Transcriptomics: An Emerging Tool for Assessing Chemical Safety

Logan J. Everett, Ph.D.

Bioinformatics Scientist

Center for Computational Toxicology & Exposure

Office of Research and Development, U.S. EPA

Research Triangle Park, North Carolina

Disclaimer

The views expressed in this presentation are those of the authors and do not necessarily reflect the views or policies of the U.S. Environmental Protection Agency. Company or product names do not constitute endorsement by US EPA.

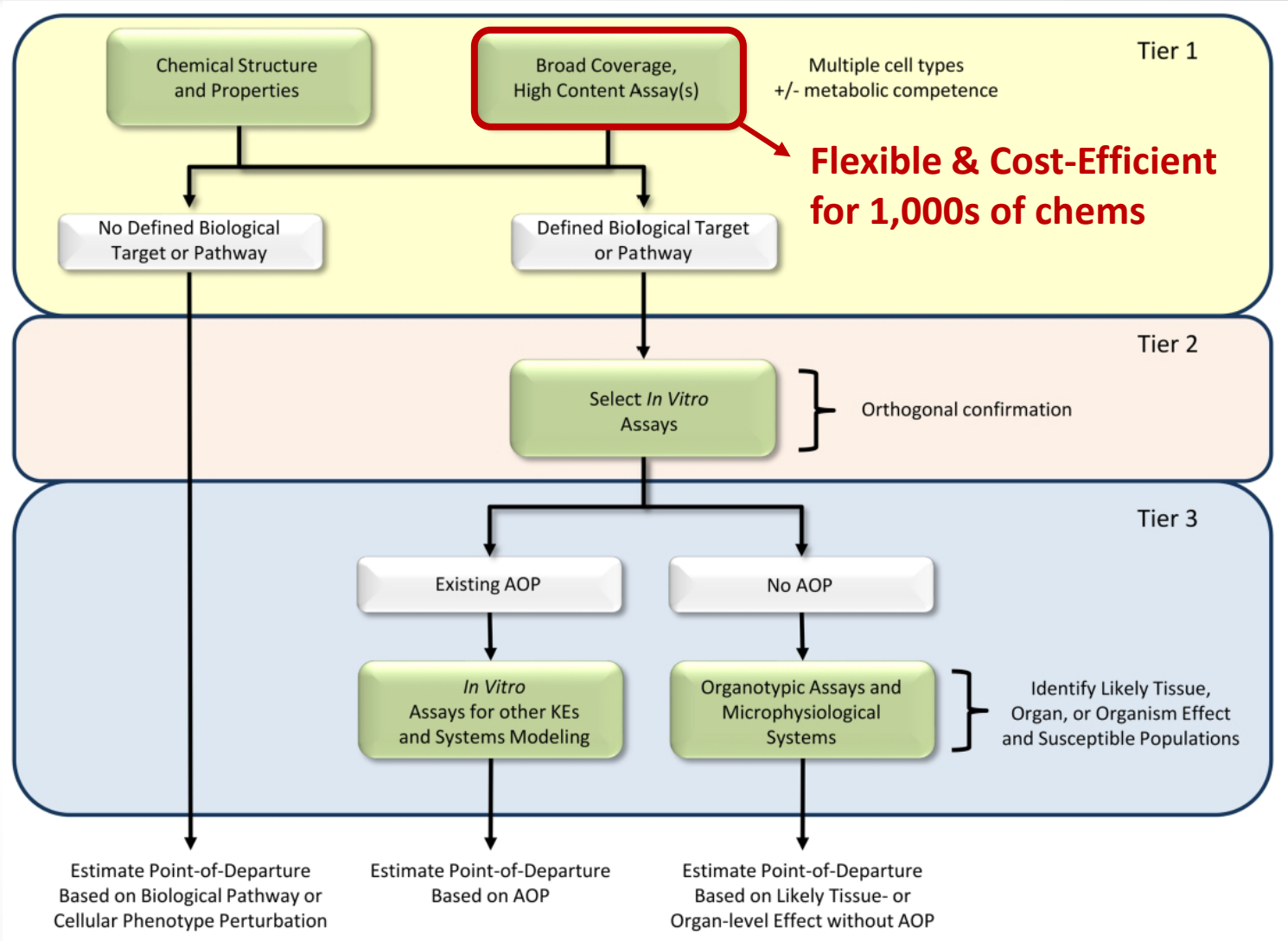
Tiered Chemical Safety Testing Strategy

Tier 1 Primary Goals:

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

Key Challenges:

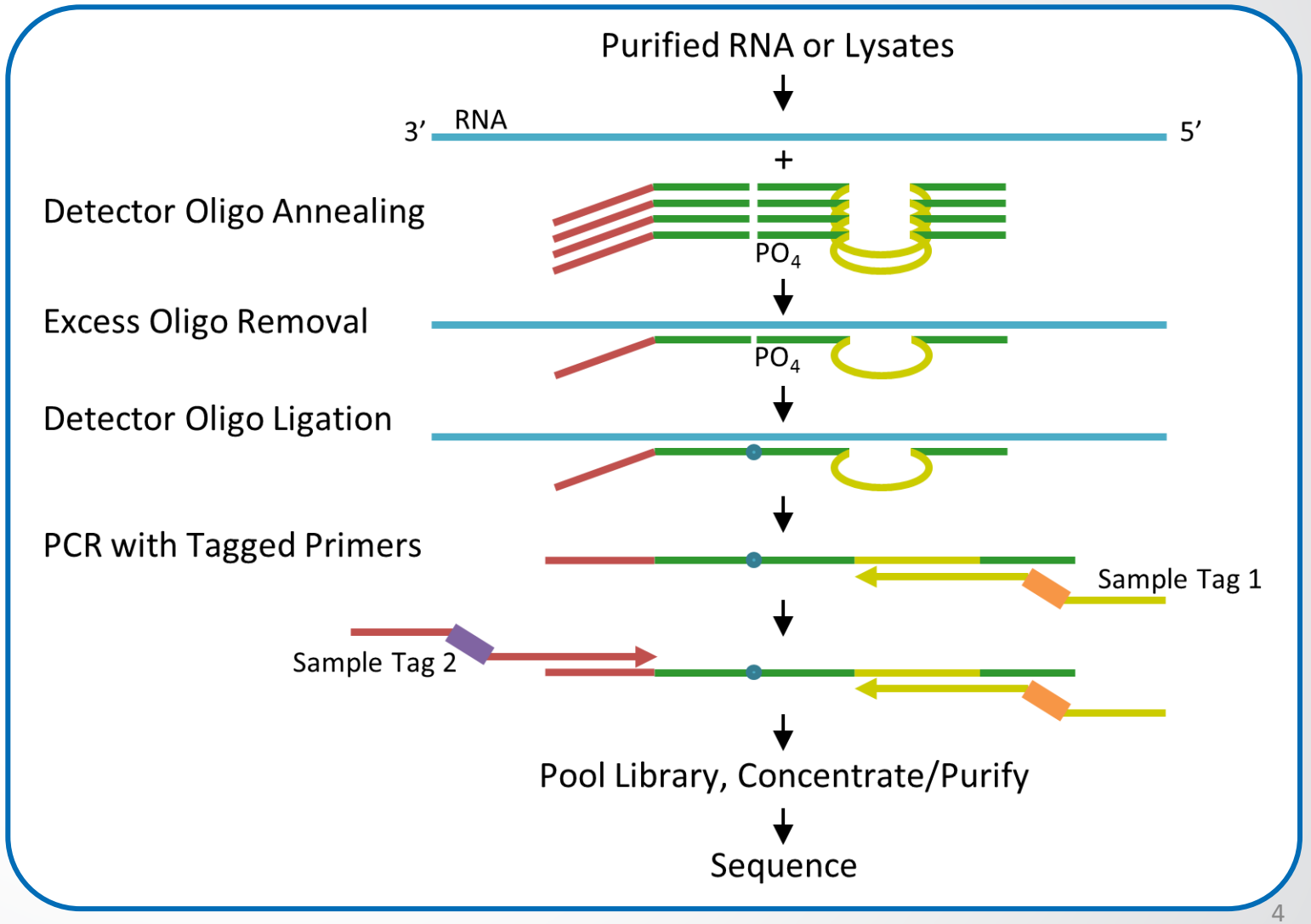
- Curve-fitting on count-based data
- Summarization at pathway/chemical level



Thomas, et al. Toxicol Sci 2019

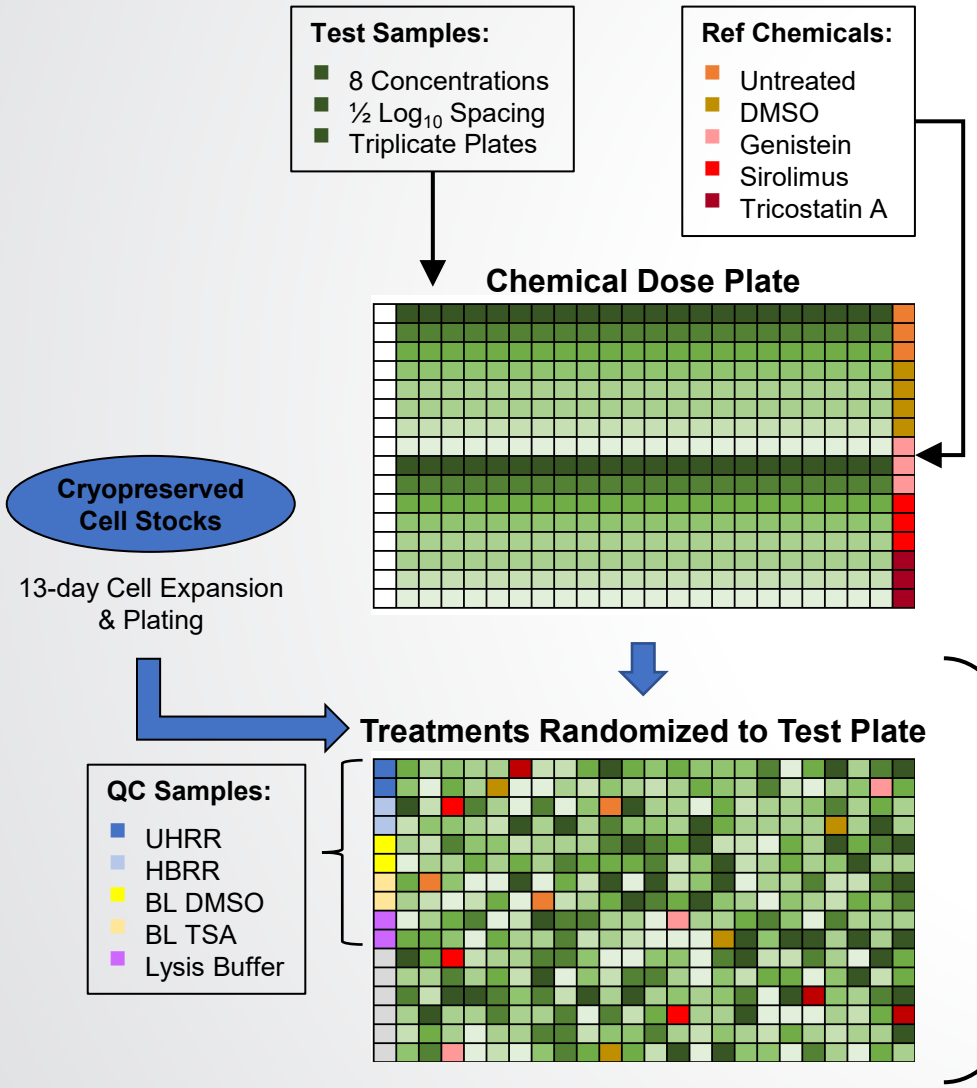
High-Throughput Transcriptomics Assay

- 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

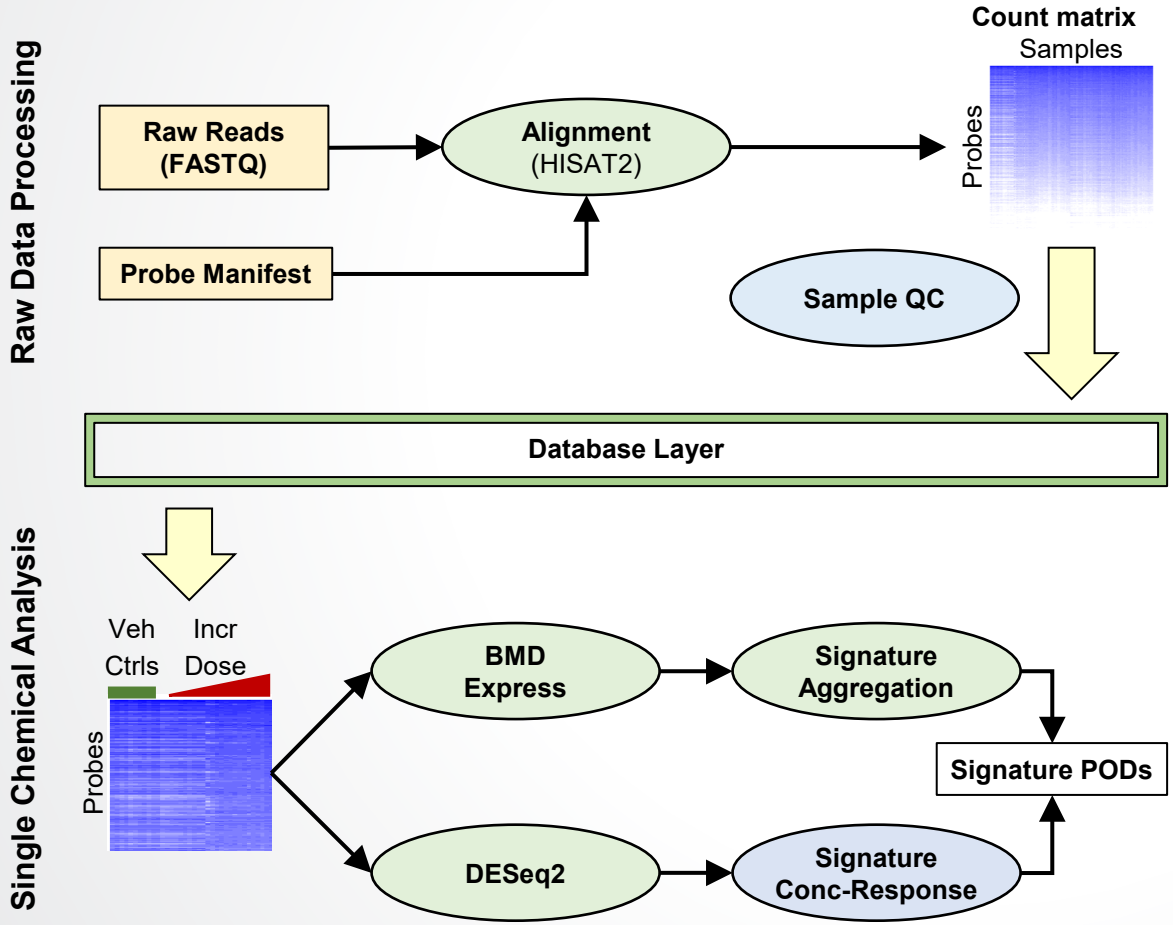
HTTr Study Design



- 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

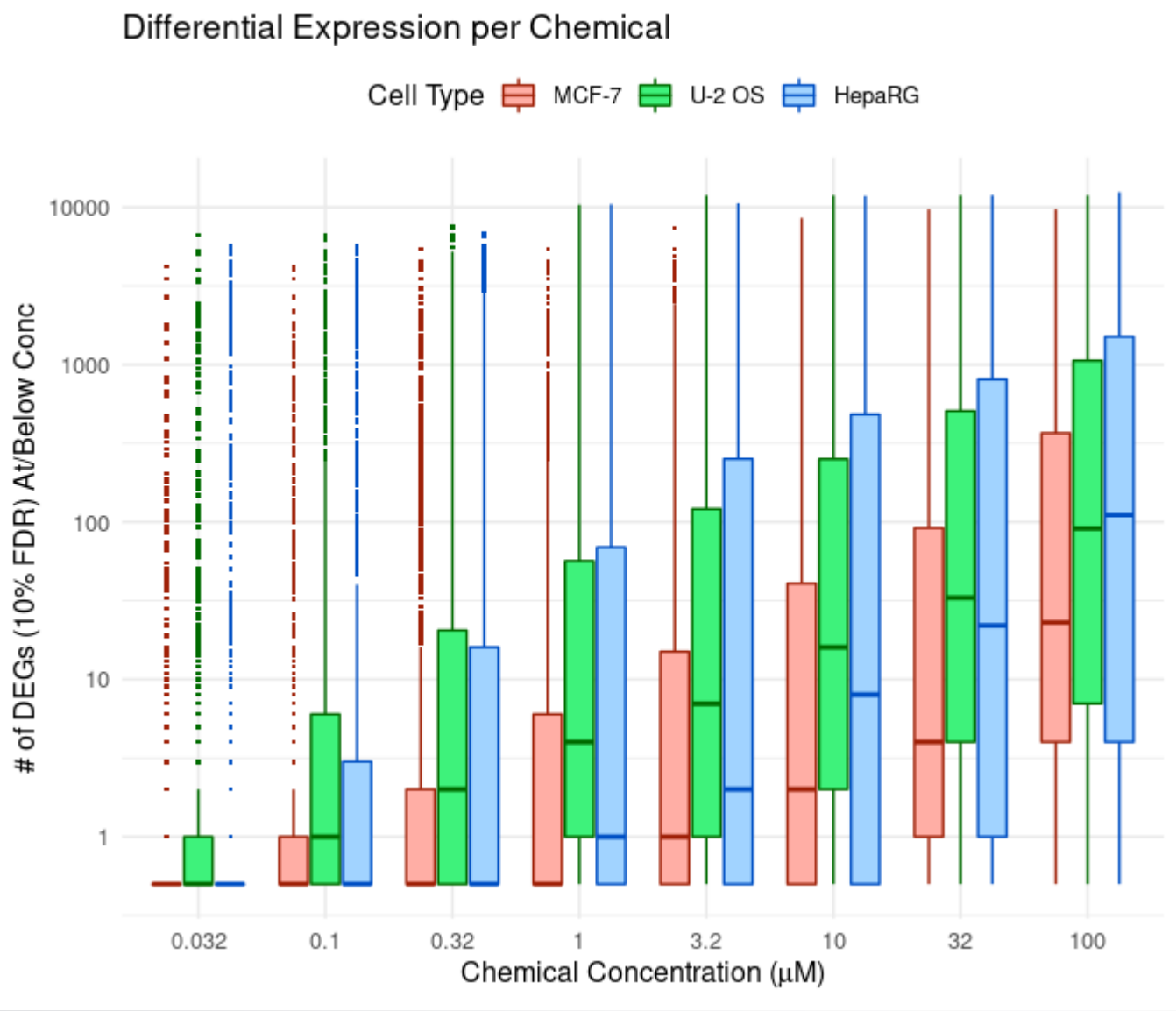


- Rapid processing for large screens
- Many data steps performed independently for each test chemical:
  - Removal of low signal probes
  - Normalization
  - DESeq2 analysis
- Exploring multiple analysis strategies for curve-fitting and signature & chemical-level summarization

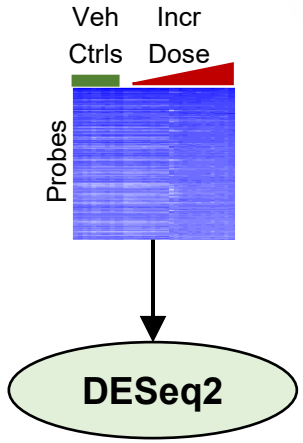
Harrill, et al. Toxicol Sci 2021

https://github.com/USEPA/httrpl_pilot

Global View of Bioactivity



Count data for single chemical (vehicle controls + 8 concs x 3 reps)



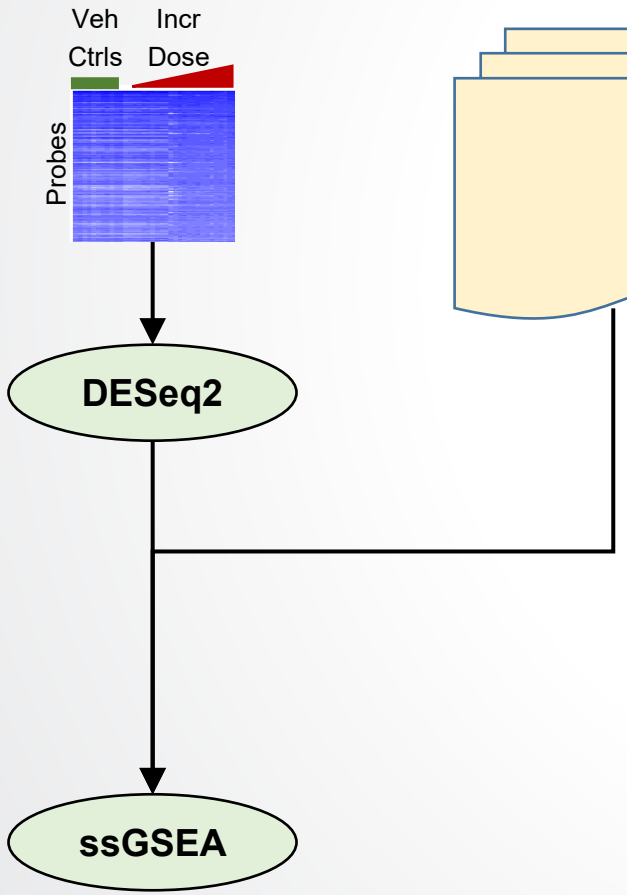
- Statistical model tailored to *-seq data
- Remove plate-level effects
- Smooths noise across depth & expression levels

(Love, et al. Genome Biol 2014)

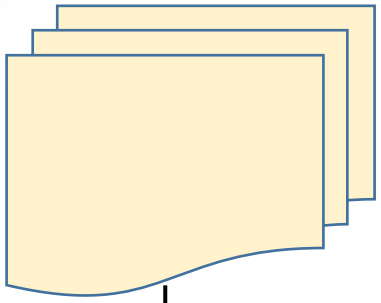
- Each boxplot shows distribution of DEG count per chemical
- **Primarily interested in transcriptional changes that:**
  - **Are coordinated across known pathways/gene sets**
  - **Fit standard curve-models across all concentrations**

Signature Scoring

Count data per chemical



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

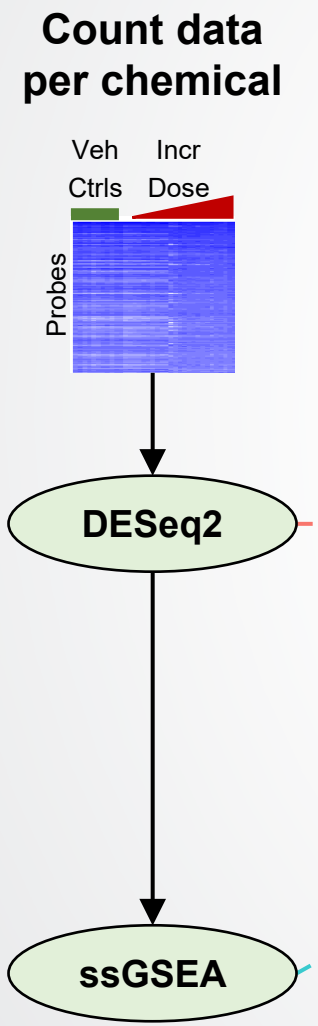


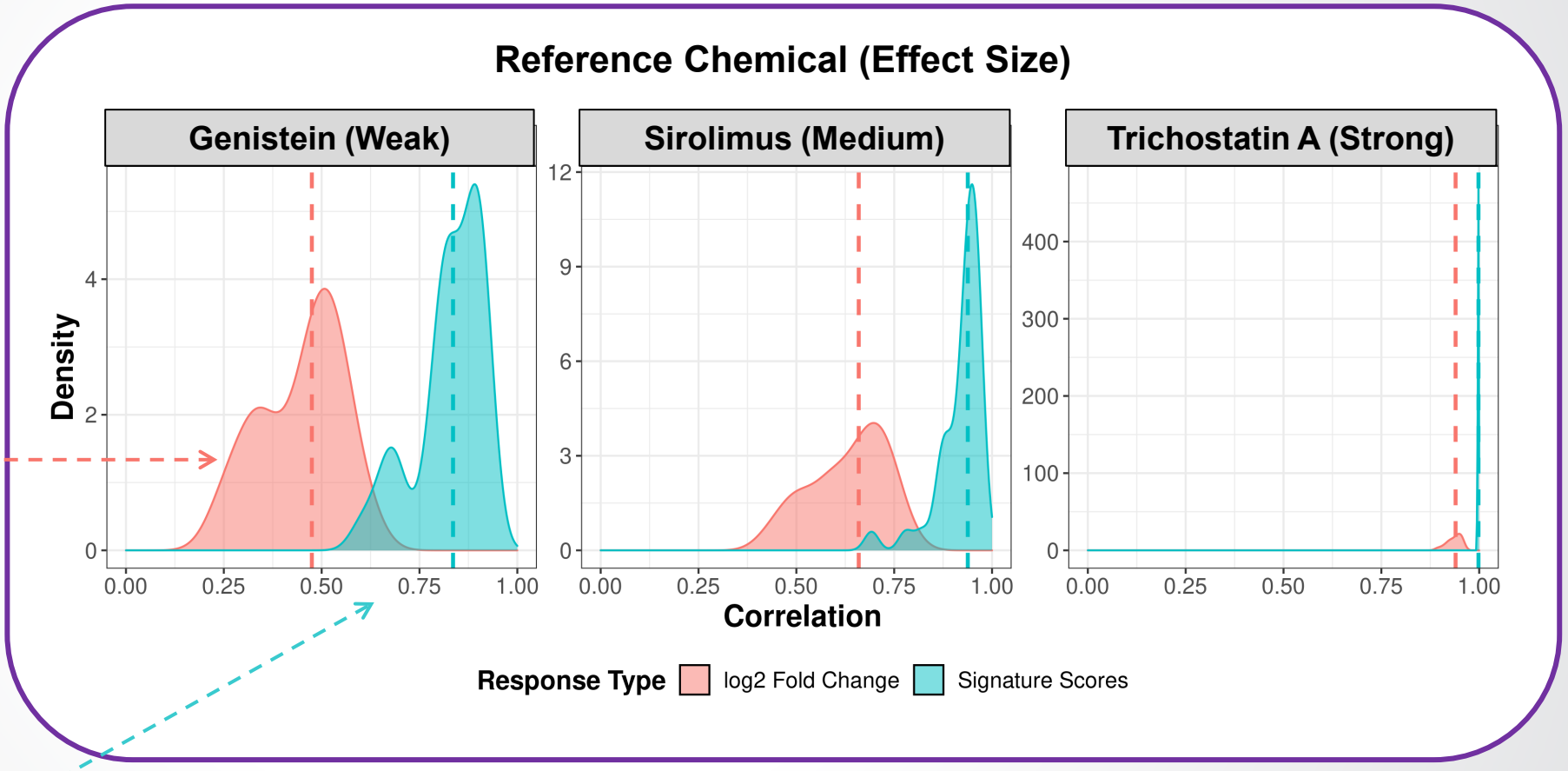
- **Bioplanet** *(Huang, et al. Front Pharmacol 2019)*
- **CMap** *(Subramanian, et al. Cell 2017)*
- **DisGeNET** *(Pinero, et al. Database 2015)*
- **MSigDB** *(Liberzon, et al. Cell Syst 2015)*

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)

Signature Scoring

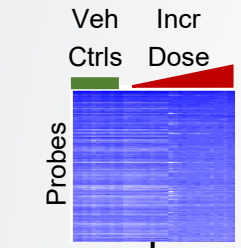




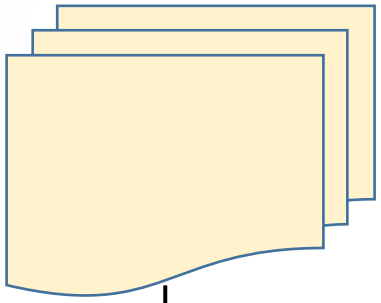
- Differential expression analysis of 3 reference chemicals replicated 37 times (MCF-7 large screen)
- Computed distribution of correlations between each replicate analysis

Signature Scoring

Count data per chemical



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

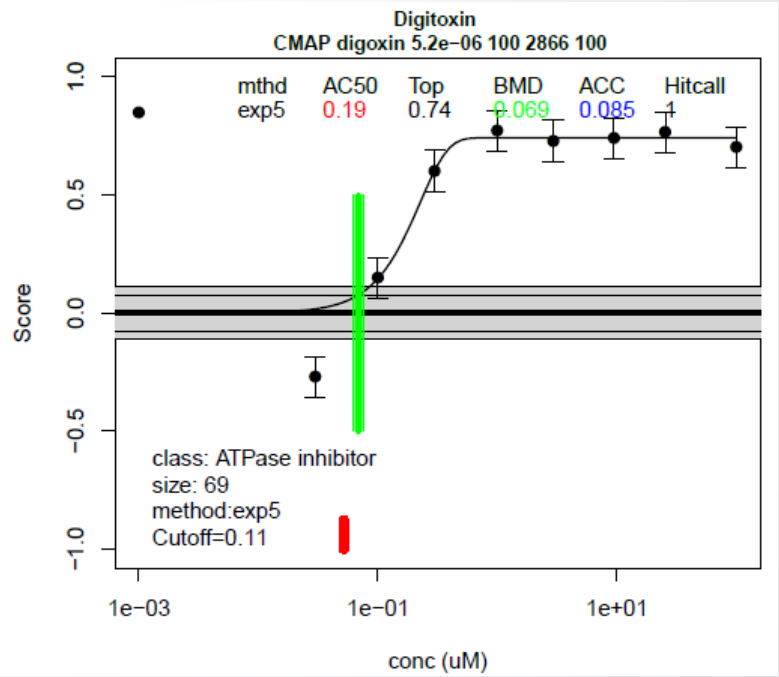


- **Bioplanet** *(Huang, et al. Front Pharmacol 2019)*
- **CMap** *(Subramanian, et al. Cell 2017)*
- **DisGeNET** *(Pinero, et al. Database 2015)*
- **MSigDB** *(Liberzon, et al. Cell Syst 2015)*

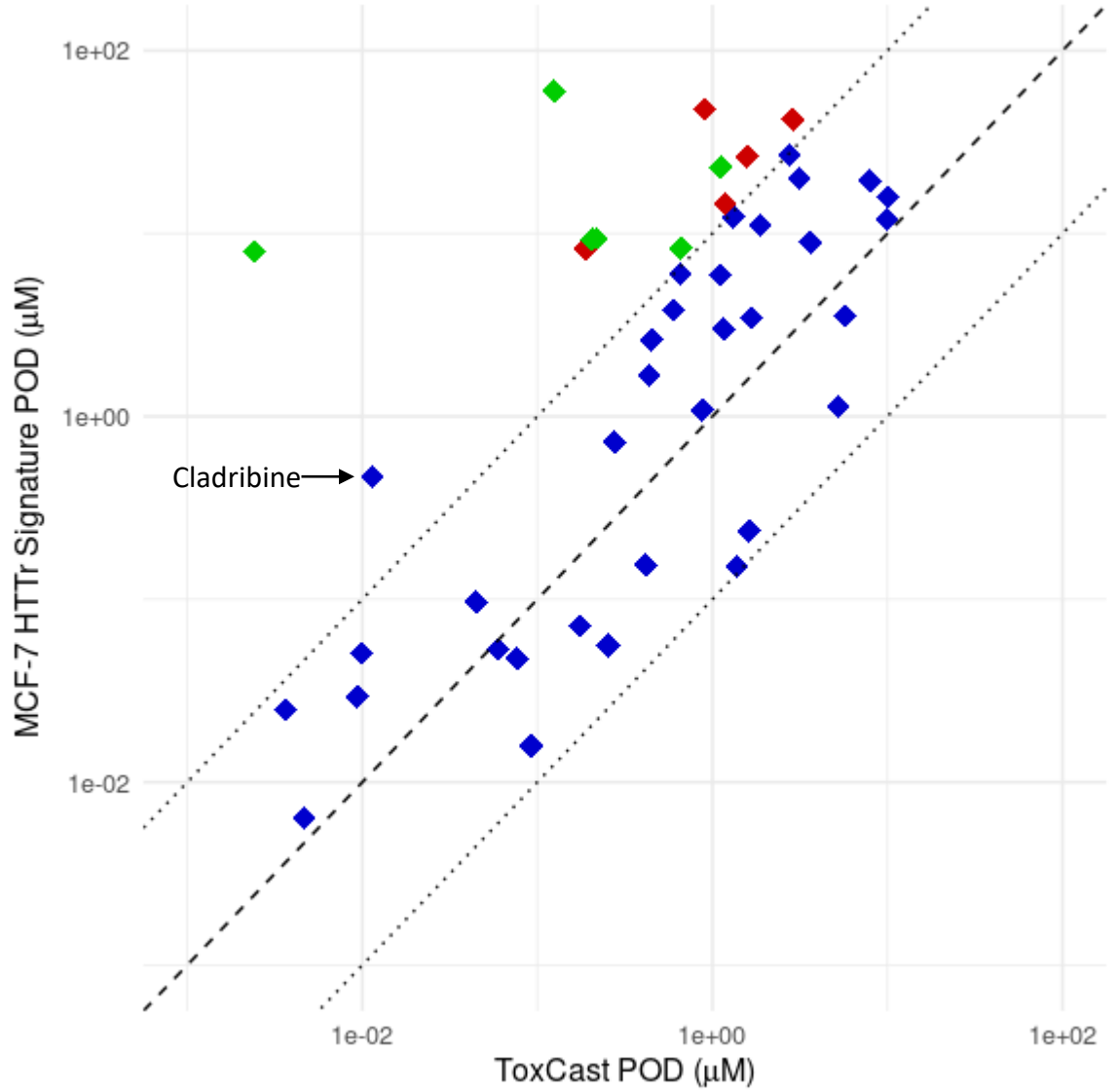
DESeq2

ssGSEA

Concentration-Response Curve Fitting (*tcplFit2*)



HTTr MCF-7 Pilot Analysis



- Pilot study of 44 well-characterized chemicals ***(Harrill, et al. Toxicol Sci, 2021)***
- 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

Conclusions

- EPA/ORD has developed reliable and cost-efficient workflow for generating HTTr data from thousands of chemicals across multiple cell lines

- Preliminary/pilot analysis demonstrates that overall results are concordant with previous assays (ToxCast/HTS) and known chemical targets
  Harrill, et al. Toxicol Sci 2021

- Upcoming research efforts will focus on:
  - Data generation in complementary cell models
  - 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

Acknowledgements

HTTr Team

Joshua Harrill

Richard Judson

Imran Shah

Woody Setzer

Derik Haggard

Beena Vallanat

Joseph Bundy

Bryant Chambers

Laura Taylor

Thomas Sheffield

Clinton Willis

CCTE Leadership

Rusty Thomas

Maureen Gwinn

John Cowden

Kimberly Slentz-Kesler

EPA Collaborators

Chris Corton

Mark Higuchi

Adam Speen

Johanna Nyffeler

HTTr Platform Selection

Matthew Martin

Agnes Karmaus

BioSpyder

Questions?

References

- Thomas RS, Bahadori T, Buckley TJ, et al. *"The Next Generation Blueprint of Computational Toxicology at the U.S. Environmental Protection Agency"*, Toxicol Sci 2019
- Yeakley JM, Shepard PJ, Goyena DE, et al. *"A trichostatin A expression signature identified by TempO-Seq targeted whole transcriptome profiling"*, PLoS ONE 2017
- Harrill J, Everett LJ, Haggard D, et al. *"High-Throughput Transcriptomics Platform for Screening Environmental Chemicals"*, Toxicol Sci 2021 *in press*
- Love MI, Huber W, and Anders S. *"Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2"*, Genome Biol 2014
- Barbie DA, Tamayo P, Boehm JS, et al. *"Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1"*, Nature 2009
- Huang R, Grishagin I, Wang Y, et al. *"The NCATS BioPlanet - An Integrated Platform for Exploring the Universe of Cellular Signaling Pathways for Toxicology, Systems Biology, and Chemical Genomics"*, Front Pharmacol 2019
- Subramanian A, Narayan R, Corsello SM, et al. *"A Next Generation Connectivity Map: L1000 Platform and the First 1,000,000 Profiles"*, Cell 2017
- Pinero J, Queralt-Rosinach N, Bravo A, et al. *"DisGeNET: a discovery platform for the dynamical exploration of human diseases and their genes"*, Database 2015
- Liberzon A, Birger C, Thorvaldsdottir H, et al. *"The Molecular Signatures Database (MSigDB) hallmark gene set collection"*, Cell Syst 2015
- Paul-Friedman K, Gagne M, Loo LH, et al. *"Utility of In Vitro Bioactivity as a Lower Bound Estimate of In Vivo Adverse Effect Levels and in Risk-Based Prioritization"*, Toxicol Sci 2020

Automated *in vitro* Chemical Screening

Cryopreserved Cell Stocks

Cell Expansion

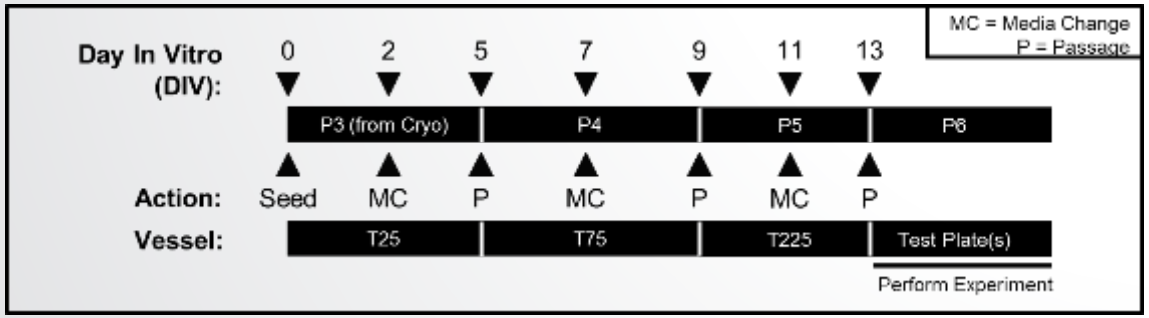
Cell Plating

BioTek MultiFlo ™ FX

Dispensing Test Chemicals

LabCyte Echo® 550 Liquid Handler

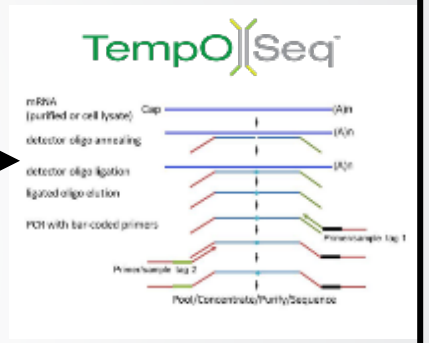
Standardized Expansion Protocol



Track 1: Targeted RNA-Seq

Generating Cell Lysates

TempO-Seq WT



Track 2: Apoptosis / Cell Viability

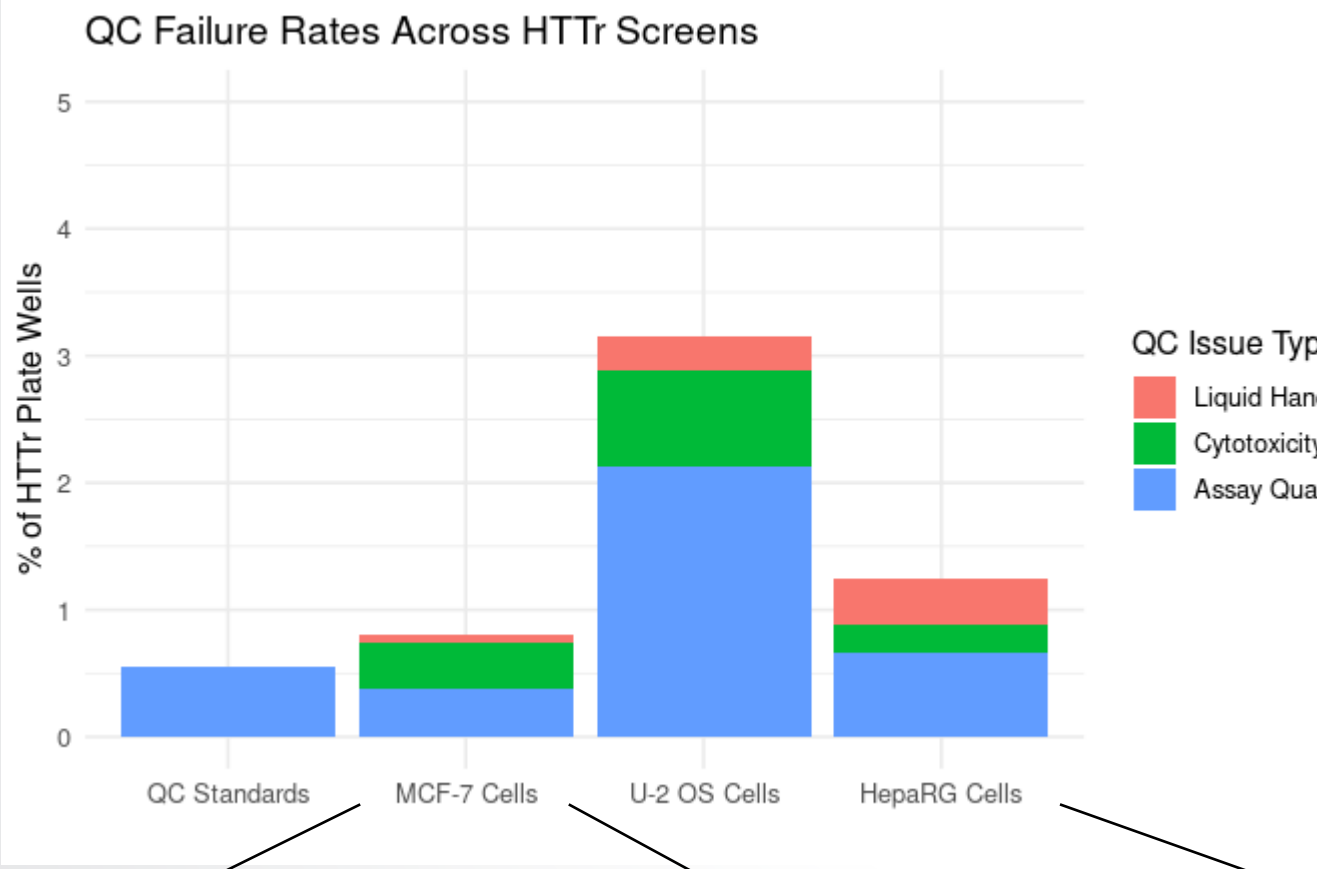
Reagent Dispensing

High Content Imaging

Perkin Elmer Opera Phenix™ High Content Screening System

Joshua Harrill

HTTr Quality Control



Acoustic dispenser logs identify problems with chemical handling

Apoptosis/cell viability assays identify cytotoxic concentrations

Bioinformatic QC checks remove:

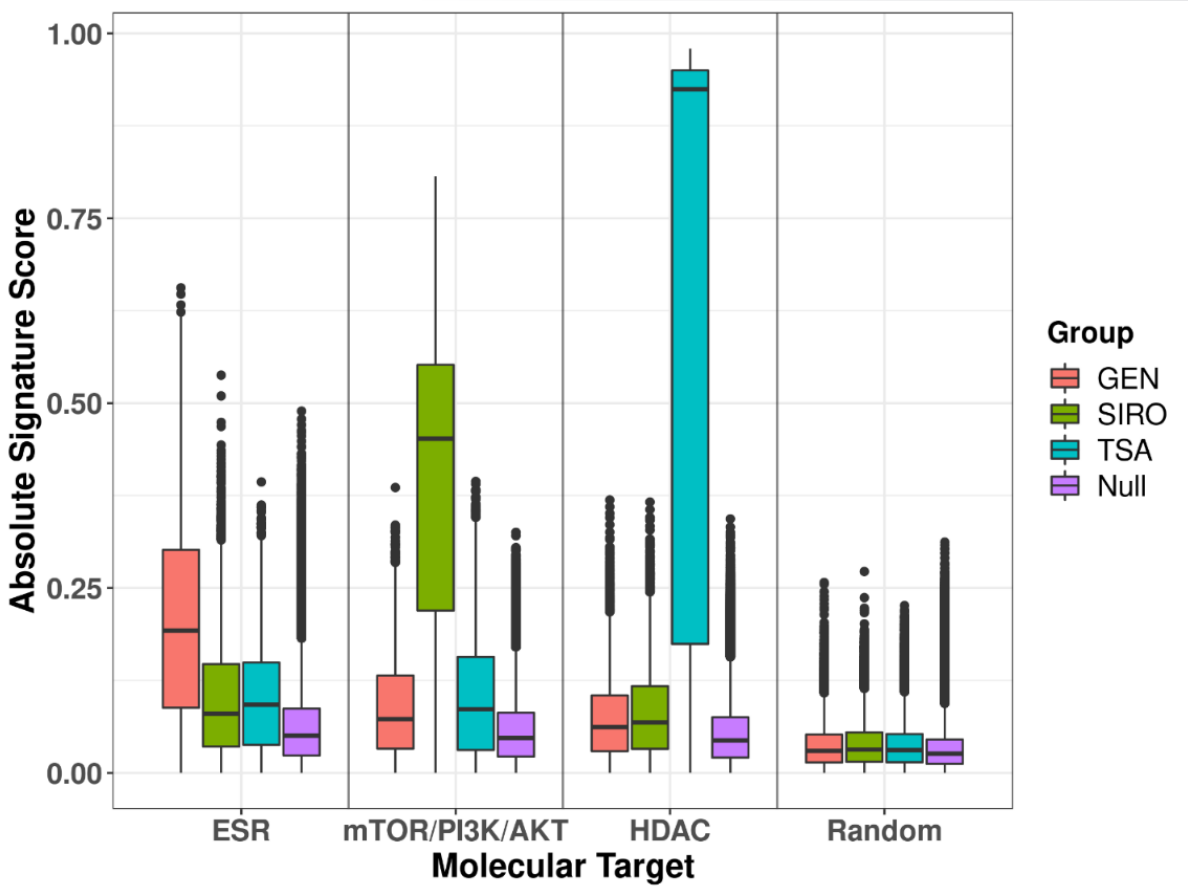
- Low read depth samples
- High rate of alignment failure
- Samples with low gene coverage
- Samples with irregular count distributions

- 44 Chemical Pilot Study
- Screened 1,577 ToxCast chemicals

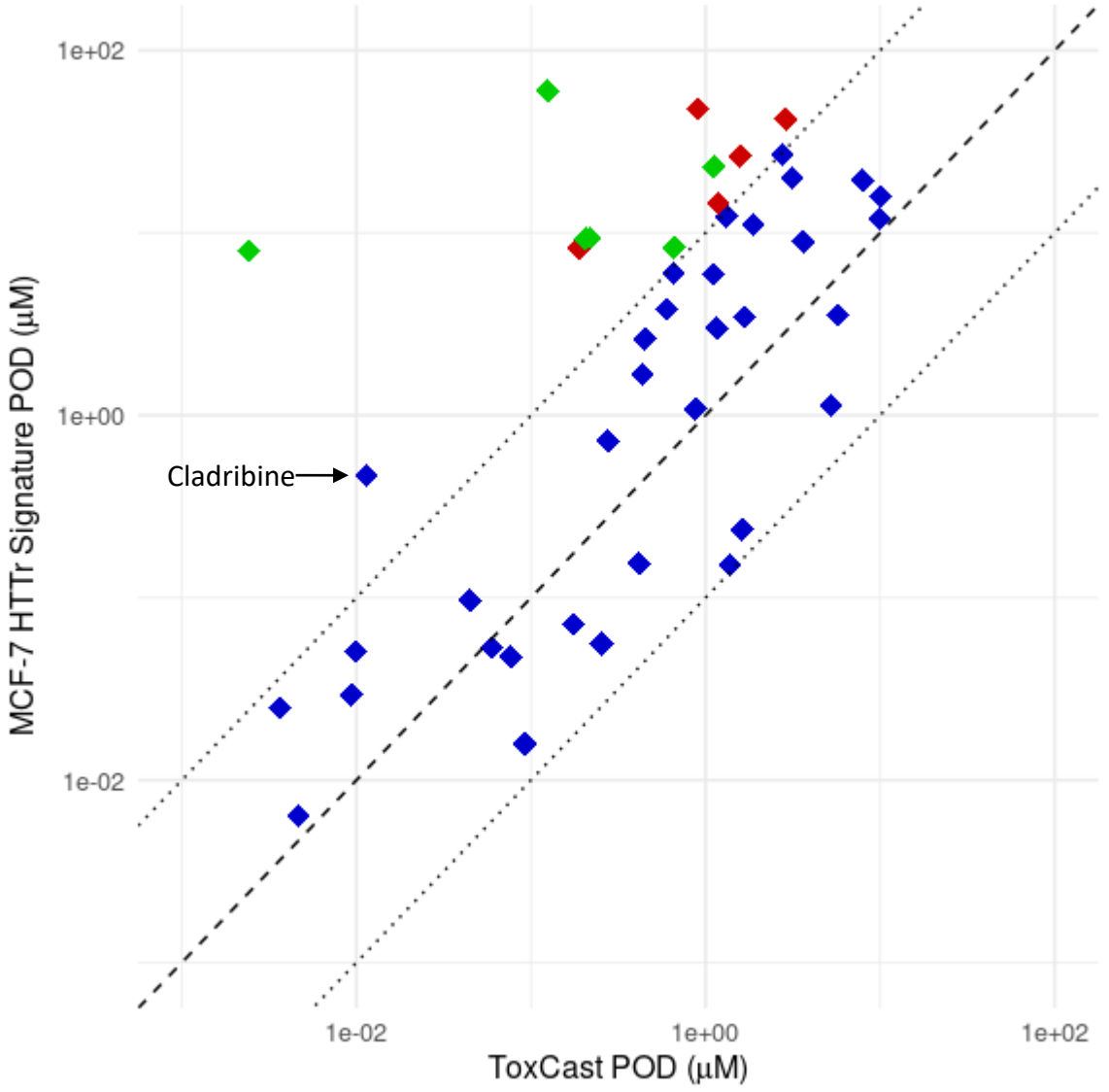
- Screened 1,201 ToxCast chemicals
- Screened 137 PFAS

Signature Scoring

- Analyzed differential expression response to 3 reference chemicals replicated 37 times throughout large screen (MCF-7)
  - GEN = Genistein (10uM)
  - SIRO = Sirolimus/Rapamycin (0.1uM)
  - TSA = Trichostatin A (1uM)
  - NULL = Signature scores derived from re-sampled log2 FC values
- Signatures were annotated for associated molecular targets
  - Random = Randomly selected gene sets with similar size to known signature gene sets
- Each reference chemical was enriched for higher scores from signature associated with correct molecular target
- Similar analysis and result found in MCF-7 pilot study ***(Harrill, et al. Toxicol Sci 2021)***



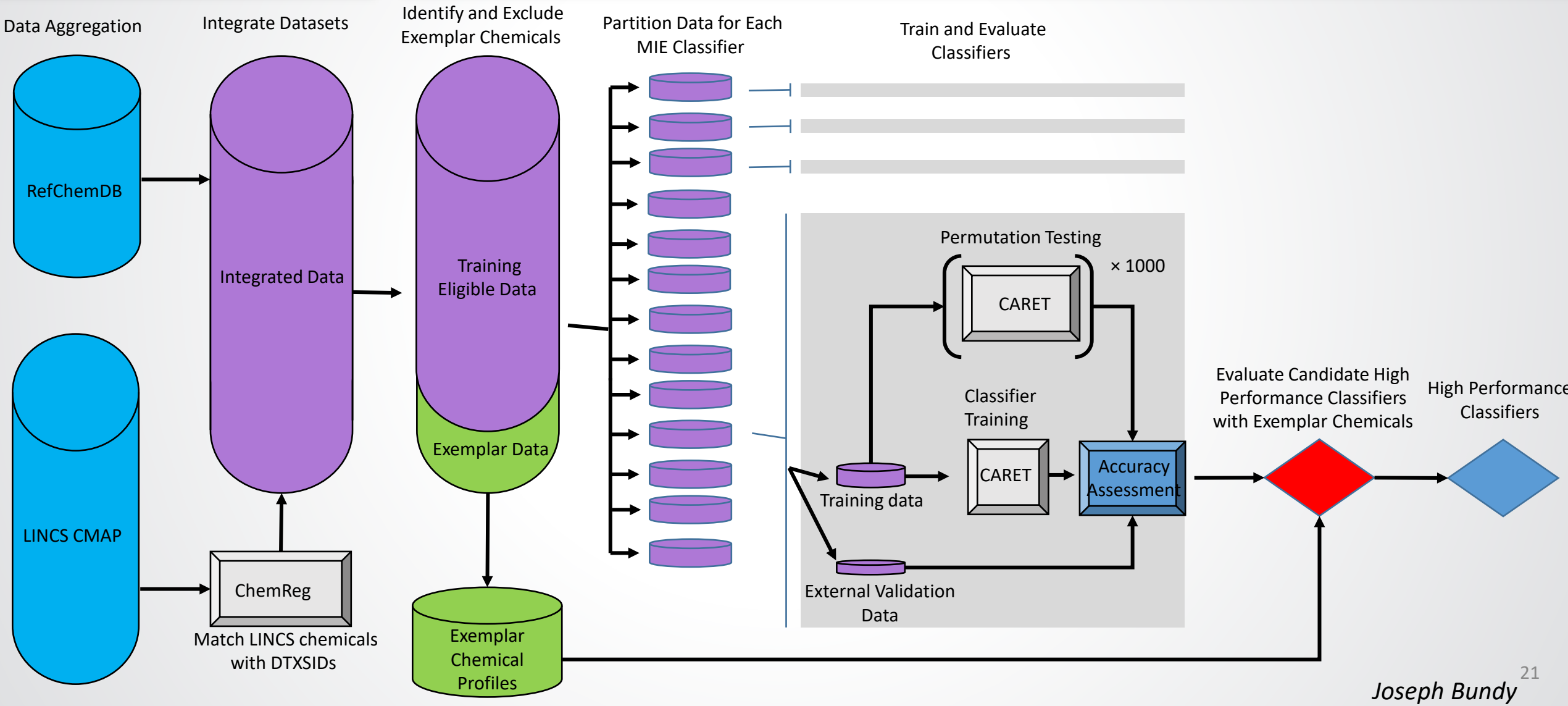
HTTr MCF-7 Pilot Analysis



- 6 chemicals with targets that have low/absent expression in MCF-7 cells
  - 3,5,3'-triiodothyronine (Thyroid Receptor)
  - Cyproconazole (pan-CYP inhibitor)
  - Butafenacil (pan-CYP inhibitor)
  - Prochloraz (pan-CYP inhibitor)
  - Imazalil (pan-CYP inhibitor)
  - Propiconazole (pan-CYP inhibitor)
- 5 chemicals show off-target hit as most potent assay in ToxCast
  - Lovastatin
  - Clofibrate
  - Maneb
  - Lactofen
  - Vinclozolin
- Cladribine is a non-specific DNA synthesis inhibitor

(Harrill, et al. Toxicol Sci, In Press)

ML Models for MIE Classification



Joseph Bundy

Stress Response Gene Signatures

Goal: Develop NAMs to characterize non-specific environmental chemicals that activate stress response pathways (SRPs)

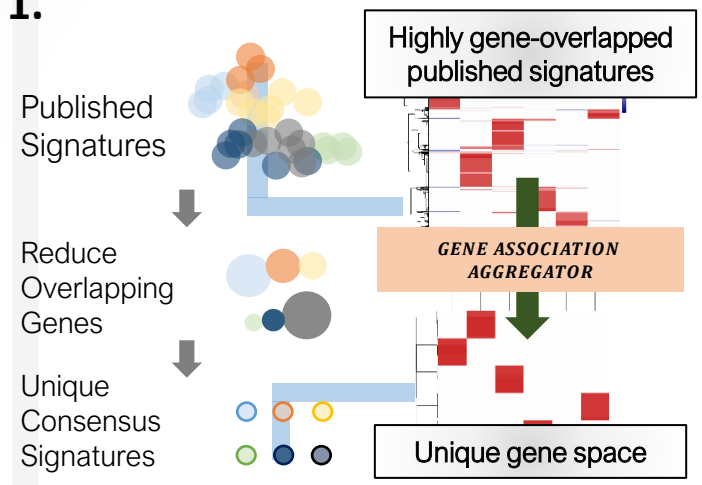
Approach: Characterize chemical hazards using HTTr data to assess SRP gene signature activity

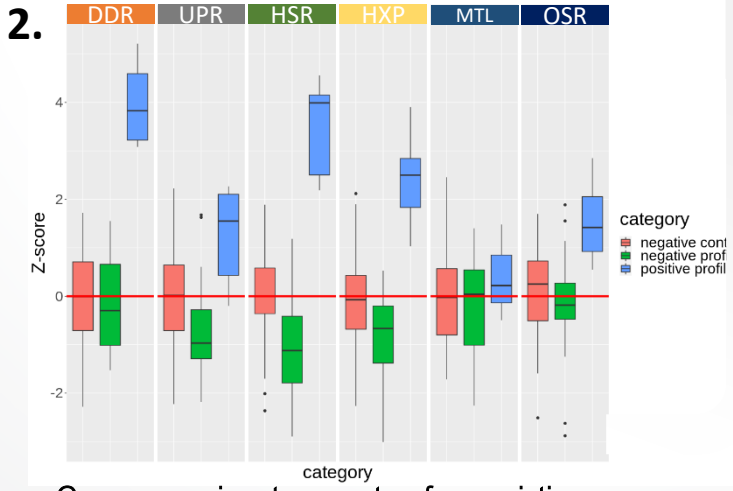
Challenges: Cross-talk in signaling networks makes it difficult to find gene signatures of SRPs

Results: We have developed *consensus* SRP signatures for accurately classifying known stressors

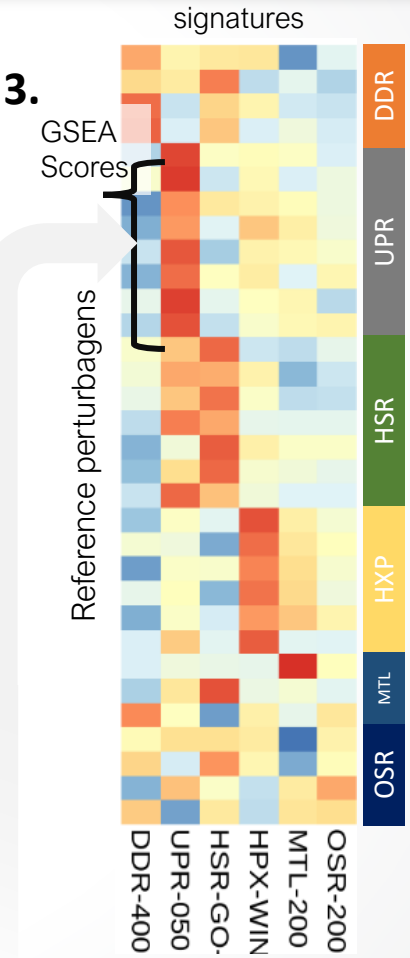
Future: Use signatures to identify cellular states involved in adaptive stress responses and "tipping points" that lead to adversity

Use crowd-sourcing strategy to build consensus signatures from published data





Consensus signatures outperform existing published signatures for SRP activity scoring



Best consensus signatures accurately classify 72% of perturbagens by score

DNA Damage Response

Unfolded Protein Response

Heat Shock Response

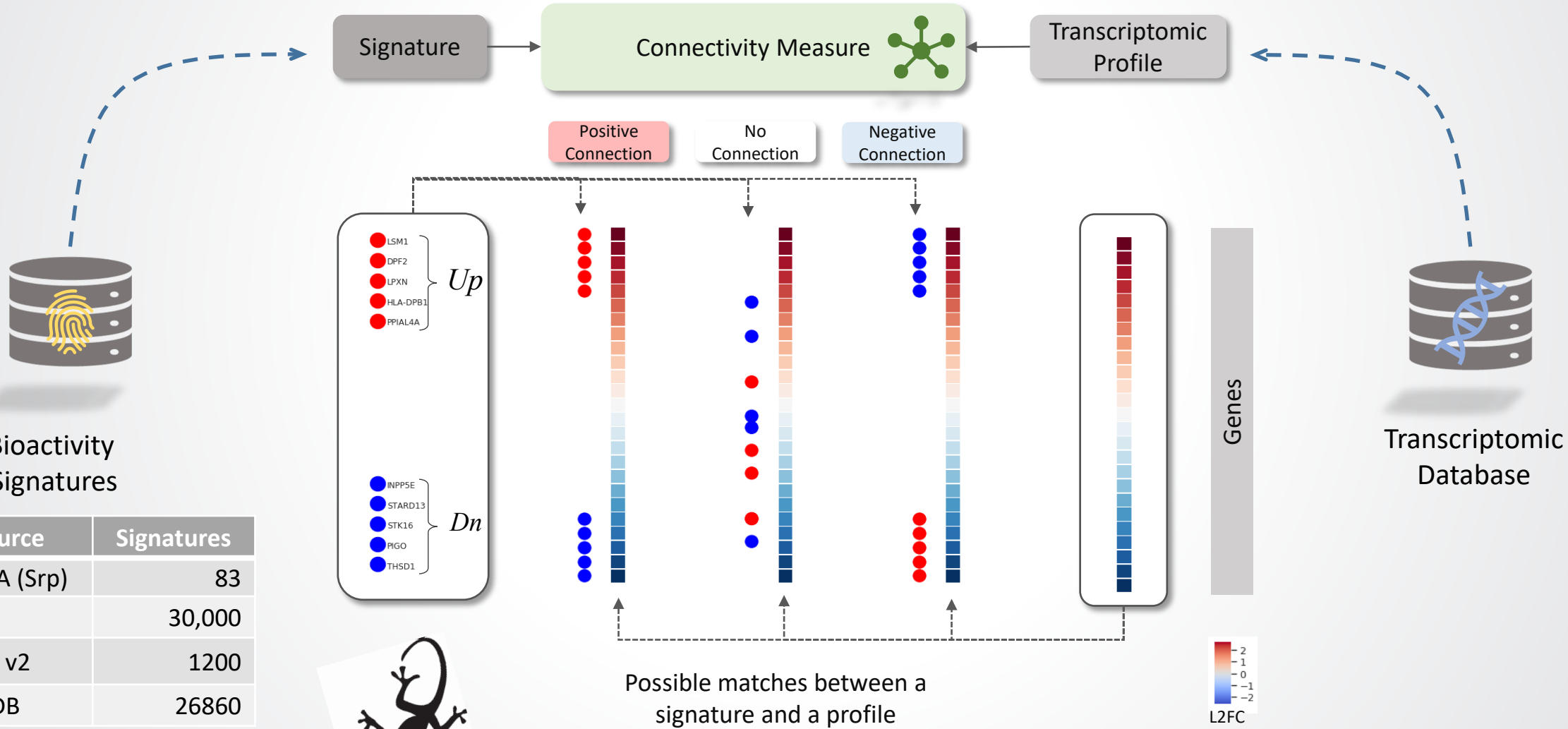
Hypoxic Response

Response to Metals

Oxidative Stress Response

Bryant Chambers & Imran Shah

Connectivity Mapping



Source	Signatures
US EPA (Srp)	83
Lincs	30,000
CMap v2	1200
MSigDB	26860

Gene Set Connectivity Toolkit (gecco) – Shah *et al.* (in prep)