EPA United States Environmental Protection Agency

High-Throughput Transcriptomics Screening of Environmental Chemicals using TempO-Seq

Joshua A. Harrill

USEPA Center for Computational Toxicology and Exposure (CCTE)

Office of Research and Development

US FDA, National Center for Toxicological Research, Division of Systems Biology
June 11th, 2021

EPA United States Environmental Protection Agency

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.

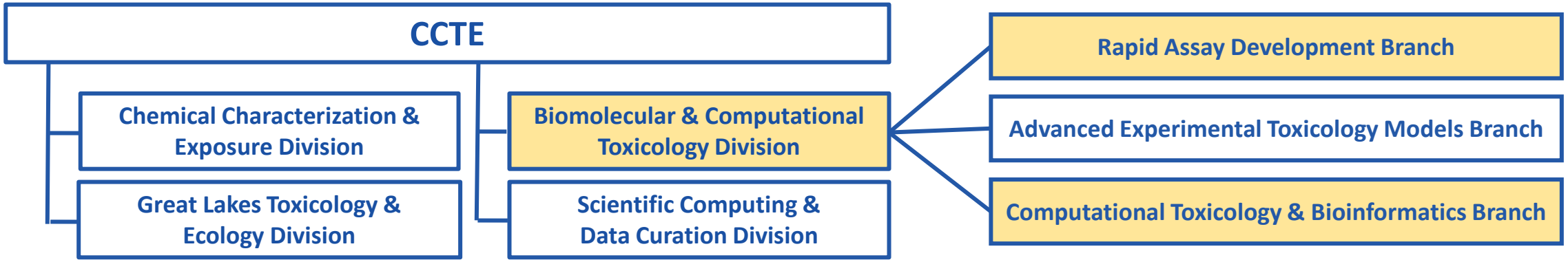
Outline

- **Background**
  - Who is CCTE?
  - What Does CCTE Do?
  - Blueprint for Computational Toxicology at USEPA
- **High Throughput Transcriptomics (HTTr)**
  - Overview of TempO-Seq Technology
  - Pilot Study in MCF-7 Cells
  - Signature Concentration-Response Modeling
  - HTTr Data Landscape
- **Applications for Molecular PODs From HTP NAMs**
- **Summary and Future Directions**

Who is CCTE?

Center for Computational Toxicology and Exposure (CCTE)

A research organization at US EPA Office of Research and Development tasked with **developing** and **applying** cutting edge innovations in methods to rapidly evaluate chemical toxicity, transport and exposure to people and environments.



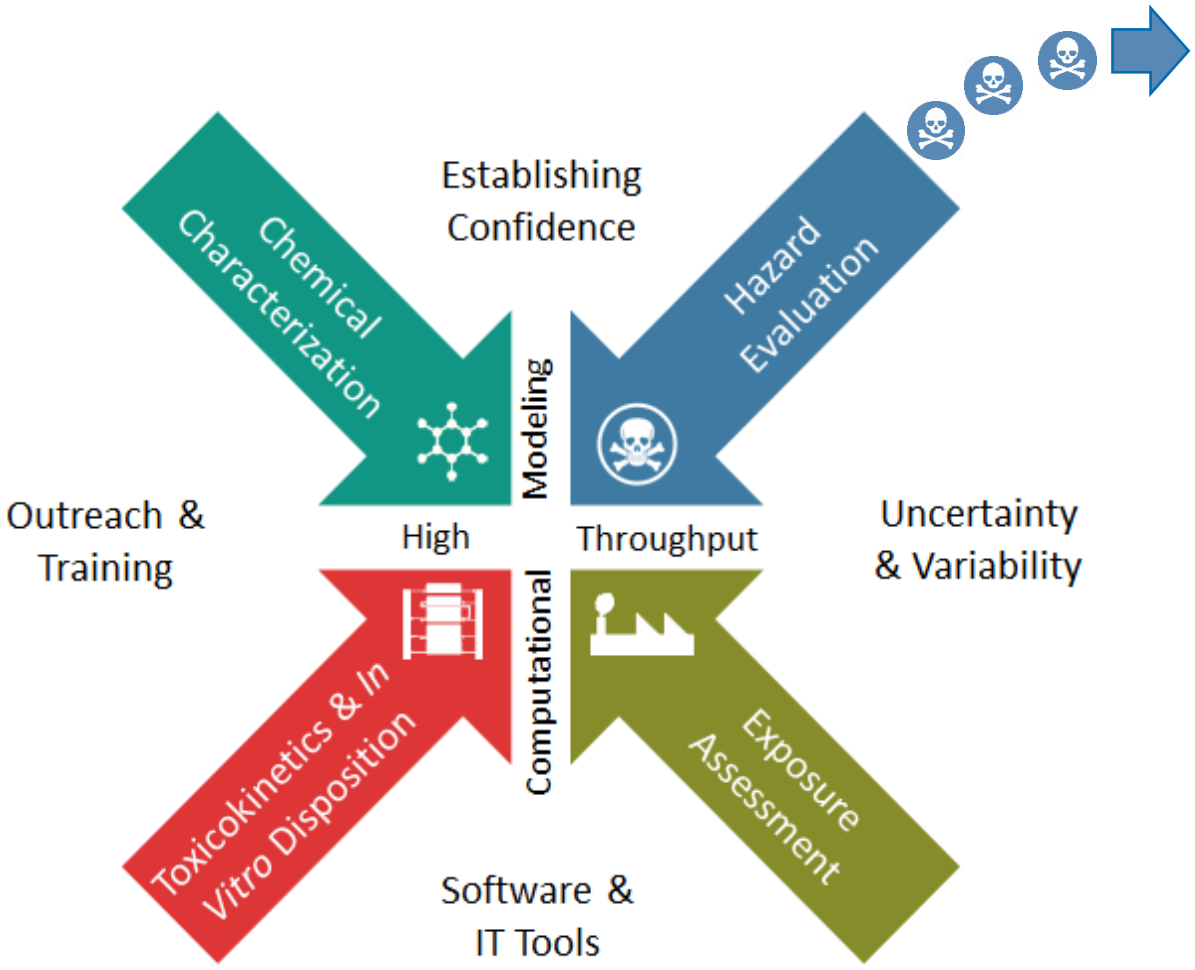
Rapid Assay Development Branch (RADB)

Develops the next generation of **high-throughput toxicity assays** to comprehensively cover the potential **molecular and phenotypic responses** resulting from chemical exposure and **fill gaps** in biological pathways and processes not addressed using existing assays.

Computational Toxicology & Bioinformatics Branch (CTBB)

Utilizes **computational and informatics approaches** to analyze and integrate data from **high-throughput toxicity assays**, complex culture models, alternative species, toxicokinetics and chemistry to **predict adverse effects of chemicals** in human and animal models.

Computational Toxicology Research Areas



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

- **ToxCast:** Used 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"

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

- **New Strategy for Hazard Evaluation:** Improve efficiency and increase biological coverage by using 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.

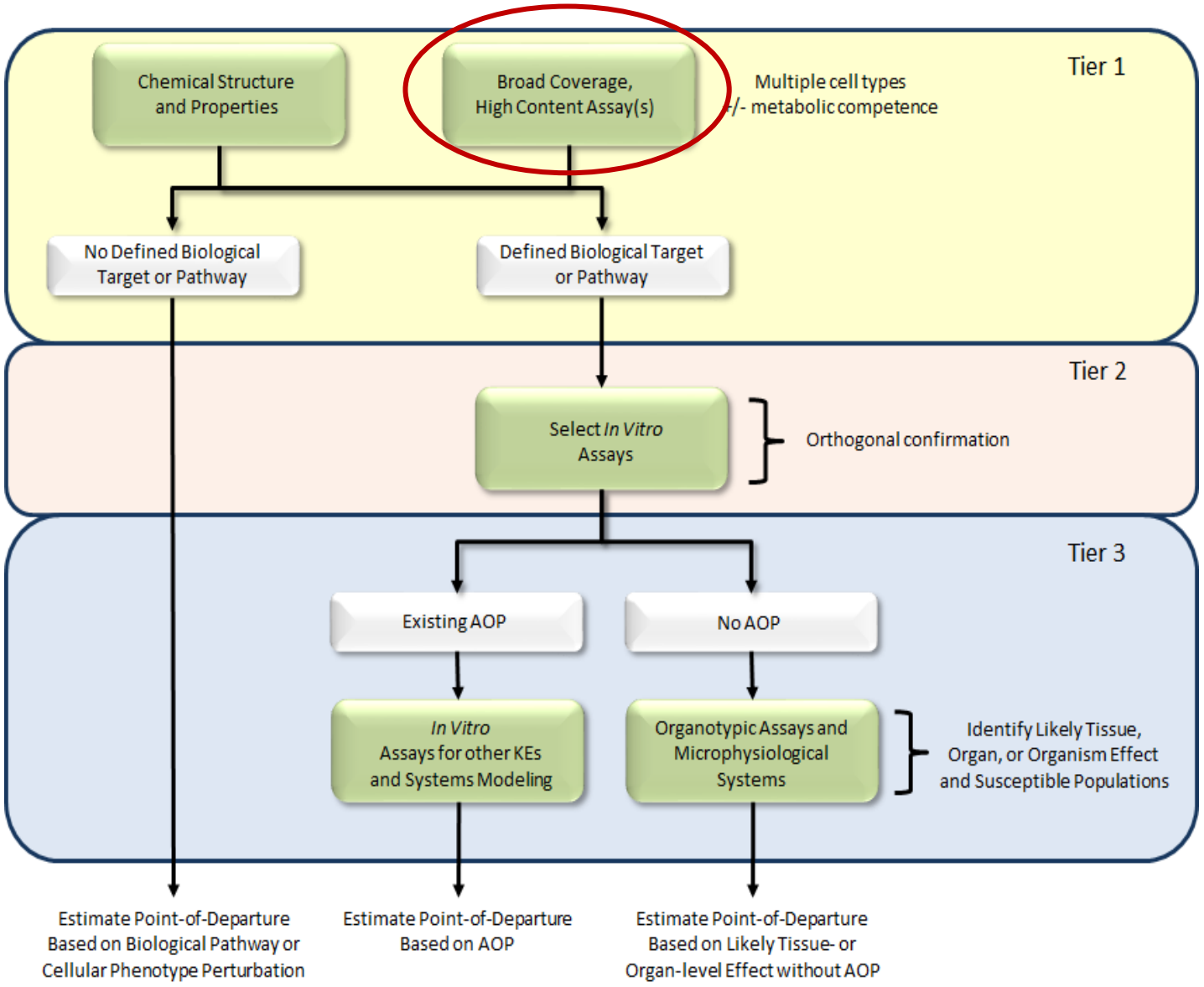
Tiered Hazard Evaluation Approach (1)

- **New Approach Methodologies (NAMs)** are any technology, methodology, approach or combination thereof that can be used to provide information on chemical hazard and risk that avoids the use of intact animals.

- NAMs are a potential means to **reduce** the use of animals in toxicity testing and **accelerate** the pace of chemical risk assessment.

- US EPA CompTox Blueprint advocates the use of **high throughput profiling (HTP) assays** as the first tier in a NAMs-based hazard evaluation approach.

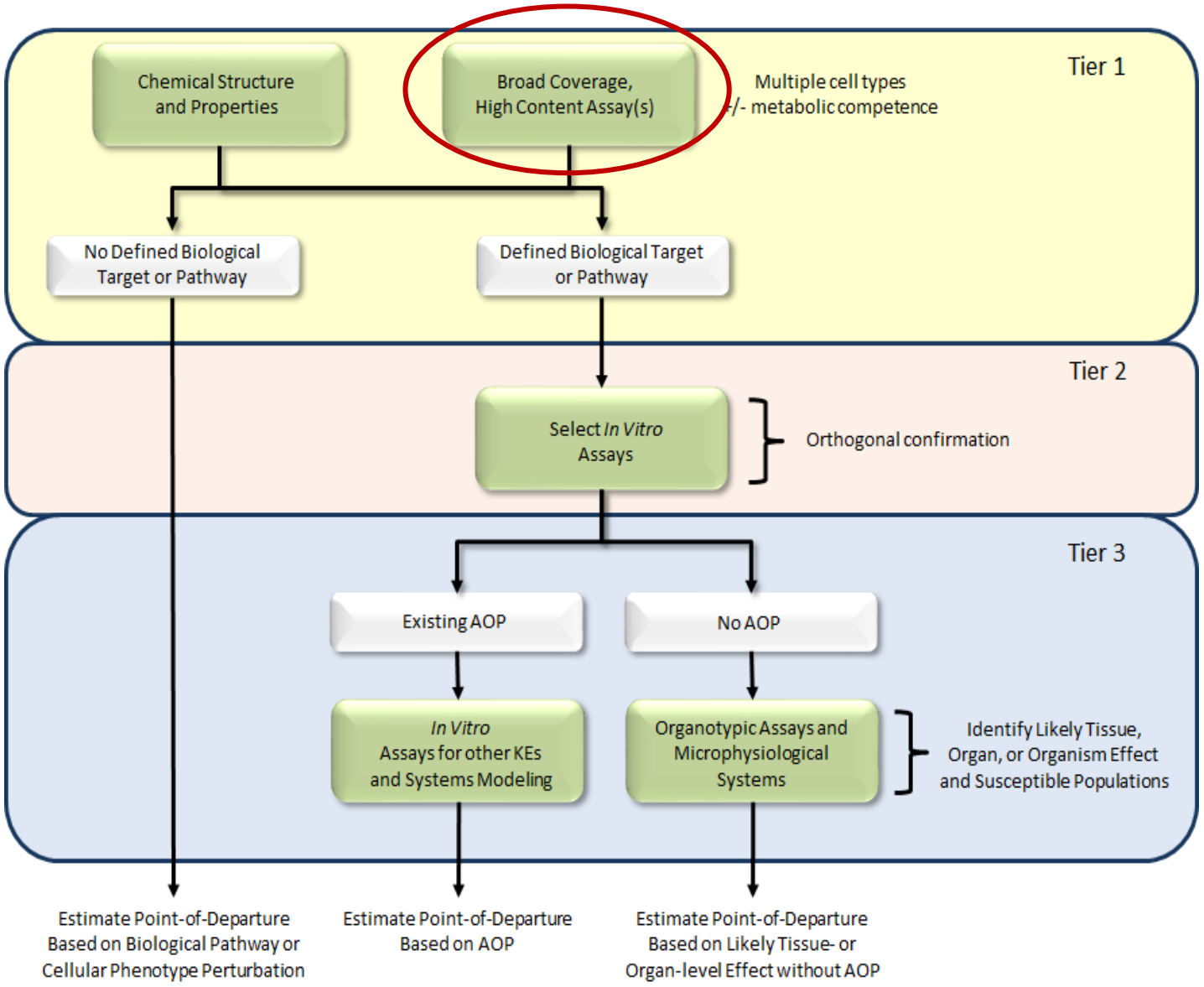
- **HTP assay criteria:**
  1. Yield bioactivity profiles that can be used for **potency estimation**, **mechanistic prediction** and evaluation of **chemical similarity**.
  2. Compatible with multiple human-derived culture models.
  3. Concentration-response screening mode.
  4. Cost-effective.



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

Tiered Hazard Evaluation Approach (2)

- To date, EPA has identified and implemented two HTP assays that meet this criteria.
- **High-Throughput Transcriptomics [HTTr]**
  - Whole Transcriptome TempO-Seq
- **High-Throughput Phenotypic Profiling [HTPP]**
  - Cell Painting
- Both methods are **complementary** to each other and can be used in many different human-derived cell types.
- EPA has established scalable laboratory and bioinformatics workflows for each assay.



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

US EPA HTTr Publications

https://doi.org/10.1016/j.cotox.2019.05.004

ELSEVIER

Current Opinion in Toxicology

Considerations for strategic use of high-throughput transcriptomics chemical screening data in regulatory decisions

Joshua Harrill[1], Imran Shah[1], R. Woodrow Setzer[1], Derik Haggard[2], Scott Auerbach[3], Richard Judson[1] and Russell S. Thomas[1]

OXFORD

SOT | Society of Toxicology

academic.oup.com/toxsci

TOXICOLOGICAL SCIENCES, 2021, 1–22

doi: 10.1093/toxsci/kfab009
Advance Access Publication Date: 4 February 2021
Research Article

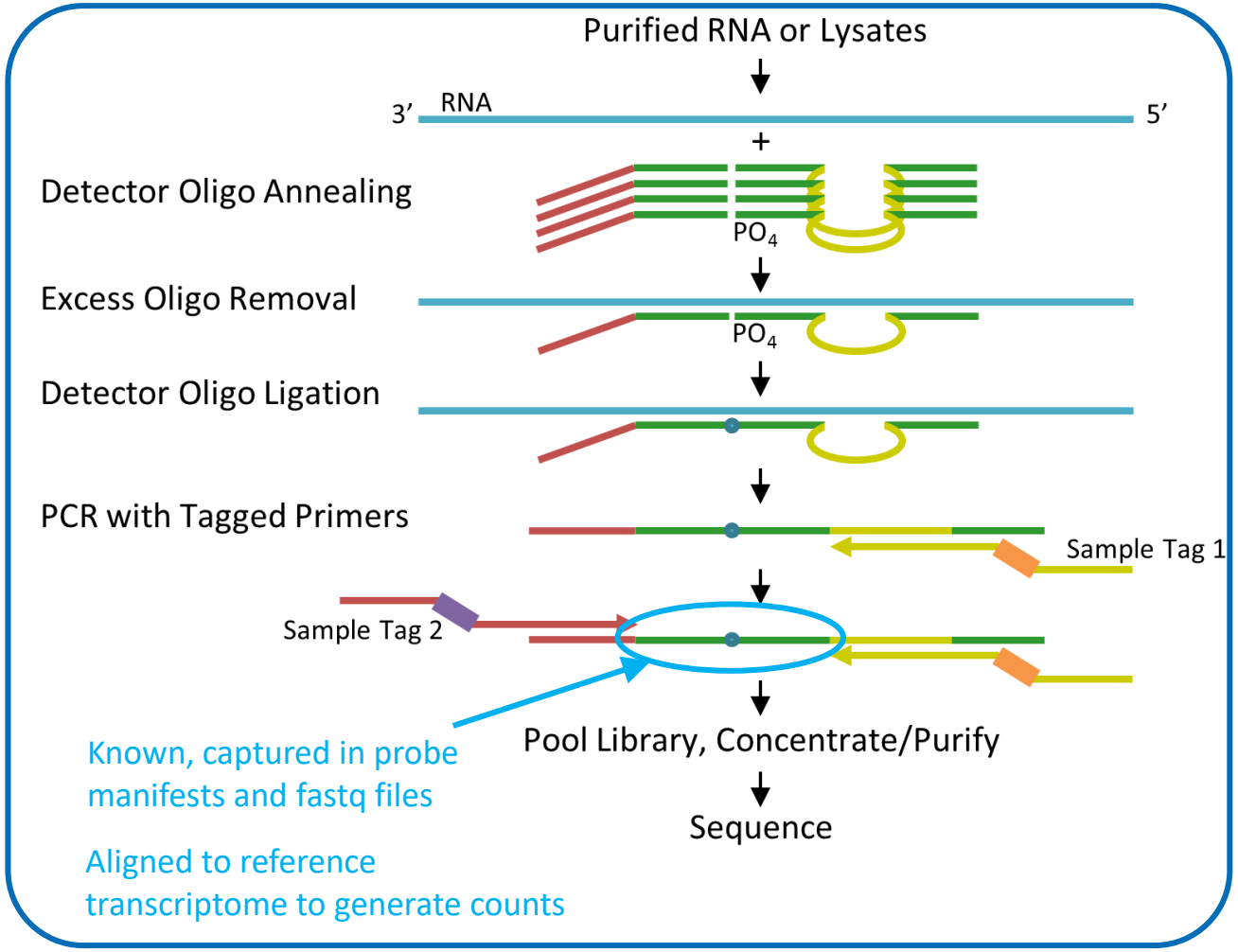
High-Throughput Transcriptomics Platform for Screening Environmental Chemicals

Joshua A. Harrill,*,1 Logan J. Everett,* Derik E. Haggard,*,† Thomas Sheffield,*,† Joseph L. Bundy,* Clinton M. Willis,*,‡ Russell S. Thomas,* Imran Shah,* and Richard S. Judson*

Templated Oligo with Sequencing Readout (TempO-Seq)

- The **TempO-Seq** human whole transcriptome assay measures the expression of greater than 20,000 transcripts.
- Requires only picogram amounts of total RNA per sample.
- Compatible with purified RNA samples or **cell lysates**.
- Lysates are barcoded according to sample identity and combined in a single library for sequencing using industry standard instrumentation.
- Scalable, targeted assay:
  - 1) specifically measures transcripts of interest
  - 2) ~50-bp reads for all genes
  - 3) requires less flow cell capacity than RNA-Seq

TempO-Seq Assay Illustration



Yeakley, et al. PLoS ONE 2017

EPA United States Environmental Protection Agency

MCF-7 Pilot Experimental Design

Parameter	Multiplier	Notes
Cell Type(s)	1	MCF7
Culture Condition	1	DMEM + 10% HI-FBS
Chemicals	44	ToxCast chemicals with mechanistic variety and some redundancy.
Time Points:	1	6 hours
Assay Formats:	2	High-Throughput Transcriptomics Cell Viability
Concentrations:	8	3.5 $\log_{10}$ units; semi $\log_{10}$ spacing
Biological Replicates:	3	Independent cultures

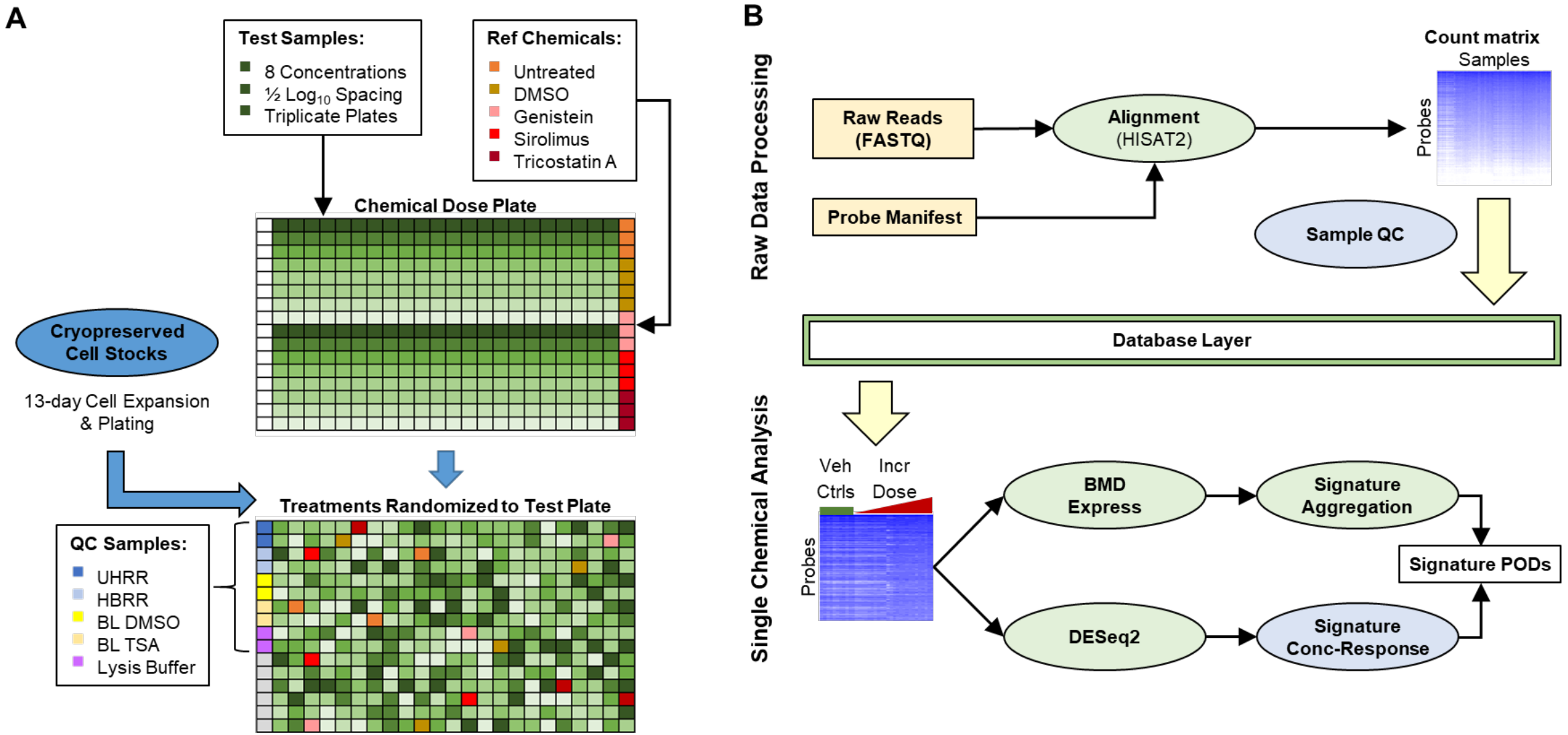
MCF-7 Pilot Chemical List

Table 1. Chemicals Used in the Study

Name	Target Annotation
Cyproterone acetate	AR antagonist
Flutamide	AR antagonist
Nilutamide	AR antagonist
Vinclozolin	AR antagonist
Amiodarone hydrochlorid	Blocks myocardial calcium, potassium and sodium channels
Cladribine	DNA synthesis inhibitor
4-Cumylphenol	ER agonist
4-Nonylphenol, branched	ER agonist
Bisphenol A	ER agonist
Bisphenol B	ER agonist
4-Hydroxytamoxifen	ER antagonist
Clomiphene citrate (1:1)	ER antagonist
Fulvestrant	ER antagonist
Cyproconazole	Ergosterol-biosynthesis inhibitor. Pan-cyp inhibitor
Imazalil	Ergosterol-biosynthesis inhibitor. Pan-cyp inhibitor
Prochloraz	Ergosterol-biosynthesis inhibitor. Pan-cyp inhibitor
Propiconazole	Ergosterol-biosynthesis inhibitor. Pan-cyp inhibitor
Atrazine	Herbicide, photosystem II inhibitor
Cyanazine	Herbicide, photosystem II inhibitor
Simazine	Herbicide, photosystem II inhibitor
Butafenacil	Herbicide, PPO inhibition
Fomesafen	Herbicide, PPO inhibition
Lactofen	Herbicide, PPO inhibition
Lovastatin	HMGCR inhibitor
Simvastatin	HMGCR inhibitor
Maneb	Inhibition of metal-dependent and sulfhydryl enzyme systems
Thiram	Inhibition of metal-dependent and sulfhydryl enzyme systems
Ziram	Inhibition of metal-dependent and sulfhydryl enzyme systems
Reserpine	Inhibition of the ATP/Mg2+ pump
Rotenone	Mitochondria (complex I inhibitor)
Pyraclostrobin	Mitochondria (complex III inhibitor)
Trifloxystrobin	Mitochondria (complex III inhibitor)
Fenpyroximate (Z, E)	Mitochondrial electron transport inhibitor
Clofibrate	PPARα agonist, upregulates extrahepatic lipoprotein lipase
Fenofibrate	PPARα agonist, upregulates extrahepatic lipoprotein lipase
Farglitazar	PPARγ agonist
Perfluorooctanoic acid (PFOA)	PPARγ, PPARα agonist
Perfluorooctanesulfonic acid (PFOS)	PPARγ, PPARα agonist
Troglitazone	PPARγ, PPARα agonist
Cycloheximide	Protein synthesis inhibitor
Bifenthrin	Sodium channel modulator
Cypermethrin	Sodium channel modulator
Tetrac	T4 synthesis inhibitor
3,5,3′-triiodothyronine	THR agonist

Harrill et al., Toxicol Sci. 2021 Apr 27;181(1):68-89.

HTTr Experimental Design and Bioinformatics Workflow



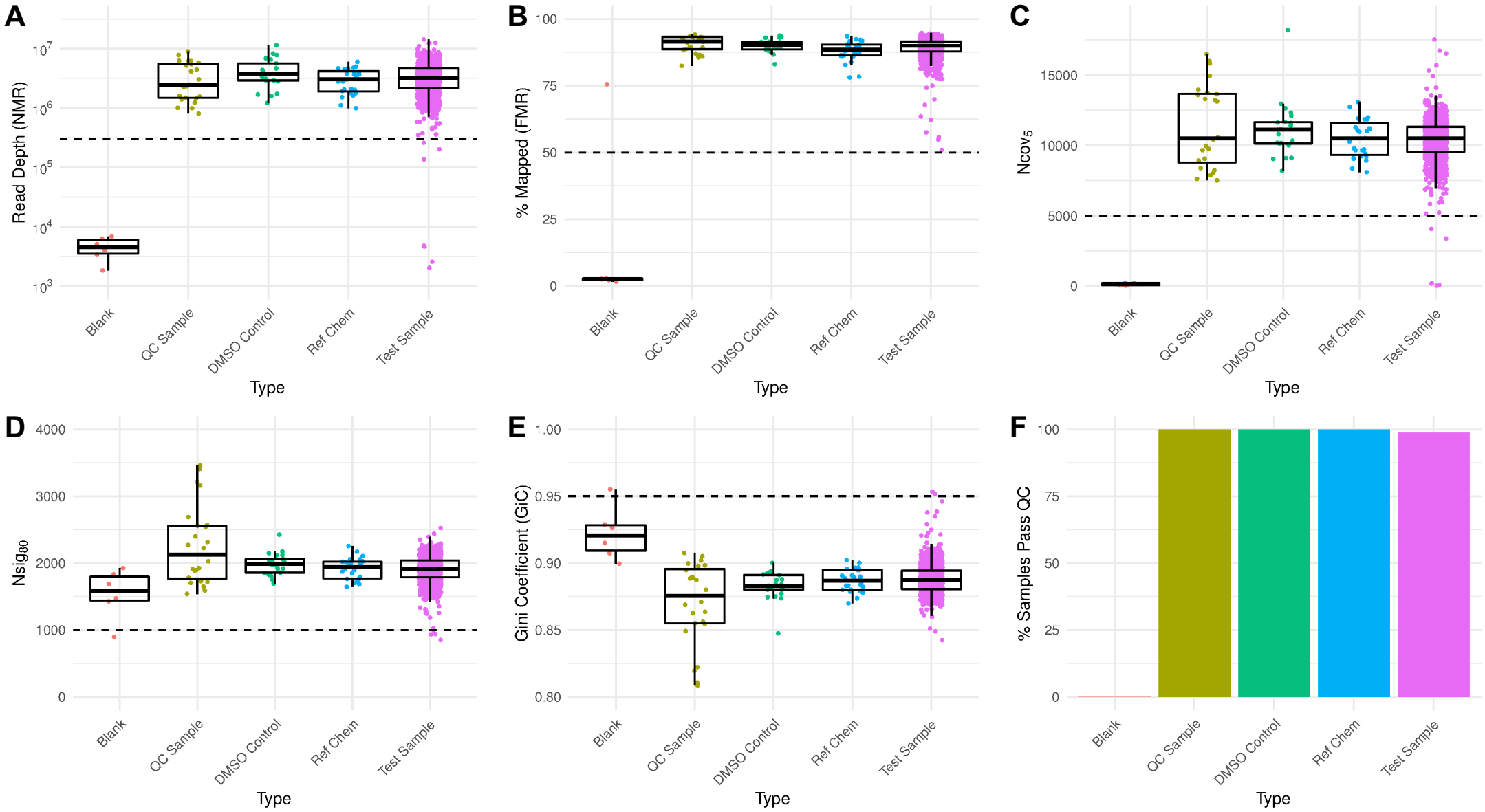
Harrill et al., Toxicol Sci. 2021 Apr 27;181(1):68-89.

HTTr Quality Control Criteria

Abbreviation	Description	Threshold	Additional Information
FrVC	Fraction of viable cells (PI-negative or Casp3/7-negative)	Reject < 50%	Highly cytotoxic conditions no longer represent molecular initiating event
NMR	Number of mapped reads, defined as sum of total read counts summed over all detected probes	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)
$Nsig_{80}$	The number of probes capturing the top 80% of signal in a sample	Reject < 1,000	
GiC	Gini coefficient computed for each sample based on the distribution of raw counts for all probes including those with 0 aligned reads	Reject > 0.95	

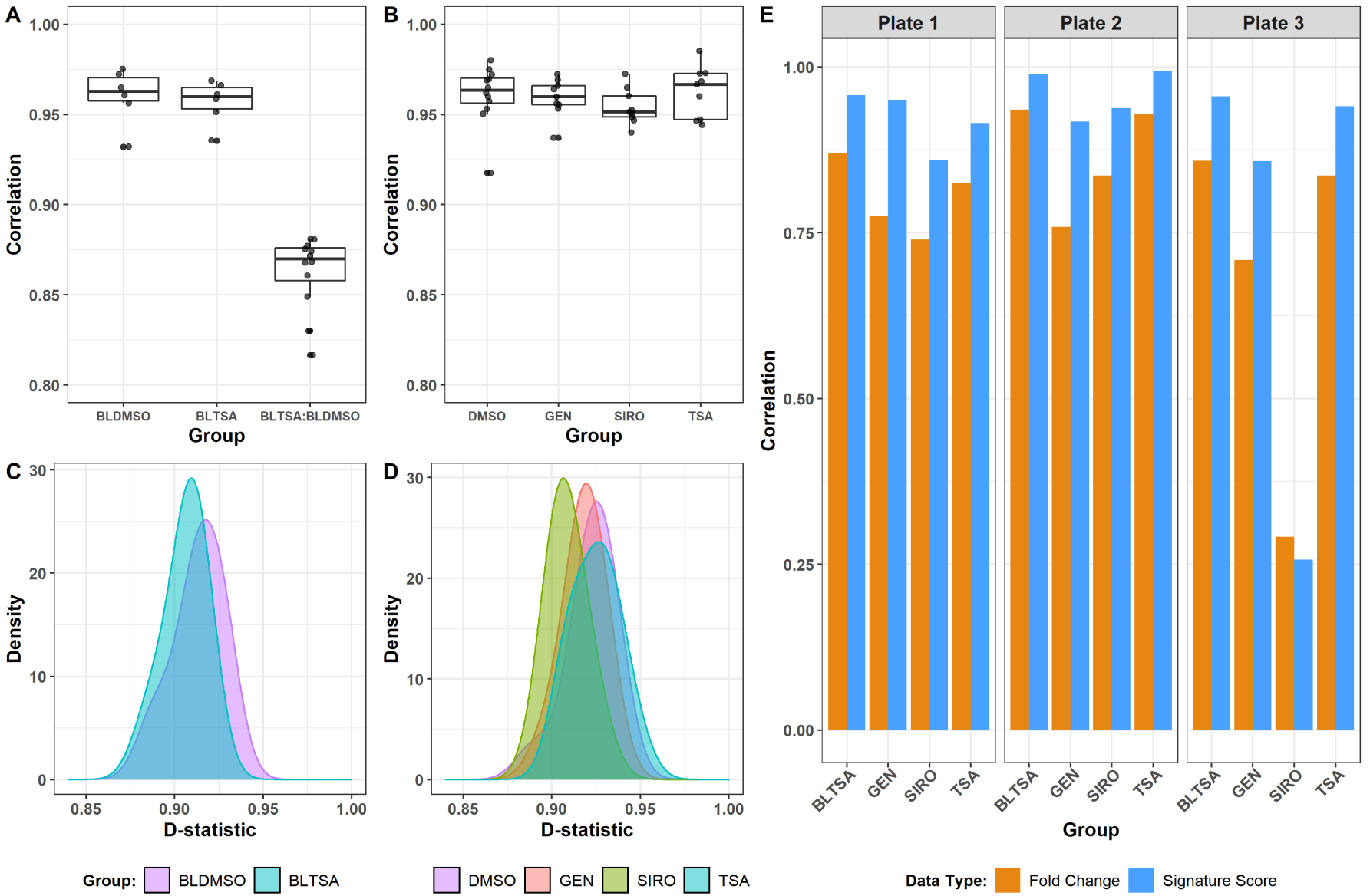
Harrill et al., Toxicol Sci. 2021 Apr 27;181(1):68-89.

HTTr Sample Quality Assessment (1)



Harrill et al., Toxicol Sci. 2021 Apr 27;181(1):68-89.

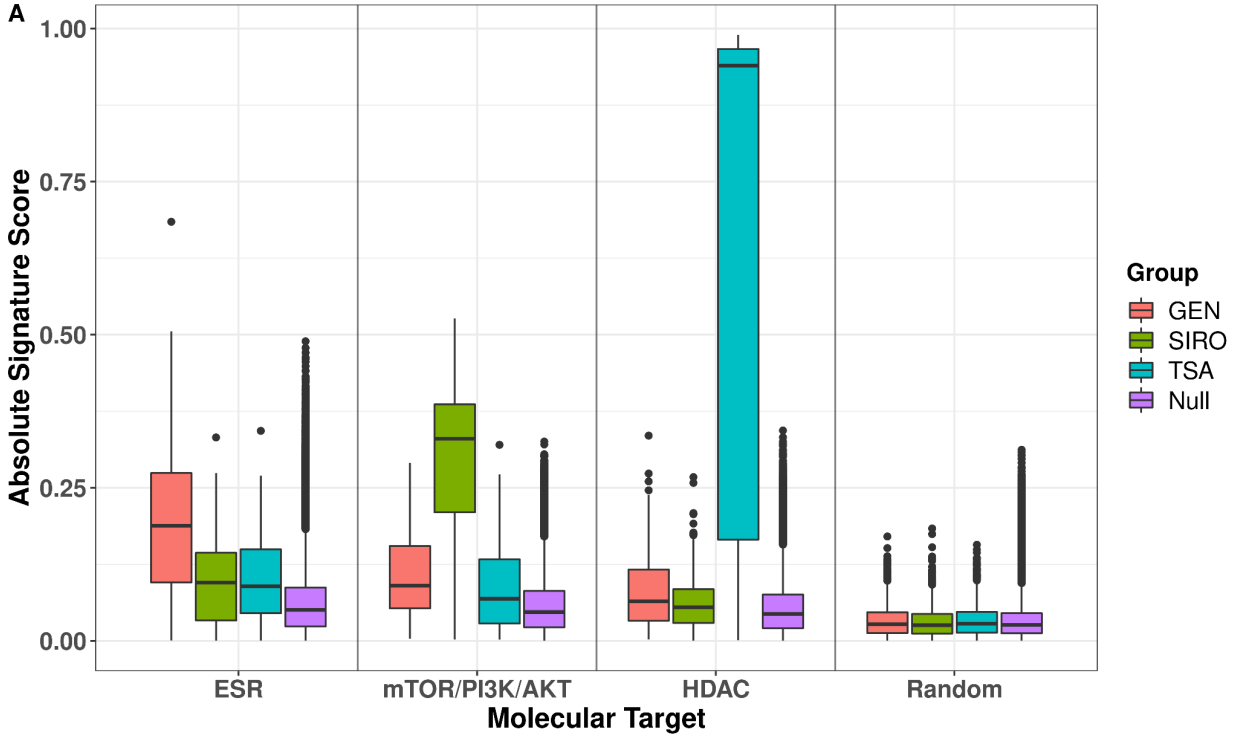
HTTr Sample Quality Assessment (2)



Harrill et al., Toxicol Sci. 2021 Apr 27;181(1):68-89.

HTTr Sample Performance Assessment

A



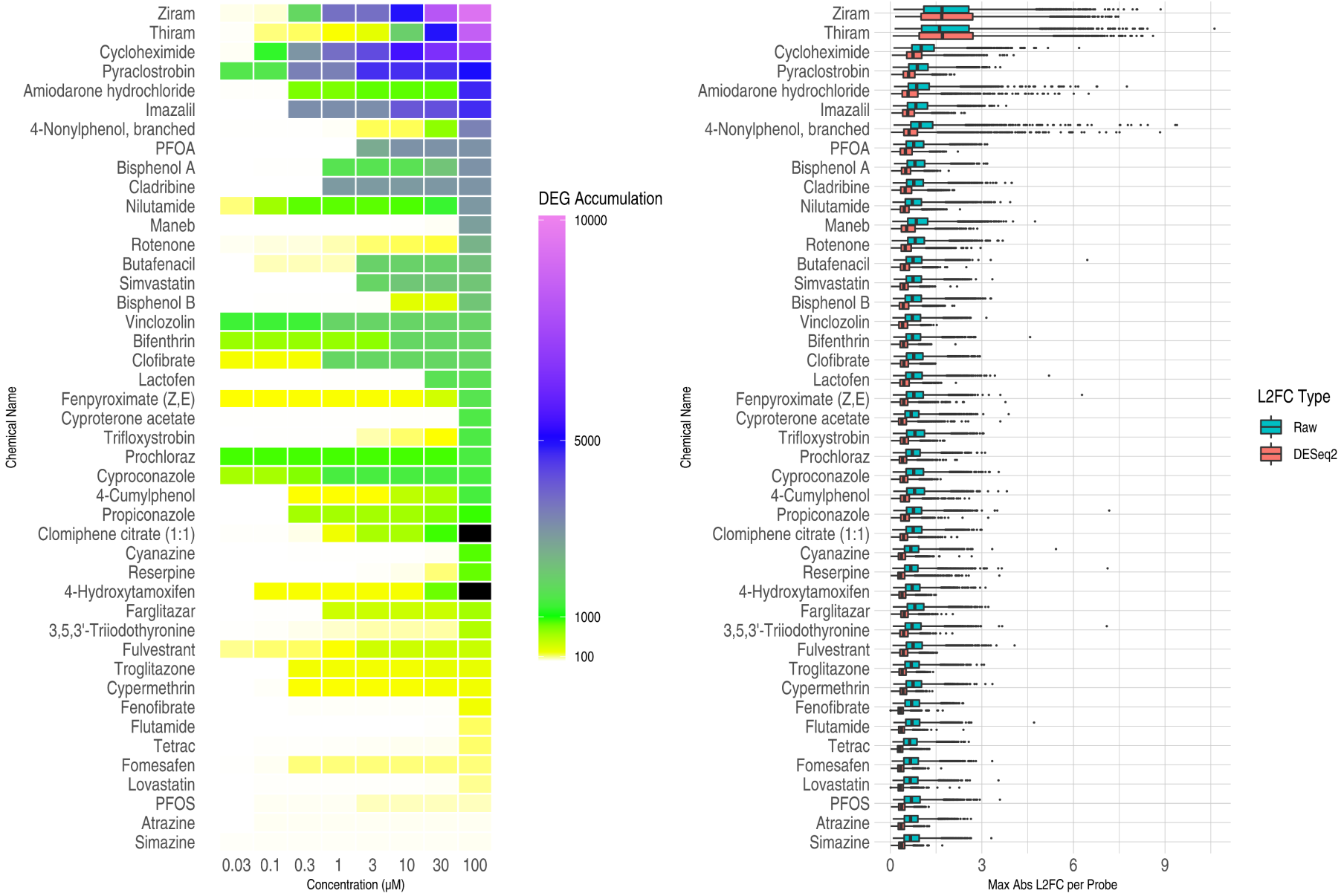
B

	Signature	Absolute Score	Direction
Genistein	RYAN_ESTROGEN_RECEPTOR_ALPHA	0.6844617	+
	CMAP equilin 1.5e-05 100 5542 100	0.5054424	+
	CMAP fulvestrant 1e-08 100 1417 100	0.4857079	-
	CMAP fulvestrant 1e-06 100 6811 100	0.4519683	-
	BHAT_ESR1_TARGETS_NOT_VIA_AKT1	0.4270213	+
Sirolimus	CMAP sirolimus 1e-07 100 5437 100	0.5265263	+
	CMAP sirolimus 1e-07 100 6409 100	0.5166319	+
	CMAP sirolimus 1e-07 100 8359 100	0.5001286	+
	CMAP wortmannin 1e-06 100 577 100	0.4990501	+
	CMAP wortmannin 1e-08 100 1423 100	0.4971645	+
Trichostatin A	CMAP vorinostat 1e-05 100 817 100	0.9899307	+
	CMAP trichostatin A 1e-07 100 8791 100	0.9869679	+
	CMAP trichostatin A 1e-07 100 6547 100	0.9854099	+
	CMAP trichostatin A 1e-06 100 8056 100	0.9852016	+
	CMAP trichostatin A 1e-07 100 7972 100	0.9849593	+

- Signature scoring using the single sample Gene Set Enrichment Analysis (ssGSEA) approach (Barbie et al. 2009)
- The “correct” target classes were identified for reference chemical treatments.

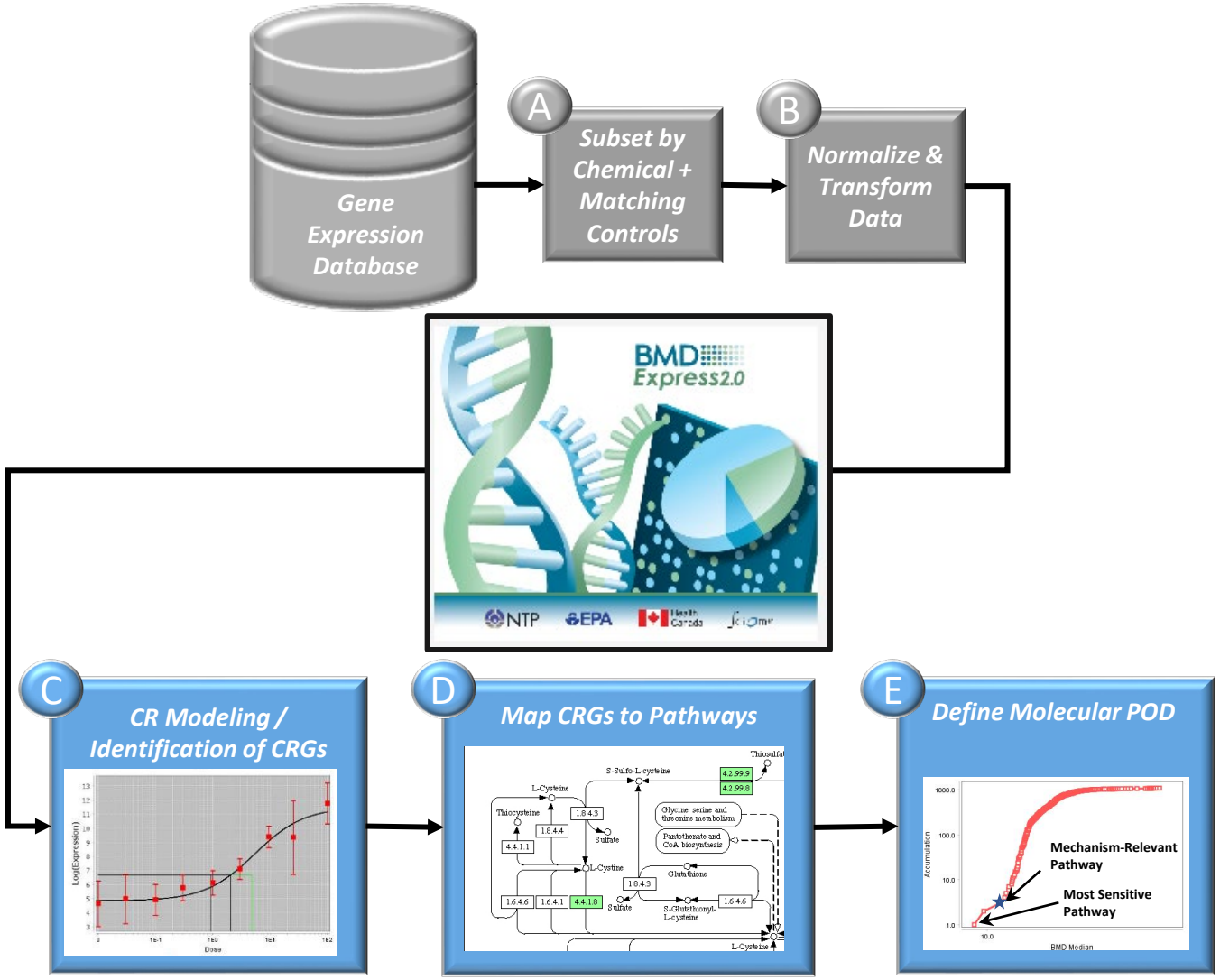
Harrill et al., Toxicol Sci. 2021 Apr 27;181(1):68-89.

HTTr Signal Strength



Harrill et al., Toxicol Sci. 2021 Apr 27;181(1):68-89.

Concentration Response Modeling: BMDExpress



Adapted from Harrill et al. (2019)

Based on National Toxicology Program Approach to Genomic Dose-Response Modeling (NTP RR 5)

BMDExpress Parameter	Criteria
Pre-filter:	|FC| > 2 at any test concentration
Models	Hill, Power, Linear, Poly2, Exponential 2|3|4|5
BMR Factor:	1.349*SD of controls (10%)
Best Model Selection:	Lowest AIC
Hill Model Flagging:	'k' < 1/3 Lowest Positive Dose Exclude Flagged Hill Models from Best Model Selection
Conc-Response Hit Criteria	(0.1*lowest conc. < BMC < highest conc.) BMC fit p-value > 0.1 BMCL / BMCU < 40
Gene Set Analysis:	≥ 3 Concentration-responsive genes ≥ 5% Gene Set Coverage
Gene Set Collections:	MSigDB (Liberzon et al. 2015) BioPlanet (Huang et al. 2019) CMAP (Subramanian et al. 2005)

Concentration-Response Modeling of Signature Scores (1)

Step 1: Inputs

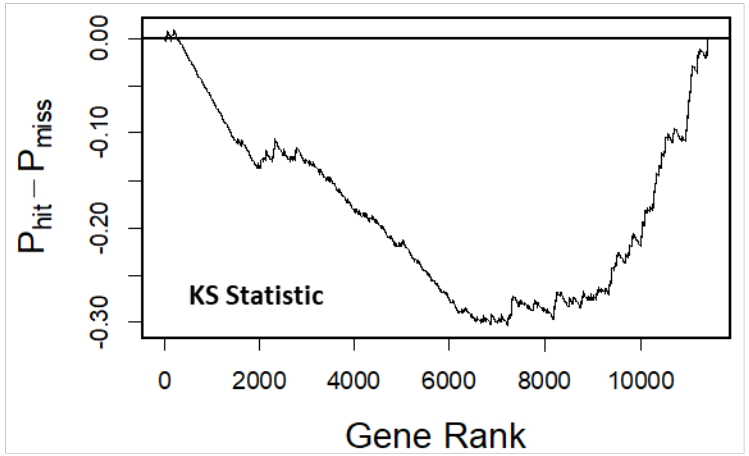
Experimental Data: Chemical_Conc × Gene matrix of $\log_2$ (fold-change) (l2fc) values.

Signature Collections: MSigDB (*Liberzon et al. 2015*), BioPlanet (*Huang et al. 2019*), CMAP (*Subramanian et al. 2005*)

Step 2: Pathway Scoring

Scores based on single sample GSEA method (Barbie et al. 2009)

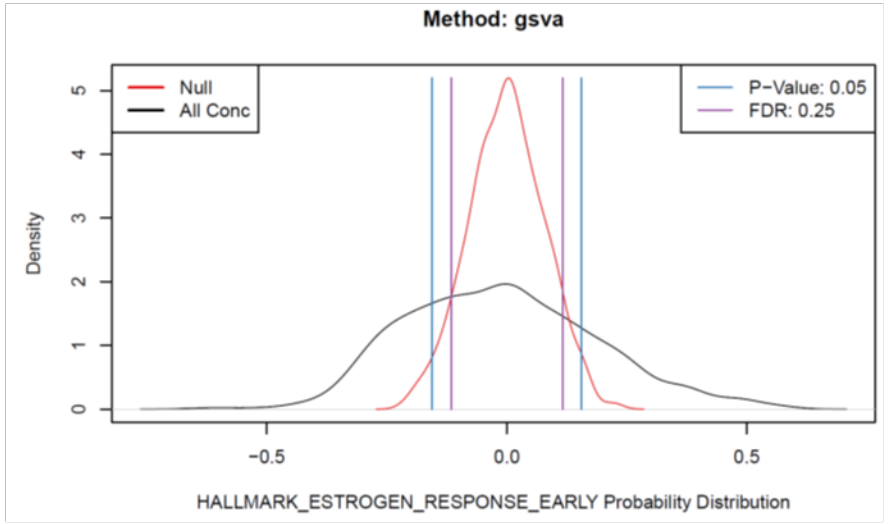
K-S statistic



Chemical_Conc × Pathway matrix of scores.

Step 3: Cut-off Estimation via NULL Modeling

- For each gene, **resample** l2fc based on the cross-sample gene distribution → breaks gene correlation
- Calculate **pathway scores for "null" data**
  - One null distribution (n = 1000 scores) / pathway

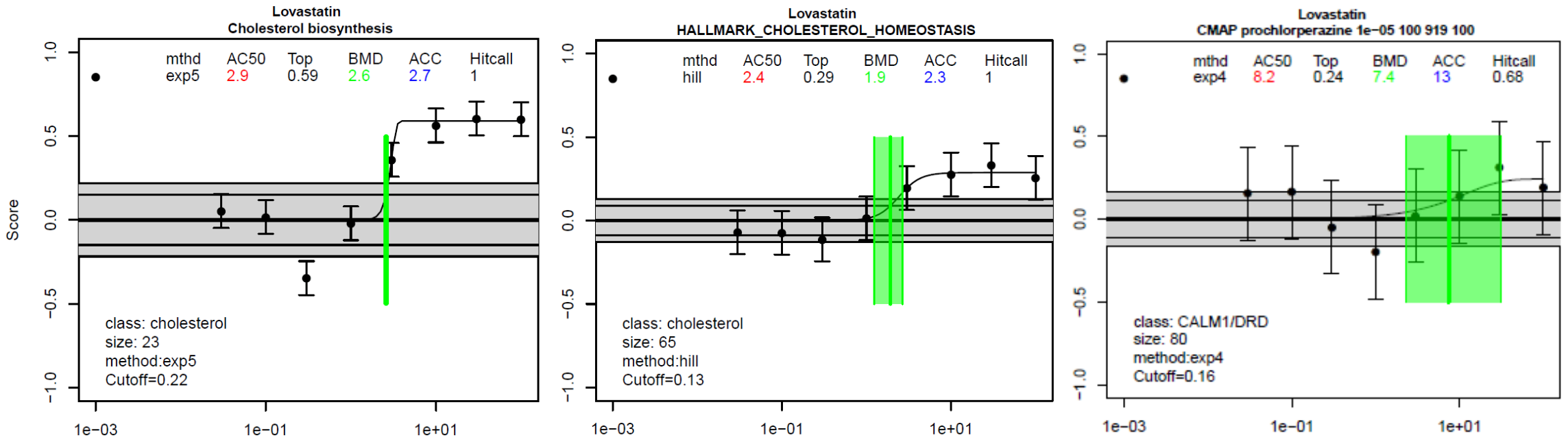


Analysis by Thomas Sheffield and Richard Judson

Concentration-Response Modeling of Signature Scores (2)

Step 4: CR Modeling

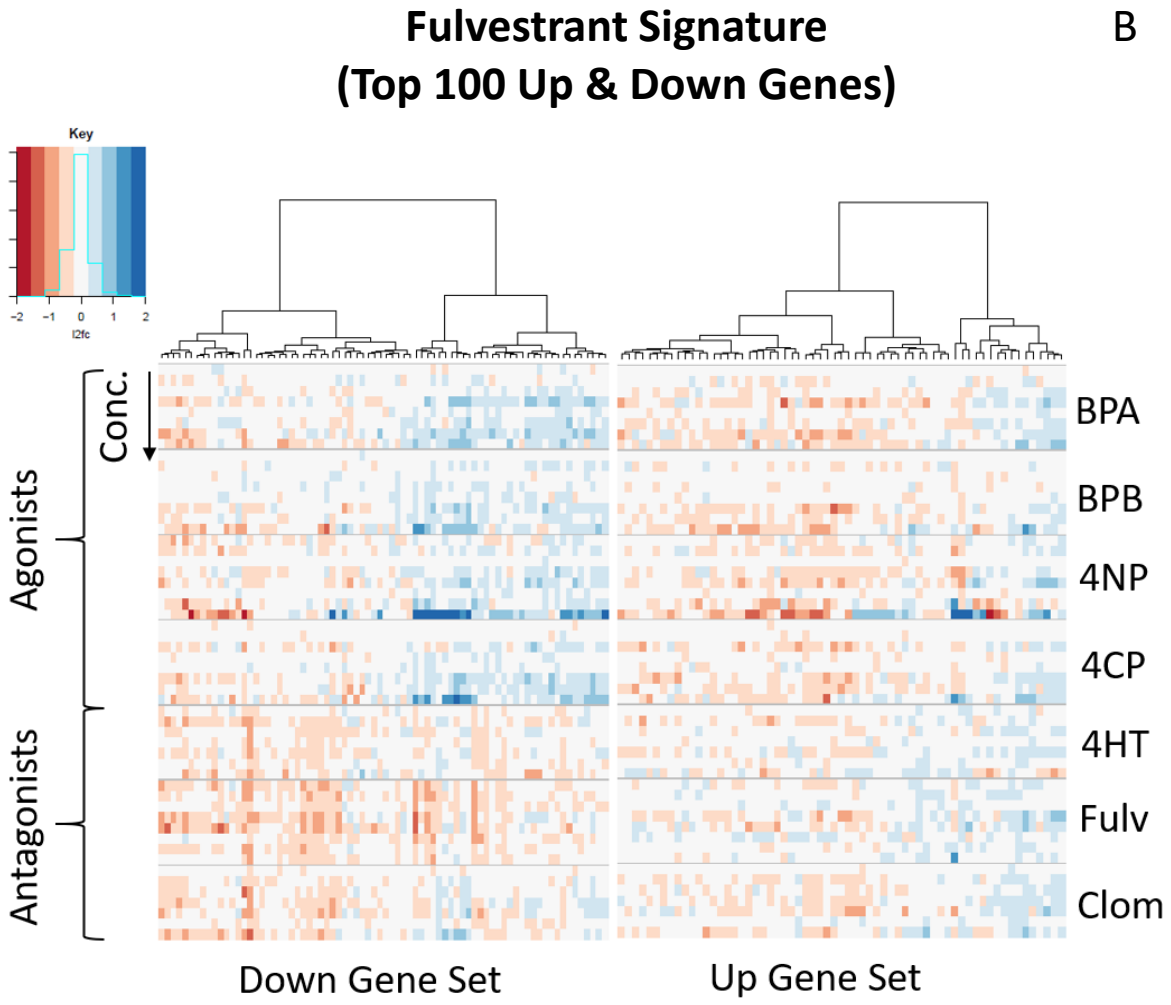
Concentration response modeling of signature scores using *tcplfit2* (https://rdrr.io/github/USEPA/CompTox-ToxCast-tcplFit2/)



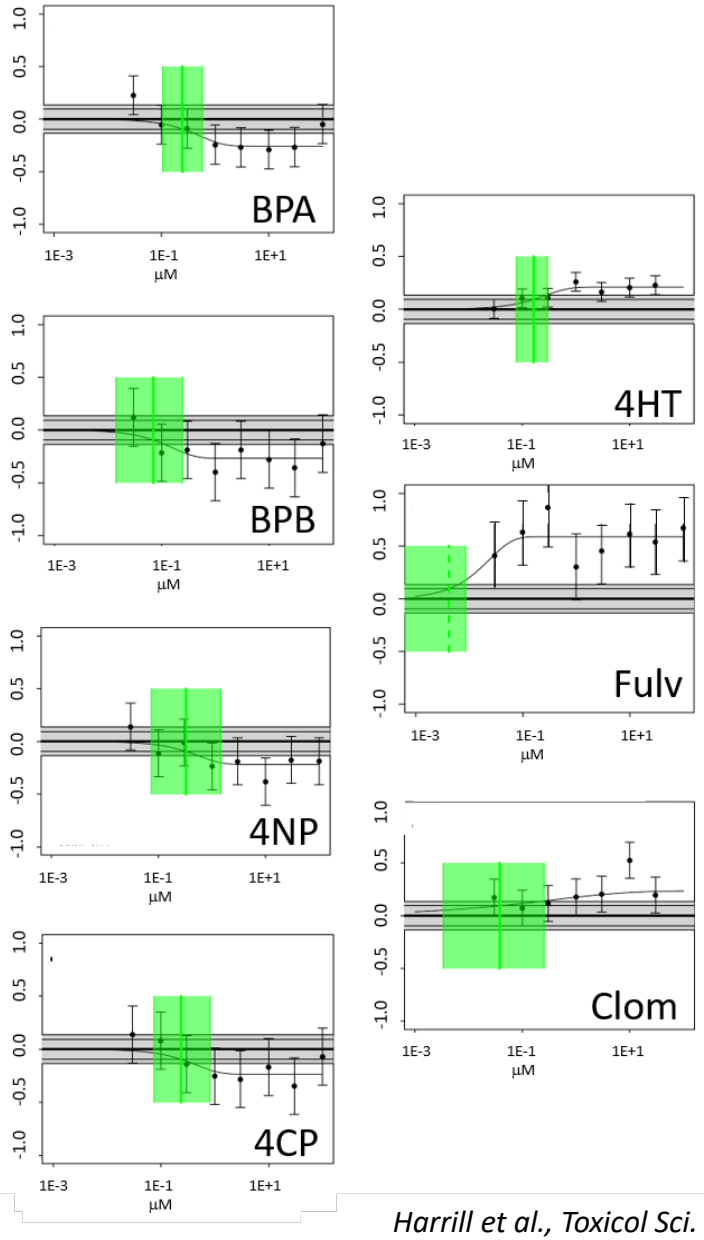
- Takes into account coordinated changes in gene expression that may not be identified using gene level fitting approaches.
- All curve forms from BMDExpress, plus constant model.
- Provides continuous hit calls for identifying high confidence and low confidence hits.

Concentration-Response Modeling of Signature Scores (3)

A

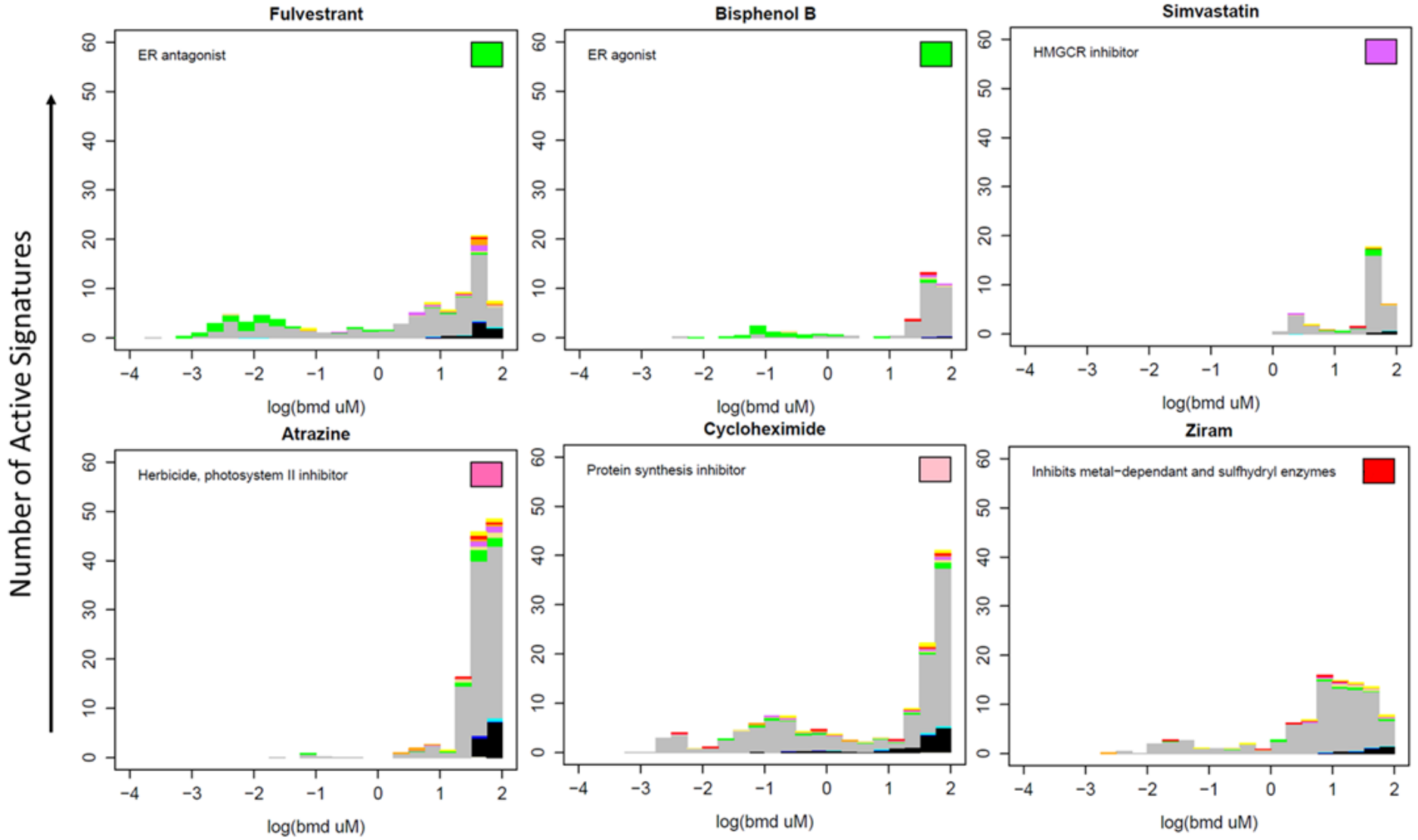


B



Harrill et al., Toxicol Sci. 2021 Apr 27;181(1):68-89.

Signature Modeling Reveals Biologically Relevant Targets as Most Sensitive

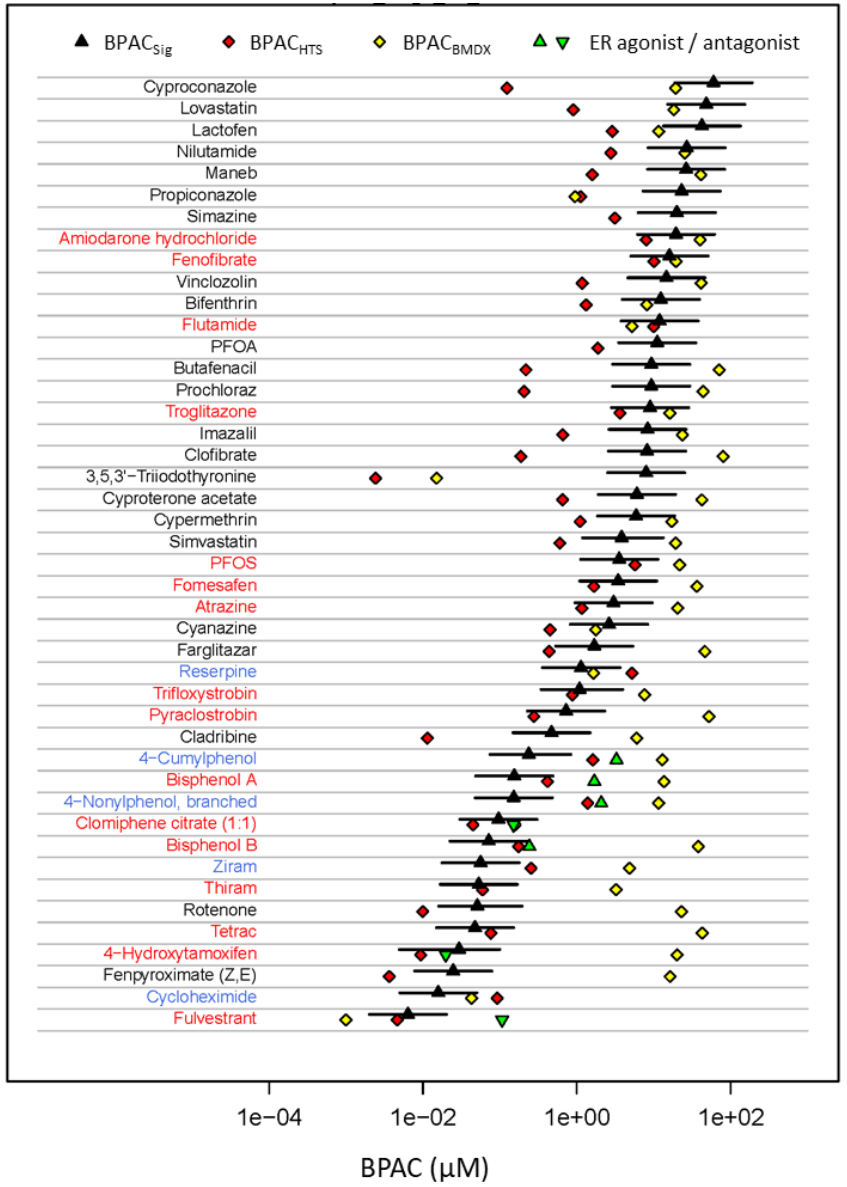


Harrill et al., Toxicol Sci. 2021 Apr 27;181(1):68-89.

Comparison of BMDExpress, Signature Modeling and ToxCast

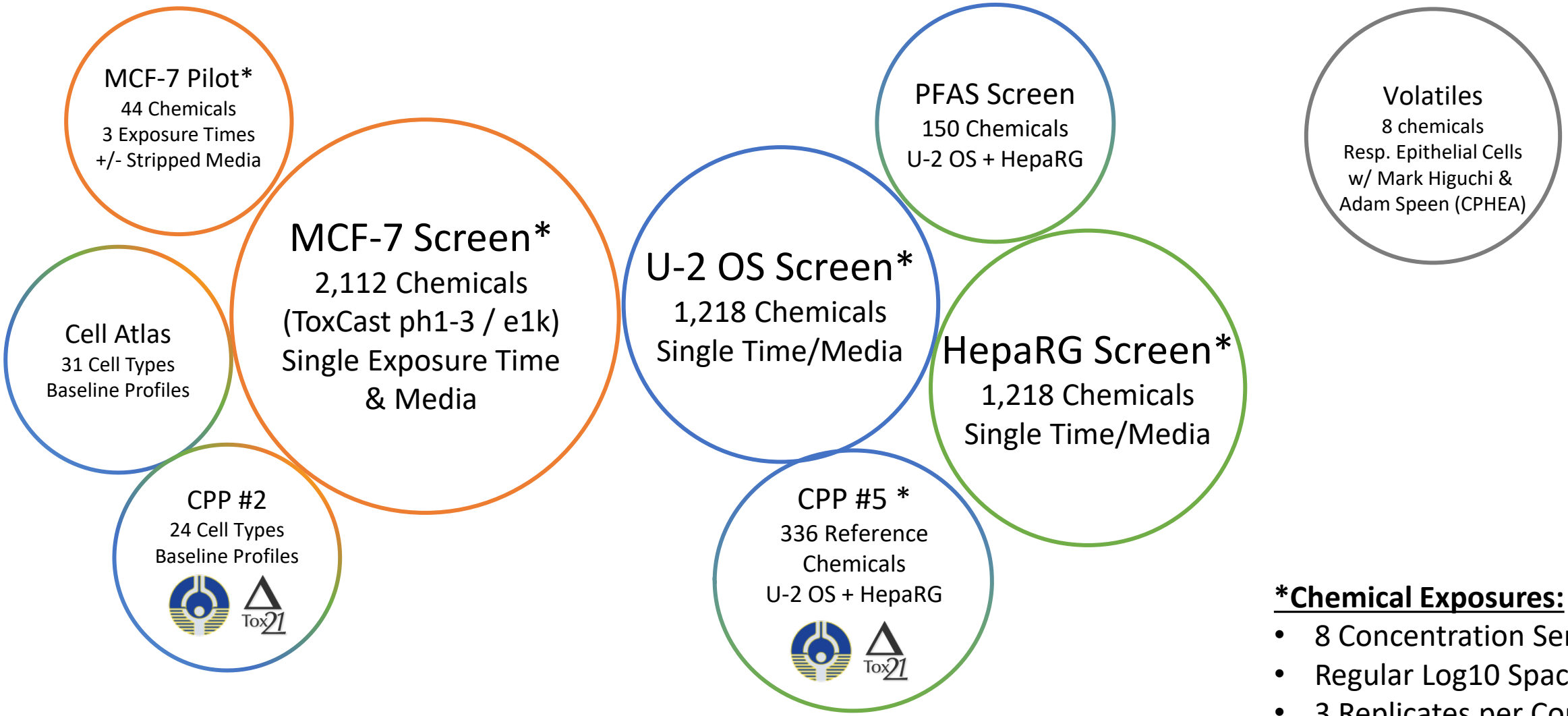
- **$BPAC_{Sig}$** → 5th lowest potency of active signatures
- **$BPAC_{BMDX}$** → Most sensitive signature / pathway
- **$BPAC_{HTS}$** → Lower 5th percentile of active AC50 values for assays that pass a series of quality filters.

- $BPAC_{HTS}$ and $BPAC_{Sig}$ are in better agreement than $BPAC_{HTS}$ and $BPAC_{BMDX}$
- In most of these cases, $BPAC_{HTS}$ is also more potent than $BPAC_{BMDX}$.
- The majority of these cases can be explained by the use of ToxCast assays for the specific target of the chemical that are not active/expressed in MCF7 cells.
  - THRA / THRB
  - CYP Assays
  - PTPN Assays



Harrill et al., Toxicol Sci. 2021 Apr 27;181(1):68-89.

High Throughput Transcriptomics (HTTr) Data Landscape



***Chemical Exposures:**

- 8 Concentration Series
- Regular Log10 Spacing
- 3 Replicates per Conc
- Randomized Plate Layout

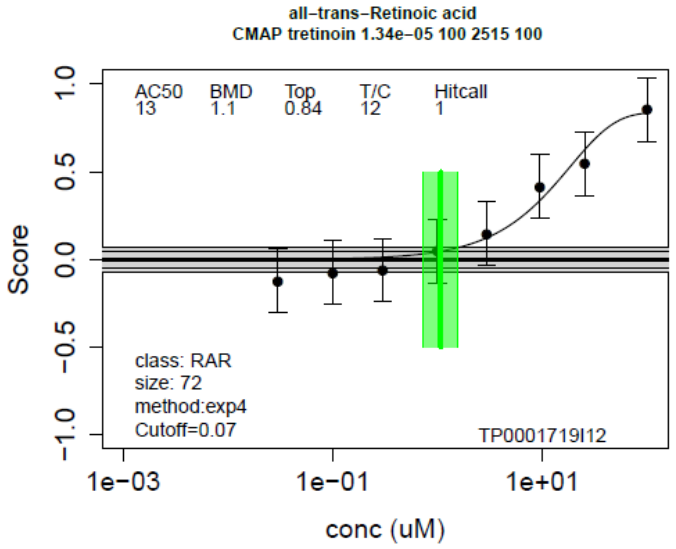
Slide courtesy of Logan Everett

Refinement of Concentration-Response Modeling Approach

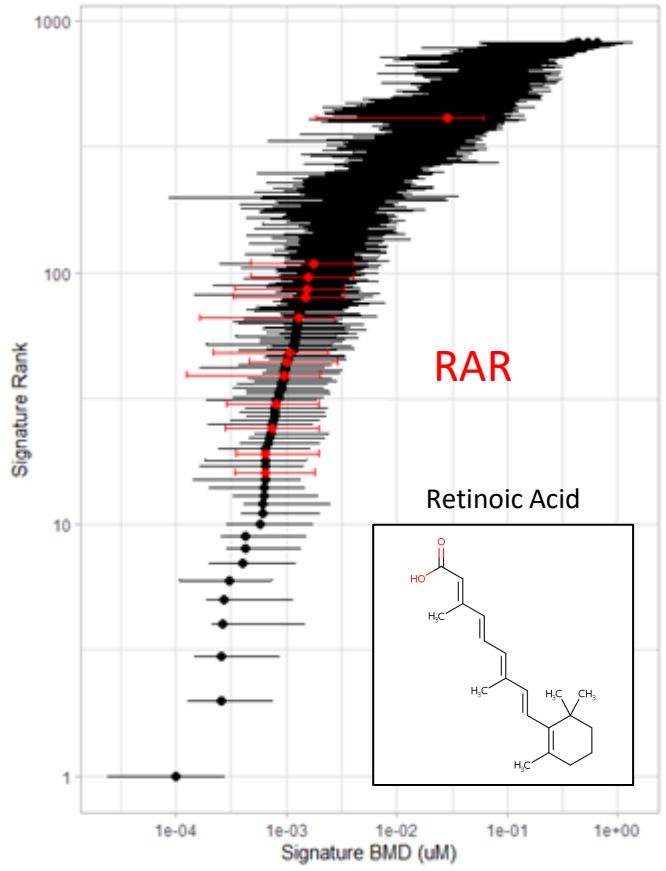
Concentration-Response Modeling (*tcplfit2*)

Signature-Level:

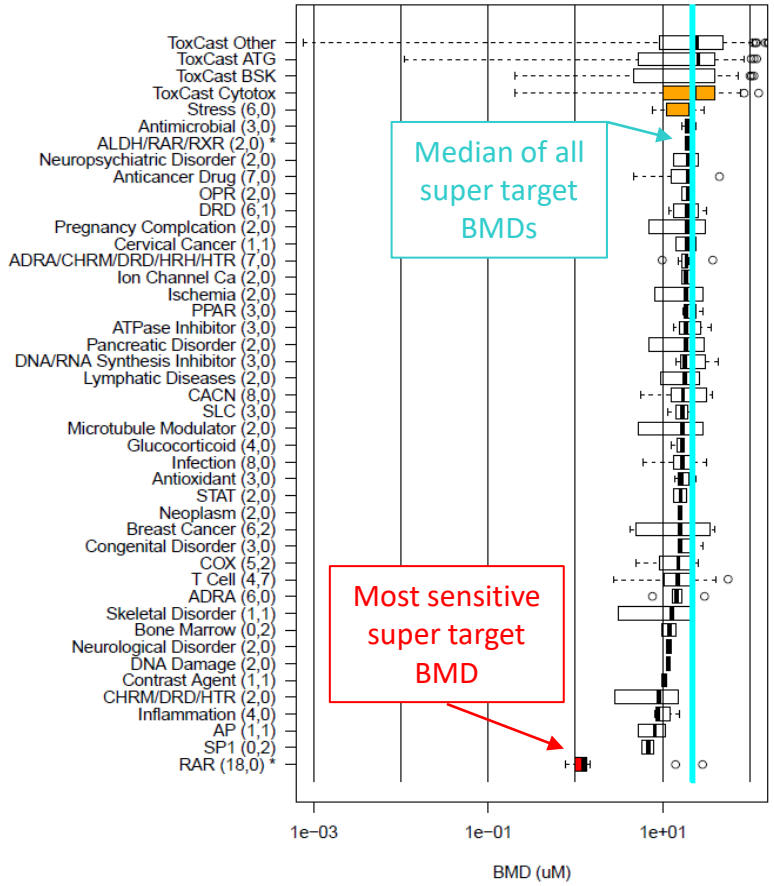
- Benchmark Dose (BMD)
- Confidence Interval on BMD
- Hit Call Probability



Ranking of Signatures



Signature Aggregation



- Aggregation of signatures can aid in biological interpretation & putative target prediction.

EPA United States Environmental Protection Agency

Applications for Molecular PODs From HTP NAMs

HTP Screening Experimental Designs

Parameter	Multiplier	Notes			
Chemicals	462	APCRA case study chemicals			
Cell Types	4	U-2 OS		HepaRG-2D	MCF-7
Assay Formats	2	HTPP	HTTr	HTTr	HTTr
Exposure Durations	Variable	24 HR	24 HR	24 HR	6 HR
Concentrations:	8	3.5 log_{10} units; ~half-log_{10} spacing			
Biological Replicates:	Variable	4	3	3	3

APCRA
ACCELERATING THE PACE OF CHEMICAL RISK ASSESSMENT

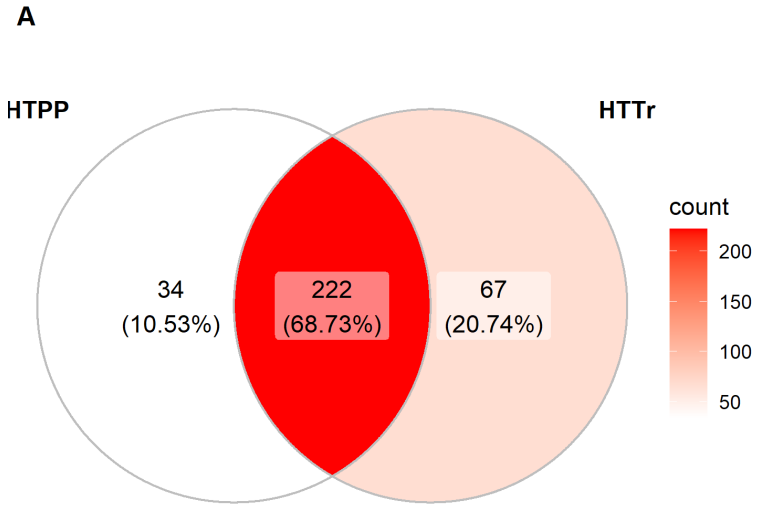
International collaboration of regulatory scientists focused on next generation chemical risk assessment including **deriving quantitative estimates of risk based on NAM-derived potency information and computational exposure estimates.**

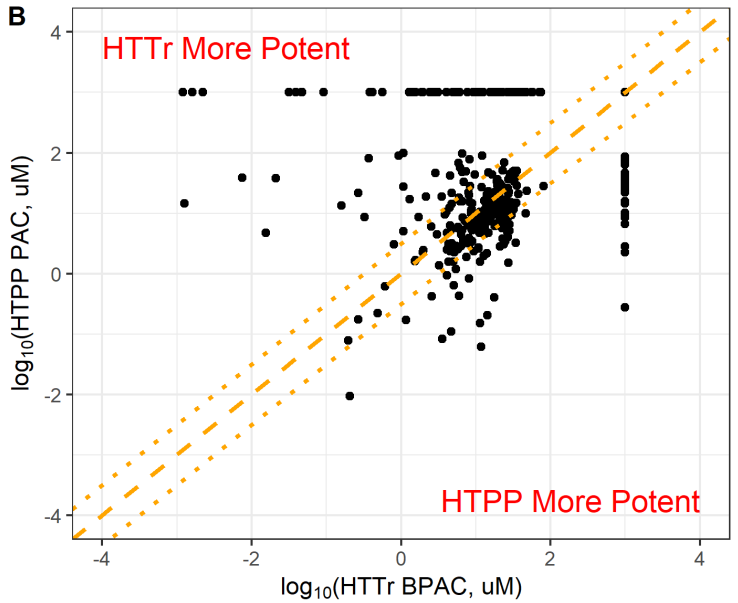
APCRA Chemicals → PK parameters necessary for *in vitro* to *in vivo* extrapolation (IVIVE)
in vivo toxicity data

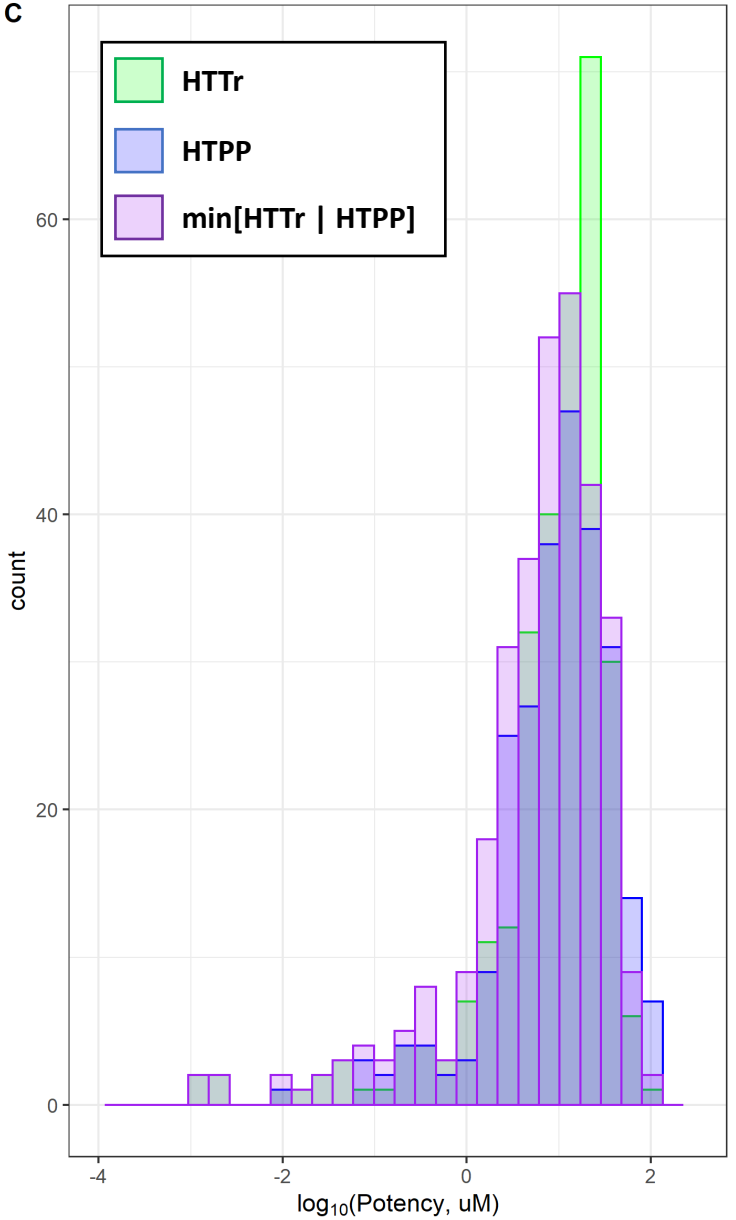
Kavlock et al. (2018)
Chem. Res. Tox; 31(5): 287-290

U-2 OS Screening Results

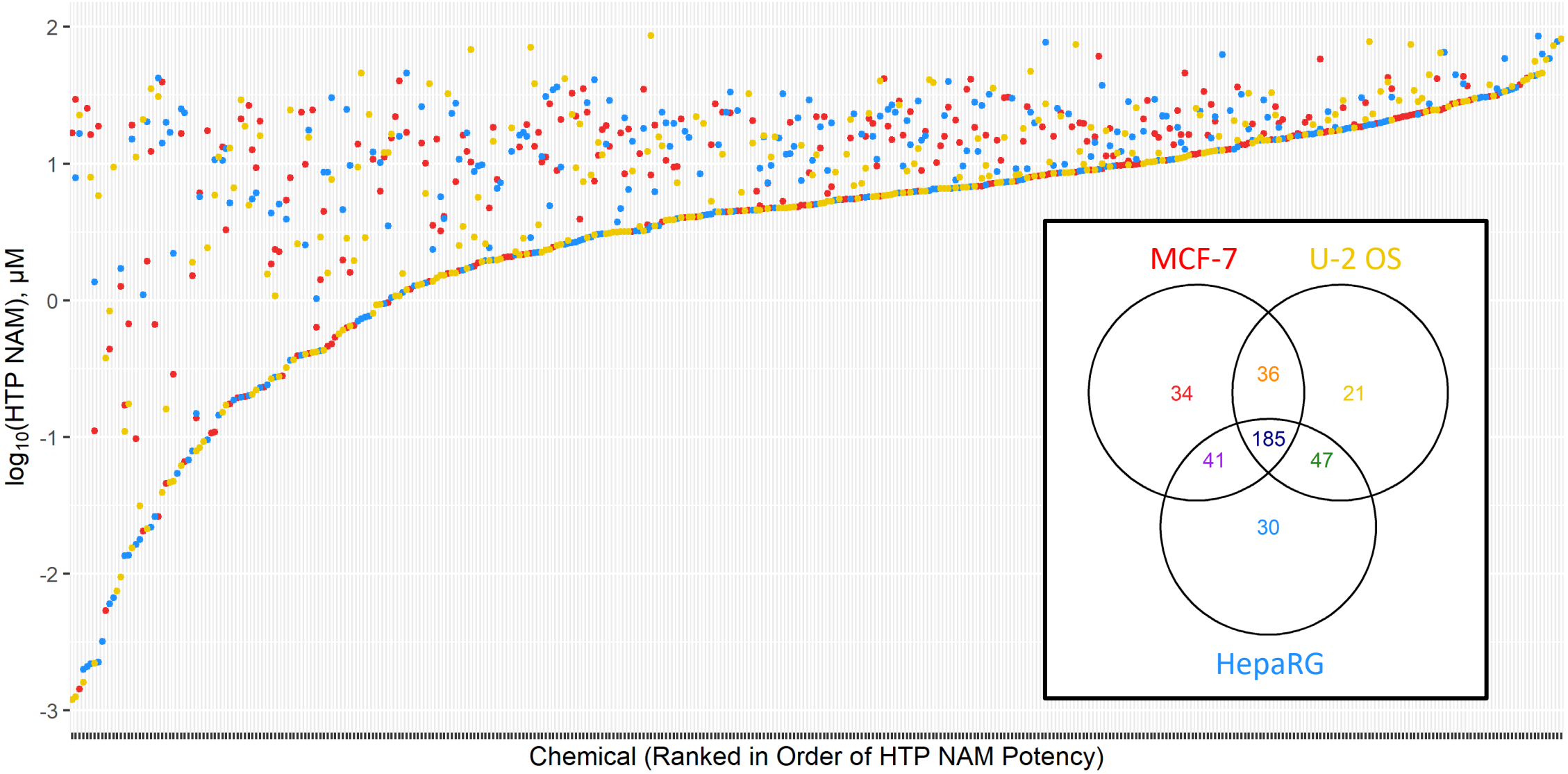
- A majority of chemicals were active in both the HTTr and HTPP assays.
- There were a larger number of chemicals active in HTTr only versus HTPP only.
- Most biological activity was observed between 1 and 10 uM.
- A few chemicals with HTTr PACs < 1 uM had HTPP BPACs > 10 uM or were inactive.







Comparison of Screening Results Across Cell Lines



- Molecular POD defined as the minimum potency observed in HTP NAM assays across three cell types.

In Vitro to *In Vivo* Extrapolation (IVIVE) Using High-Throughput Toxicokinetic (httk) Modeling

Predicted exposure

Exposure predictions (EPA ExpoCast)

- Systematic Empirical Evaluation of Models (SEEM) version 3
- Inferred from human biomonitoring data, production volume and use categories (industrial / consumer use)

New approach methodologies (NAMs)

Toxicological threshold of concern (TTC)

HTP Potency Estimate (µM)

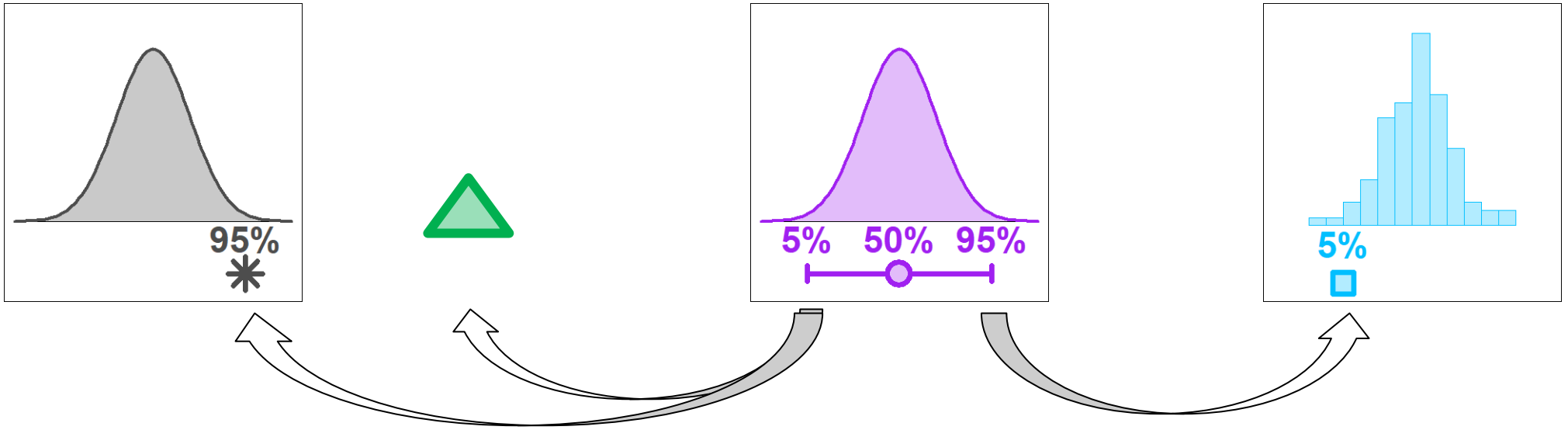
***In vitro*-to-*in vivo* extrapolation (IVIVE)**
high-throughput toxicokinetics (httk)

HTP AED (mg/kg bw/day)

***in vivo* point-of-departure**

Database of *in vivo* effect values (EPA – ToxValDB)

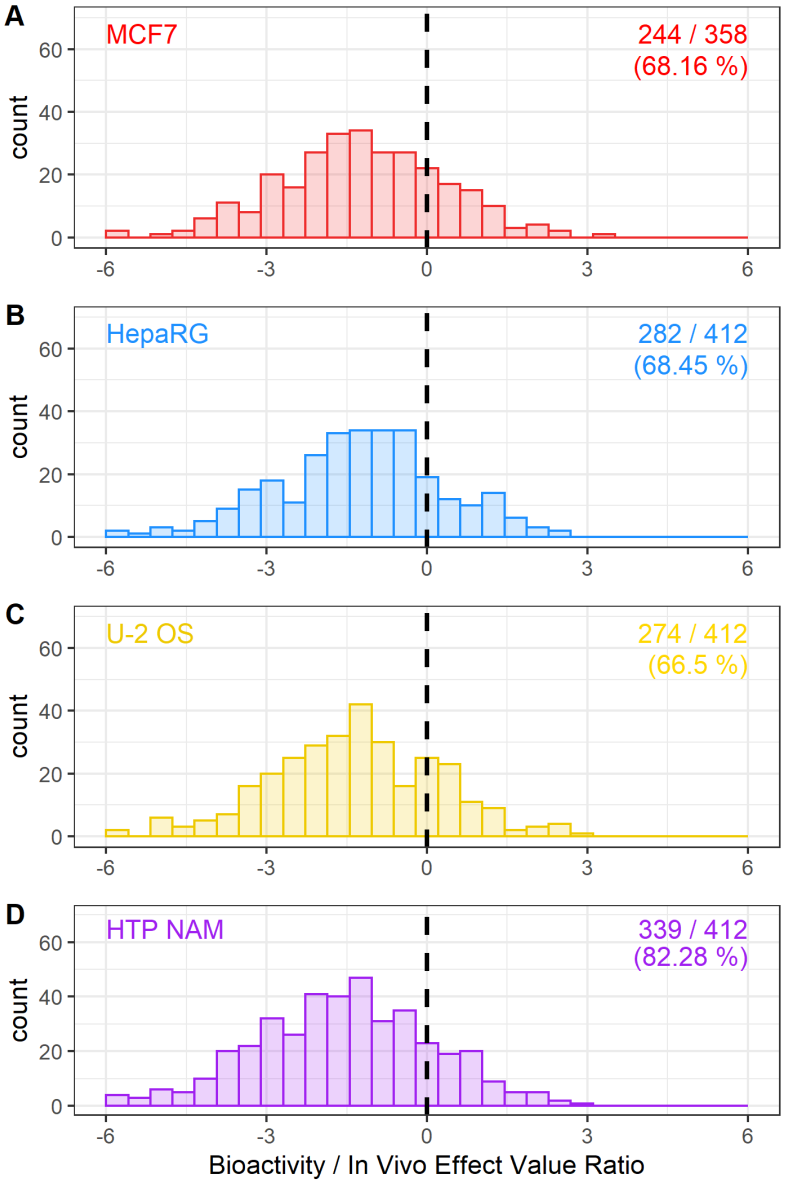
- Mammalian species
- oral exposures
- Various study types
- NOEL, LOEL, NOAEL, LOAEL
- mg/kg/day

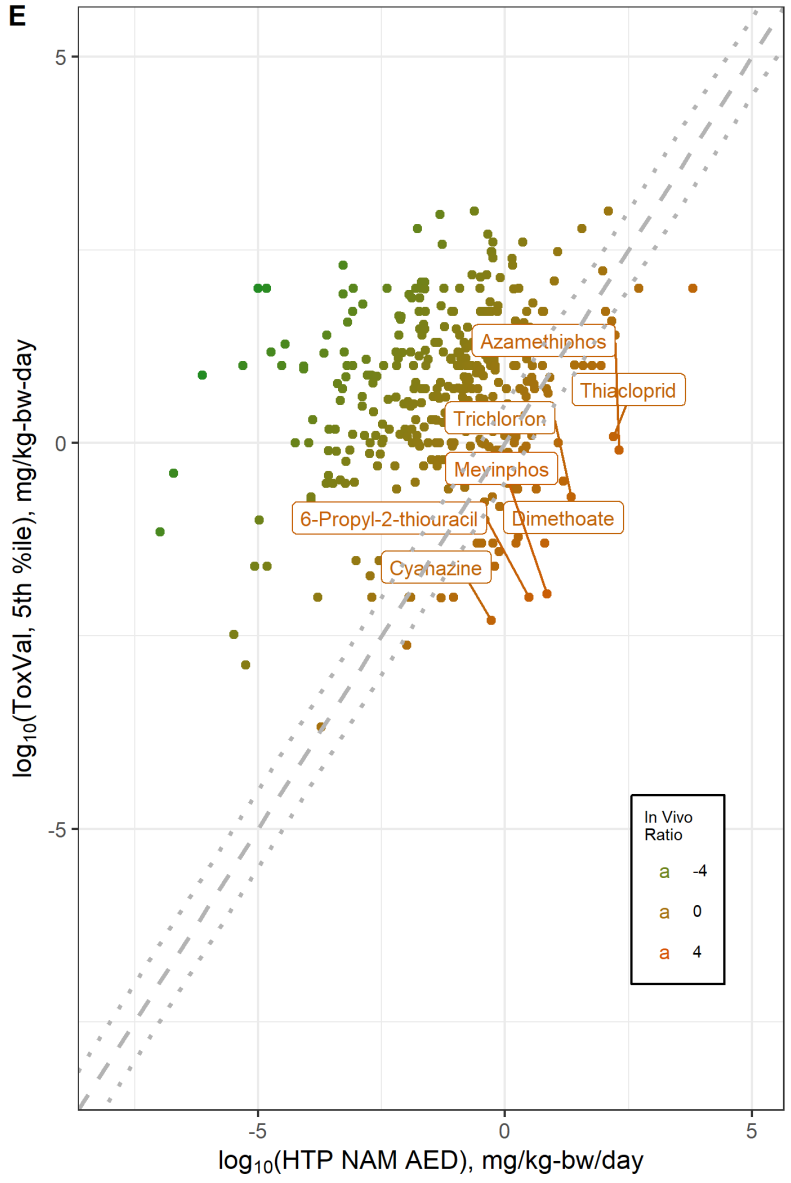


POD: point-of-departure
AED: administered equivalent dose

Bioactivity / *In Vivo* Effect Value Ratio Analysis

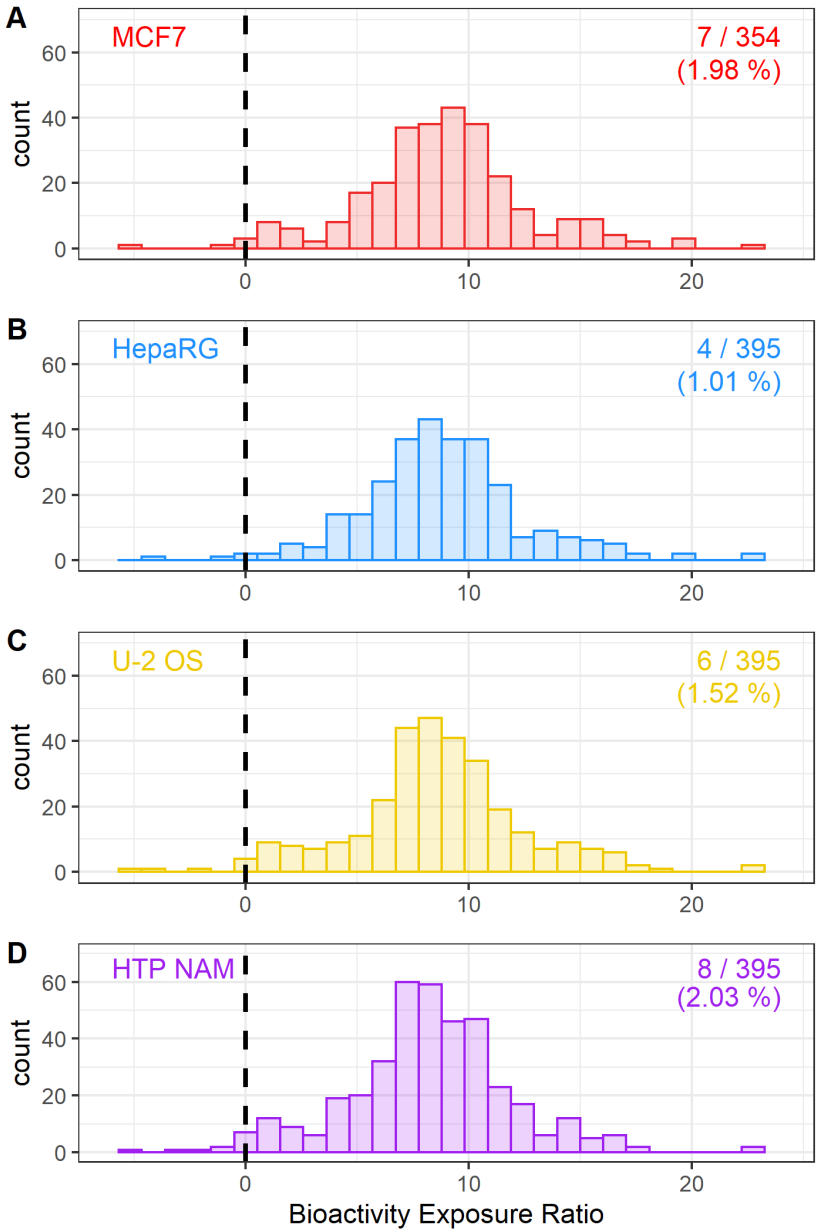
- **Negative ratios** indicate that AEDs derived from HTP NAMs molecular PODs are **conservative** surrogates for traditional *in vivo* PODs.
- When cell lines are considered individually, **~66-68%** of chemicals had negative ratios.
- When considered in combination, the number and percentage of chemicals with negative ratios **increased (82.3 %)**.
- Paul-Friedman et al. (2020)[a]:
  - Using ToxCast, **89 %** of APCRA chemicals had negative ratios.
- Positive ratios observed for several organophosphate and carbamate pesticides.

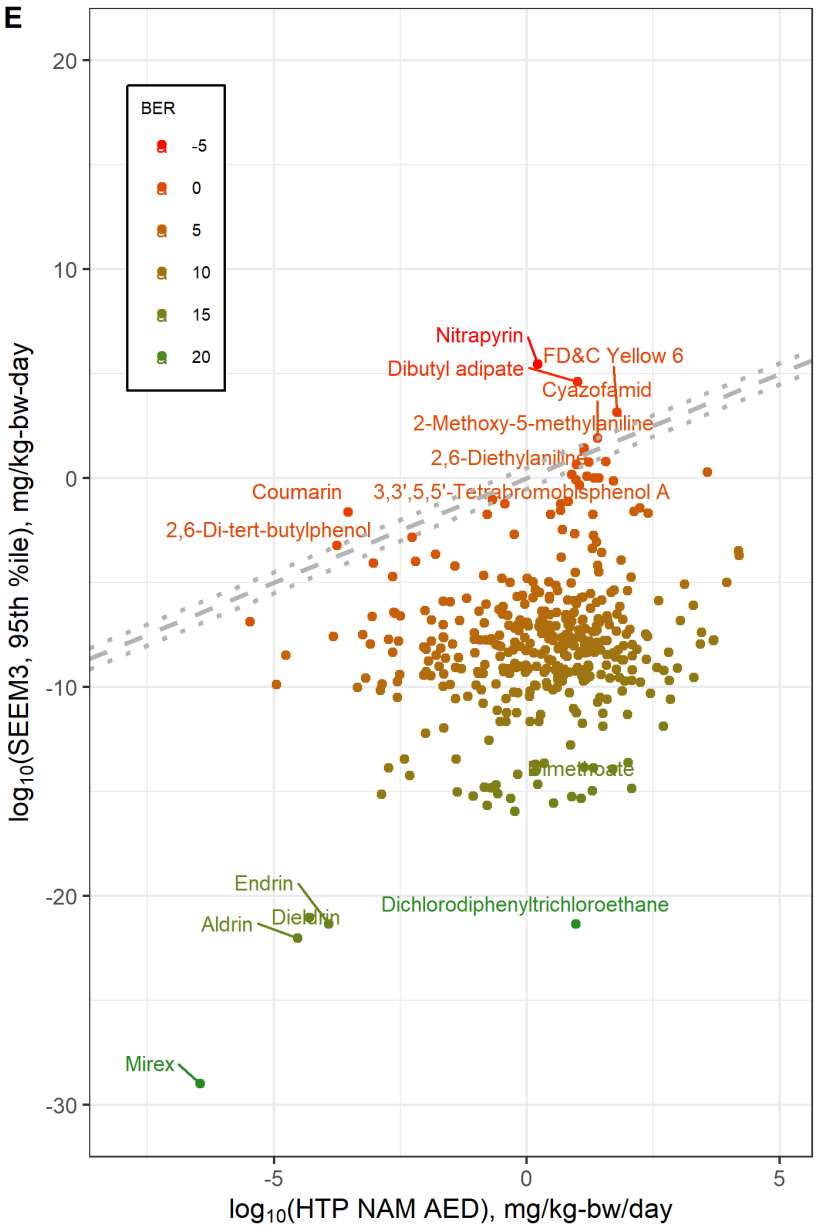




Bioactivity Exposure Ratio (BER) Analysis

- **Negative ratios** indicate a potential for human exposure to chemicals in a range that is bioactive in vitro.
- When cell lines are considered individually, **~1-2%** of chemicals had negative ratios.
- When considered in combination, the percentage of chemicals with negative ratios **did not appreciably change**.
- Positive ratios observed for several chemicals found in consumer products.
- Most extreme negative ratios associated with banned or limited use organochlorine pesticides.





Summary and Conclusions

- **High-Throughput Profiling:** Developed experimental designs and scalable laboratory workflows for high-throughput transcriptomics (and high-throughput phenotypic profiling) of environmental chemicals that can be used in multiple human-derived cell types.
- **Potency Estimation:** Developed high-throughput concentration-response modeling workflows to identify thresholds for perturbation of gene expression (e.g. BPACs).
- **Comparison to ToxCast:** BPACs from HTTr were comparable to BPACs from ToxCast HTS assays.
- **IVIVE:** Potency estimates can be converted to administered equivalent doses (AEDs) using high-throughput toxicokinetic modeling.
- **Bioactivity to *In Vivo* Effect Value Ratio Analysis:** AEDs derived from HTP assays were conservative compared to traditional PODs a majority of the time. Performance improved to ~80% when results from multiple cell types were considered in combination.
- **Bioactivity to Exposure Ratio (BER) Analysis:** AEDs derived from HTP assays were compared to high-throughput exposure predictions. There were very few chemicals where AEDs were within the range of exposure predictions.

Future Directions

- **Expand chemical space**
  - Screen additional chemicals in accordance with programmatic needs / goals.
- **Expand biological space**
  - Continue screening a subset of chemicals through many biologically diverse cell lines.
- **Refine signature concentration-response modeling approach**
  - Reduce redundancy in signature collection.
  - Continued curation of target annotation.
- **Refine methods for putative target prediction & confirmation**
  - Integration with other NAM's data streams.
  - Machine learning approaches.
  - Bioactivity confirmation within tiered hazard evaluation framework.

Acknowledgements

Office of Research and Development (ORD)
Center for Computational Toxicology and Exposure (CCTE)

- Clinton Willis
- Johanna Nyffeler
- Rick Brockway
- Megan Culbreth
- Dan Hallinger
- Terri Fairley

- Ann Richard
- Kathy Coutros
- Maureen Gwinn
- Russell Thomas

- Logan Everett
- Imran Shah
- Richard Judson
- Derik Haggard
- Thomas Sheffield
- Joseph Bundy
- Woody Setzer

- Katie Paul Friedman
- John Wambaugh

National Toxicology Program
U.S. Department of Health and Human Services

- Scott Auerbach

BioSpyder™

- Jo Yeakley
- Bruce Seligmann
- Joel McComb
- Pete Shepherd
- Milos Babic
- Dalia Gonzalez
- Kyle LeBlanc
- Garrett McComb