High-Throughput Transcriptomics for Chemical Bioactivity Screening and Tiered Hazard Evaluation

Joshua A. Harrill, PhD

US EPA, Center for Computational Toxicology and Exposure (CCTE)

109 TW Alexander Drive

Research Triangle Park, NC 27709

E-mail: harrill.joshua@epa.gov

ORCID: 0000-0003-4317-6391

Conflict of Interest Statement

- The views expressed in this presentation are those of the presenter and do not necessarily represent the views or policies of the US Environmental Protection Agency, nor does mention of trade names or products represent endorsement for use.
- The presenter has no conflict of interest regarding the materials in this presentation.

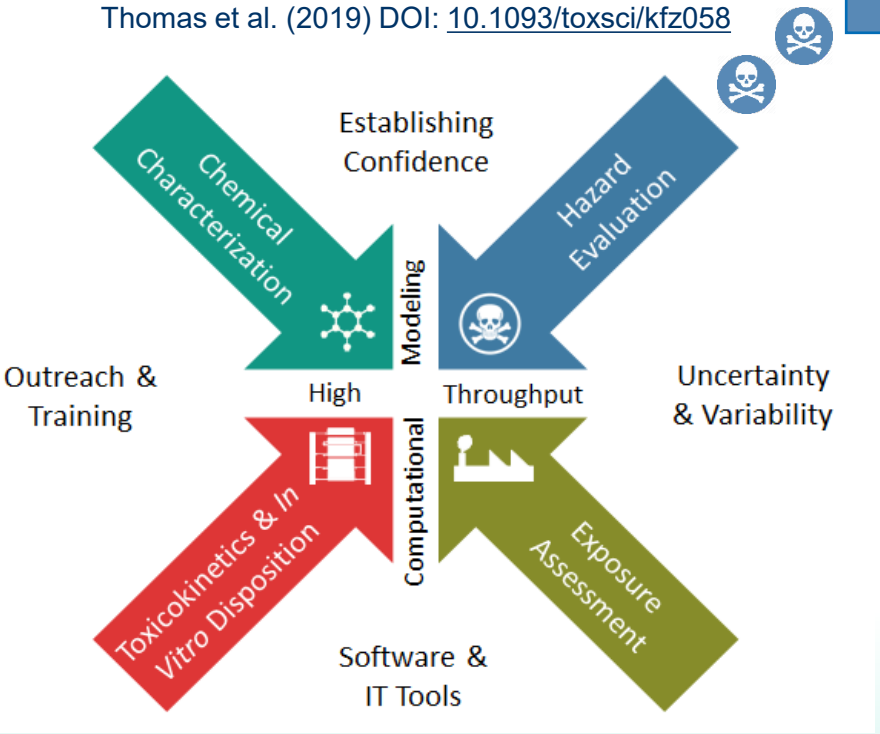
Objectives

- Broad overview of the Next Generation Blueprint of Computational Toxicology at US EPA → emphasis on the role of **transcriptomics**.

- Provide information on technological and analytical innovations that support **high-throughput transcriptomics (HTTr)** chemical screening.
  - Targeted RNA-Seq technology.
  - Novel bioinformatics workflows and associated open-source tools.
  - Transcriptomic reference materials.
  - International effort to develop omics reporting frameworks.

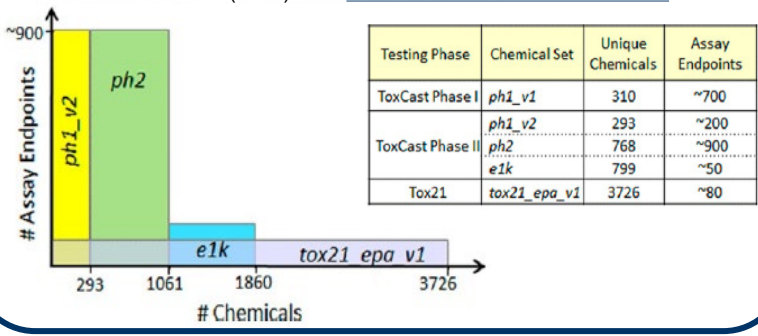
Computational Toxicology Research Areas at EPA

The NexGen Blueprint of CompTox at US EPA
Thomas et al. (2019) DOI: 10.1093/toxsci/kfz058



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

Richard et al. (2016) DOI: 10.1021/acs.chemrestox.6b00135



Testing Phase	Chemical Set	Unique Chemicals	Assay Endpoints
ToxCast Phase I	*ph1_v1*	310	~700
ToxCast Phase II	*ph1_v2*	293	~200
	ph2	768	~900
	e1k	799	~50
Tox21	*tox21_epa_v1*	3726	~80

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

New Strategy for Hazard Evaluation: Improve efficiency and increase biological coverage by using non-targeted profiling assays that cast the broadest net possible for capturing the potential molecular and phenotypic responses of human cells to chemical exposures.

NAMs-Based Tiered Hazard Evaluation Approach

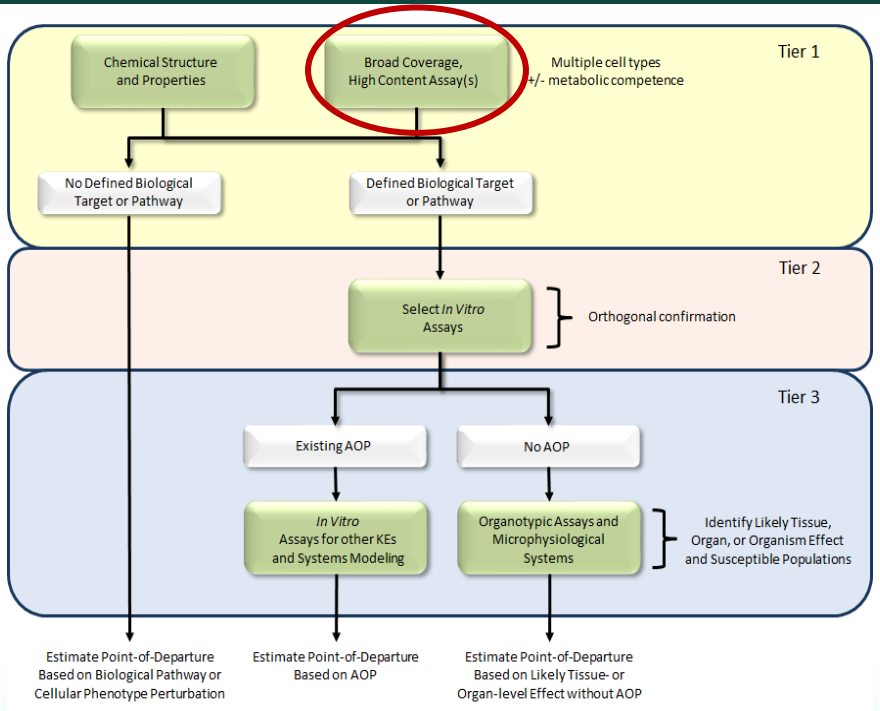
High throughput profiling (HTP) assays are proposed as the first tier in a NAMs-based hazard evaluation approach.

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.

To date, EPA has identified and implemented two HTP assays that meet this criteria.

- **High-Throughput Transcriptomics [HTTr]**
- High-Throughput Phenotypic Profiling [HTPP]



The NexGen Blueprint of CompTox at US EPA
Thomas et al. (2019) DOI: 10.1093/toxsci/kfz058

Templated Oligo with Sequencing Readout (TempO-Seq)

The **TempO-Seq** human whole transcriptome assay measures the expression of greater than 20,000 transcripts.

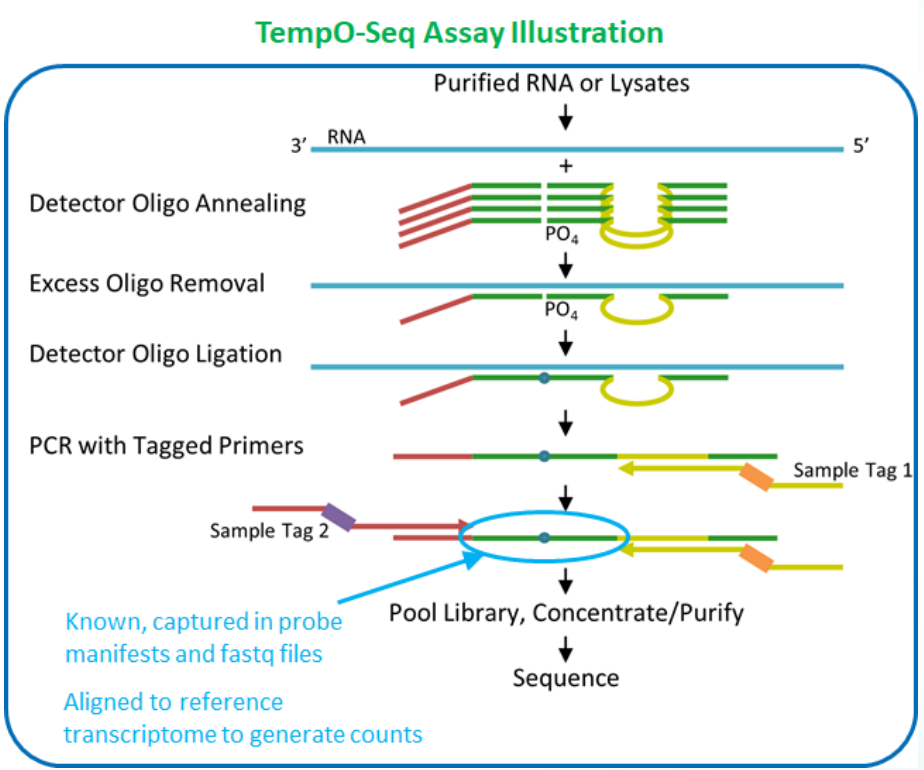
Requires only picogram amounts of total RNA per sample.

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 instruments.

Scalable, targeted assay:

- 1) specifically measures transcripts of interest
- 2) ~50-bp reads for all targeted genes
- 3) requires less flow cell capacity than RNA-Seq



Yeakley et al. (2017) DOI: 10.1371/journal.pone.0178302

Chemical Screening in MCF7 Cells Using HTTr

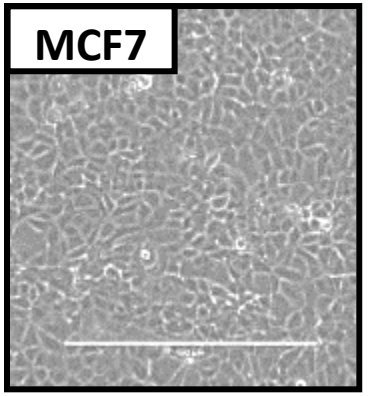
High-Throughput Transcriptomics Platform for Screening Environmental Chemicals

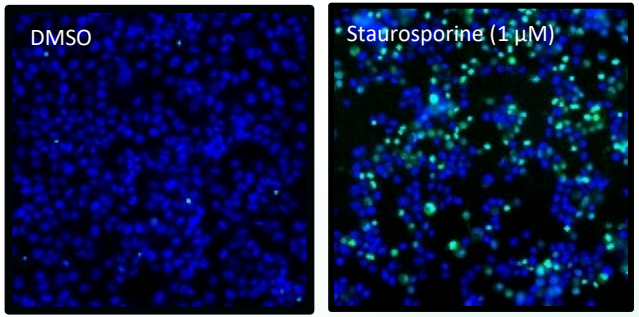
Joshua A. Harrill,*,1 Logan J. Everett,* Derik E. Haggard,*,† Thomas Sheffield,*,† Joseph L. Bundy,* Clinton M. Willis,*,‡ Russell S. Thomas,* Imran Shah,* and Richard S. Judson*

TOXICOLOGICAL SCIENCES, 2021, 1–22

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

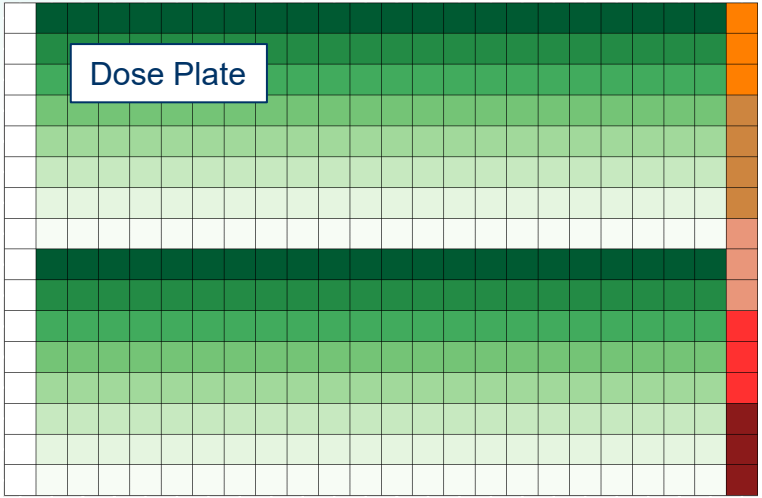
Parameter	MCF7 Pilot	MCF7 Screen	Notes
Cell Type(s)	1	1	MCF7
Assay Formats:	2	2	High-Throughput Transcriptomics Cell Viability
Culture Condition	1	1	DMEM + 10% HI-FBS
Chemicals	44	1784	ToxCast chemicals
Time Points:	1	1	6 hours
Concentrations:	8	8	3.5 $\log_{10}$ units; semi $\log_{10}$ spacing
Biological Replicates:	3	3	Independent cultures



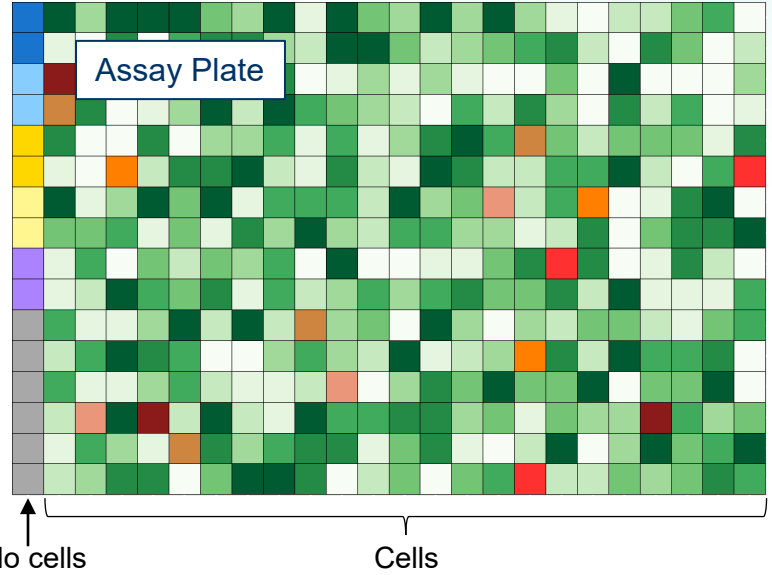


CellEvent Caspase 3/7

Experimental Design for HTTr



LabCyte® Echo 550 Acoustic Dispenser



= Test chemicals in 8-point dilution series

= Vehicle controls (DMSO)

= No treatment controls

= Reference chemical #1 (ex. Genistein, 10 μM)

= Reference chemical #2 (ex. Sirolimus, 0.1 μM)

= Reference chemical #3 (ex. Trichostatin A, 1 μM)

Reference Treatments

Used to track assay performance inclusive of cellular response

= Reference RNAs

= Reference Lysates

= Lysis Buffer Blanks

Reference Samples

= Reserved for Sequencing Vendor

Used to track assay performance independent of chemical treatments and responsivity of culture.

Use of Reference Samples in HTTr Screening

- Reference samples are intended to provide objective evaluation(s) of the **technical performance** of an ‘omics assay…**NOT the biological response** of an *in vitro* test system.
  - Use reference treatments for this latter purpose.

- Processed in parallel with test samples → they should be subject to the same manipulations and assay conditions as test samples.

- Implemented in a manner that facilitates monitoring of consistency of transcriptomics assay results generated within studies, across studies, across laboratories and over time.

Reference Samples: History of Use for HTTr

"Early days" (2017-2020) at US EPA:

Name	Description	Observations
Reference Pair #1 (purified RNA)	Takara UHRR (636690) Takara HBRR (636530)	• Comparable to Microarray Quality Control Consortium (MAQC) reference samples (doi: 10.1038/nbt1239). • Finite resource sourced from distinct individuals. • Not optimal for evaluating performance of cell-lysate compatible transcriptomics assays.
Reference Pair #2 (bulk lysates)	**MCF7 Cells** DMSO (0.5%) Treated TSA (1 µM) Treated	• Generated at US EPA • Fewer genes detected compared to Reference Pair #1. • Range of FC values smaller than Reference Pair #1.

US EPA perceived a need to develop **replenishable** human-derived transcriptomics reference samples that are:

- Compatible with multiple assay technologies.
- Available as both purified RNA and cell lysates.
- Yield reproducible fold-change profiles across production batches.

Engineering of Transcriptomics Reference Samples

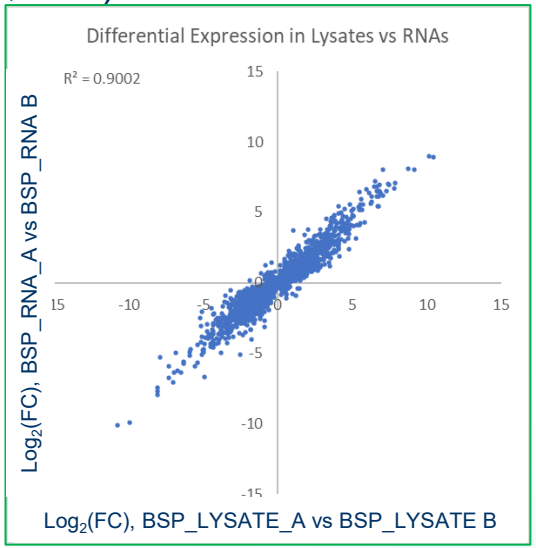
- Paired reference samples were prepared by combining the genetic material from different human-derived cell lines cultured under different conditions.
- Formulated to mimic the performance characteristics of MAQC samples.
- Prepared as both purified RNAs and cell lysates (*BioSpyder, Inc.*)

Sample	# of Genes Detected [a]
BSP_RNA_A	13,962
BSP_RNA_B	13,779
BSP_LYSATE_A	14,919
BSP_LYSATE_B	14,565

Sample Pair	# of Genes in Common [a]
RNA_A & RNA_B	12,881
LYSATE_A & LYSATE_B	13,546

[a] Whole transcriptome TempO-Seq @ 8M mapped reads. Genes with count > 5 considered "detected"

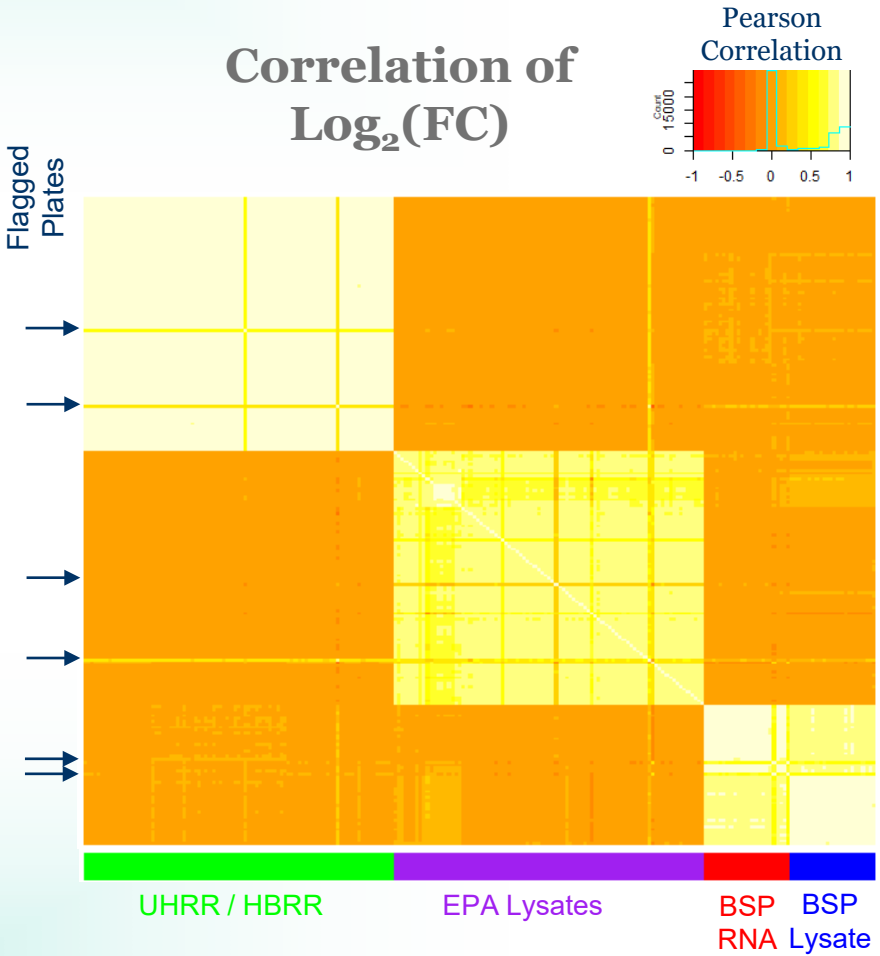
Similar numbers of detected genes in engineered reference samples compared to MAQC or Takara samples.

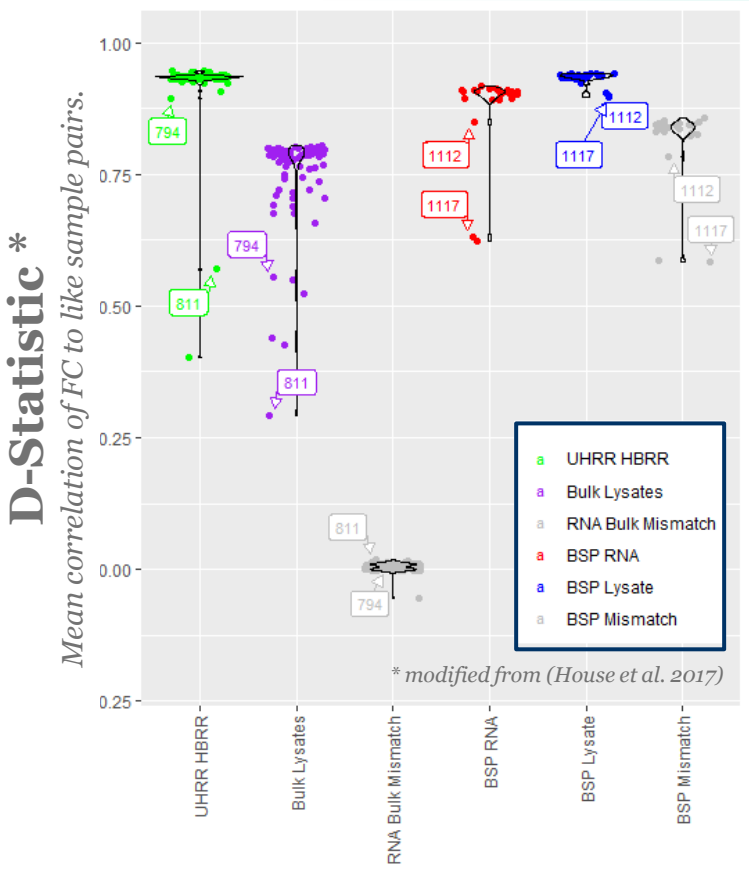


** DESeq2-moderated fold changes*

Expression profiles of BioSpyder RNA and lysates are highly correlated.

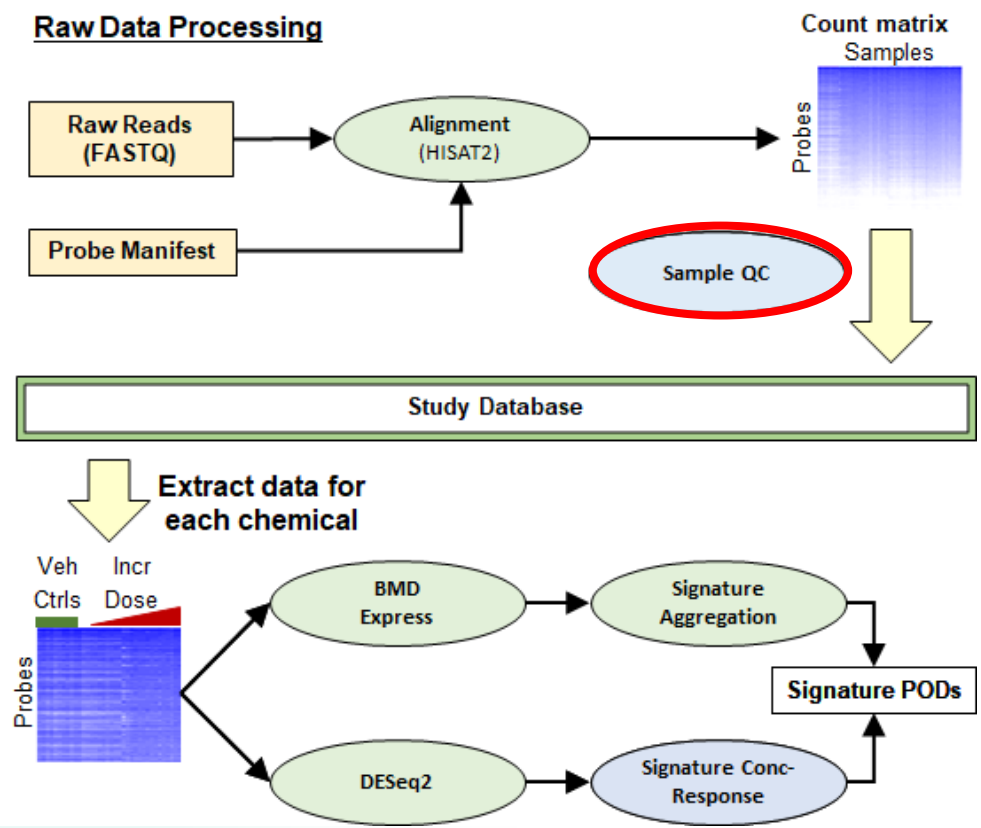
Evaluating HTTr Assay Performance





Plates with potential performance issues flagged for additional scrutiny

HTTr Bioinformatics Pipeline



Primary Goals:

- Reproducible & open source
  - github.com/USEPA/httrpl_pilot
  - github.com/USEPA/CompTox-httrpathway
- Automate and efficiently execute computationally intensive steps.
- Focus on concentration-response modeling and **molecular point-of-departure (mPOD)** determination.
- Store analysis results in a queryable database structure (MongoDB).

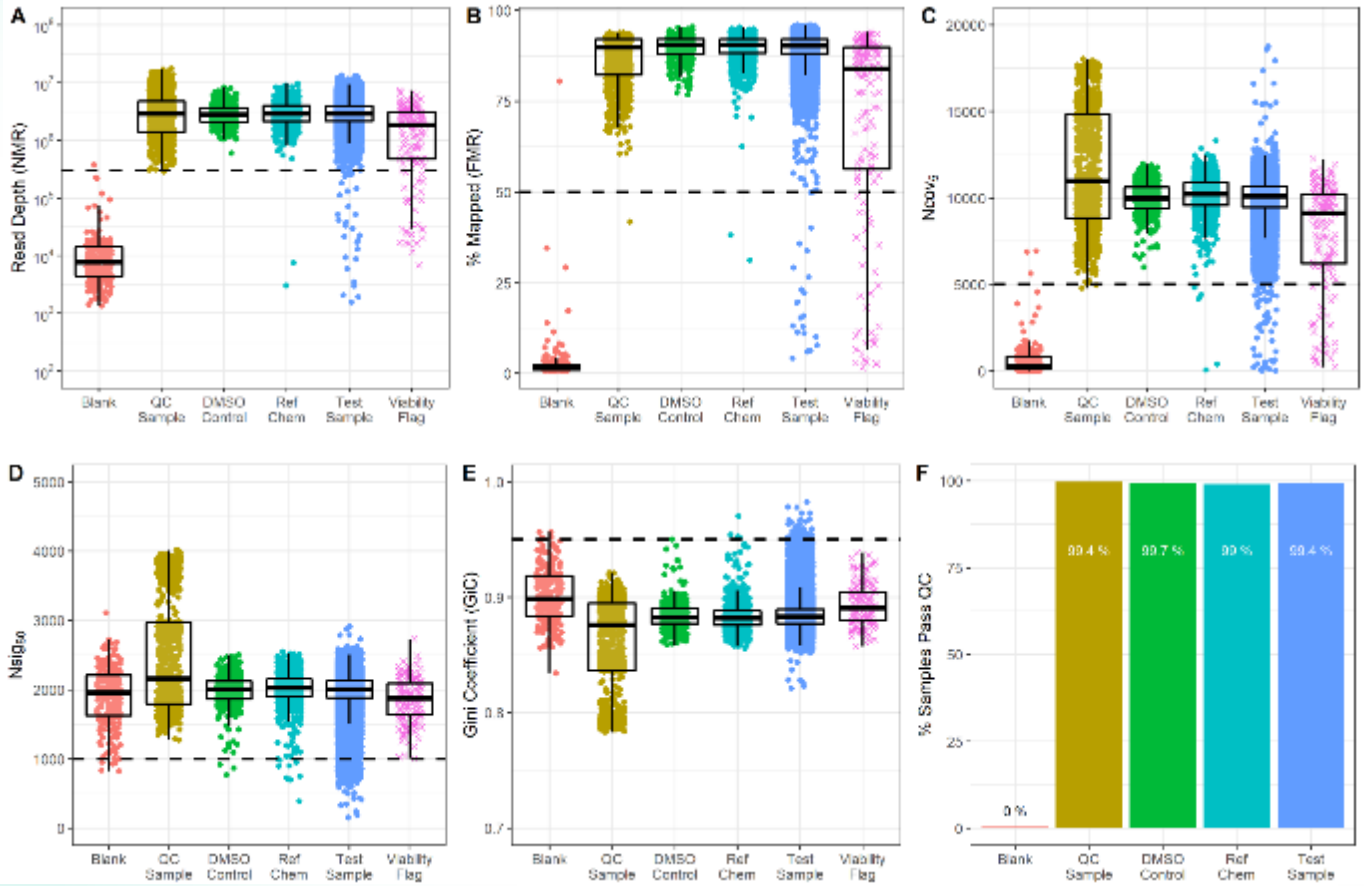
Harrill et al. (2021) DOI: 10.1093/toxsci/kfab009

HTTr Quality Control Metrics

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	

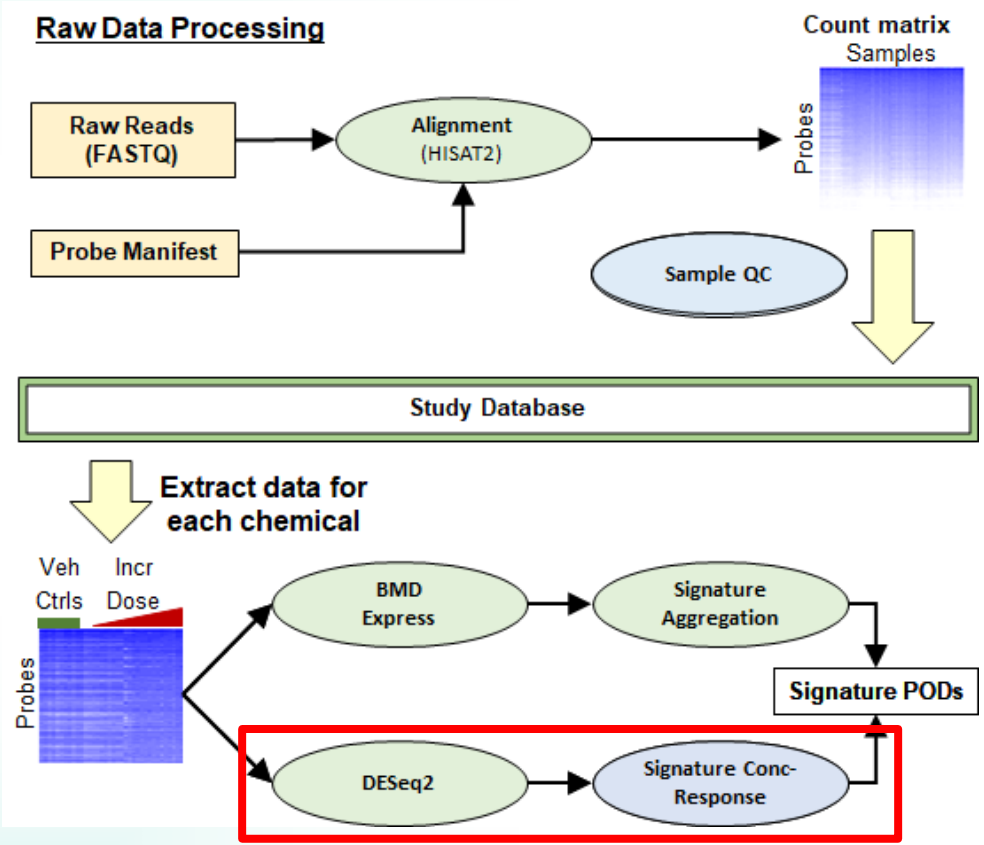
Harrill et al. (2021) DOI: 10.1093/toxsci/kfab009

HTTr QC Results – MCF7 Screen

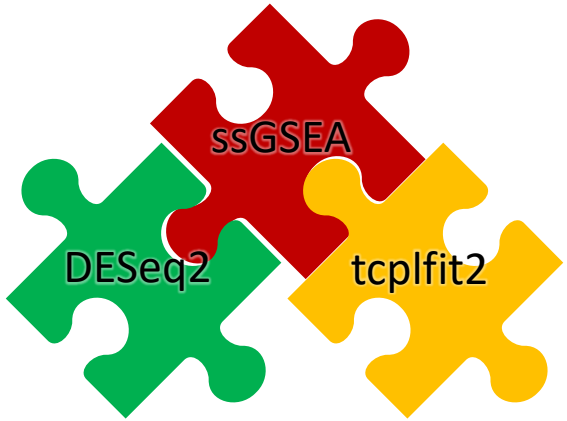


- The screen contained 32,886 TempO-Seq samples.
- None of the lysis buffer blank samples passed the QC criteria.
- > 99% of test samples were of acceptable quality based on QC criteria.
- In some cases, samples flagged for viability did not fail other QC criteria.

Signature Concentration-Response Modeling



Signature Concentration-Response Modeling



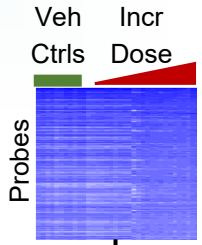
github.com/USEPA/CompTox-httrpathway

(Richard Judson)

Method intended to address coordinated changes in expression in genes belonging to the same gene set / signaling pathway.

Signature Scoring Procedure

Count data per chemical



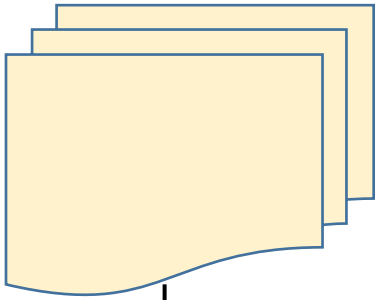
Estimate fold-changes for all genes

DESeq2

Compute signature scores from all gene expression changes

ssGSEA

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

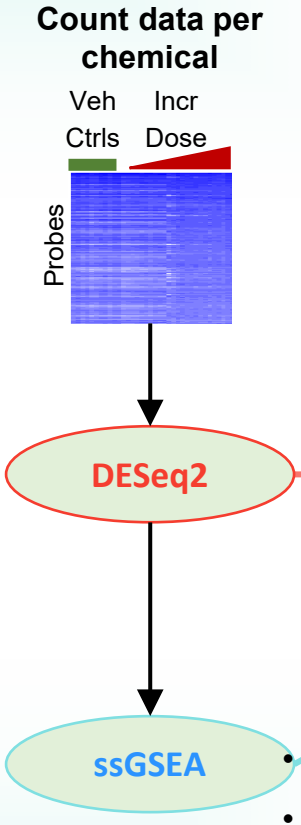


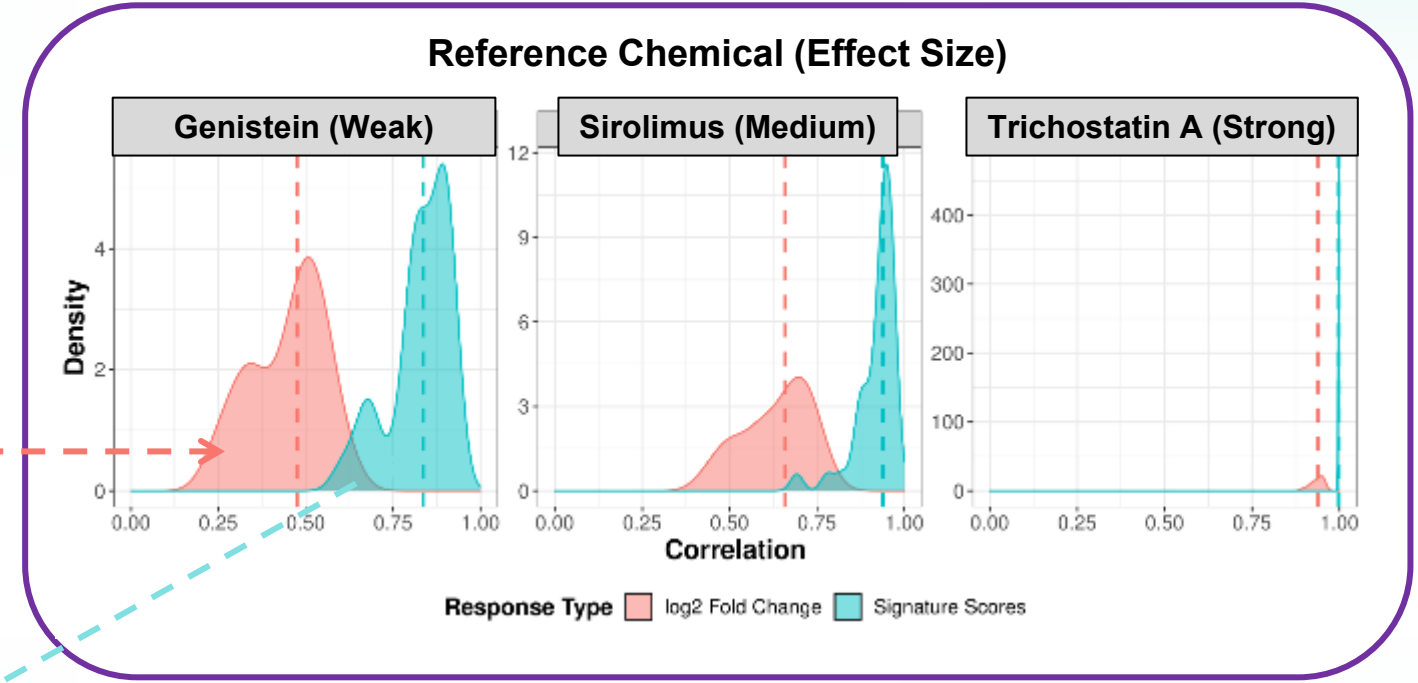
- **Bioplanet** *(Huang, et al. Front Pharmacol 2019)*
- **CMap** *(Subramanian, et al. Cell 2017)*
- **DisGeNET** *(Pinero, et al. Database 2015)*
- **MSigDB** *(Liberzon, et al. Cell Syst 2015)*

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

- Score coordinated responses at each concentration
- Test for multiple genes in a signature enriched among most extreme fold-changes

Signature Scoring of Reference Treatments

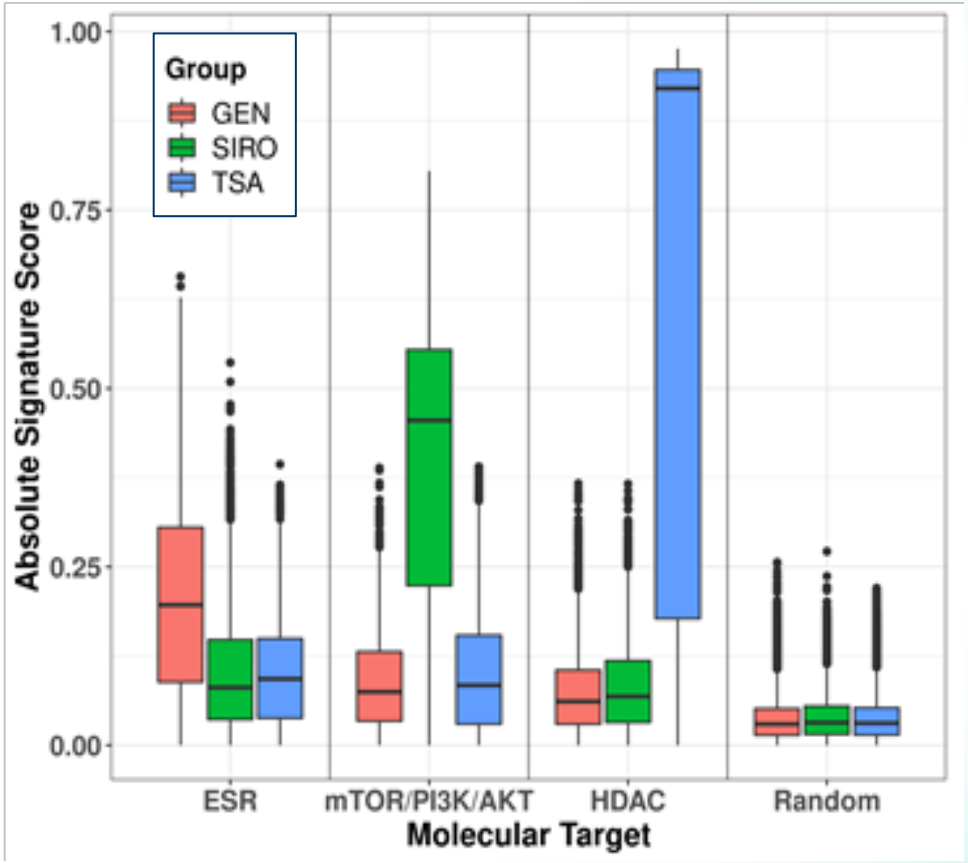




- Differential expression analysis of 3 reference chemical exposures repeated 37 times (MCF7)
- Computed distribution of correlations between each repeat analysis
- **Signature scores have higher reproducibility than fold-changes, especially for weaker effect sizes**

Signature Scoring Identified Expected Biology

- Reference treatments produced higher absolute signature scores for signatures associated with primary mechanisms of action.
- The expected biology was identified!
- Reference treatments did not produce higher absolute signature scores in a set of synthetic “random” signatures.



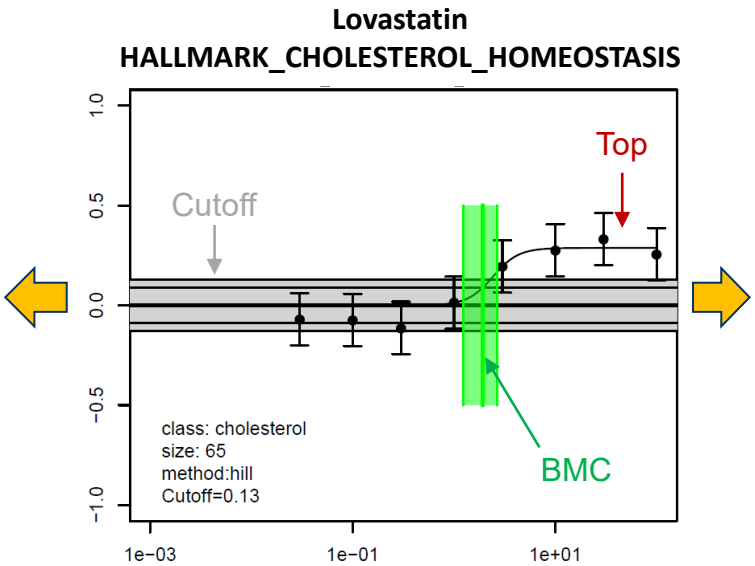
Signature Score Concentration-Response Modeling

Concentration response modeling of signature scores using ***tcplfit2*** (github/USEPA/CompTox-ToxCast-tcplFit2/)

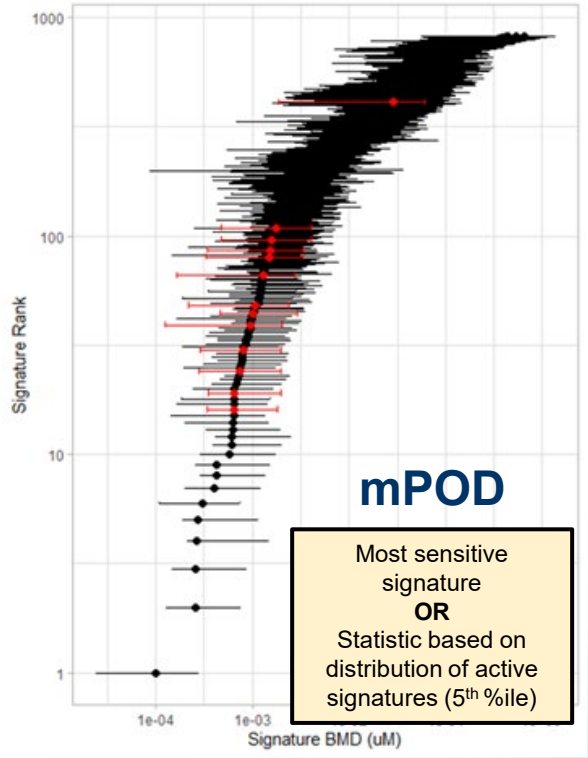
Signed, Scaled Area Under the Curve (ssAUC)

[3-$\log_{10}$(BMC)]
X
| Top / Cutoff |
X
Sign (Top)

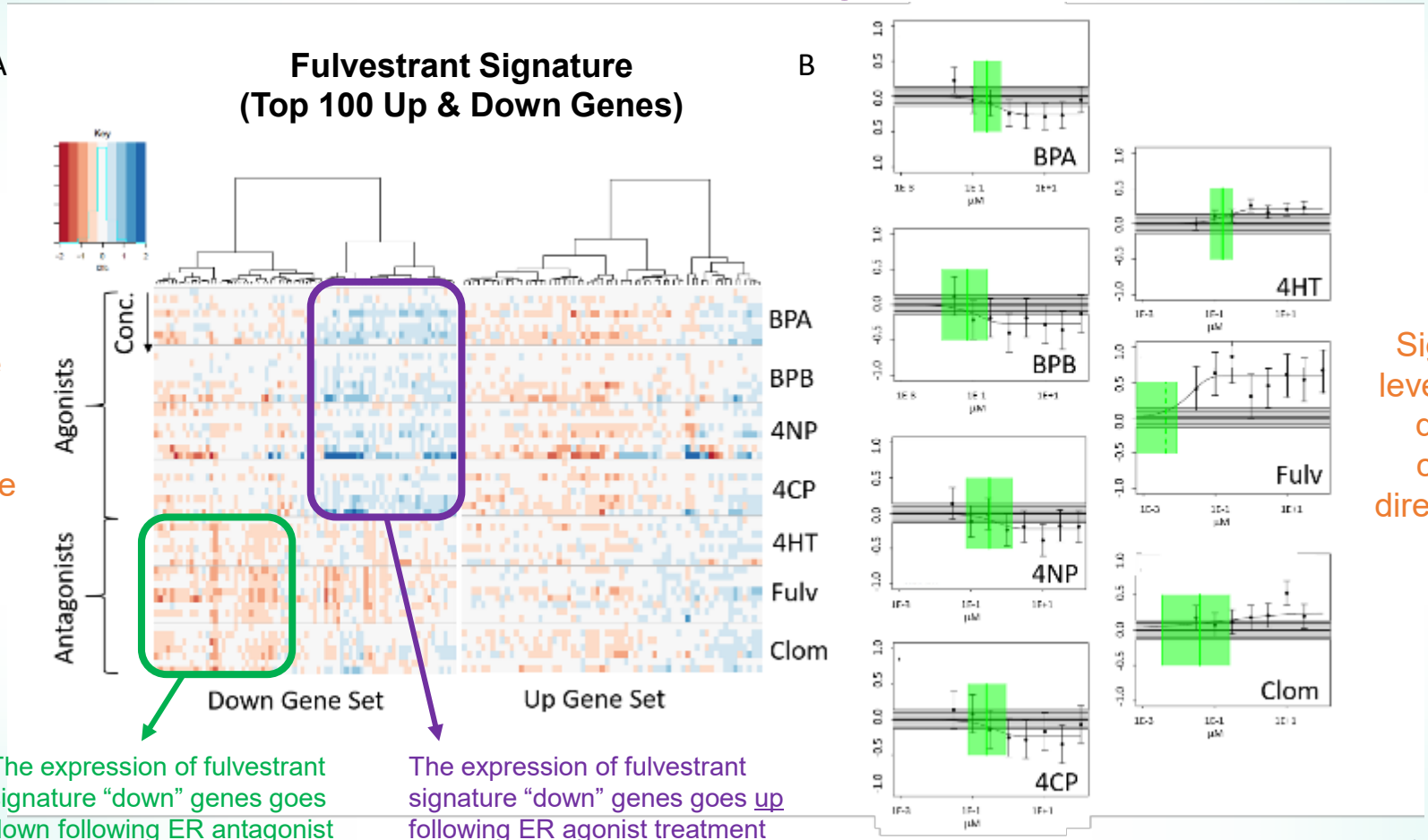
Used to discern mechanism



Ranking of Active Signatures



Signature Scores - Directionality of Response



These gene level data are noisy!

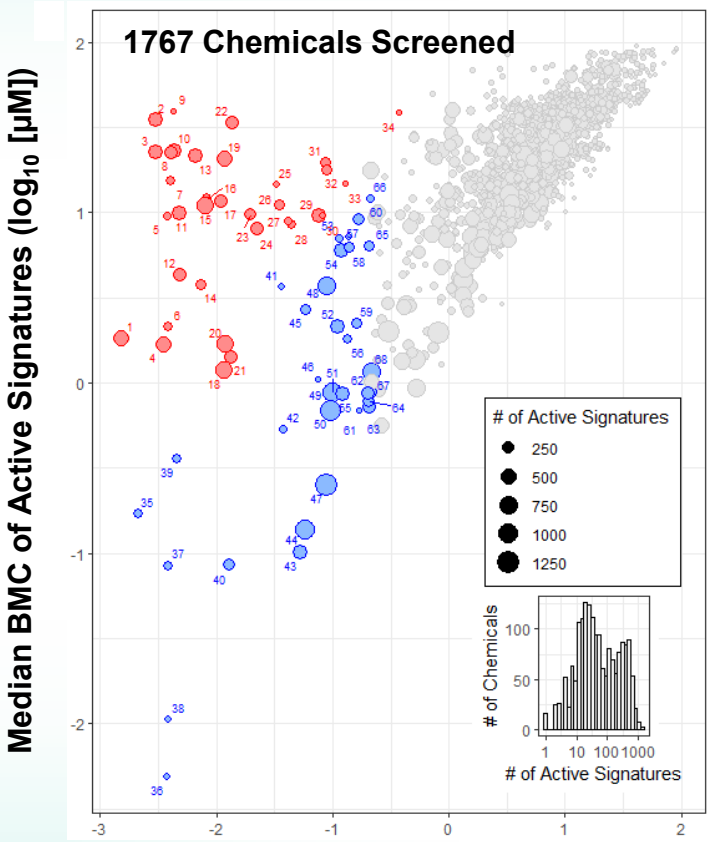
Signature level results display correct directionality !

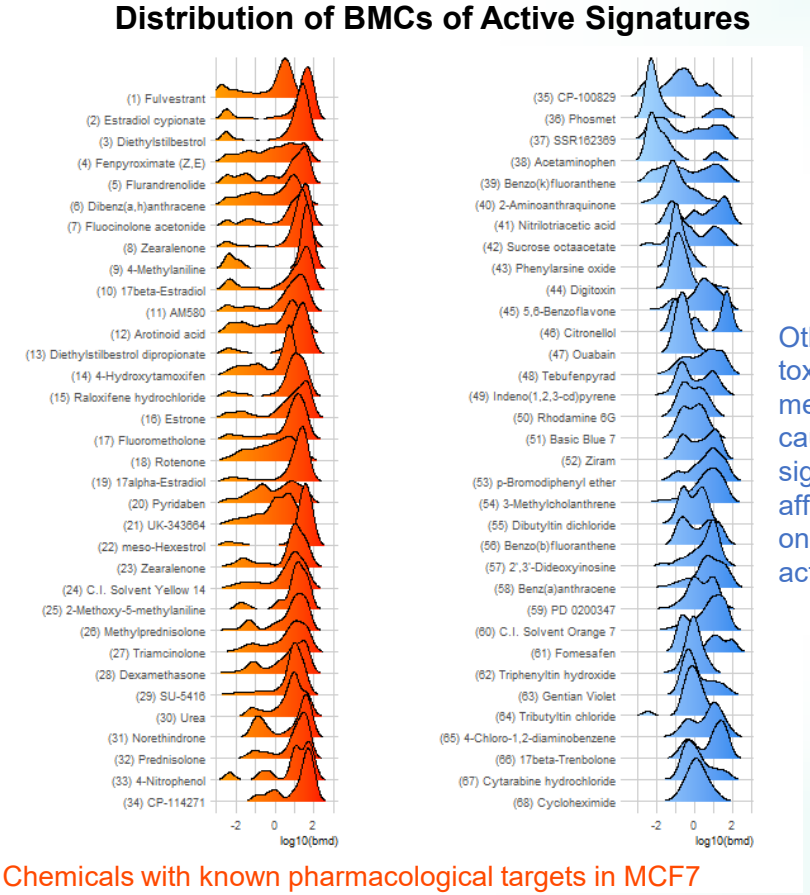
The expression of fulvestrant signature “down” genes goes down following ER antagonist treatment

The expression of fulvestrant signature “down” genes goes up following ER agonist treatment

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

HTTr Screening Results

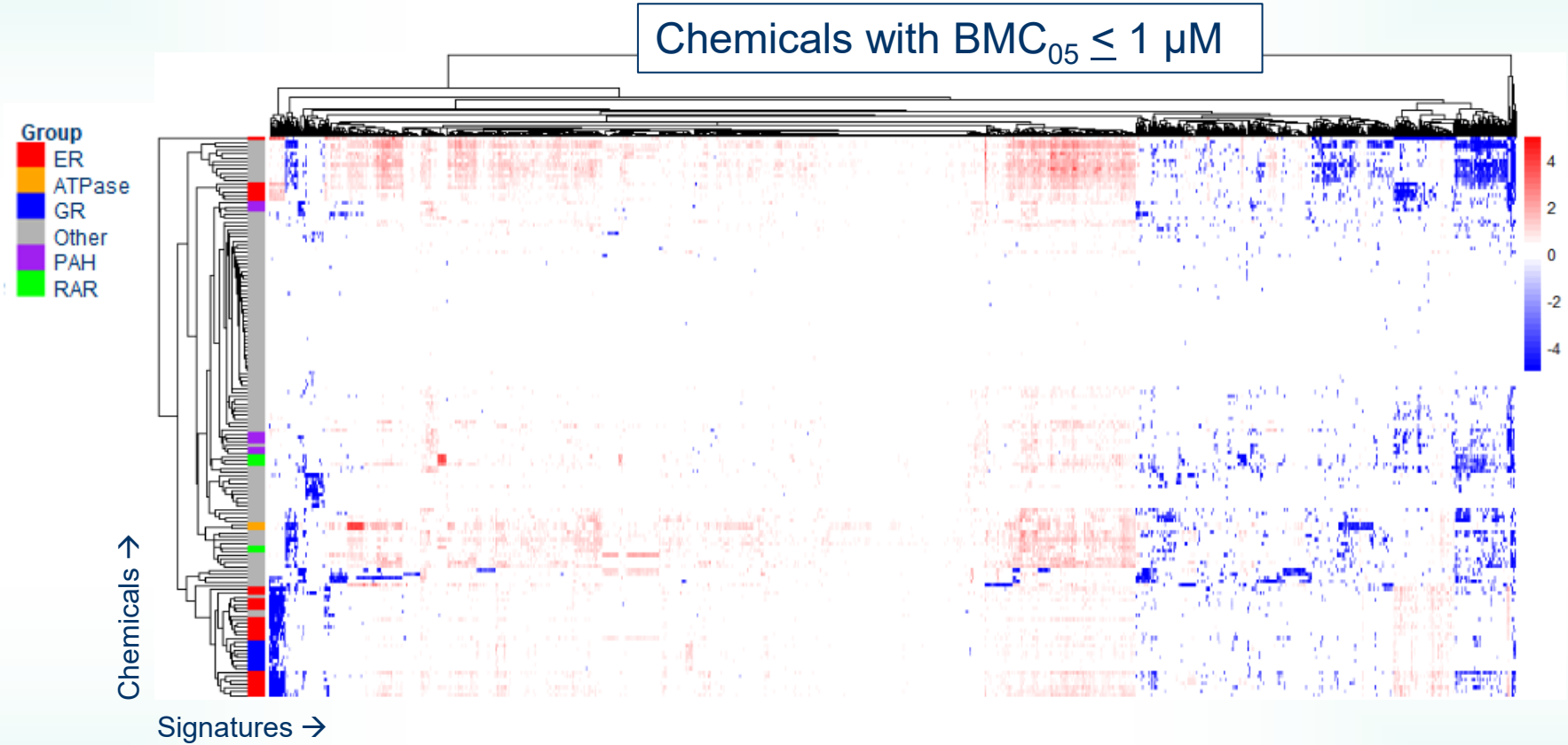




Other potent toxicants (organo-metallics, dyes, etc) cause many signatures to be affected near the onset of biological activity.

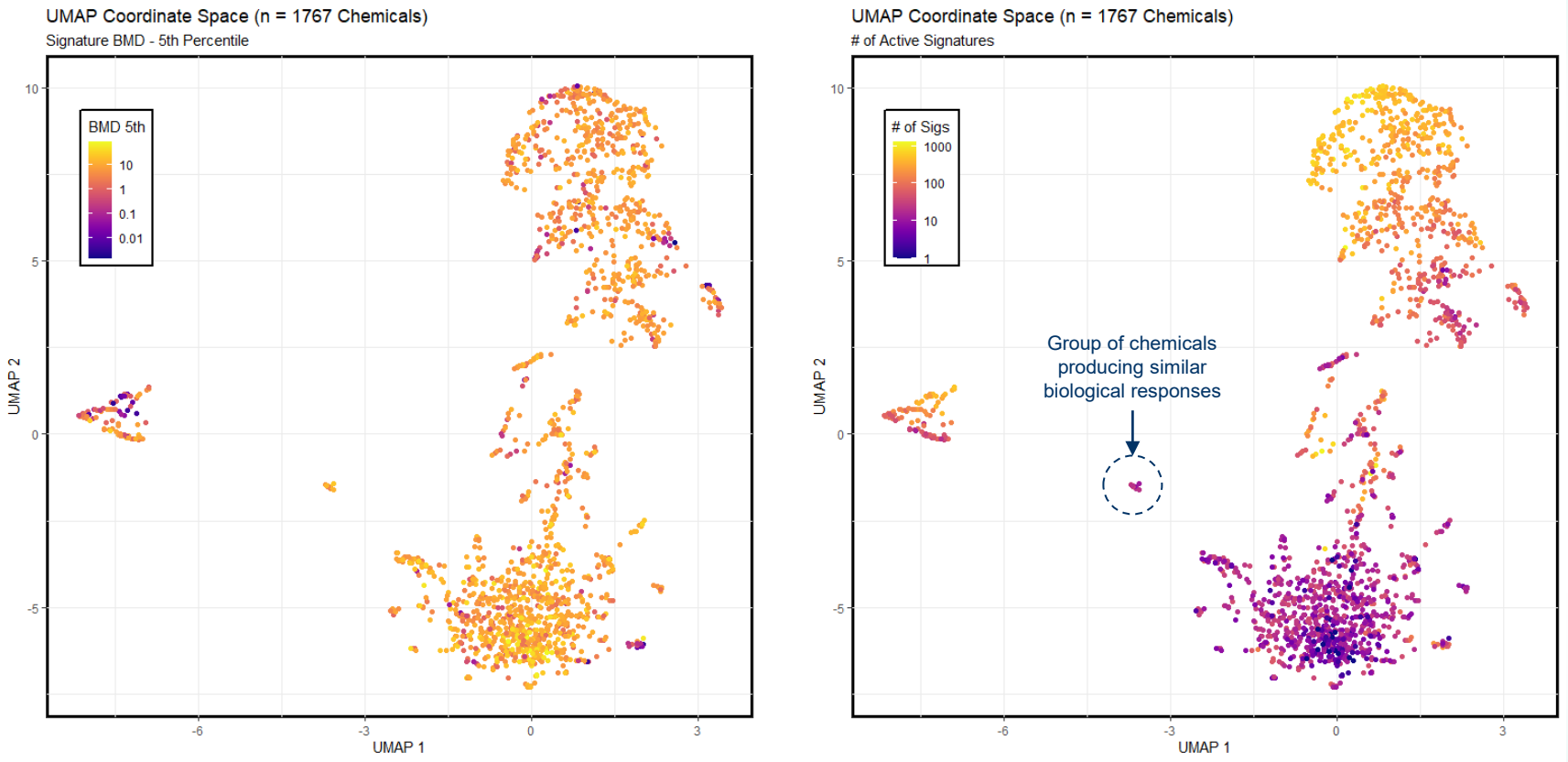
Chemicals with known pharmacological targets in MCF7 cells show an “early wave” of biological activity.

Mechanistic Clustering Using Signature ssAUC



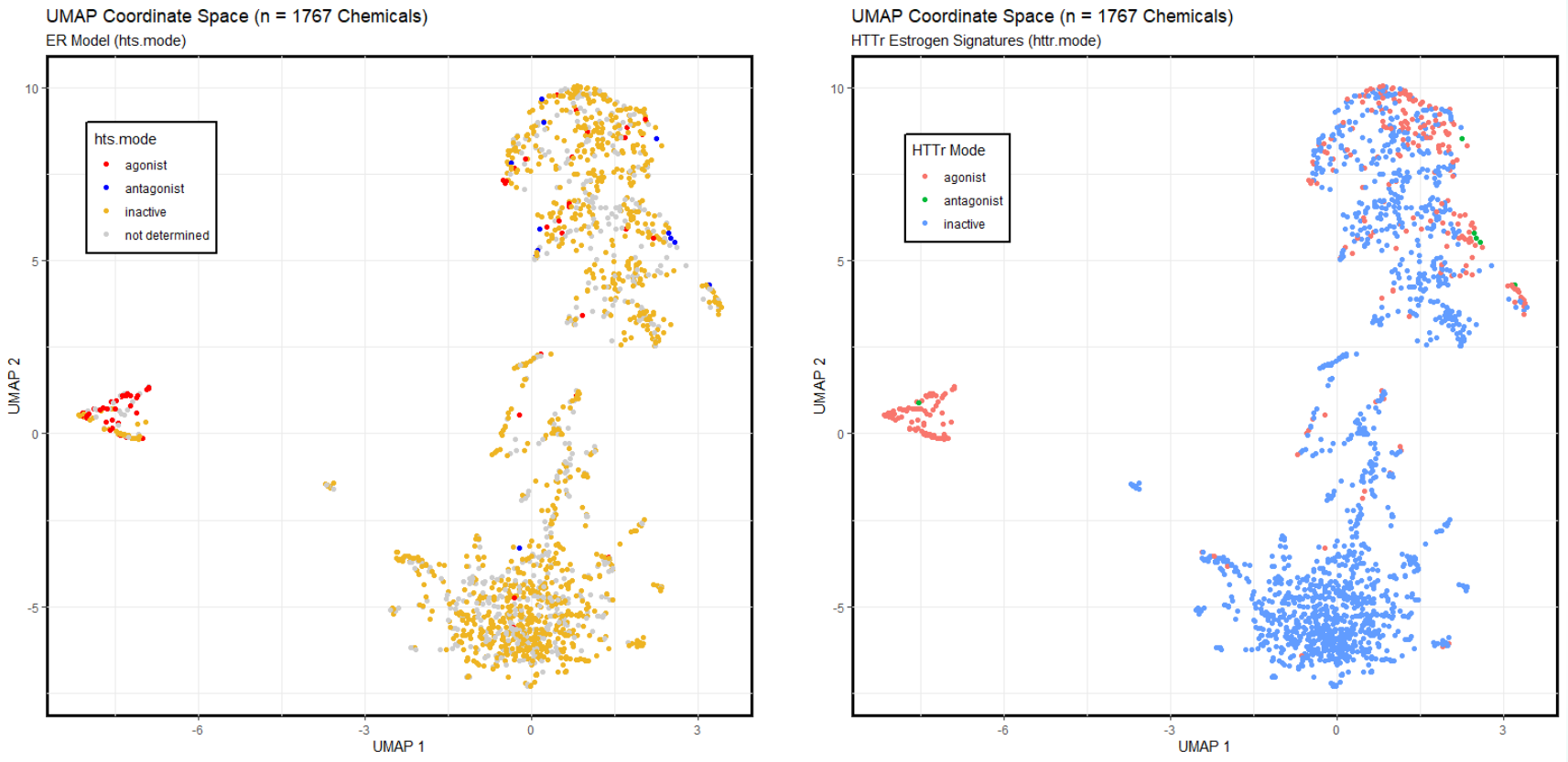
Potent chemicals with known specificity for a molecular target expressed in MCF7 cells tend to cluster together when using signature scores (ssAUC) as the response metric.

Exploring Similarities in Chemical Response (1)



Data visualization tools (UMAP) help identify chemicals that produce similar responses.

Exploring Similarities in Chemical Response



Similar classification of chemicals as ER agonists or ER antagonists using HTS or HTTr signatures.

OECD Omics Reporting Framework Project
(2018 – Present)

To develop frameworks for the standardisation of reporting of 'omics data generation and analysis, to ensure that all of the information required to understand, interpret and reproduce an 'omics experiment and its results are available.

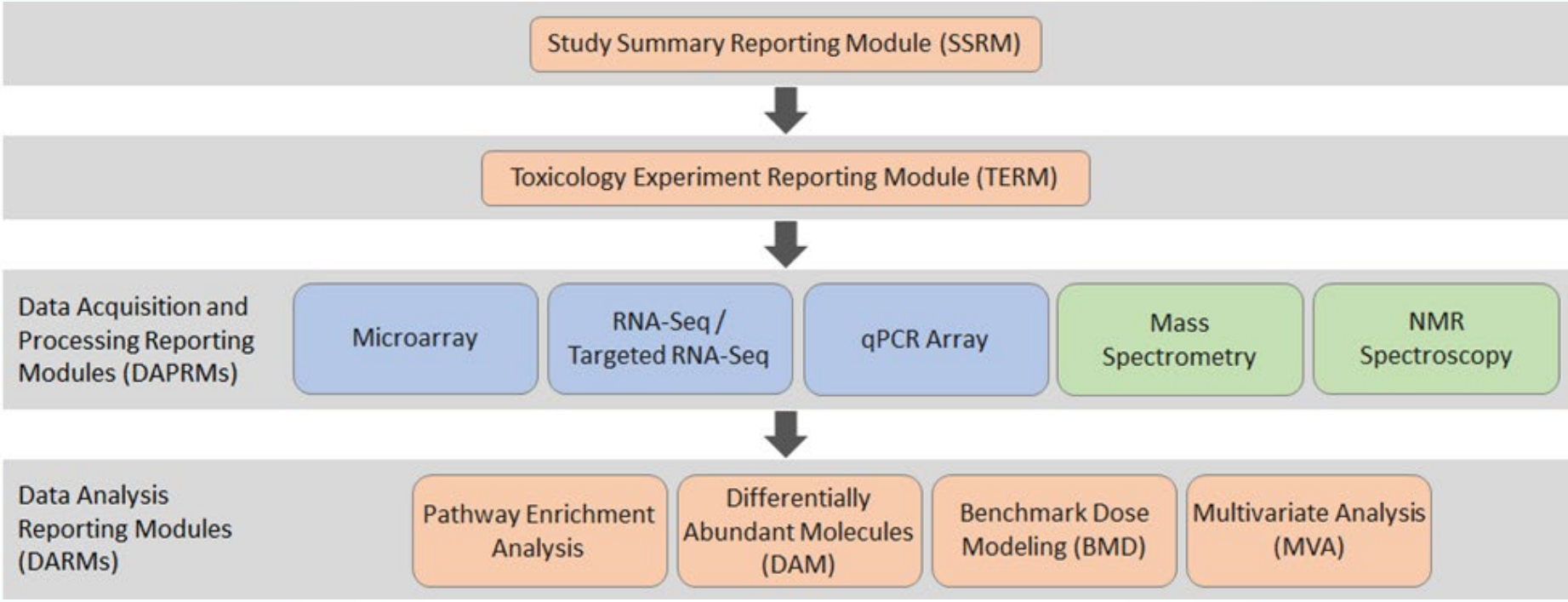
Purpose: to ensure that sufficient information is available to enable an evaluation of the quality of the experimental data and interpretation, and support reproducibility.

NOT to stipulate the methods of data analysis or interpretation….**Rather**, provide guidance on reporting of information that fosters transparency and reproducibility.

Project Name	Project Leads
Metabolomics Reporting Framework (MRF)	Mark Viant (Univ. Birmingham, UK)
Transcriptomics Reporting Framework (TRF)	Joshua Harrill (US EPA) Carole Yauk (University of Ottawa) Matt Meier (Health Canada)
OECD Secretariat	Magda Sachana

Harmonized **O**ECD **O**mics **R**eporting **F**ramework (OORF)

Modular Structure of OORF



- Both TRF and MRF
- TRF only
- MRF only

Each module has:

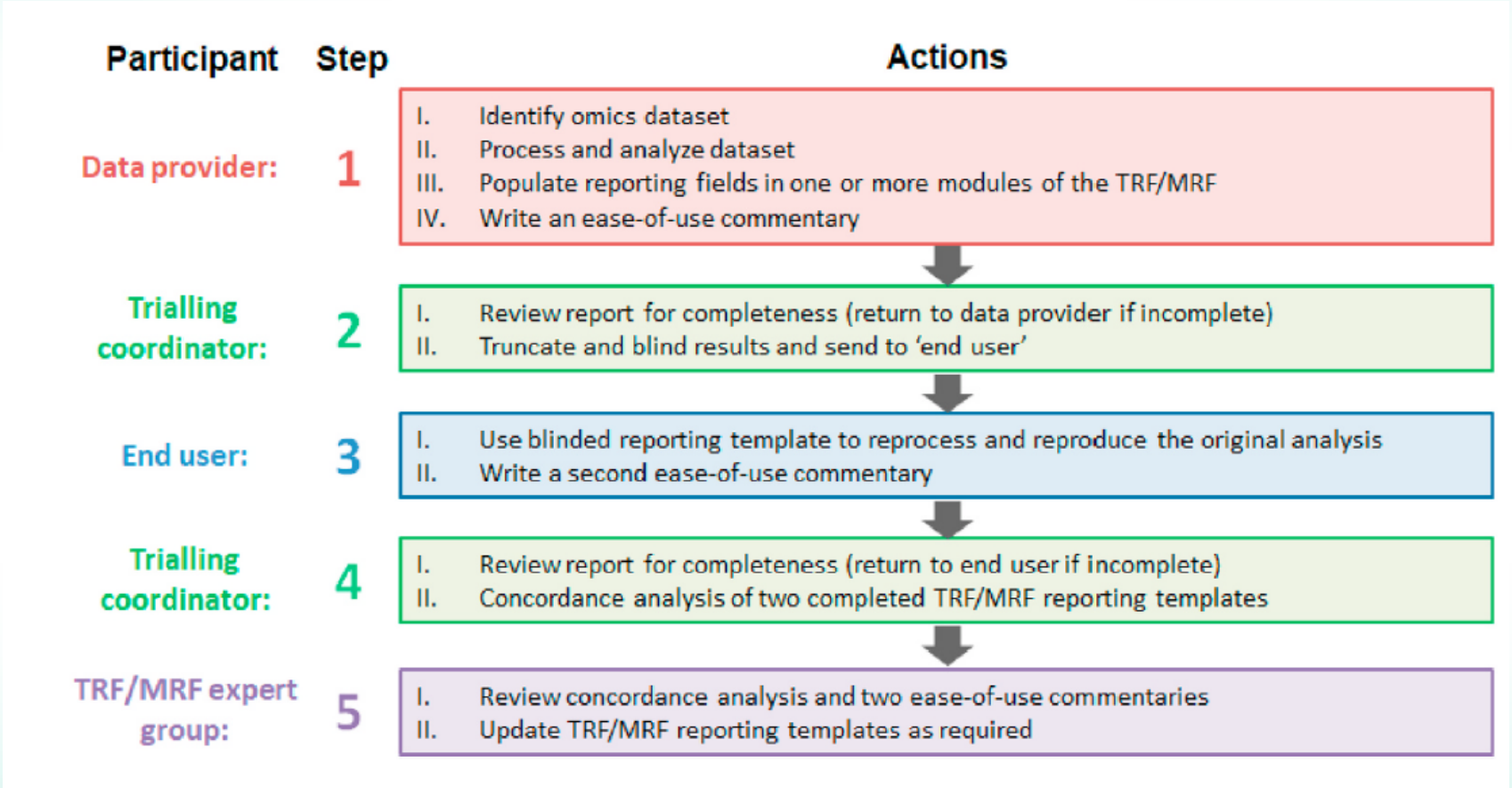
1) a reporting template (Excel)
2) a narrative guidance

OORF Reporting Templates

1	**8. Data Analysis Reporting Module (DARM) for Detection of Enriched Biological Pathways**			*INSTRUCTIONS TO DATA PROVIDERS: Use Column D for data entry. Text in italics are instructions (e.g. reporting tips) for the data submitter for individual reporting fields. Text in bold italics are instructions for the data submitter that are applicable to all reporting elements in a section. For additional clarification regarding a reporting field, please refer to the corresponding guidance document for this reporting template.*	
2	Red = Required				
3	Blue = Optional				
4					
5	REPORTING CATEGORY	REPORTING ELEMENT	REQUIRED / OPTIONAL	INPUT	NOTES AND INSTRUCTIONS FOR DATA PROVIDERS
6	8.1. Software Documentation	**Software and Documentation**			
7		Software	Required		
8		Operating System	Required		Specify if it is a web-based application and provide the URL.
9		Additional Libraries used	Required		
10		Software Availability	Optional		
11	8.2. Description of Data Used as	**Data Description**			
12		Data used as input	Required		the enrichment analysis; e.g., a list of differentially abundant molecules; genes by
13		Methods used to produce input	Required		expressed genes or differentially abundant metabolites for each contrast in the
14		Pre-filtering of input data	Required		cutoffs, statistical tests, etc). If no prefiltering was done, report "NA".
15		Pre-processing and/or normalization	Required		transformation of counts; normalizing to counts per million; calculation of
16		Background set(s) used	Required		Report the background set of genes/metabolites, and how it was established.
17	8.3. Contrasts for Which	**Contrasts**			
18		Contrasts	Required		compared (e.g., treatment groups versus control).
19	8.4. Database of Pathways or Gene	**Biological Entity or Biological Set**			
20		Biological Entity or Biological Set	Required		Report the annotation source (e.g., GO, KEGG, WikiPathways, MSigDB, IPA, etc).
21		Species Name	Required		Report the species name.
22		Version or Date of Biological Set	Required		additional information on the origin of the annotations (i.e., the date the
23	Enriched Pathways	**Data Description**			
24		Statistical Test Performed to Identify	Required		
25		Statistical Threshold Applied	Required		
26		Multiple Testing Correction Method	Required		
27		Additional Filtering	Required		
28	8.6. Outputs	**Outputs and Supporting Files**			
29		Output and Supporting Files	Required		supplementary files associated with the analysis described in this reporting module.
30					

- Prompts data providers to report details of their experiment.
- Links to narrative guidance with descriptions of what type of information to enter in each field.

Refining the OORF with Paired Trials



OECD Omics Reporting Framework Project

Regulatory Toxicology and Pharmacology 125 (2021) 105020

Contents lists available at ScienceDirect

Regulatory Toxicology and Pharmacology

journal homepage: www.elsevier.com/locate/yrtph

Commentary

Progress towards an OECD reporting framework for transcriptomics and metabolomics in regulatory toxicology

Joshua A. Harrill [a,1,*], Mark R. Viant [b,c,**,1], Carole L. Yauk [d,1,***], Magdalini Sachana [e], Timothy W. Gant [f], Scott S. Auerbach [g], Richard D. Beger [h], Mounir Bouhifd [i], Jason O'Brien [j], Lyle Burgoon [k], Florian Caiment [l], Donatella Carpi [m], Tao Chen [h], Brian N. Chorley [a], John Colbourne [b,c], Raffaella Corvi [m], Laurent Debrauwer [n,o], Claire O'Donovan [p], Timothy M. D. Ebbels [q], Drew R. Ekman [r], Frank Faulhammer [s], Laura Gribaldo [m], Gina M. Hilton [t], Stephanie P. Jones [j], Aniko Kende [u], Thomas N. Lawson [c], Sofia B. Leite [m], Pim E.G. Leonards [v], Mirjam Luijten [w], Alberto Martin [i,2], Laura Moussa [x], Serge Rudaz [y,z,aa], Oliver Schmitz [ab], Tomasz Sobanski [i], Volker Strauss [s], Monica Vaccari [ac], Vikrant Vijay [h], Ralf J.M. Weber [b,c], Antony J. Williams [a], Andrew Williams [ad], Russell S. Thomas [a], Maurice Whelan [m]

https://www.oecd.org/chemicalsafety/testing/omics.htm

- Genesis and progress towards the OORF detailed in this publication.
- Early draft available at OECD omics website.
- Review and approval by OECD Extended Advisory Group for Molecular Screening and Toxicogenomics (EAGMST) in Q3 2022!

Summary

- **Innovations in High-Throughput Transcriptomics Screening:**
  - Targeted RNA-Seq assay.
  - Scalable laboratory workflows for chemical exposure and lysate generation.
  - Reproducible, open-source data analysis pipeline(s).
  - Improvements to ToxCast pipeline (*tcpl*) concentration-response modeling software.
  - Novel approach for signature level concentration-response analysis.
  - Data visualization techniques for exploring chemical / biological response similarity.

- **Development of transcriptomics reference materials.**

- **Reporting frameworks for toxicology studies involving omics technologies.**

References

1. Thomas RS, et al. The Next Generation Blueprint of Computational Toxicology at the U.S. Environmental Protection Agency. Toxicol Sci. 2019 Jun 1;169(2):317-332. doi: 10.1093/toxsci/kfz058. PMID: 30835285
2. Richard AM, et al. ToxCast Chemical Landscape: Paving the Road to 21st Century Toxicology. Chem Res Toxicol. 2016 Aug 15;29(8):1225-51. doi: 10.1021/acs.chemrestox.6b00135. Epub 2016 Jul 20. PMID: 27367298.
3. Yeakley JM, et al. A trichostatin A expression signature identified by TempO-Seq targeted whole transcriptome profiling. PLoS One. 2017 May 25;12(5):e0178302. doi: 10.1371/journal.pone.0178302. PMID: 28542535
4. Harrill JA, et al. High-Throughput Transcriptomics Platform for Screening Environmental Chemicals. Toxicol Sci. 2021 Apr 27;181(1):68-89. doi: 10.1093/toxsci/kfab009. PMID: 33538836.
5. MAQC Consortium. The MicroArray Quality Control (MAQC) project shows inter- and intraplatform reproducibility of gene expression measurements. Nat Biotechnol. 2006 Sep;24(9):1151-61. doi: 10.1038/nbt1239. PMID: 16964229.
6. Huang R, 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 Apr 26;10:445. doi: 10.3389/fphar.2019.00445. PMID: 31133849.
7. Subramanian A, et al. A Next Generation Connectivity Map: L1000 Platform and the First 1,000,000 Profiles. Cell. 2017 Nov 30;171(6):1437-1452.e17. doi: 10.1016/j.cell.2017.10.049. PMID: 29195078.

References

1. Piñero J, et al. DisGeNET: a discovery platform for the dynamical exploration of human diseases and their genes. Database (Oxford). 2015 Apr 15;2015:bav028. doi: 10.1093/database/bav028. PMID: 25877637.
2. Liberzon A, et al. The Molecular Signatures Database (MSigDB) hallmark gene set collection. Cell Syst. 2015 Dec 23;1(6):417-425. doi: 10.1016/j.cels.2015.12.004. PMID: 26771021.
3. Barbie DA, et al. Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. Nature. 2009 Nov 5;462(7269):108-12. doi: 10.1038/nature08460. Epub 2009 Oct 21. PMID: 19847166.
4. Sheffield T, et al. tcplfit2: an R-language general purpose concentration-response modeling package. Bioinformatics. 2021 Nov 15:btab779. doi: 10.1093/bioinformatics/btab779. Epub ahead of print. PMID: 34791027.
5. Harrill JA, et al. Progress towards an OECD reporting framework for transcriptomics and metabolomics in regulatory toxicology. Regul Toxicol Pharmacol. 2021 Oct;125:105020. doi: 10.1016/j.yrtph.2021.105020. Epub 2021 Jul 29. PMID: 34333066.

Acknowledgements

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

Logan Everett
Imran Shah
Richard Judson
Katie Paul Friedman
Derik Haggard
Joseph Bundy
Laura Taylor
Bryant Chambers
Jesse Rogers
Thomas Sheffield
Woody Setzer
Johanna Nyffeler
Clinton Willis
Megan Culbreth
Gabrielle Byrd
Felix Harris
Dan Hallinger
Terri Fairley

BioSpyder

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

Omics Reporting Framework Contributors

Leadership Team

Joshua Harrill
Carole Yauk
Mark Viant
Matthew Meier
Magda Sachana

EAGMST Co-chairs

Rusty Thomas
Maurice Whelan

Tim Gant
Scott Auerbach
Rick Beger
Mounir Bouhifd
Jason O-Brien
Lyle Burgoon
Florian Caiment
Donatella Carpi
Tao Chen
Brian Chorley
John Colbourne
Raffaella Corvi
Claire O'Donovan
Laurent Debrauwer
Timothy Ebbels
Drew Ekman
Frank Faulhammer
Laura Gribaldo
Gina Hilton
Stephanie Jones
Aniko Kende
Thomas Lawson
Sofia Batista-Leite
Pim Leonards
Mirjam Luijten
Alberto Martin
Laura Moussa
Serge Rudaz
Oliver Schmitz
Tomasz Sobanski
Volker Strauss
Monica Vaccari
Vikrant Vijay
Ralf Weber
Antony Williams
Andrew Williams