High-Throughput Transcriptomics (HTTr) for Toxicological Testing

Joshua A. Harrill, Ph.D.

Toxicologist

Logan J. Everett, Ph.D.

Bioinformatician

Biomolecular and Computational Toxicology Division
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 presenters 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.

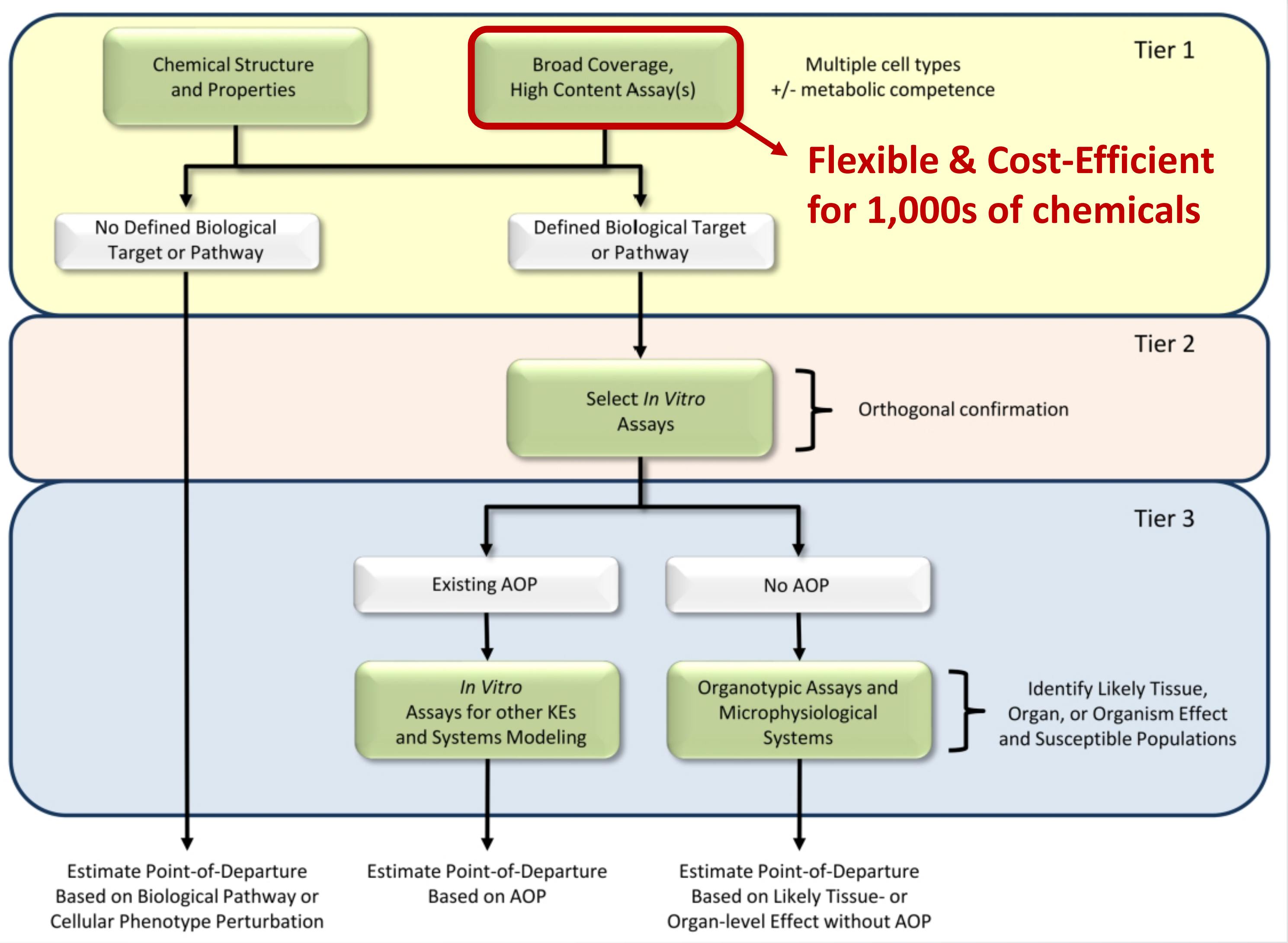
Tiered Chemical Safety Testing Strategy

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



Tier 1

Chemical Structure and Properties

Broad Coverage, High Content Assay(s)

Multiple cell types +/- metabolic competence

Flexible & Cost-Efficient for 1,000s of chemicals

No Defined Biological Target or Pathway

Defined Biological Target or Pathway

Tier 2

Select *In Vitro* Assays

Orthogonal confirmation

Tier 3

Existing AOP

No AOP

In Vitro Assays for other KEs and Systems Modeling

Organotypic Assays and Microphysiological Systems

Identify Likely Tissue, Organ, or Organism Effect and Susceptible Populations

Estimate Point-of-Departure Based on Biological Pathway or Cellular Phenotype Perturbation

Estimate Point-of-Departure Based on AOP

Estimate Point-of-Departure Based on Likely Tissue- or Organ-level Effect without AOP

Thomas, et al. Toxicol Sci 2019

Automated *in vitro* Chemical Screening

Cryopreserved Cell Stocks

Cell Expansion

Cell Plating

BioTek MultiFlo ™ FX

Dispensing Test Chemicals

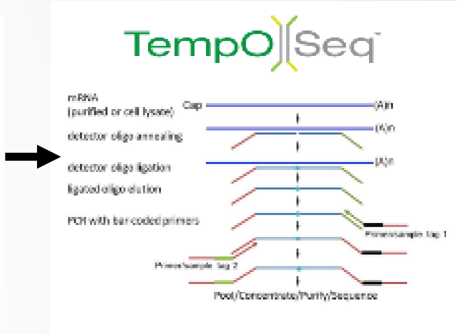
LabCyte Echo® 550 Liquid Handler

6 or 24 Hour Exposures

Track 1: Targeted RNA-Seq

Generating Cell Lysates

TempO-Seq WT



Track 2: Apoptosis / Cell Viability

Reagent Dispensing

High Content Imaging

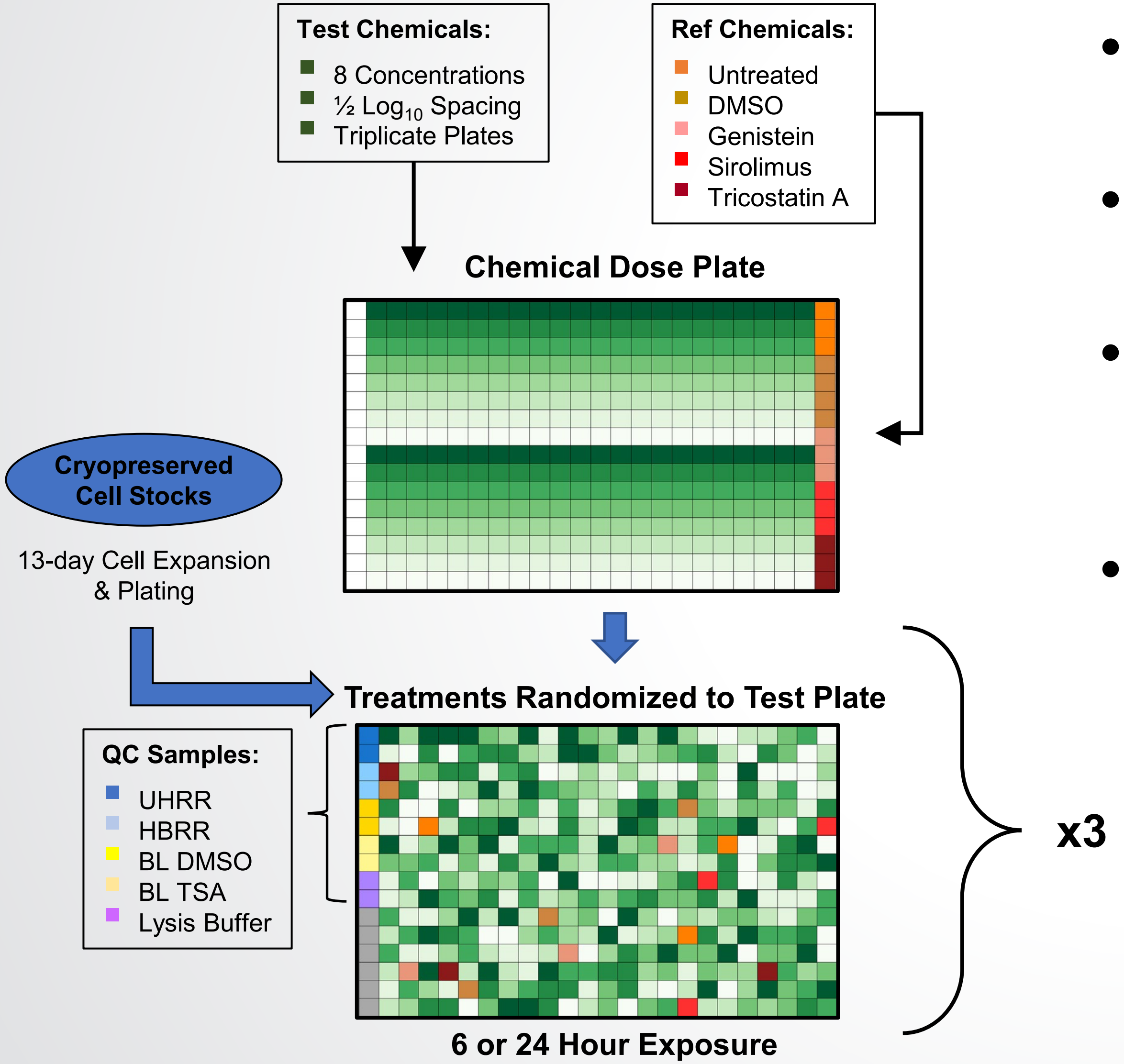
Perkin Elmer Opera Phenix™ High Content Screening System

Standardized Expansion Protocol

MC = Media Change
P = Passage

Day In Vitro (DIV):	0	2	5	7	9	11	13
	P3 (from Cryo)		P4		P5		P6
Action:	Seed	MC	P	MC	P	MC	P
Vessel:	T25		T75		T225		Test Plate(s)
							Perform Experiment

HTTr Study Design



Test Chemicals:
- 8 Concentrations
- ½ Log_{10} Spacing
- Triplicate Plates

Ref Chemicals:
- 🟧 Untreated
- 🟨 DMSO
- 🟥 Genistein
- 🟥 Sirolimus
- 🟥 Tricostatin A

Chemical Dose Plate

Cryopreserved Cell Stocks

13-day Cell Expansion & Plating

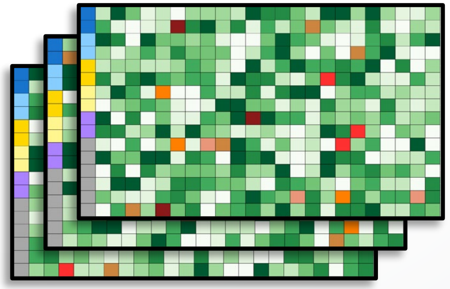
Treatments Randomized to Test Plate

QC Samples:
- 🟦 UHRR
- 🟦 HBRR
- 🟨 BL DMSO
- 🟨 BL TSA
- 🟪 Lysis Buffer

6 or 24 Hour Exposure

x3

- 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:

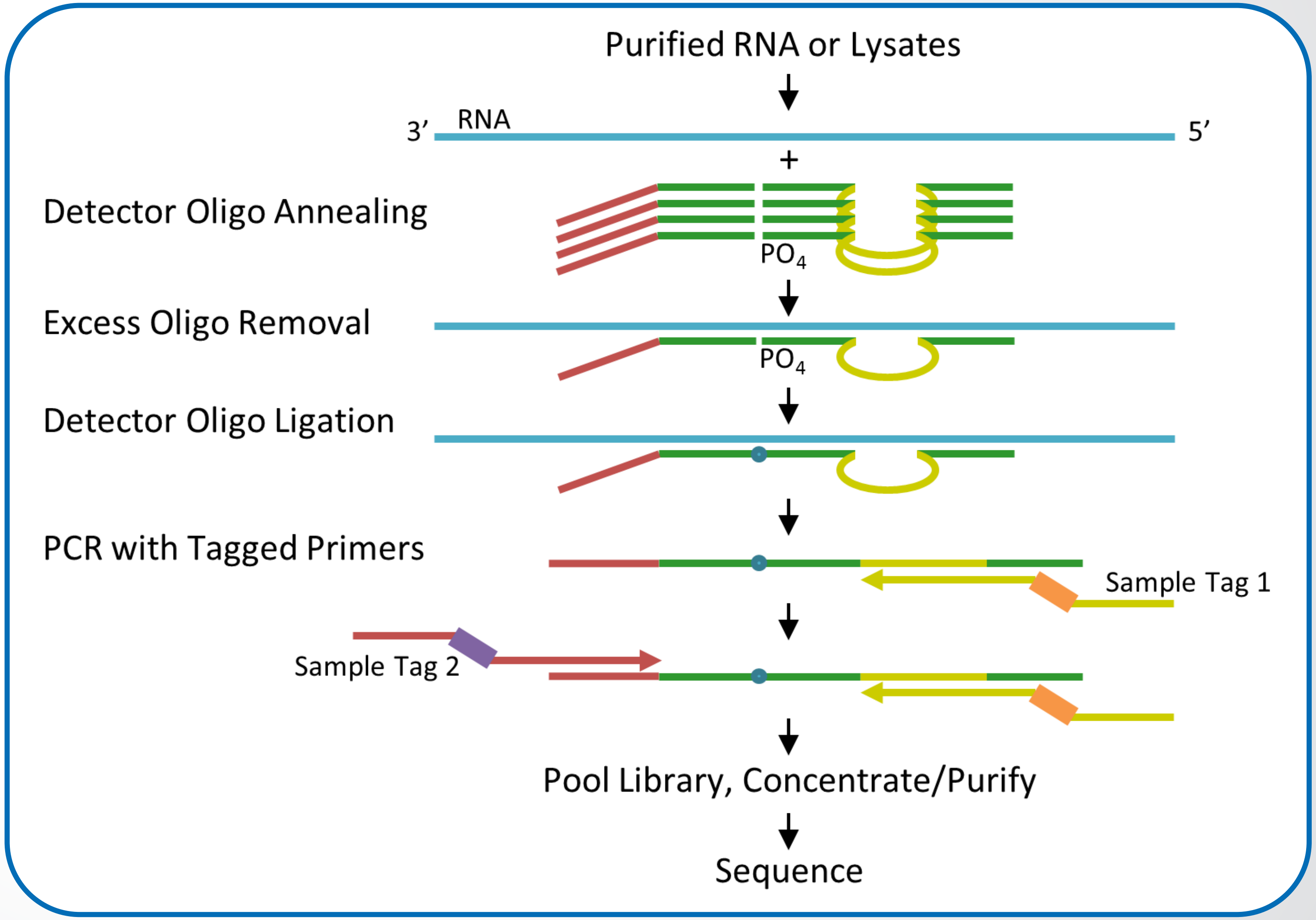


- ➢ Randomized independently
- ➢ Separate cell culture batches

Harrill, et al. Toxicol Sci 2021

TempO-seq Assay

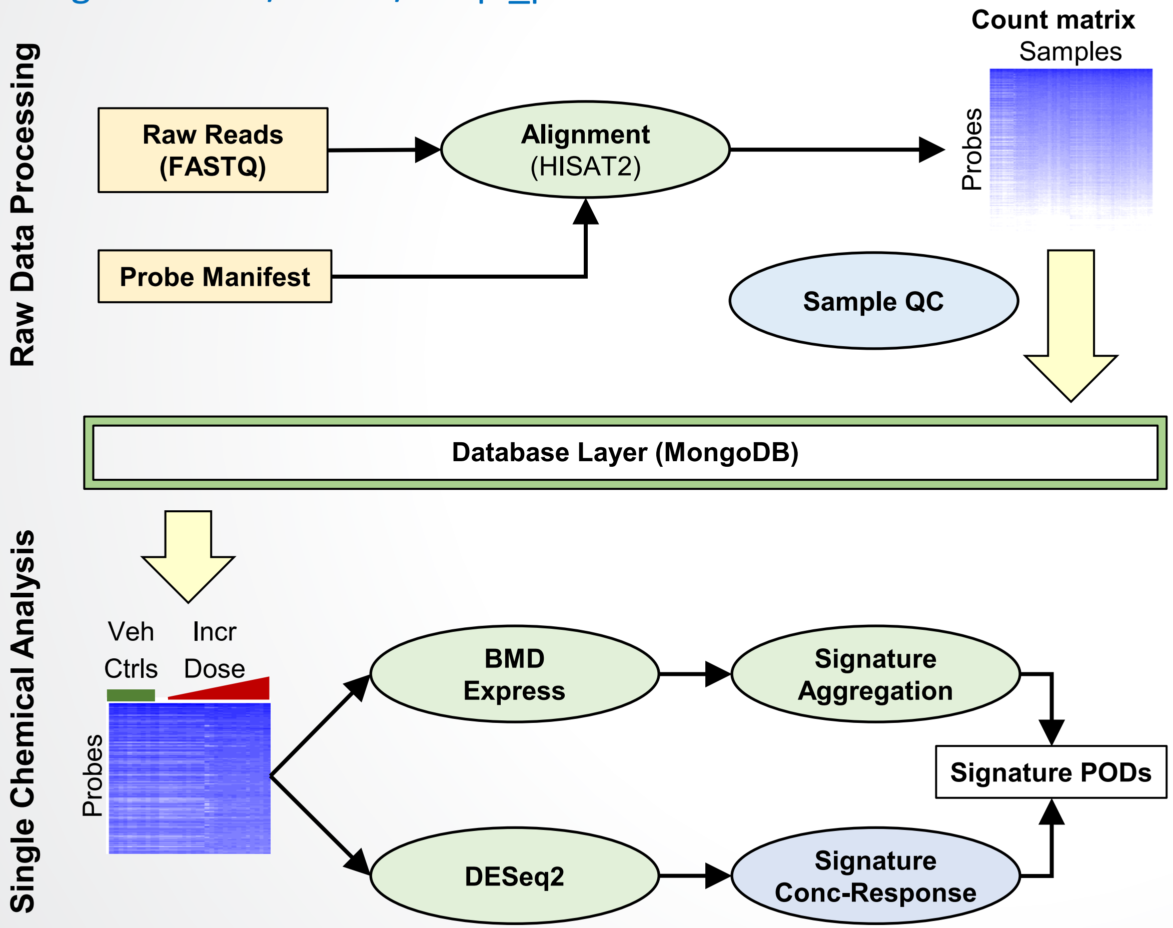
- Targeted RNA-seq (TempO-seq) enables high-throughput profiling of **cell lysates** or purified RNA
- Probe set for whole human transcriptome targets ~21,000 human genes
- Captures majority of signal with much lower sequencing depth (~3M reads with attenuation)
- Barcoding and pooling allows multiplexing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



Yeakley, et al. PLoS ONE 2017

HTTr Bioinformatics Pipeline

github.com/USEPA/httrpl_pilot

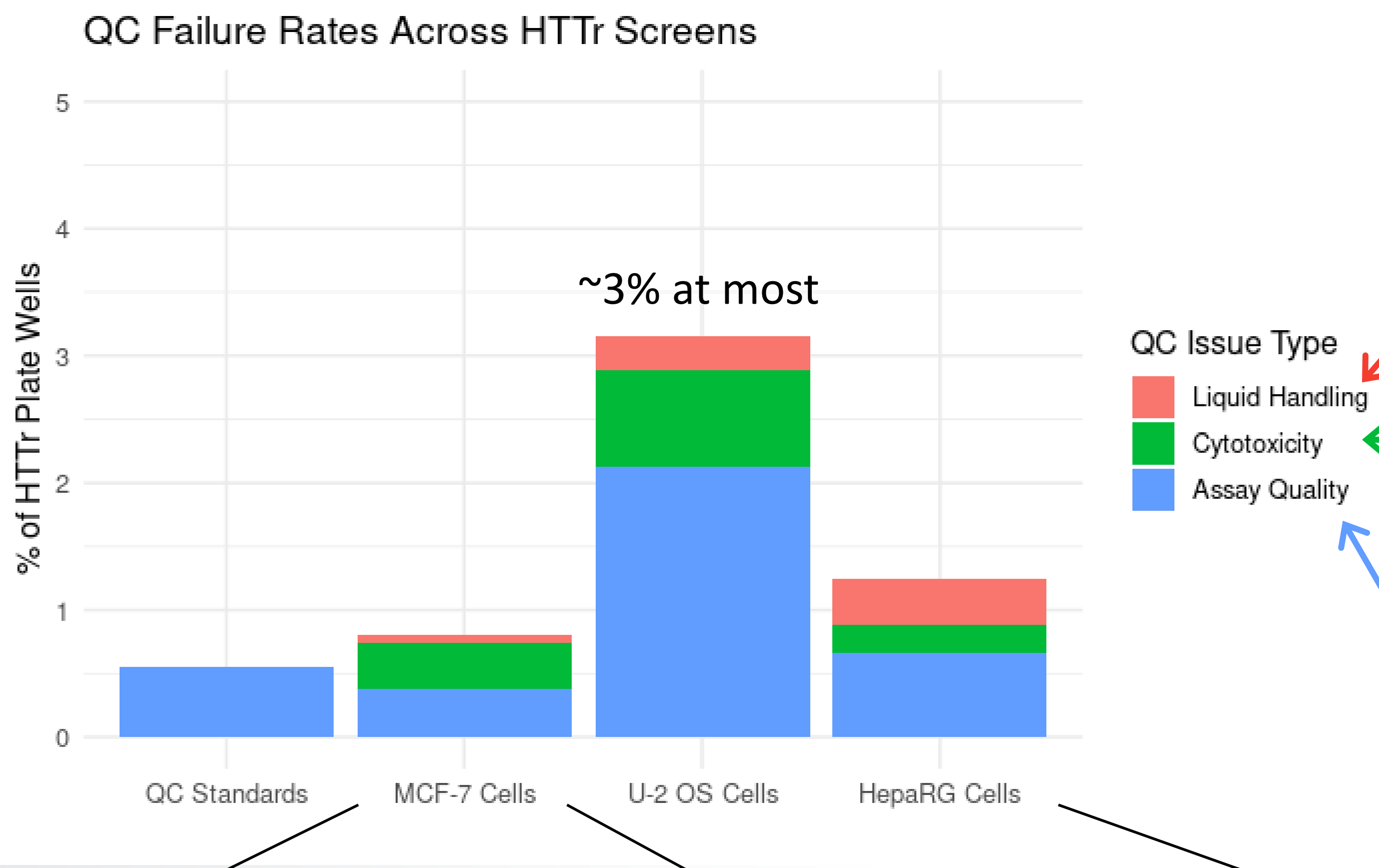


github.com/USEPA/CompTox-httrpathway

- 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 Quality Control



Acoustic dispenser logs identify problems with chemical handling

Apoptosis/cell viability assays identify >50% cytotoxic concentrations

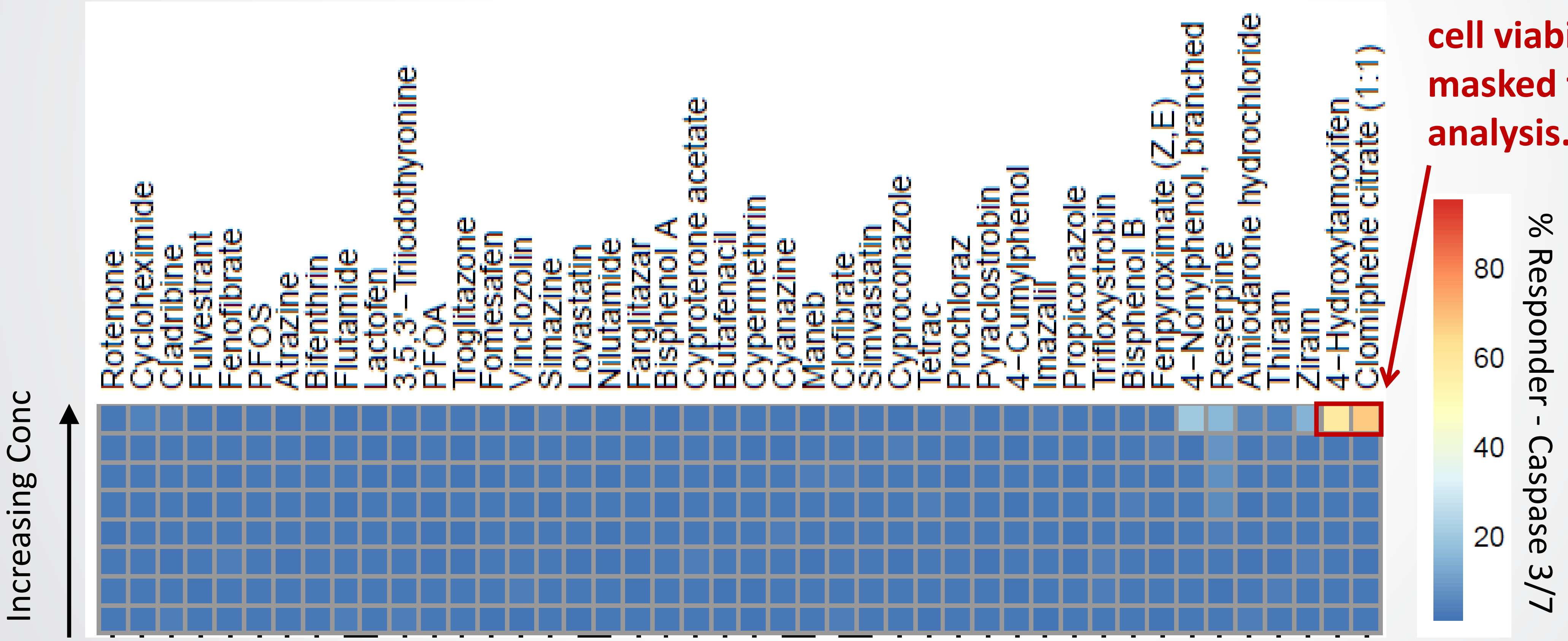
Bioinformatic QC checks remove:

- 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

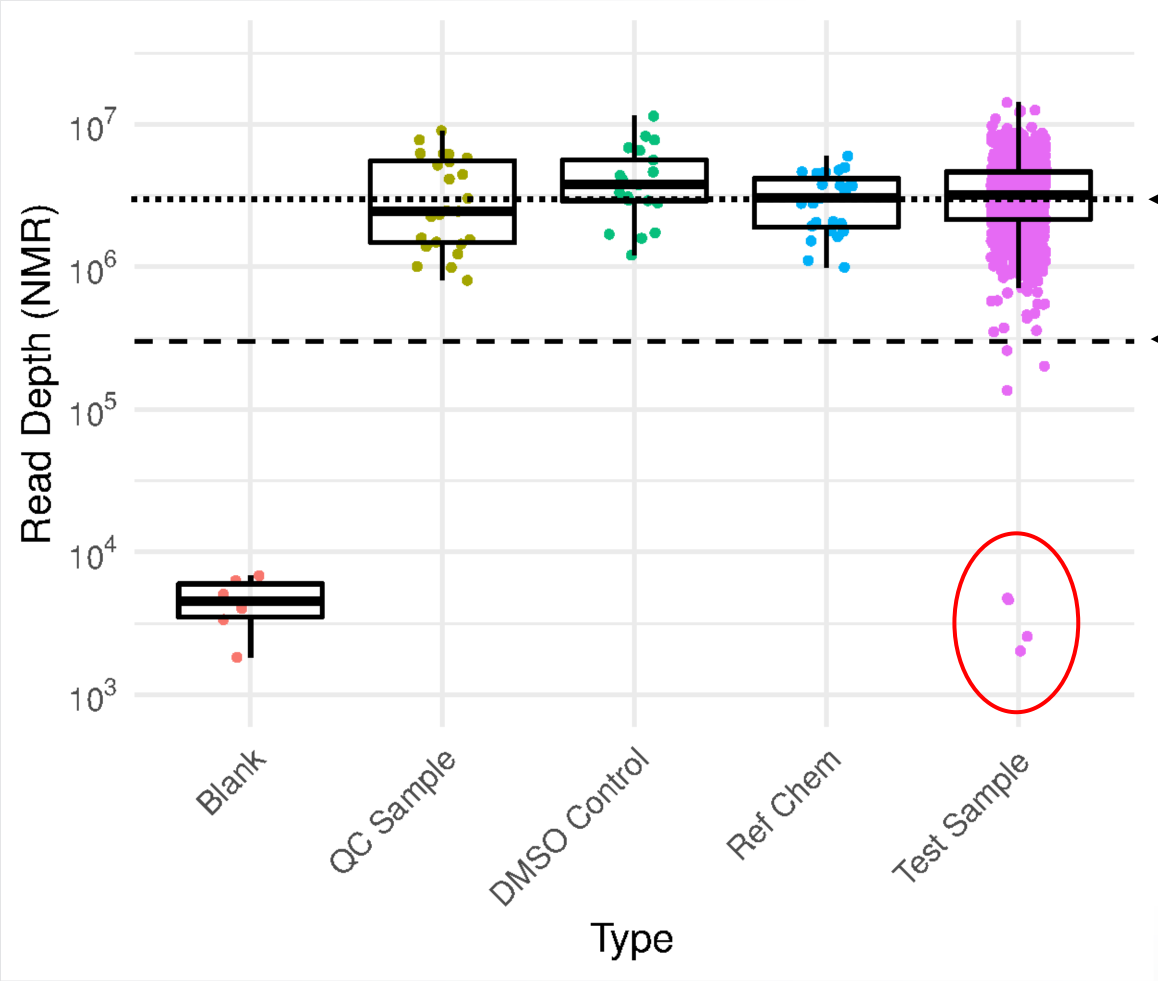
- Screened 1,201 ToxCast chemicals
- Screened 137 PFAS

Cell Viability



Conditions causing cell viability loss >50% masked from further analysis.

QC Metrics: Read Depth



Depends on probe set, cell type, attenuation

Target Depth = 3M Reads

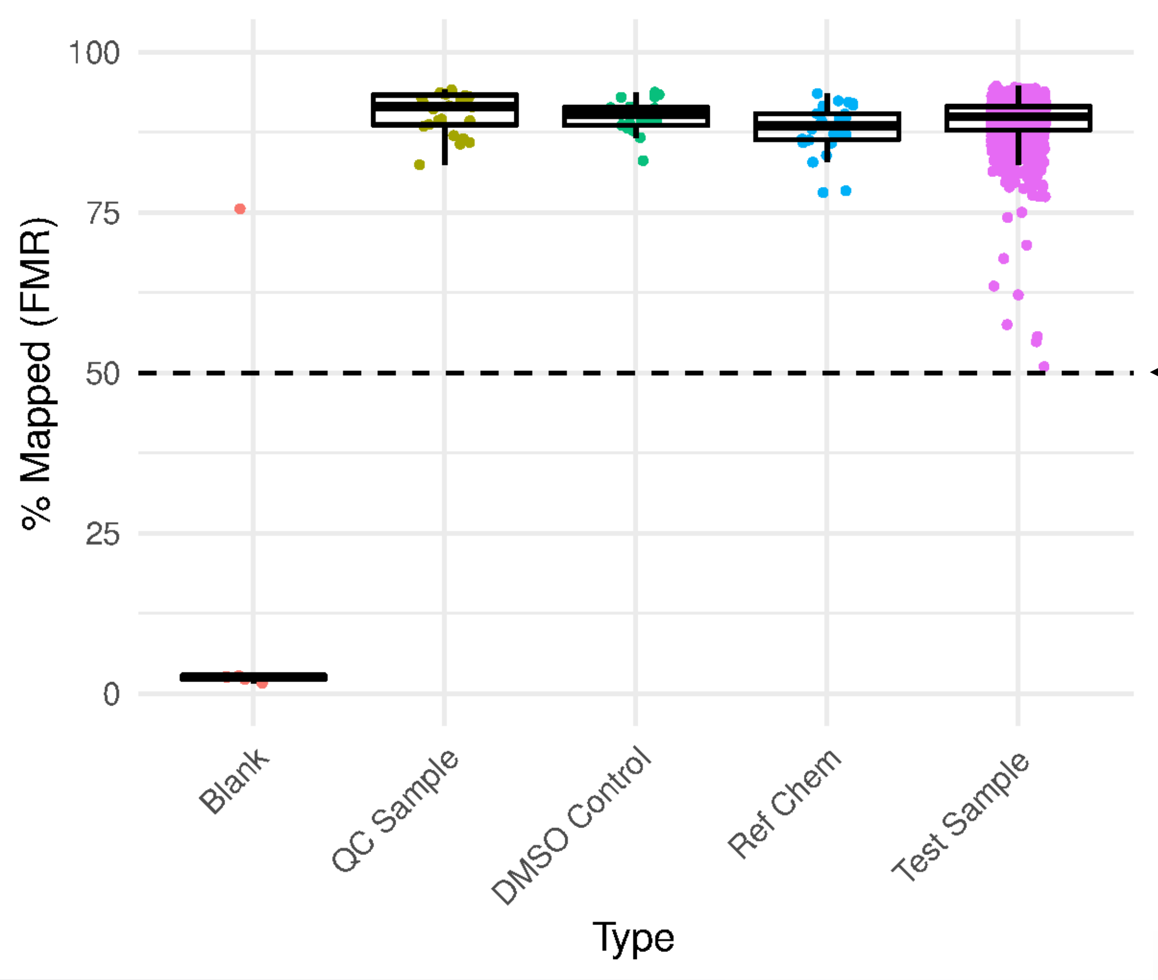
Threshold = 10% of Target Depth

More variability for cell lysates than purified RNA

Reasons for low read depth:

- **Cytotoxicity**
- **Sample degradation**
- **Low input**
- **Assay failure**

QC Metrics: Mapping Rate



- **Each read mapped to known probe sequences**
- **Only uniquely mapped reads used for analysis**

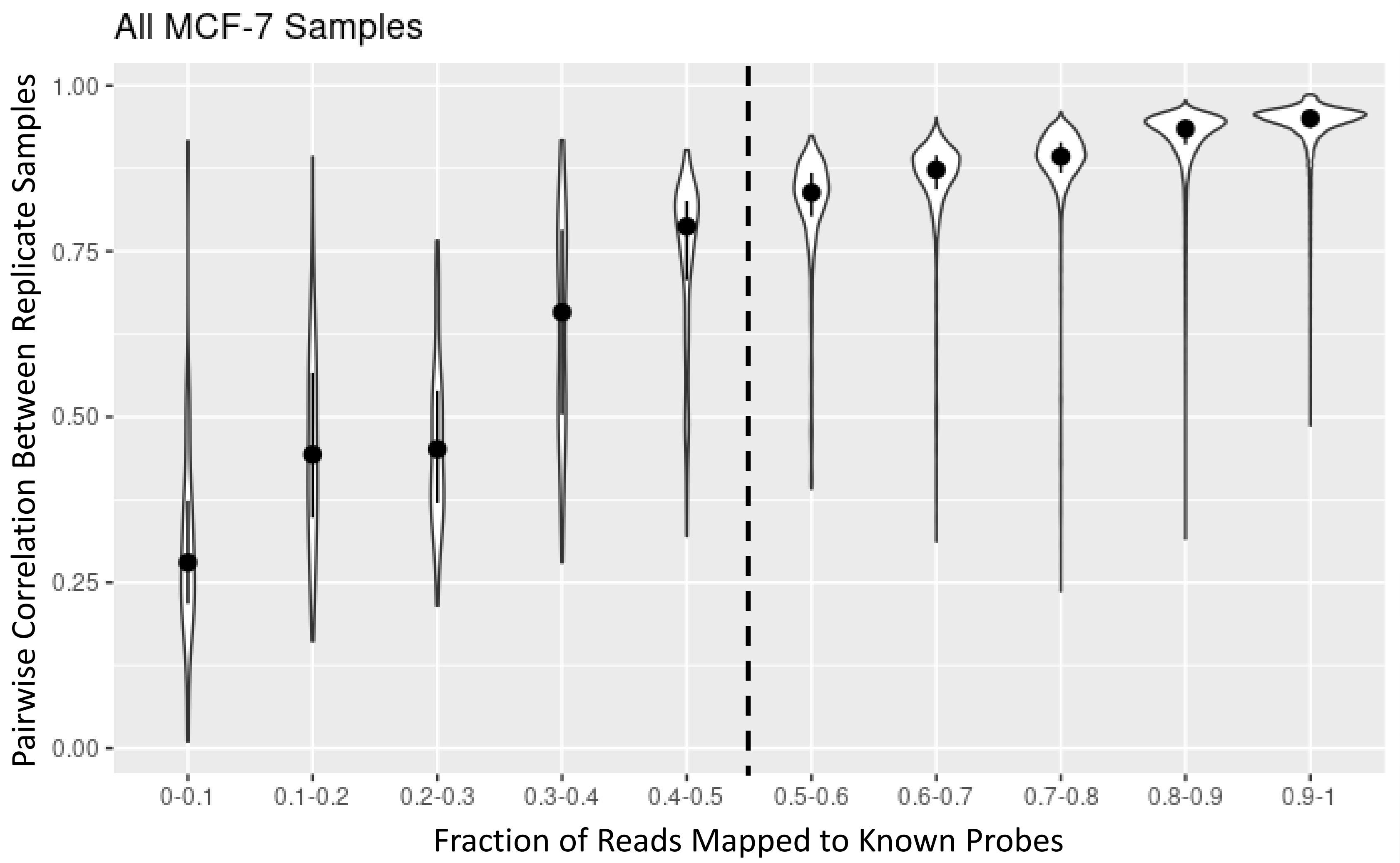
Threshold = 50% Mapping Rate

May depend on media/lysate condition, cell type

Reasons for low mapping rate:

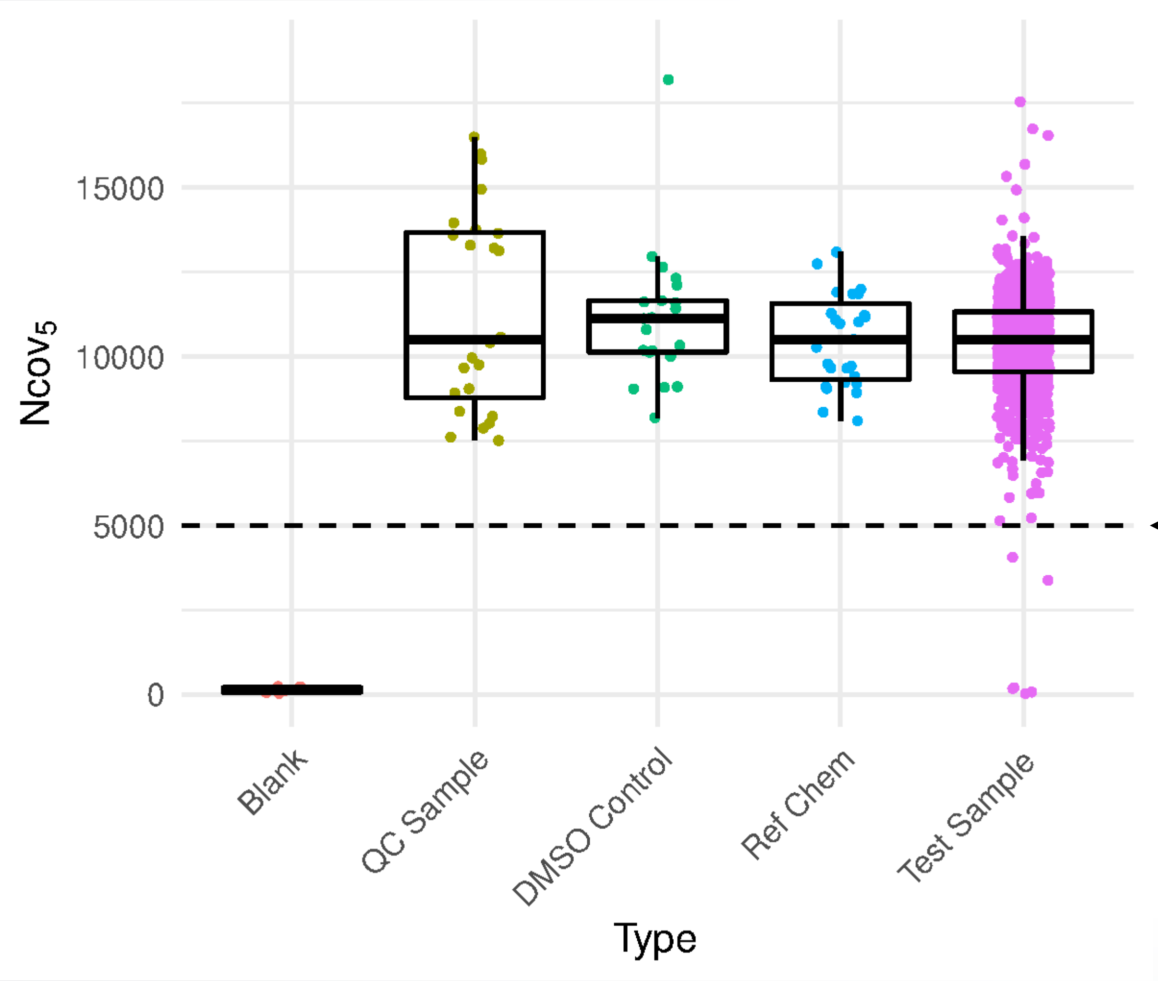
- **Cytotoxicity**
- **Sample degradation**
- **Low input**
- **Assay failure**

QC Metrics: Mapping Rate



- Replicate correlation drops off when <50% of reads mapped uniquely to probe sequences
- Lower mapping rate leads to lower depth
- May also indicate sample quality issues (e.g. RNA degradation or incomplete cell lysis)

QC Metrics: Transcriptome Coverage



$Ncov_5$ = # of probes with at least 5 reads

Threshold = 5,000 Probes (MCF-7)

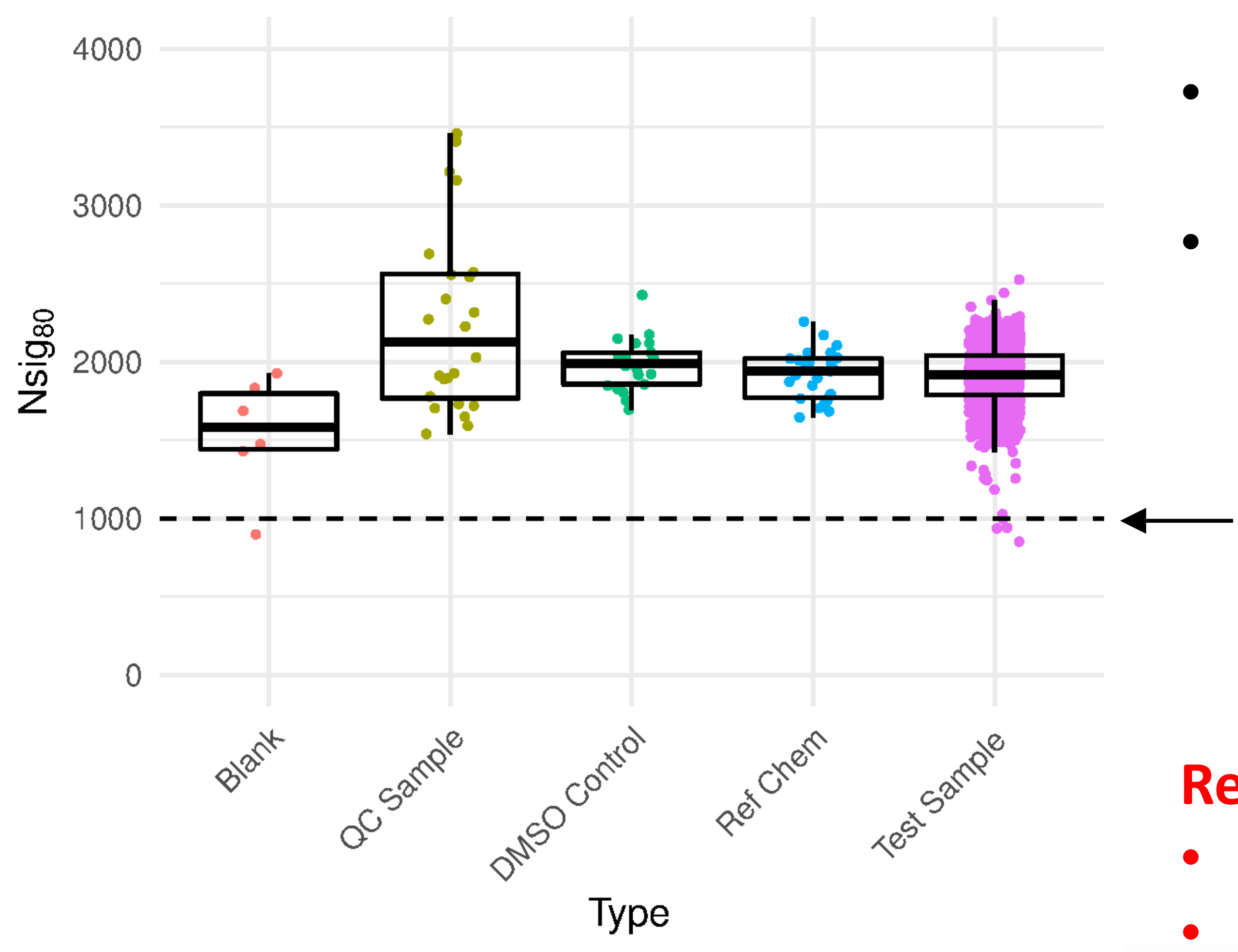
Based on "outer fence" principle (Tukey, 1976)

Re-evaluated on new cell types, probe sets, and attenuation strategies

Reasons for low coverage samples:

- **Low read depth**
- **Sample degradation**
- **Low input**
- **Assay failure**

QC Metrics: Signal Distribution



- **Nsig_{80} = # of probes capturing top 80% of signal**
- **Low values = reads highly concentrated among small number of probes**

Threshold = 1,000 Probes (MCF-7)

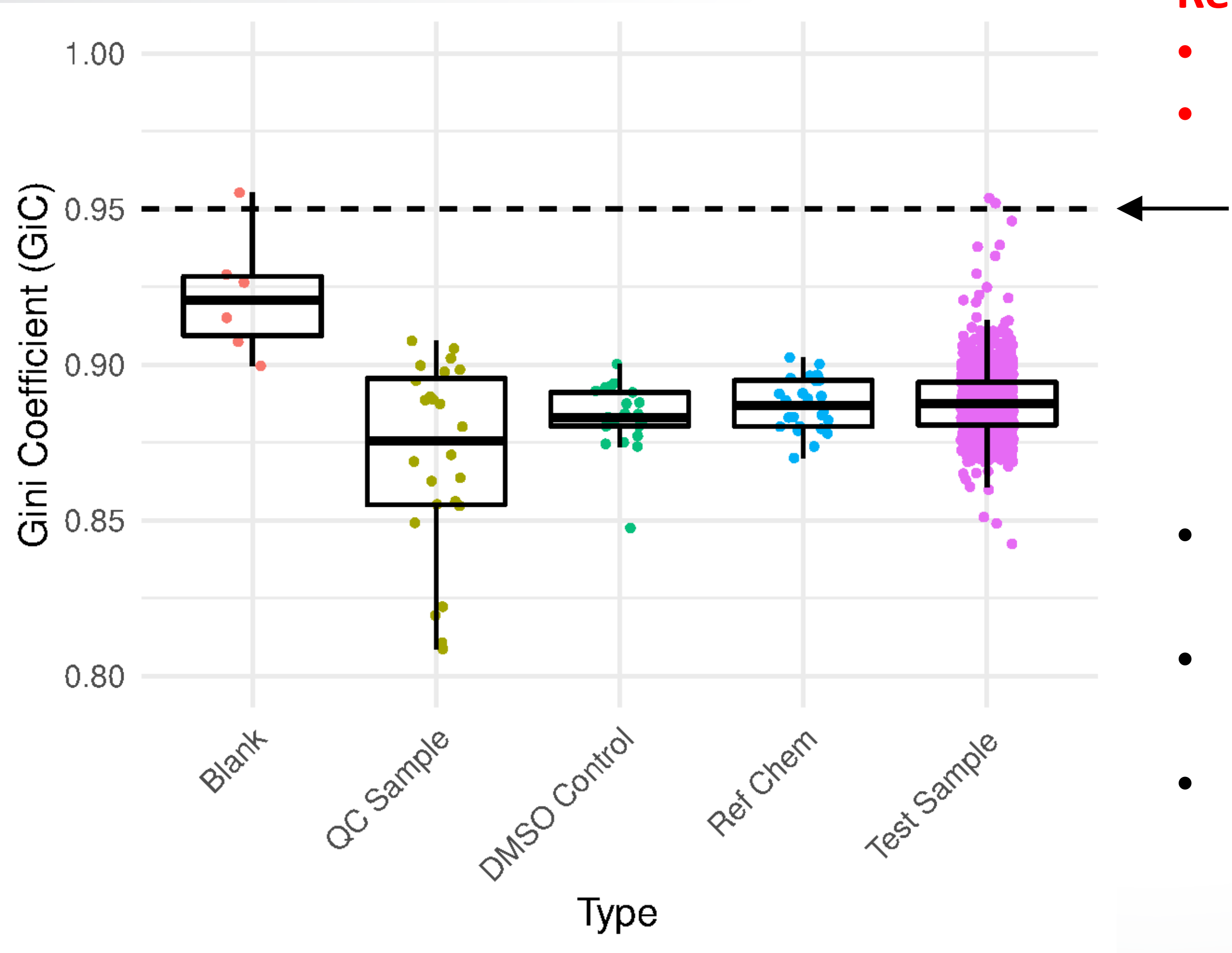
Based on "outer fence" principle (Tukey, 1976)

Should be re-evaluated on new cell types, probe sets, and attenuation strategies

Reasons for low values:

- **Sample degradation**
- **Low input**
- **Assay failure**

QC Metrics: Signal Distribution



Reasons for high values:

- **Sample degradation**
- **Low input**

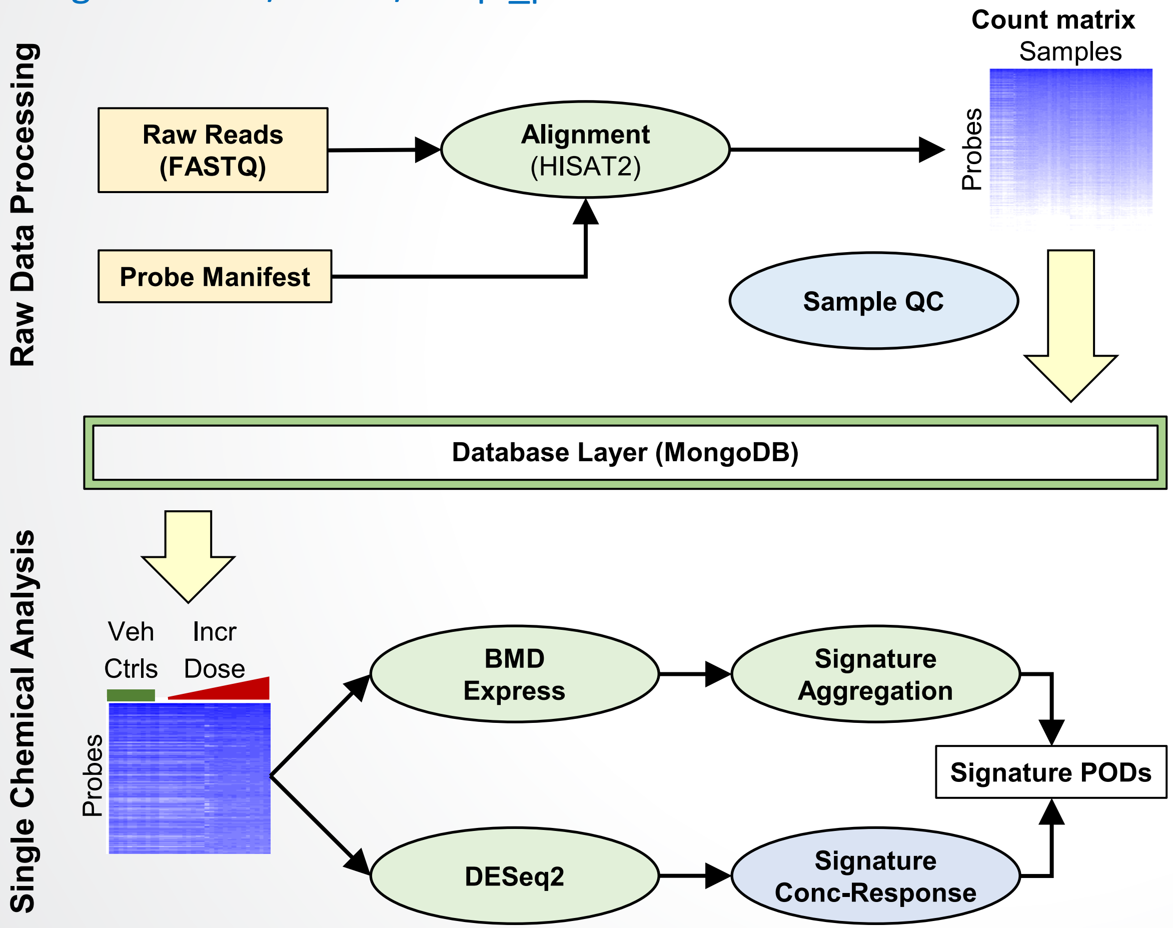
Threshold = 0.95

Based on "outer fence" principle (Tukey, 1976)
Should be re-evaluated on new cell types, probe sets, and attenuation strategies

- **Gini coefficient = measure of inequality or skewness in a distribution**
- **High values = most reads coming from few probes (Max 1: All reads from 1 probe)**
- **Lower values = closer to uniform distribution of reads across all probes (Min 0, not expected for expression data)**
- **Expect samples from same cell type to be similar**

HTTr Bioinformatics Pipeline

github.com/USEPA/httrpl_pilot



github.com/USEPA/CompTox-httrpathway

- 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

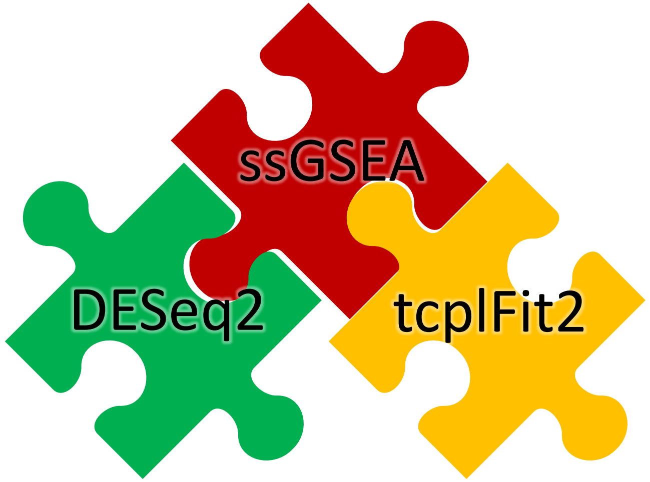
Harrill, et al. Toxicol Sci 2021

3 Flavors of Bioinformatic Approaches

Gene-First Approaches

- BMDExpress (NTP)
- tcplFit2 (CCTE)
- BIFROST (Unilever)

Signature Conc-Response



github.com/USEPA/CompTox-httrpathway

(Richard Judson)

Latent Variable Methods

- Many possible tools, e.g. PLIER, WGCNA
- Identify latent variables in data that capture primary response patterns
- Annotate biological relevance of LVs by gene components
- Perform curve-fitting on LVs
  - Fewer features to fit
  - Compatible with BMDExpress & tcplFit2

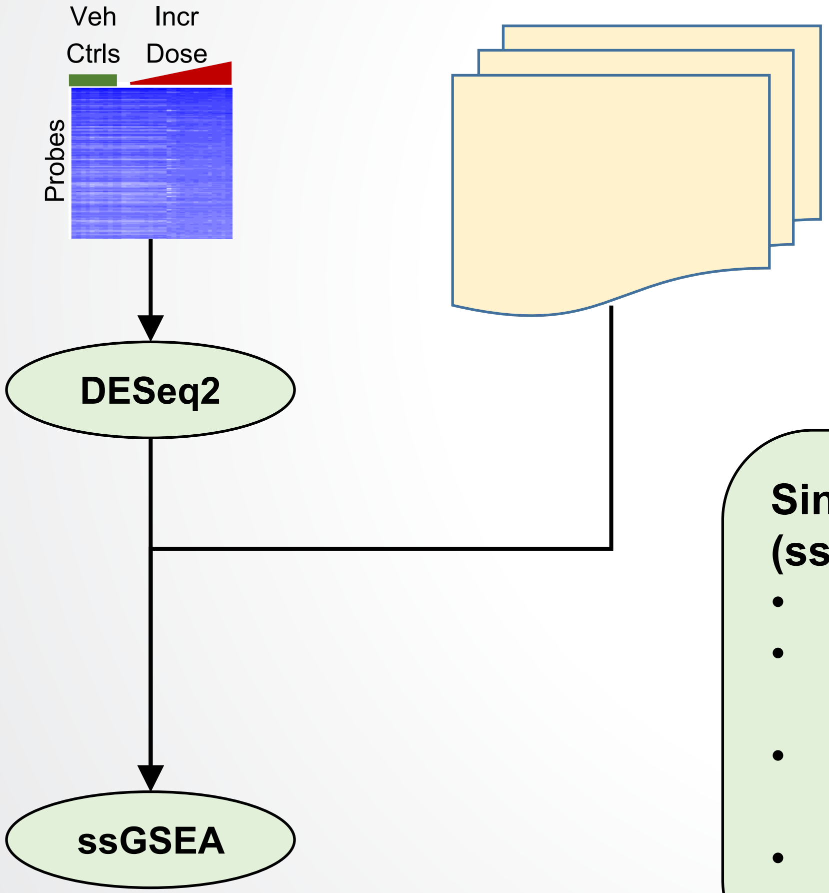
Improved integration through HTTr pipeline & database development

Signature Scoring

Count data per chemical

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

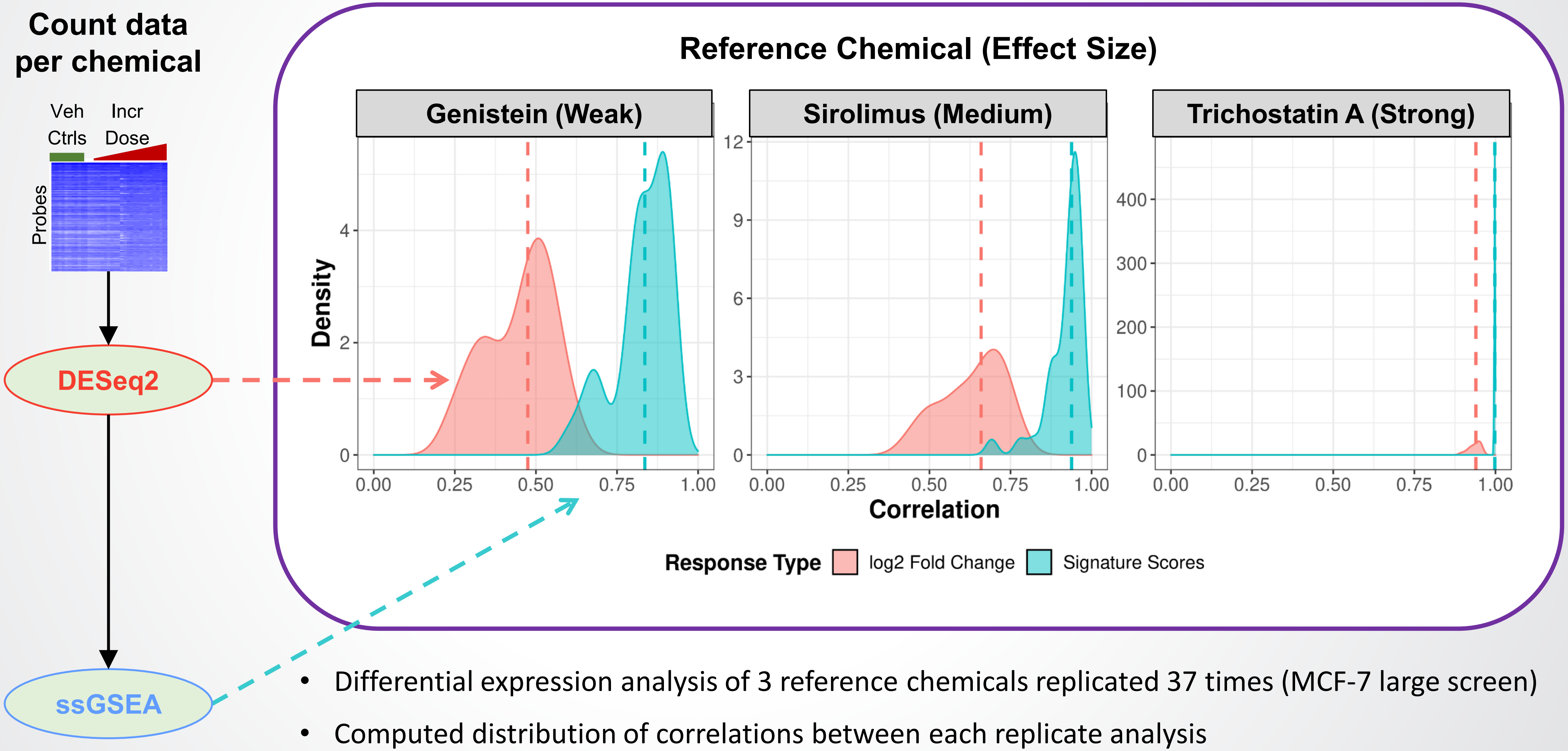
- **Bioplanet** *(Huang, et al. Front Pharmacol 2019)*
- **CMap** *(Subramanian, et al. Cell 2017)*
- **DisGeNET** *(Pinero, et al. Database 2015)*
- **MSigDB** *(Liberzon, et al. Cell Syst 2015)*



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: $\text{mean(gene set log2FC)} - \text{mean(background log2FC)}$

Signature Scoring

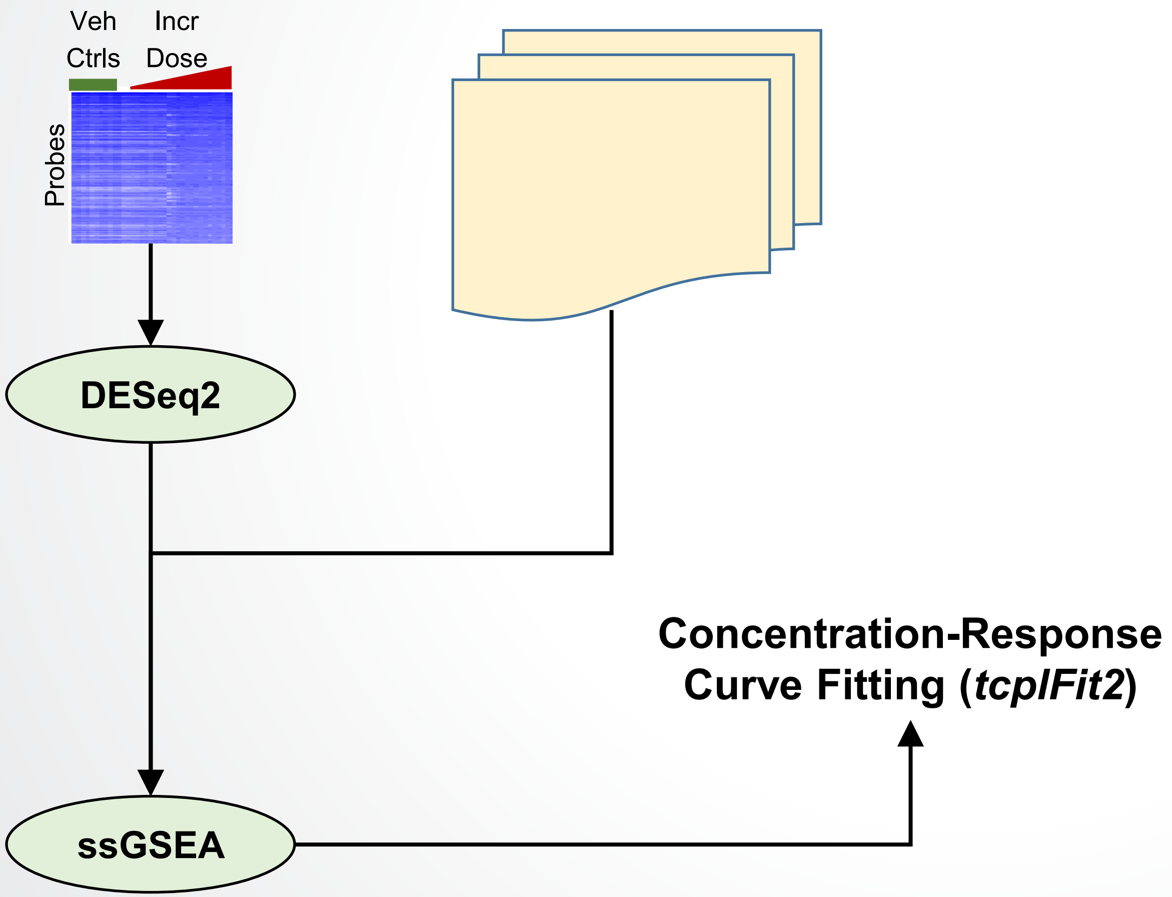


- Differential expression analysis of 3 reference chemicals replicated 37 times (MCF-7 large screen)
- Computed distribution of correlations between each replicate analysis
- **Signature scores have higher reproducibility than fold-changes, especially for weaker effect sizes**

Signature Scoring

Count data per chemical

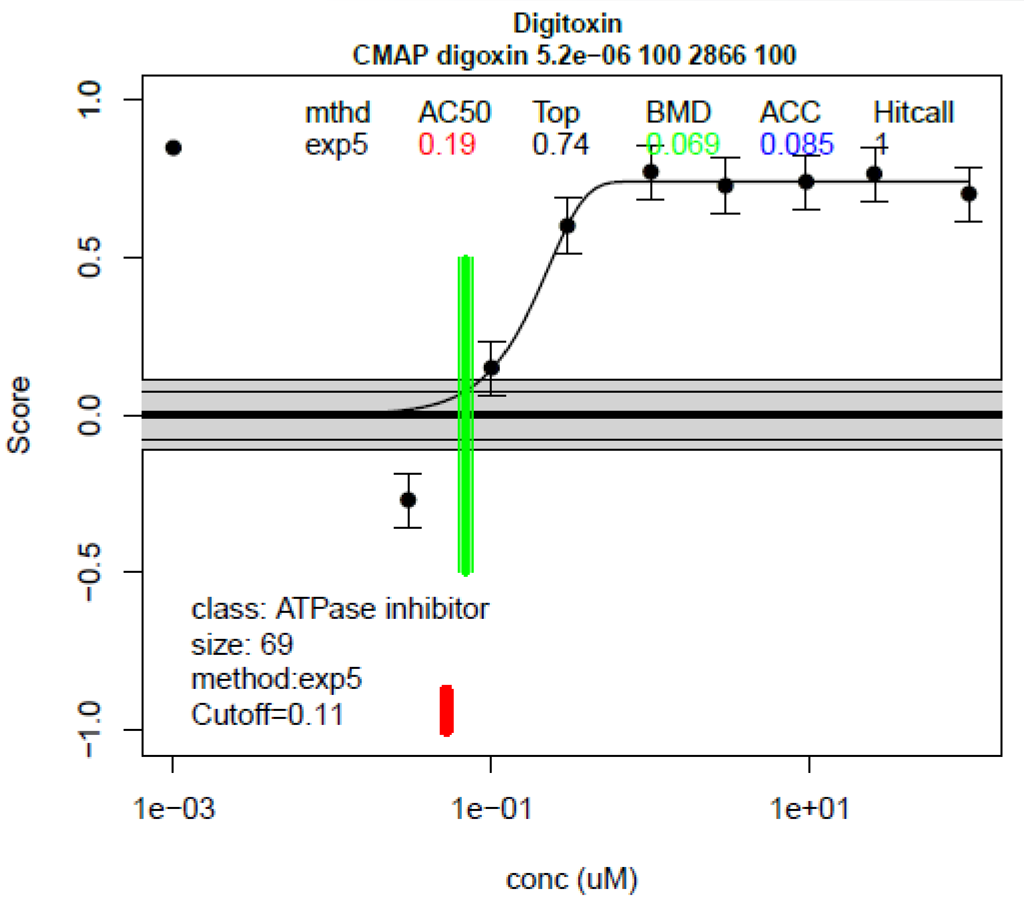
Catalog of signatures with toxicological relevance, annotated for known molecular targets



DESeq2

ssGSEA

Concentration-Response Curve Fitting (*tcplFit2*)



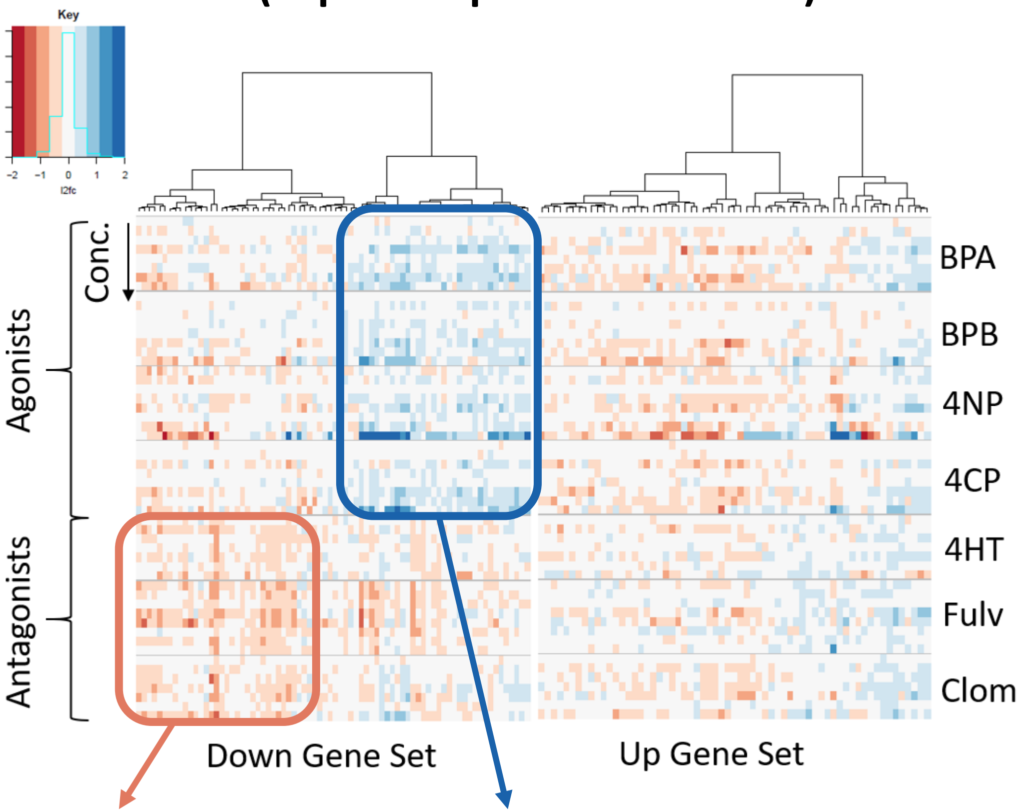
class: ATPase inhibitor
size: 69
method:exp5
Cutoff=0.11

Richard Judson

Directionality of Signature Scores

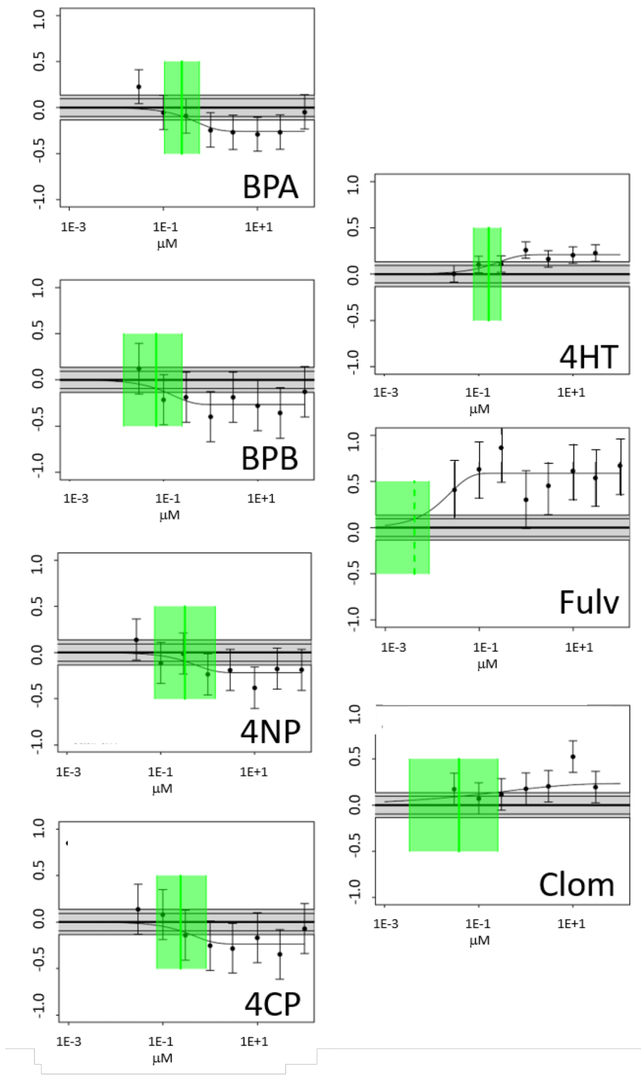
Fulvestrant Signature from CMap (Top 100 Up & Down Genes)

Gene level data are noisy!



The expression of some fulvestrant signature “down” genes goes <u>down</u> following ER antagonist treatment

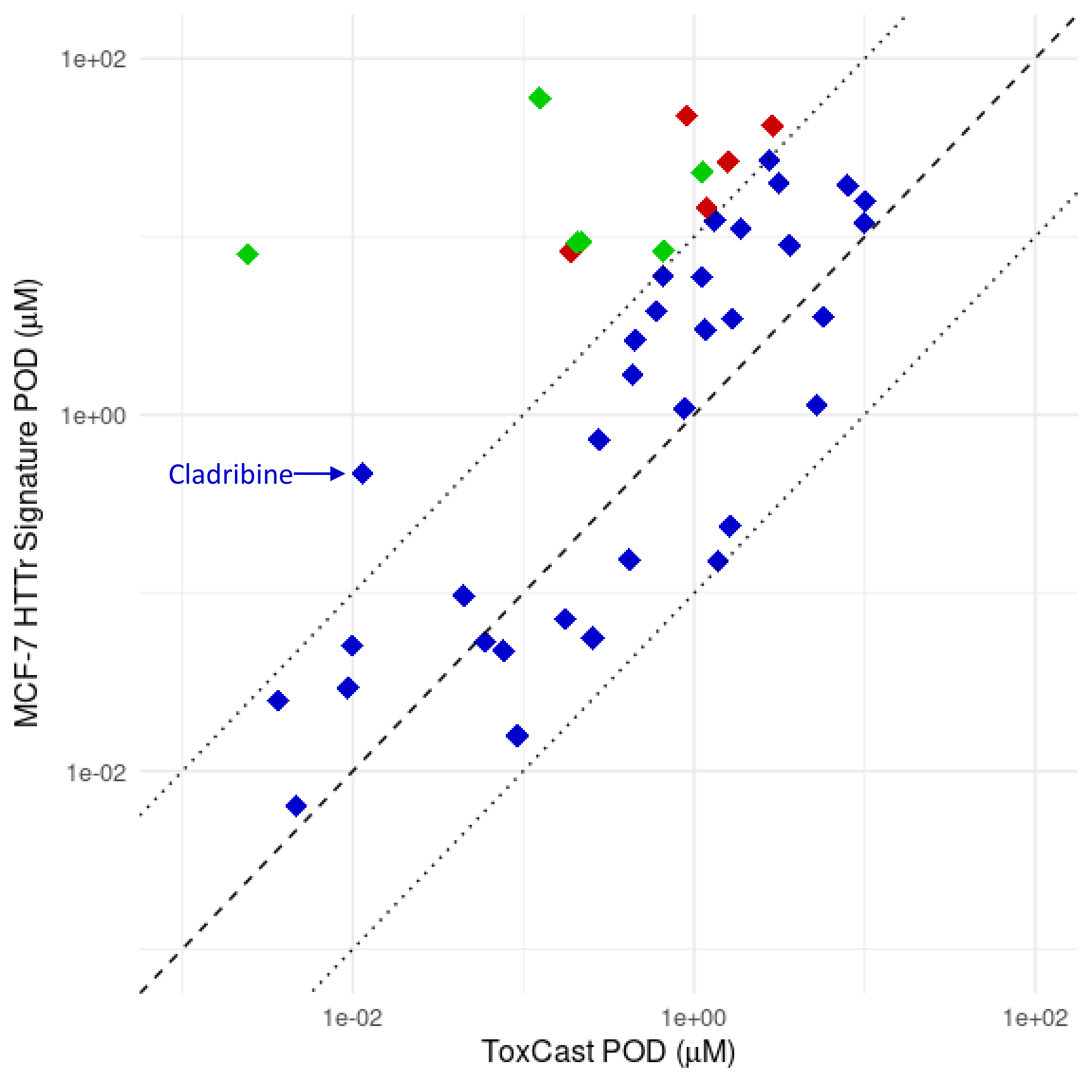
The expression of some fulvestrant signature “down” genes goes <u>up</u> following ER agonist treatment



Signature level results display correct directionality!

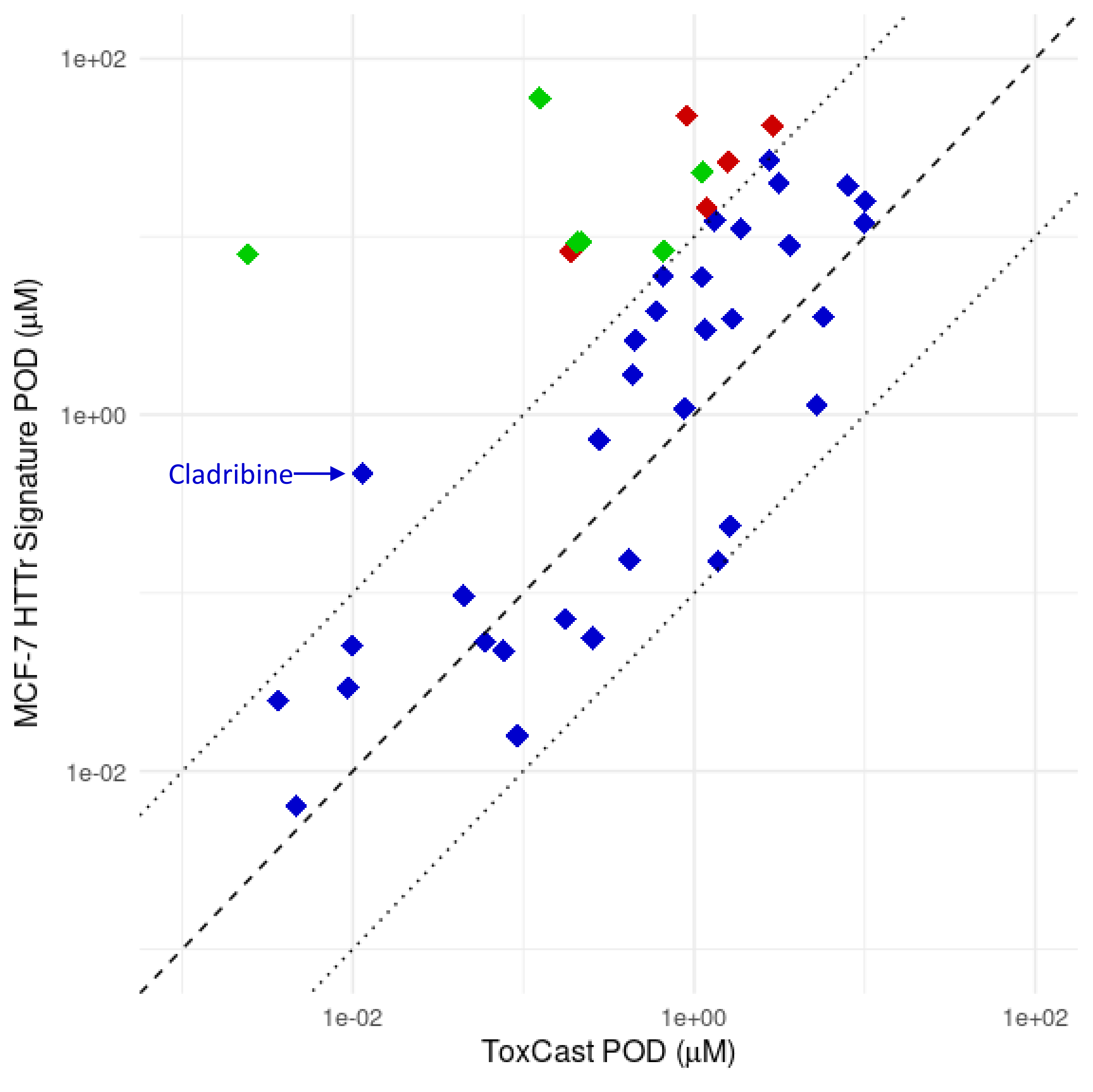
Harrill et al. (2021) DOI: 10.1093/toxsci/kfab009

HTTr MCF-7 Pilot Analysis



- Pilot study of 44 well-characterized chemicals in MCF-7 cells, 6h exposure ***(Harrill, et al. Toxicol Sci, 2021)***
- Compared HTTr-derived PODs to previous ToxCast HTS assay results *(multiple cell types, assays, and exposure lengths)* (***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

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 where most potent assays in ToxCast do not match known target(s)
  - Lovastatin
  - Clofibrate
  - Maneb
  - Lactofen
  - Vinclozolin
- Cladribine (2-chloro-2’-deoxyadenosine) is a DNA synthesis inhibitor
- All other PODs within 1 order of magnitude

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

- Ongoing research efforts focused on:
  - Methods to summarize signature-level/overall PODs from high-dimensional data
  - Inference of underlying mechanism (e.g. Connectivity Mapping)
  - Comparative evaluation of methods on simulated/synthetic conc-response data

Acknowledgements

Questions?

Harrill.Joshua@epa.gov

Everett.Logan@epa.gov

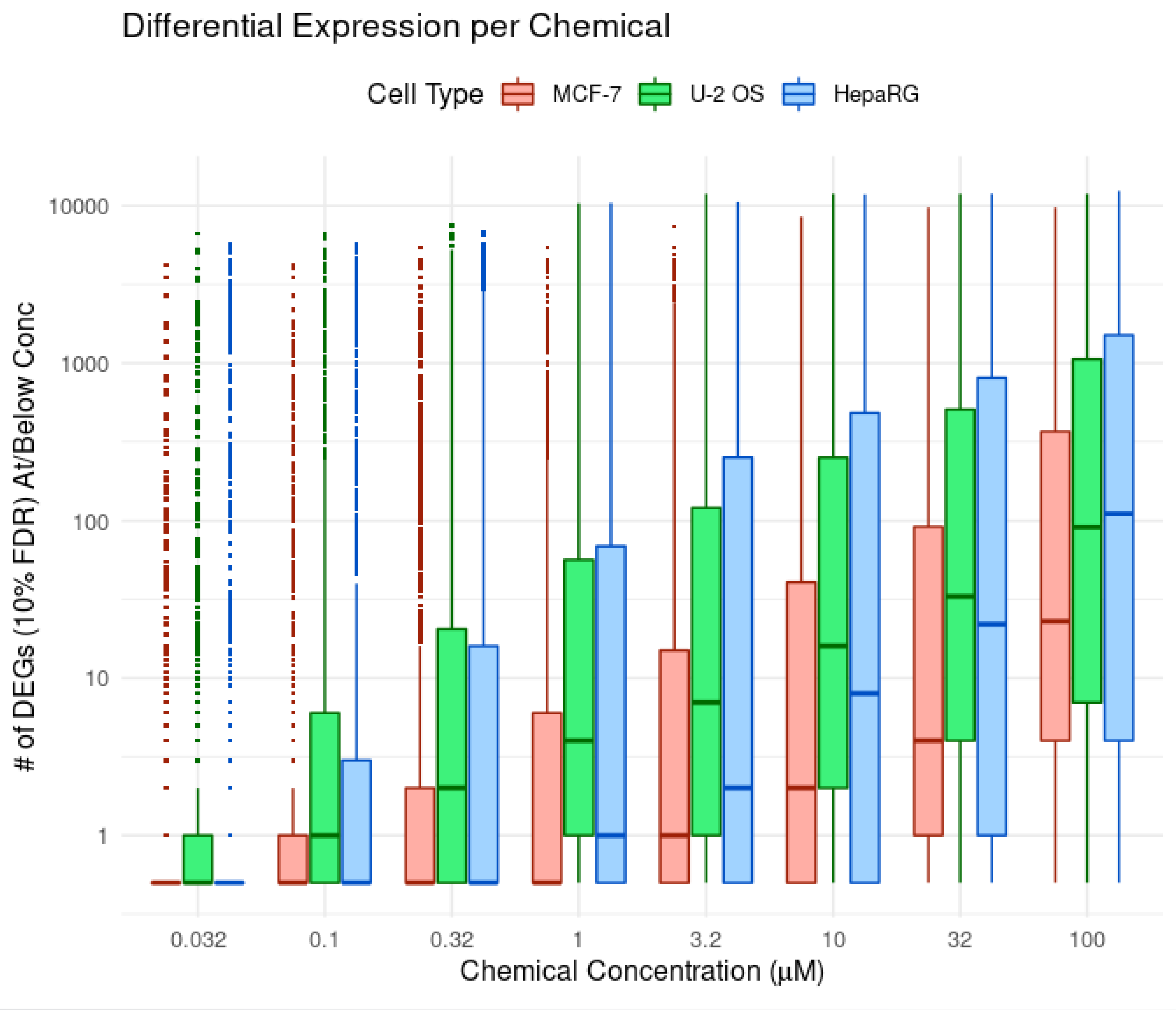
Richard Judson
Imran Shah
Woody Setzer
Derik Haggard
Beena Vallanat
Joseph Bundy
Bryant Chambers
Jesse Rogers
Laura Taylor
Clinton Willis
Thomas Sheffield

CCTE Leadership

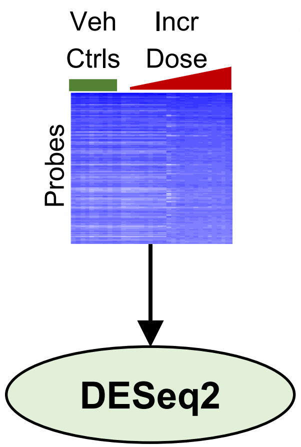
Rusty Thomas
Sid Hunter
Andrew Watkins
John Cowden
Kimberly Slentz-Kesler

Center for Computational Toxicology and Exposure
Biomolecular and Computational Toxicology Division

Global View of Bioactivity



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

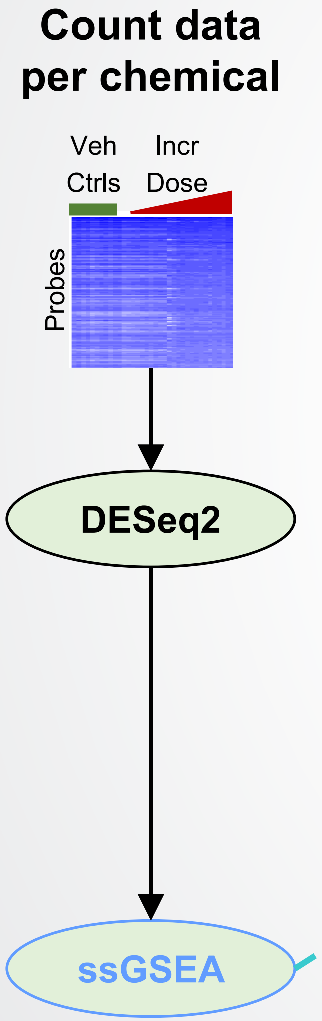


- Statistical model tailored to *-seq data
- Remove plate-level effects
- Smooths noise across depth & expression levels

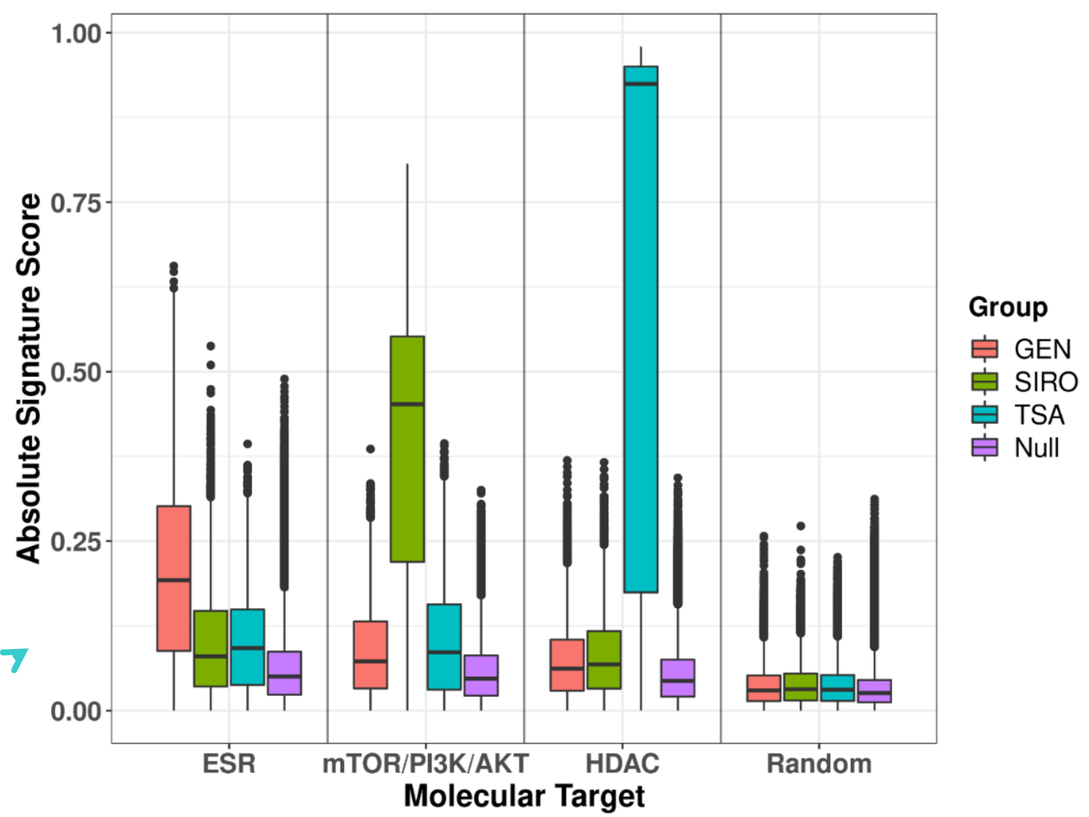
(Love, et al. Genome Biol 2014)

- Each boxplot shows distribution of Differentially Expressed Gene (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

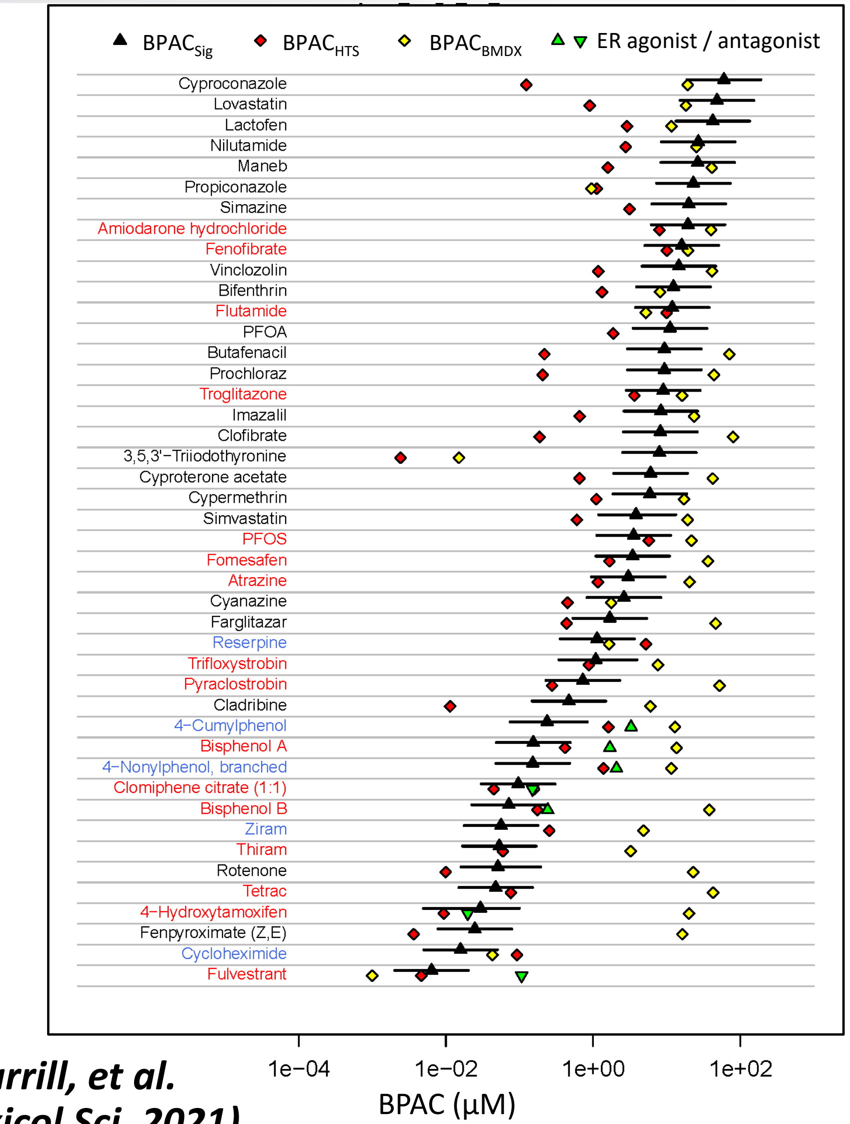


Reference Chemical Signature Specificity



- **Higher scores for signatures corresponding to known targets of each reference chemical**

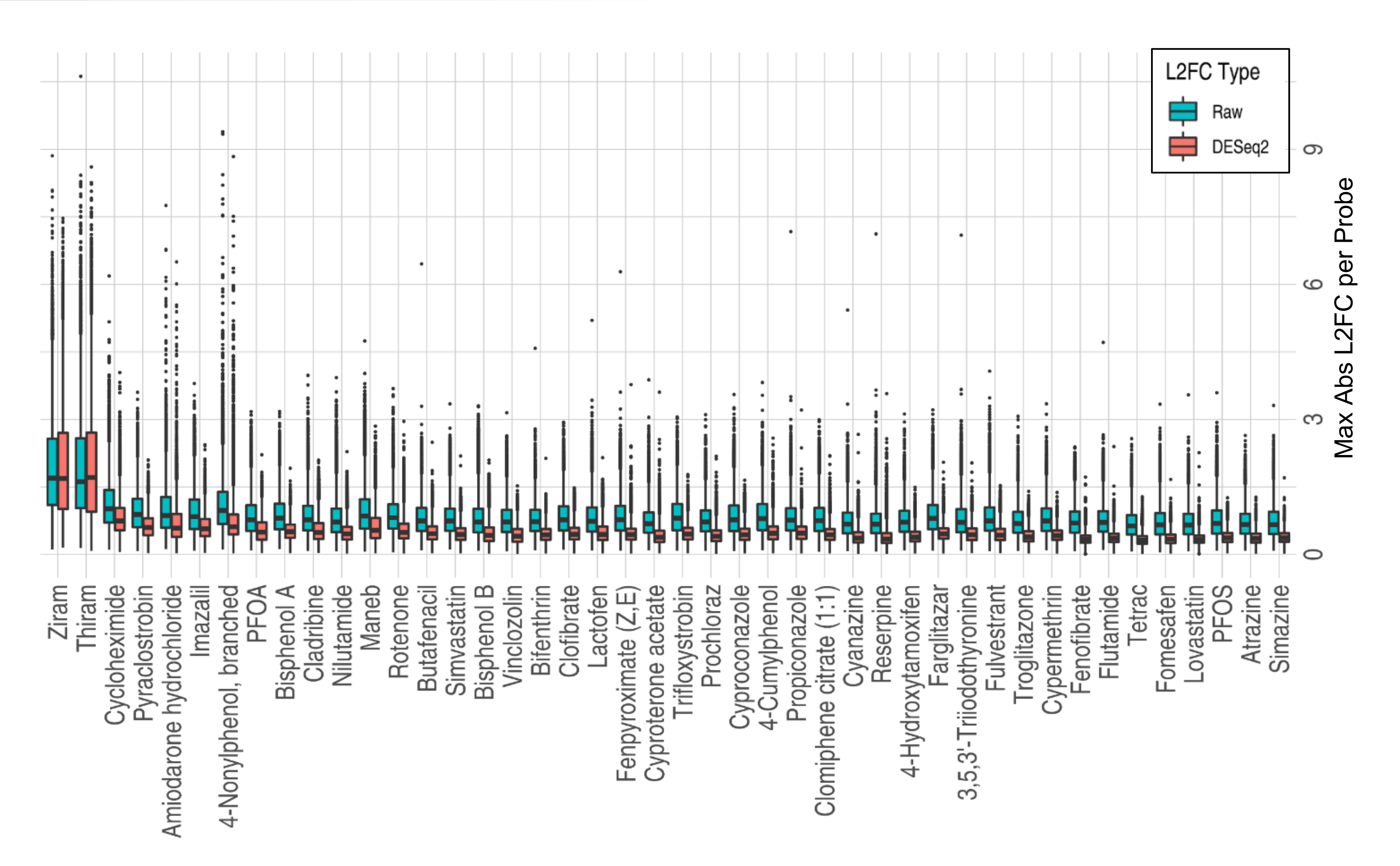
HTTr MCF-7 Pilot Analysis



(Harrill, et al. Toxicol Sci, 2021)

- Also calculated BPAC/PODs using NTP approach with BMDExpress2 ***(NTP Research Report 5, 2018; Phillips, et al. 2019)***
- BPAC$_{BMDX}$ (◇) tended to be higher and less concordant with ToxCast PODs
  - Poor signal:noise at gene-level is likely cause
- ***We continue to use BMDExpress for other transcriptomics applications and continue to explore this issue***

HTTr MCF-7 Pilot Analysis

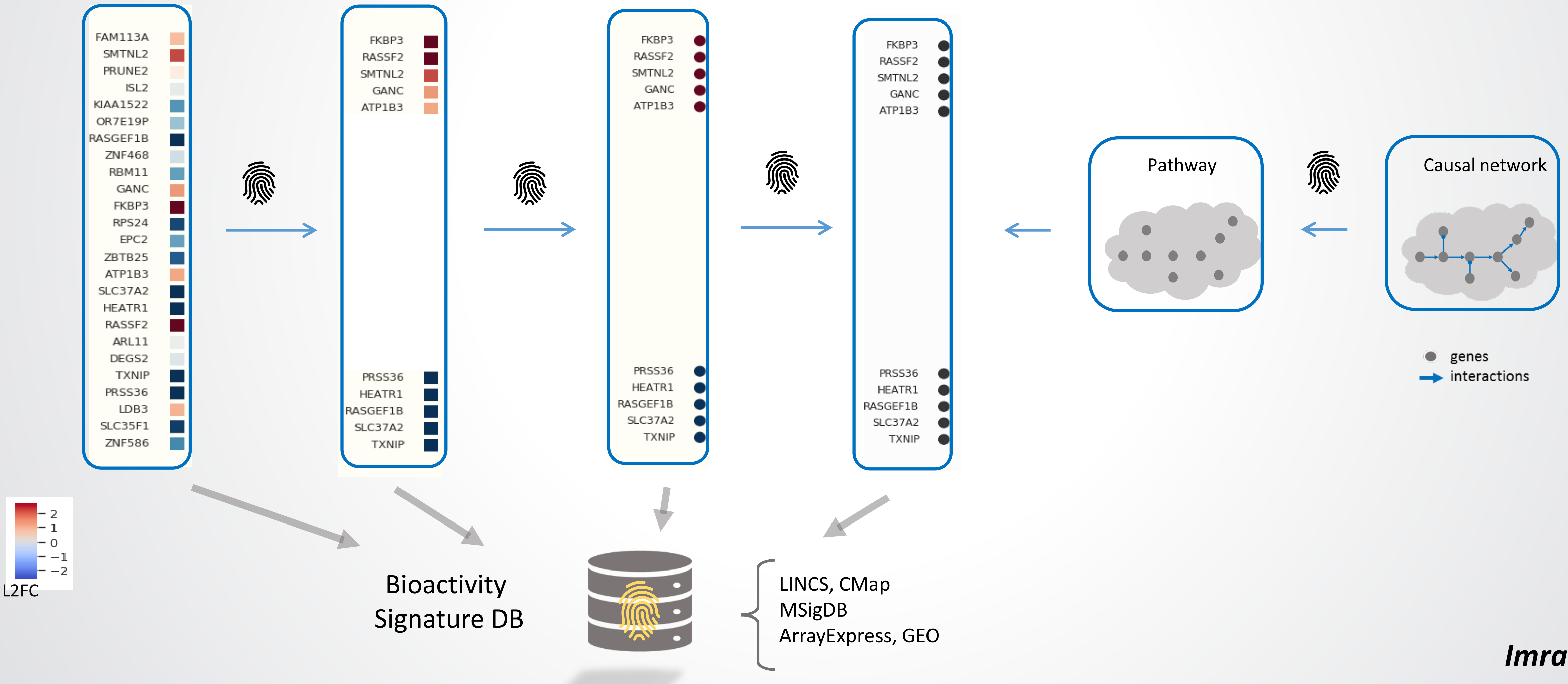


(Harrill, et al. Toxicol Sci, 2021)

- Majority of differential expression is weak (2-4x) for most chemical treatments
  - DESeq2 dampens these further in most cases
- Consistent with previous studies using MCF-7 cells
- Lower effect size results in lower signal:noise
- Signature-level scores (e.g. GSEA) may perform better than probe-level when this is the case

“Fingerprinting” bioactivity via gene signatures

Reference Profiles | Top N Up-/Dn-genes | Directional Signature | Signature | Pathway | Network

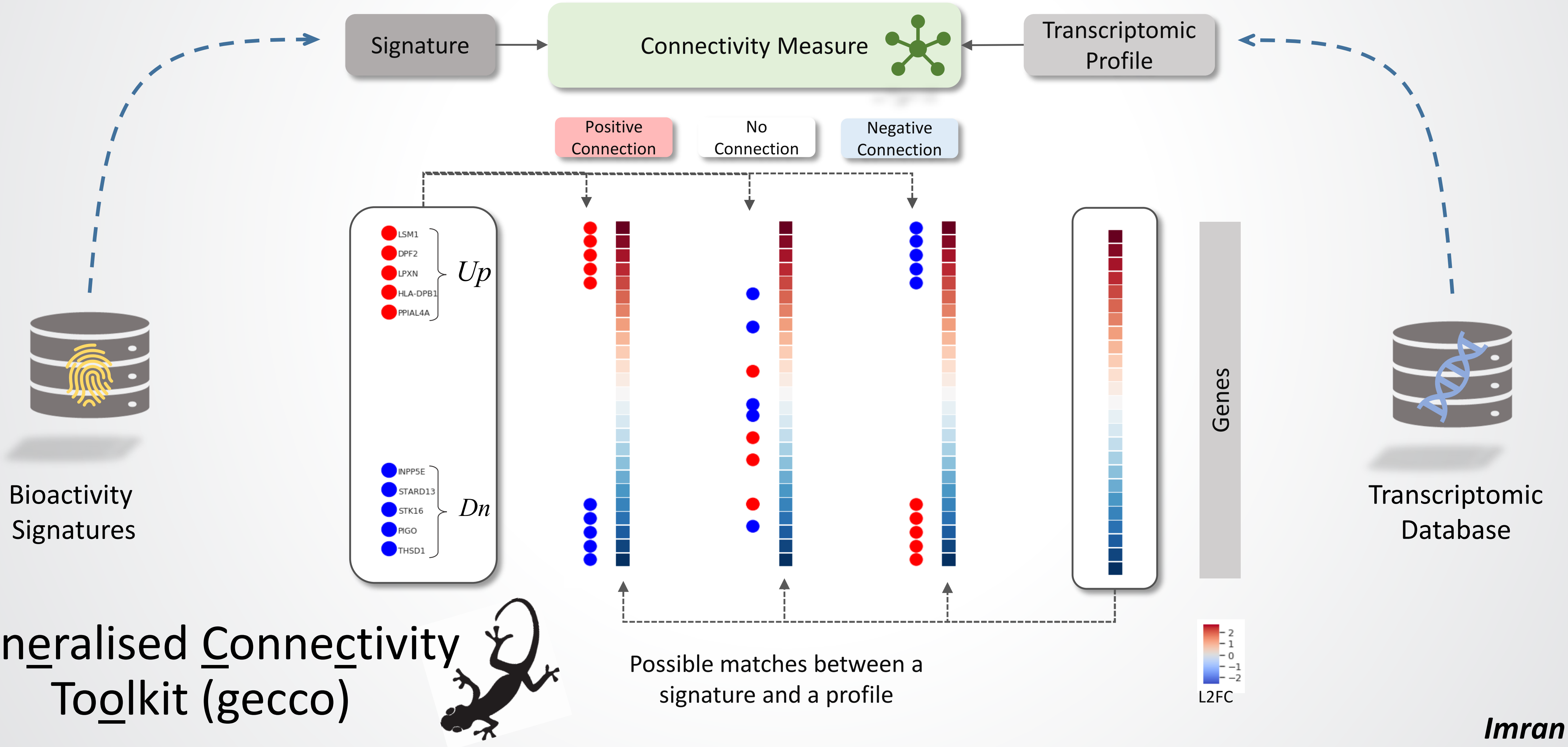


Bioactivity Signature DB

LINCS, CMap
MSigDB
ArrayExpress, GEO

Imran Shah

Connectivity-mapping with gene signatures



Possible matches between a signature and a profile

Generalised Connectivity Toolkit (gecco)

Imran Shah