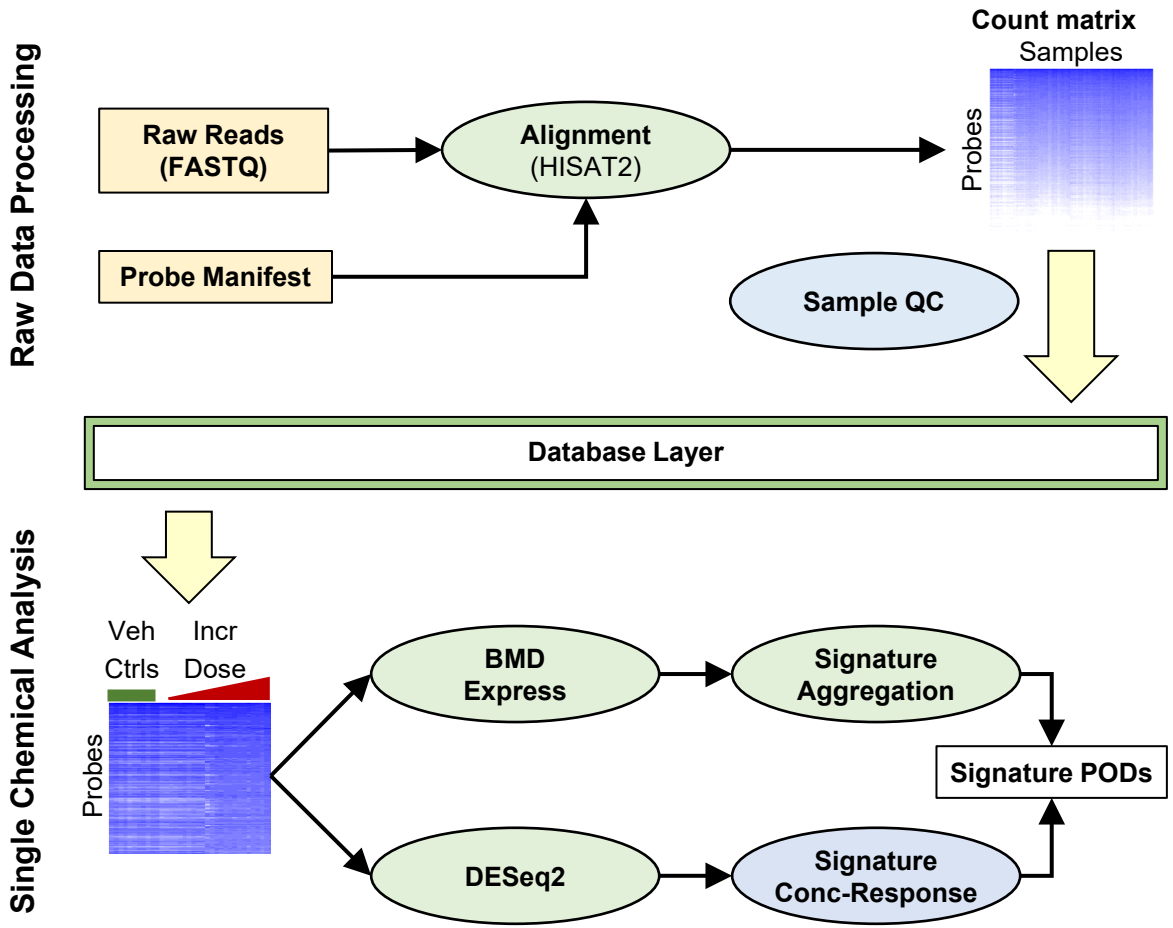
HTTr Study Design



- High-throughput *in vitro* screens performed on 384 well plates
- Standardized dilution series for every test sample
- Multiple QC and reference chemicals included on every plate to track assay performance
- Triplicate Test Plates:
  - Randomized independently
  - Separate cell culture batches

Harrill, et al. Toxicol Sci 2021

HTTr Bioinformatics Pipeline



- Rapid processing for large screens
- Many data steps performed independently for each test chemical:
  - Removal of low signal probes
  - Normalization
  - DESeq2 analysis
- Exploring multiple analysis strategies for curve-fitting and signature & chemical-level summarization

Harrill et al., Toxicol Sci 2021

HTTr MCF-7 Screen: Experimental Design

Parameter	Multiplier	Notes
Cell Type(s)	1	MCF-7
Culture Condition	1	DMEM + 10% HI-FBS [a]
Chemicals	2,112[b]	ToxCast ph1, ph2 Nominated chemicals from e1k / ph3
Time Points:	1	6 hours
Assay Formats:	2	TempO-Seq HCI Cell Viability & Apoptosis
Concentrations:	8	3.5 $\log_{10}$ units; ~half-$\log_{10}$ spacing
Biological Replicates:	3	--

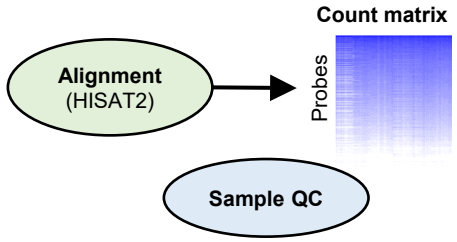
Reference **Samples** and Reference **Chemicals**:

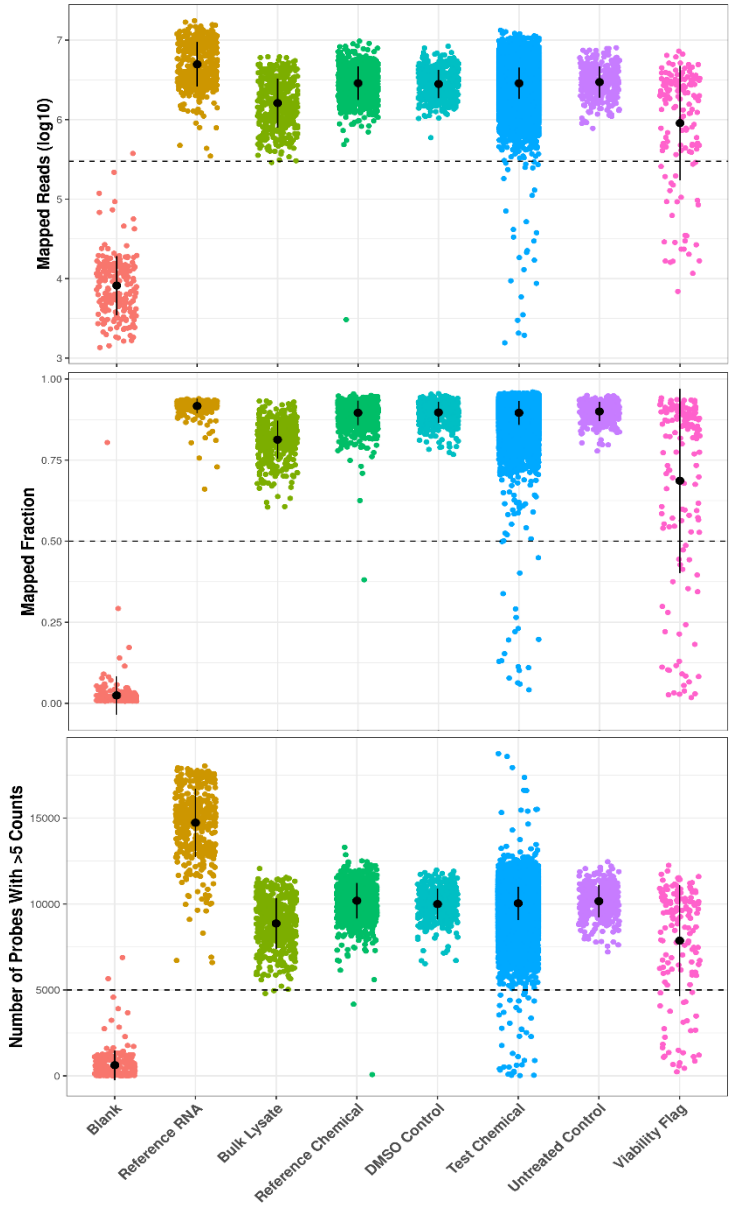
- **Reference RNA** – UHRR and HBRR
- **Bulk Lysate Preparations** – DMSO vehicle control and Trichostatin A
- **Reference Chemicals** – Genistein, Sirolimus, and Trichostatin A

[a] MCF7 cells cultured in DMEM + 10% HI-FBS was selected

[b] Due to reagent error, one experimental block was removed leaving **1577 unique chemicals** across 37 triplicate test plates

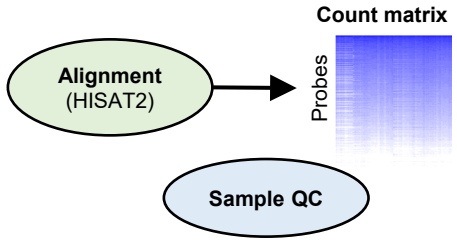
MCF-7 Screen Sample Quality



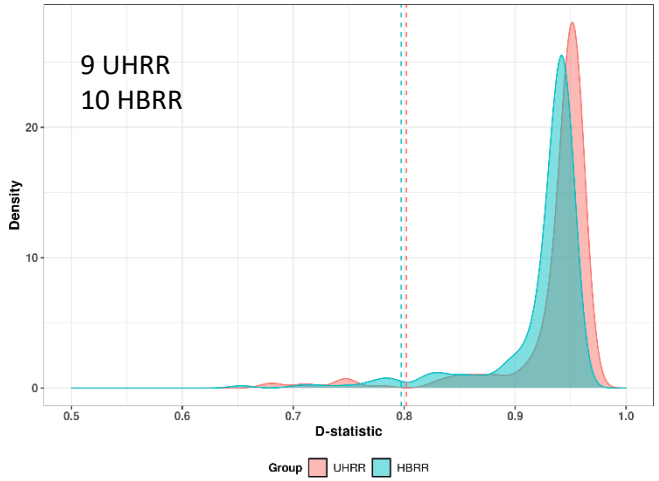


- Five alignment-based quality control metrics are estimated from raw count data
- Sample below thresholds are removed from analysis
- The parallel cell viability/apoptosis assay is used to remove samples due to cytotoxicity
- A total of **~98%** of all samples passed initial QC

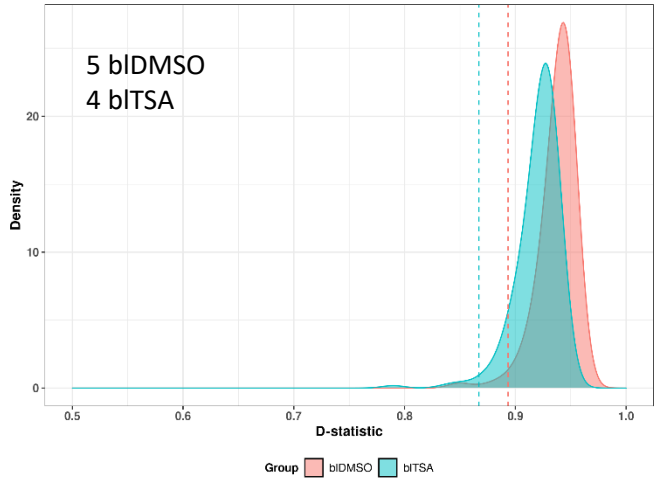
MCF-7 Reference Count Reproducibility



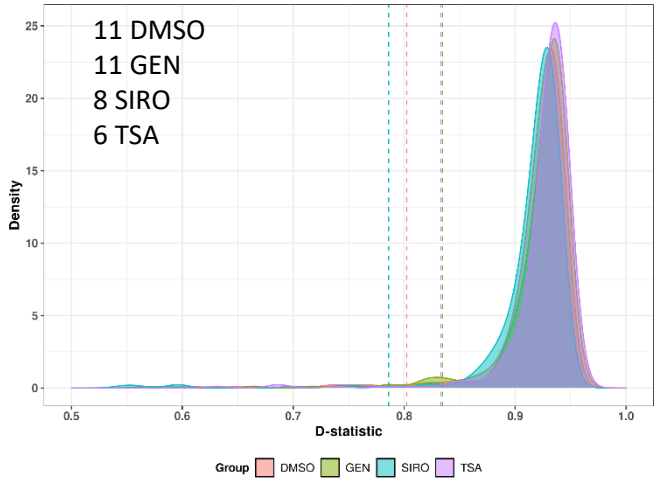
Reference RNA



Bulk Lysate



Reference Chemicals



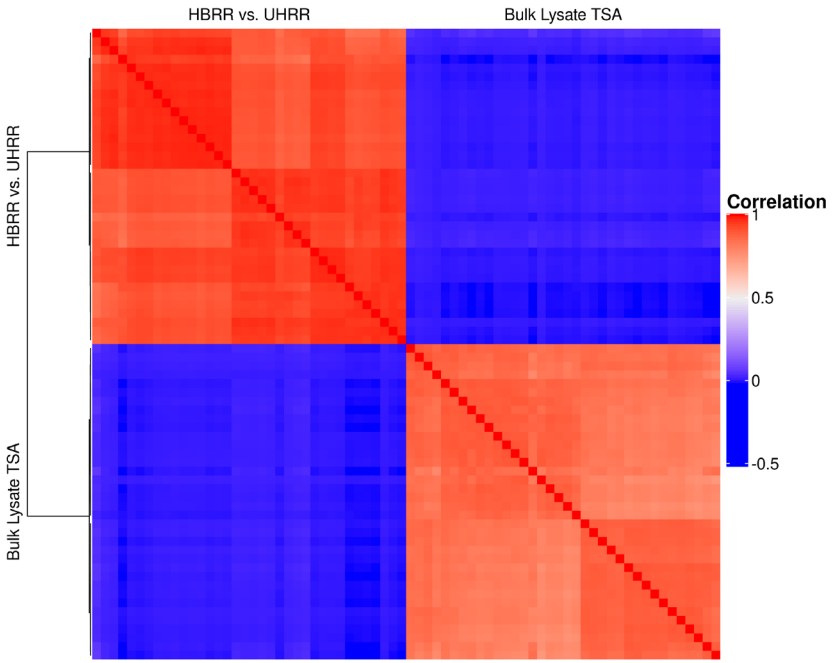
D-Statistic for Outlier Identification (*House et al., Front Genet 2017*)

- Counts were converted to $\log_2$ counts-per-million (CPM)
- D-statistic calculated as the average correlation of a reference sample or chemical against all other replicate wells of the same sample type

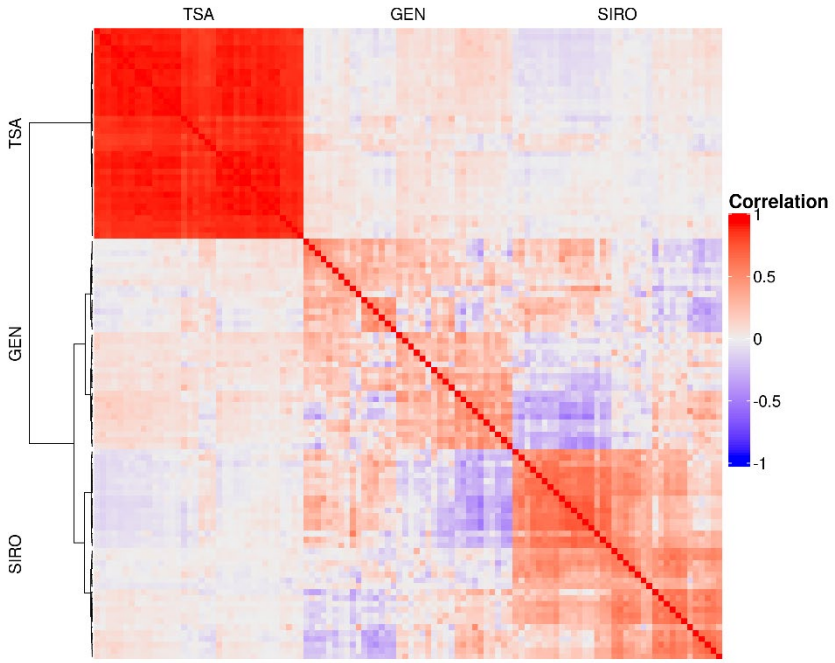
- Computed distribution of the D-statistic for each reference sample and chemical type
- Outliers defined as 3 SDs below median D-statistic
- Count-level quality metrics alongside the added D-statistic approach demonstrated **96.8%** of all reference samples and chemicals passed quality control

HTTr Fold Change Estimation

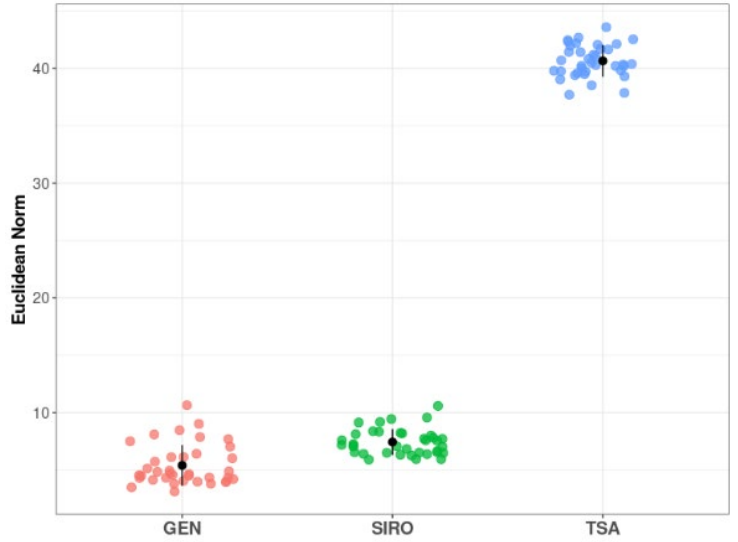
Reference Samples



Reference Chemicals



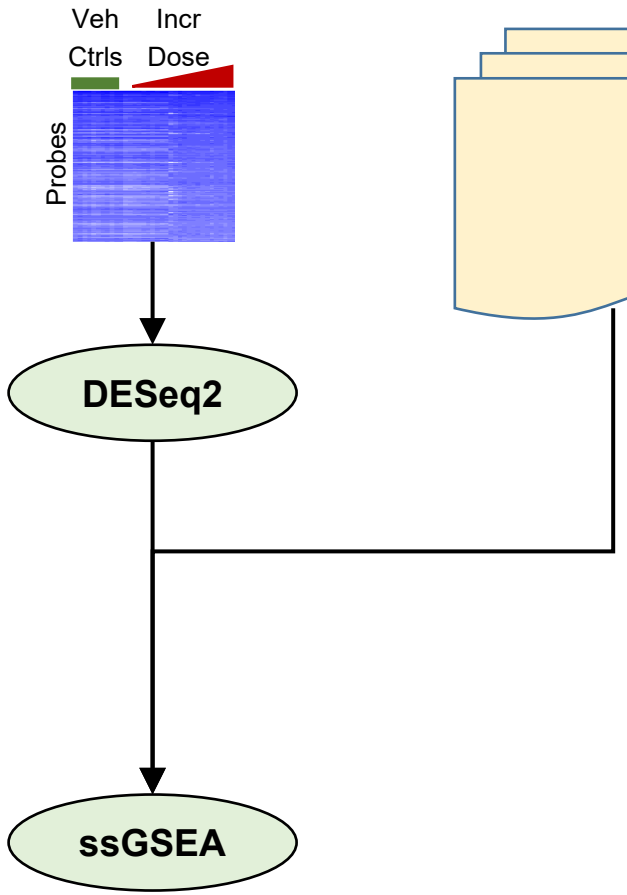
Transcriptional Signal Strength



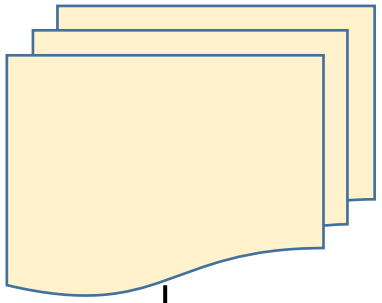
- Estimated moderated fold changes using DESeq2 with default parameters and including plate
- Determined correlation of DESeq2 moderated $\log_2$ FC values for QC samples (left) and reference chemicals (right)
- Compared correlation in $\log_2$ FC to transcriptional signal strength of reference chemicals

HTTr Signature Scoring

Count data per chemical



Catalog of signatures with toxicological relevance, annotated for known molecular targets

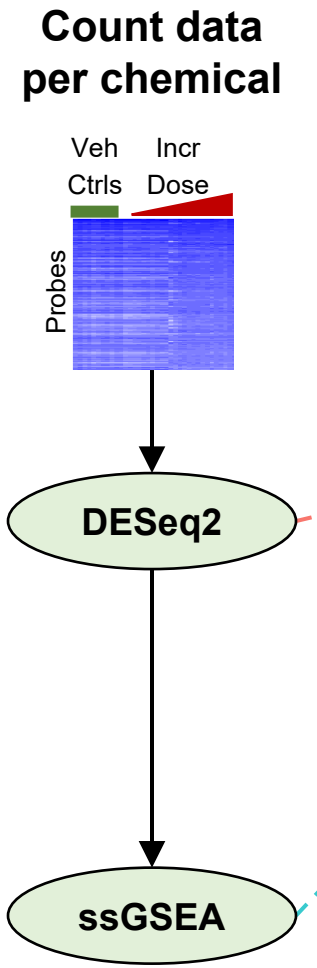


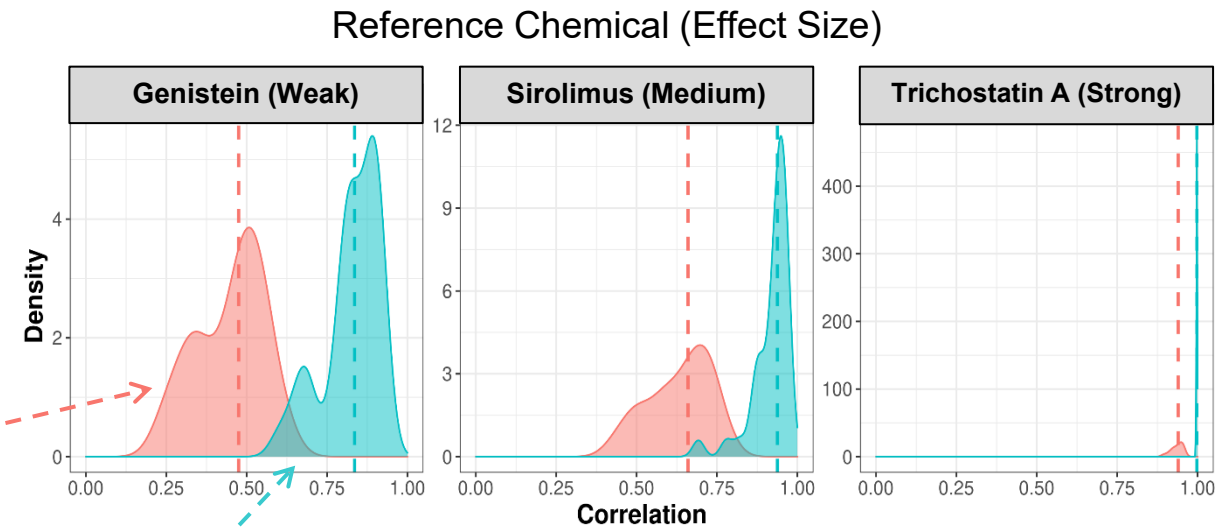
- **Bioplanet** *(Huang, et al. Front Pharmacol 2019)*
- **CMap** *(Subramanian, et al. Cell 2017)*
- **DisGeNET** *(Pinero, et al. Database 2015)*
- **MSigDB** *(Liberzon, et al. Cell Syst 2015)*

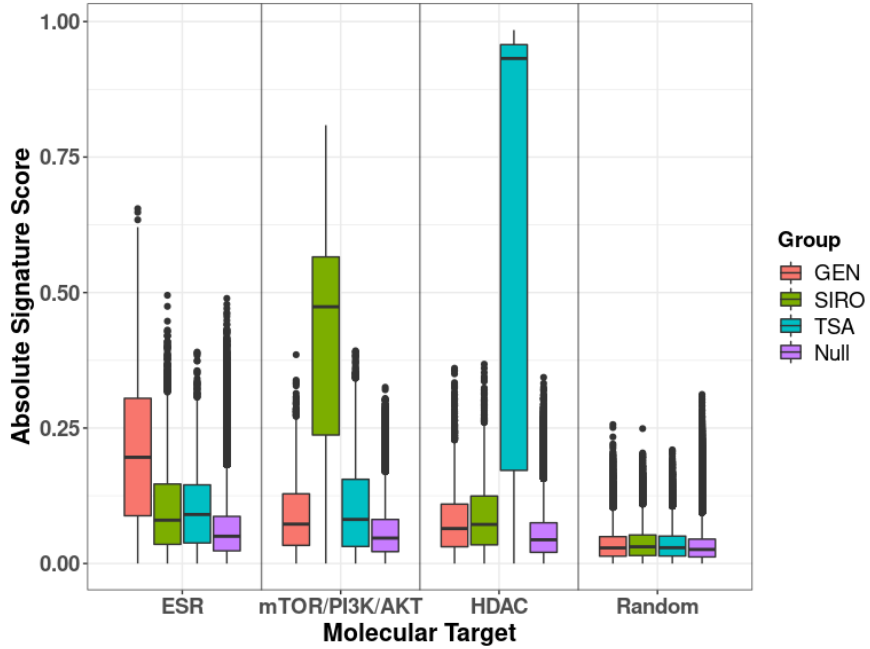
Single-Sample Gene Set Enrichment Analysis (ssGSEA) ***(Barbie et al., Nature 2009)***

- Score coordinated responses at each concentration
- Use moderated log2 FC values from DESeq2 as input (no thresholds)
- Null distributions constructed by resampling log2 FC values from whole screen
- Alternate scoring function:
  mean(gene set log2FC) – mean(background log2FC)

MCF-7 Screen Reference Chemical Signature Scores



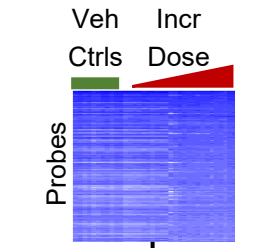




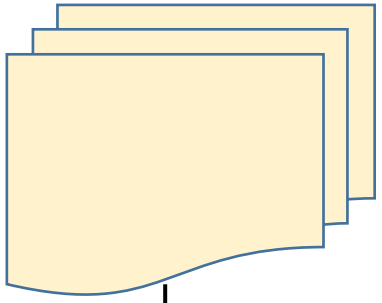
- Differential expression analysis of 3 reference chemicals replicated 37 times (MCF-7 large screen)
- Computed distribution of correlations between each replicate analysis
- Categorized signature scores based on relevant molecular target for each reference chemical

HTTr Signature Scoring

Count data per chemical



Catalog of signatures with toxicological relevance, annotated for known molecular targets

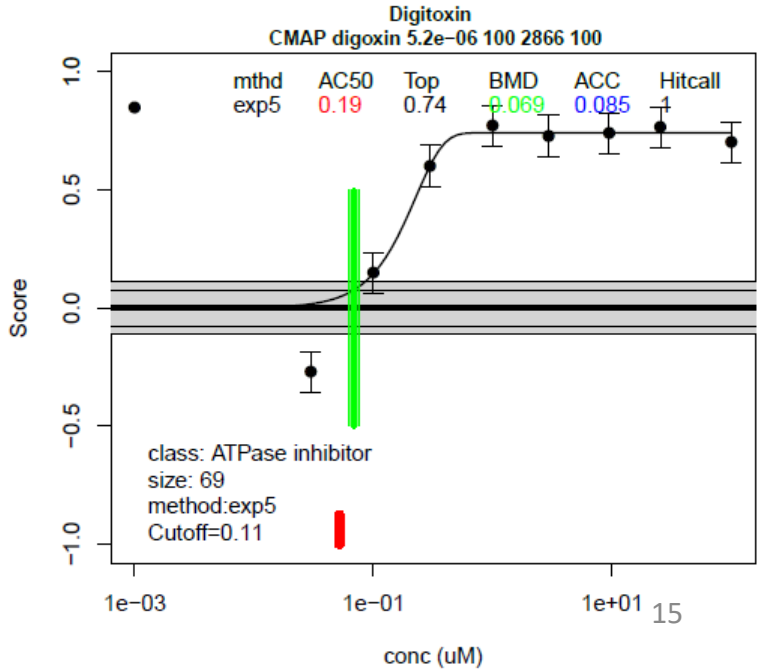


- **Bioplanet** *(Huang et al., Front Pharmacol 2019)*
- **CMap** *(Subramanian et al., Cell 2017)*
- **DisGeNET** *(Pinero et al., Database 2015)*
- **MSigDB** *(Liberzon et al., Cell Syst 2015)*

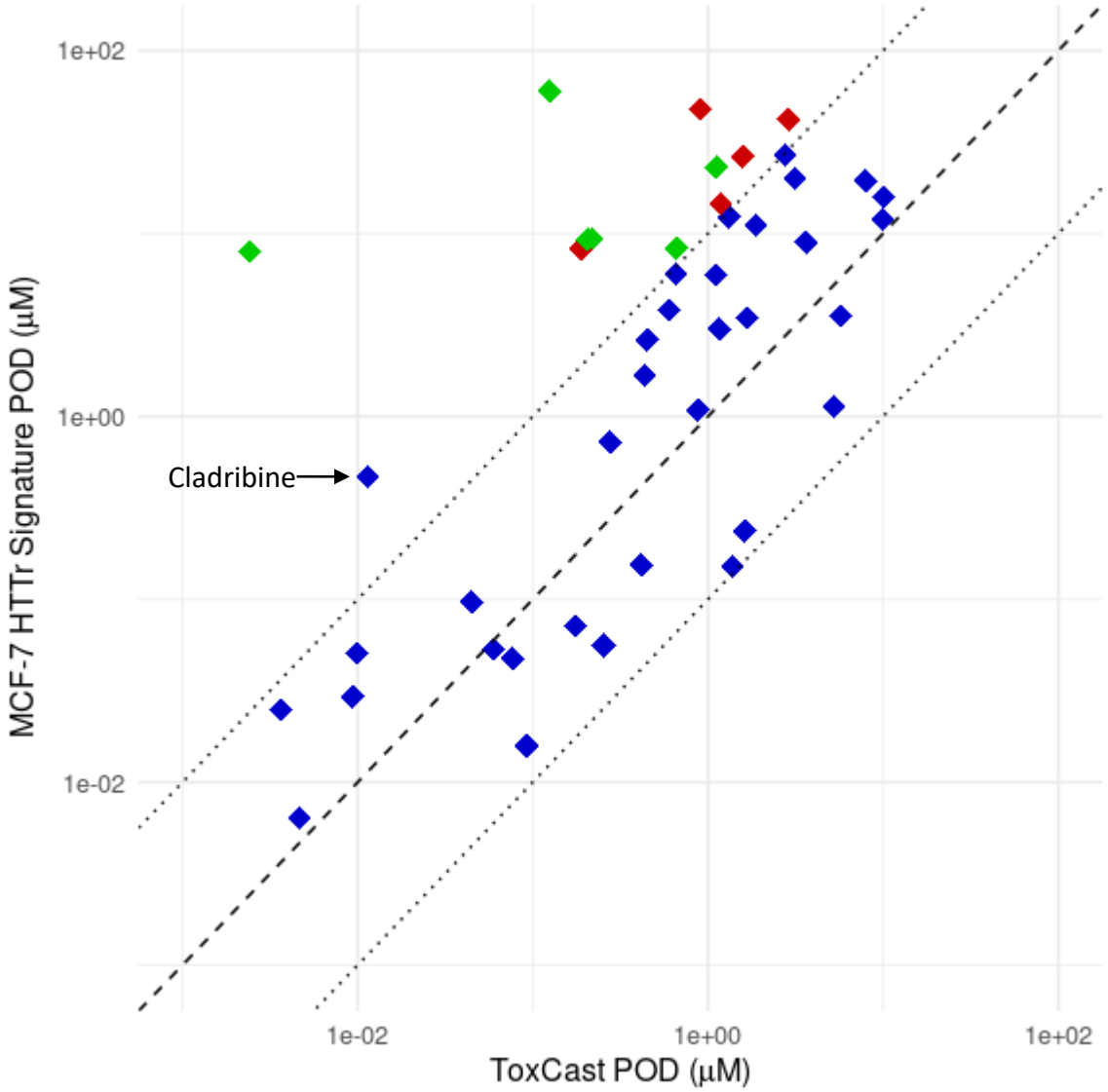
DESeq2

ssGSEA

Concentration-Response Curve Fitting (*tcplFit2*)



MCF-7 Pilot Point of Departure Analysis



- Pilot study of 44 well-characterized chemicals ***(Harrill et al., Toxicol Sci, In Press)***
- Compared HTTr-derived PODs from MCF-7 cells to previous ToxCast HTS assay results (***Paul-Friedman et al,. Toxicol Sci 2020)***
- Signature-based POD are highly concordant with ToxCast results for the majority of test chemicals in pilot study
  - 6 chemicals with targets that have low/absent expression in MCF-7 cells
  - 5 chemicals show off-target hit as most potent assay in ToxCast
  - Cladribine is a non-specific DNA synthesis inhibitor

Summary and Future Directions

- CCTE has developed reliable and cost-efficient workflow for generating HTTr data from thousands of chemicals across multiple cell lines
- Correlation and reproducibility of reference samples and chemicals in a large MCF-7 screen demonstrate the experimental design and TempO-Seq HTTr platform to be robust
- Functional analysis of reference chemicals shows the benefit of signature-level analysis compared to probe-level and fold-change estimates, with signature scores reflecting the biological targets of the reference chemicals
- Preliminary/pilot analysis demonstrates that overall results are concordant with previous assays (ToxCast/HTS) and known chemical targets

- Future research efforts focus on:
  - Data generation in complementary cell models (e.g. HepaRG and U2OS screens)
  - Validation by orthogonal assays
  - Methods to summarize signature-level/overall PODs from high-dimensional data
  - Predictive models of MIEs/pathways relevant to toxicity
  - Coupling HTTr-derived PODs with HTTK/IVIVE work to predict *in vivo* safety levels

Acknowledgments

Center for Computational Toxicology & Exposure

- Richard Judson
- Logan Everett
- Imran Shah
- Beena Vallanat
- Woody Setzer
- Thomas Sheffield
- Joseph Bundy
- Bryant Chambers
- Rusty Thomas

Questions?

Orion and surrounding nebulae. B Everett Jordan Dam, April 3rd, 2021.

HTTr QC Metrics: Overview

Abbreviation	Description	Threshold	Additional Information
FrVC	Fraction of viable cells (**assay varies by cell type/study**)	Reject < 50%	Highly cytotoxic conditions no longer represent molecular initiating event
NMR	# of uniquely mapped reads	Reject < 300,000	Threshold =10% of target depth
FMR	Fraction of uniquely mapped reads	Reject < 50%	Majority of reads must align to a single probe sequence
$Ncov_5$	The number of probes with at least 5 uniquely mapped reads	Reject < 5,000	Based on Tukey's Outer Fence (3*IQR) of all viable samples cultured on each plate (test samples, vehicle controls, and reference chemical treatments)
$Nsig_{80}$	# of probes capturing the top 80% of signal in a sample	Reject < 1,000	
GiC	Gini coefficient computed on count vector for each sample	Reject > 0.95	

Adapted from MCF7 Pilot Manuscript (Harrill, et al., 2021) – larger screens also include QC flags for errors on LabCyte Echo indicating problems with chemical dispensing; these data streams not well standardized or fully captured