Determining points of departure from multivariate experimental data

British Society for Toxicology
April 2021
Cardiff, UK

Imran Shah
Center for Computational Toxicology & Exposure
Office of Research and Development

The views expressed in this presentation are those of the author[s] and do not necessarily reflect the views or policies of the U.S. Environmental Protection Agency.

The views expressed in this presentation are mine and do not necessarily reflect the views or policies of the U.S. Environmental Protection Agency.

&

I have no conflict of interest with the work included in this presentation.

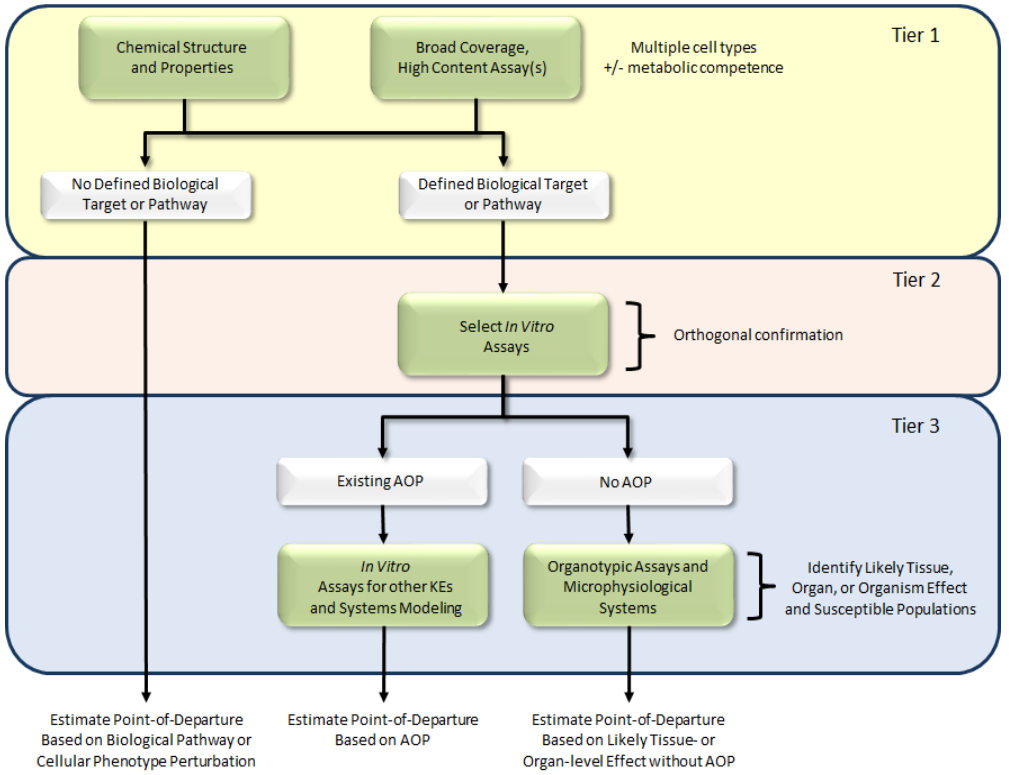
Outline

- New approach methodologies (NAMs)
- High-throughput transcriptomics (HTTr)
- Multivariate analysis via connectivity mapping with signatures
- NAM-based point of departure using signature concentration-response analysis

Tiered Hazard Evaluation

- New approach methodologies (NAMs) are any technology, methodology, approach, or combination of methods that can provide information about chemical hazard and point of departure (POD) without using whole animals
- NAMs based on high-throughput profiling (HTP) assays are flexible, portable and cost-efficient platforms to comprehensively evaluate the potential effects of chemical exposure
- Two types of HTP assays:
  - High-throughput transcriptomics (HTTr)
  - High-throughput phenotypic profiling (HTPP)
- We are evaluating HTPP and HTTr to develop NAMs
  - For hazards: identify putative targets and pathways
  - For POD: estimate *in vitro* POD associated with hazards

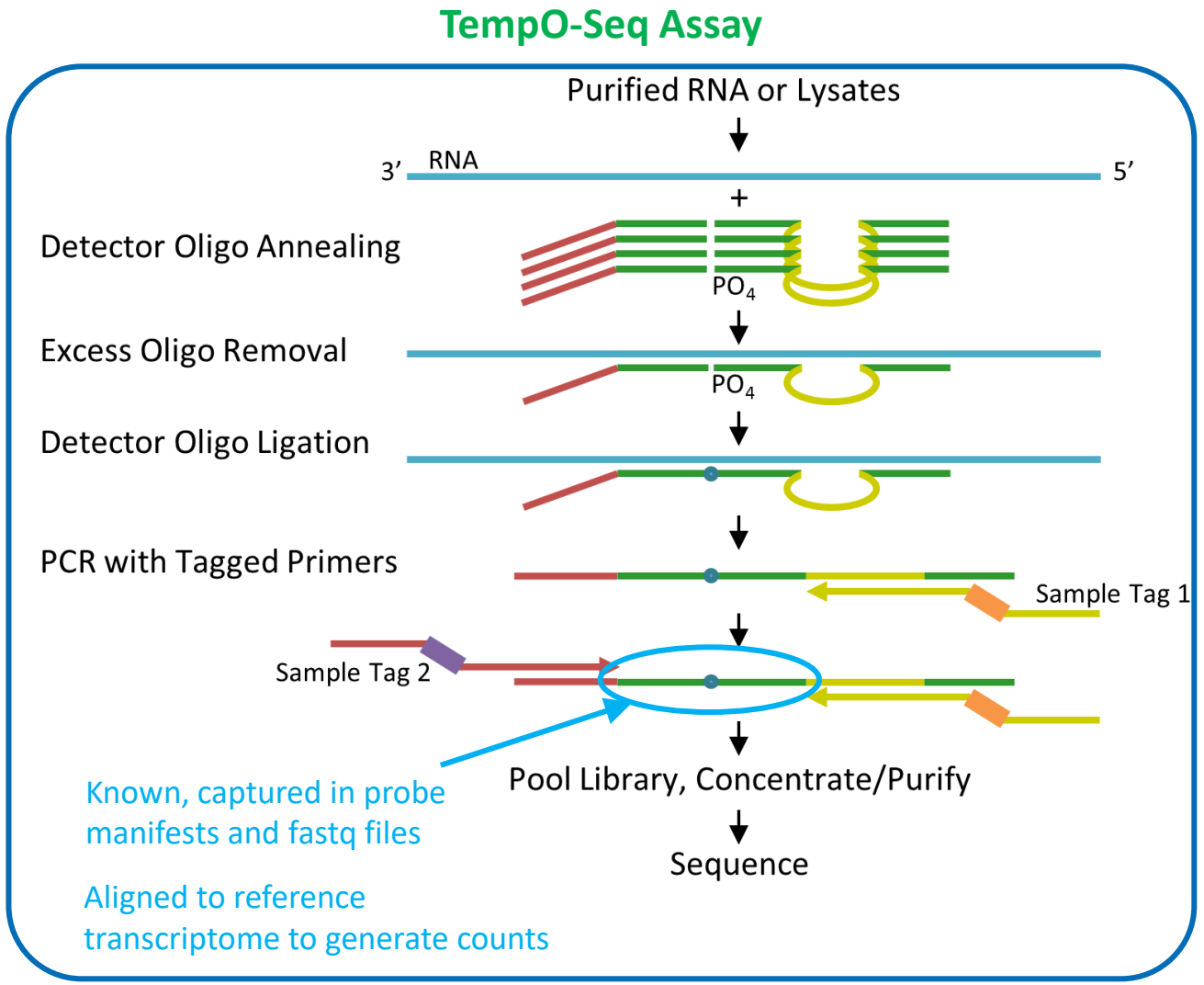
"CompTox Blueprint"



Thomas *et al.*, 2019

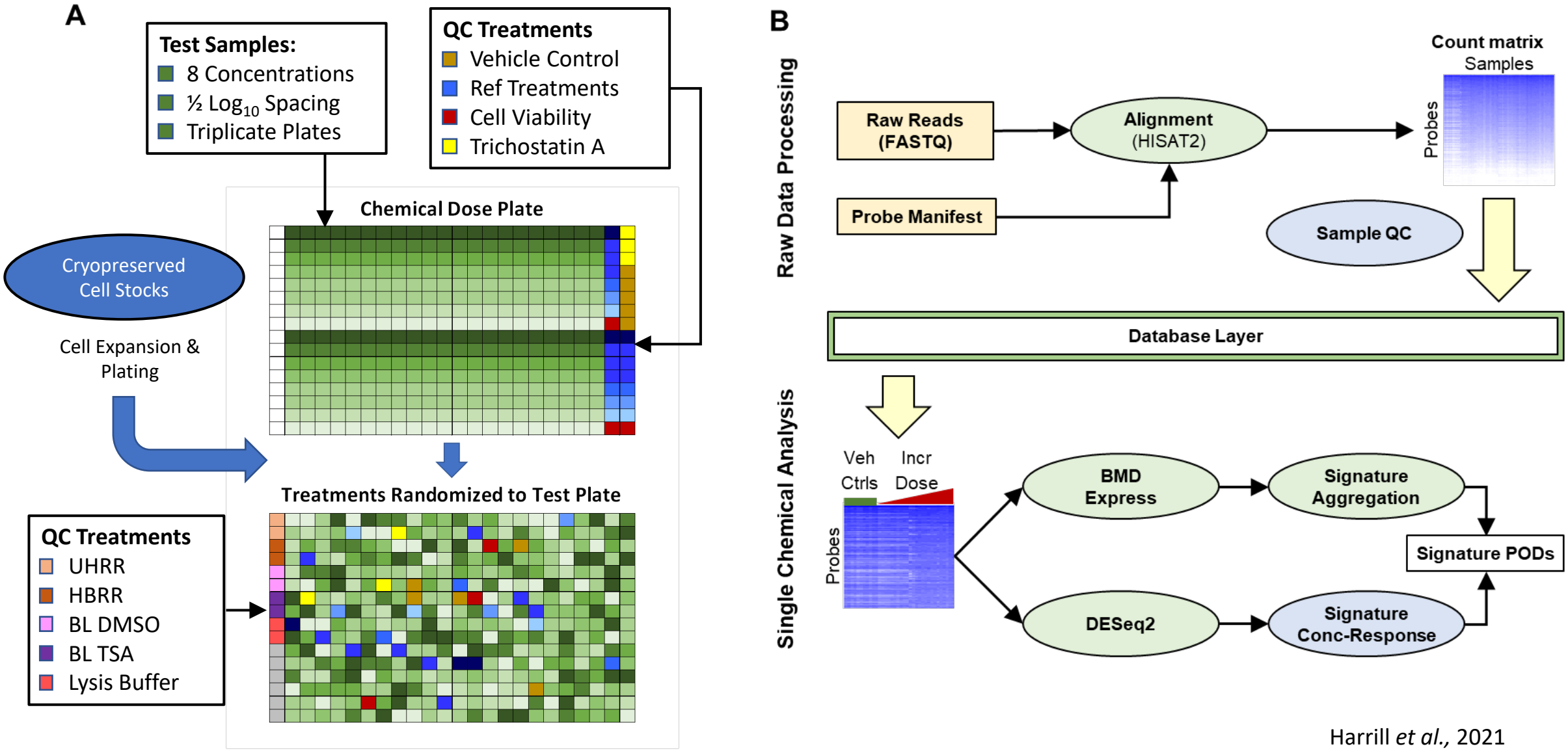
Templated Oligo with Sequencing Readout (TempO-Seq)

- 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



Yeakley, et al. PLoS ONE 2017

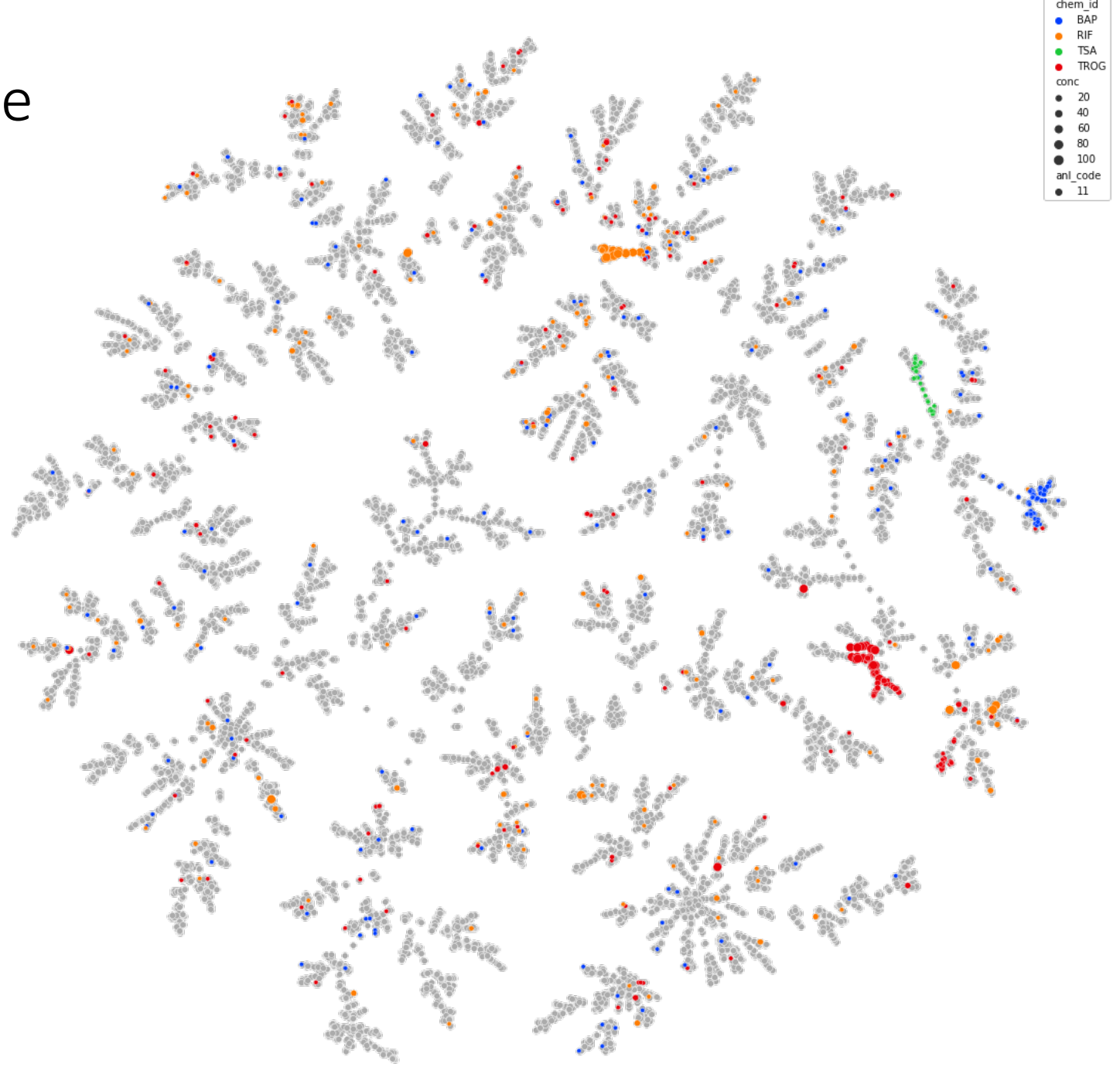
HTTr Experimental Design and Bioinformatics Workflow



Harrill *et al.*, 2021

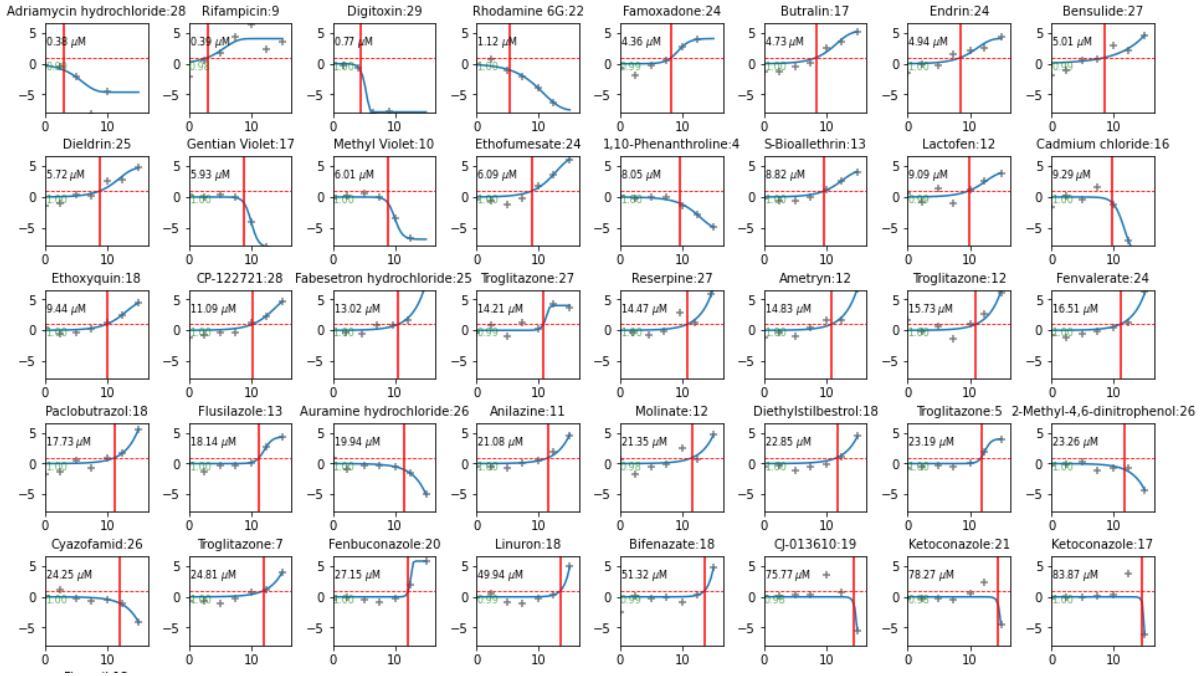
We are generating large-scale HTTr data sets on environmental chemicals

- For example, HepaRG cells treated with 1,366 chemicals 8 concentrations (0.01-100µM) for 24h
- Four reference chemicals: benzo(a)pyrene (●), rifampicin (●), trichostatin A (●) and troglitazone (●)
- Data processed by EPA HTTr pipeline to produce 11,551 L2FC profiles x ~12,250 genes
- Figure shows clustering of all 11,551 transcriptomic fingerprints
- Over 100,000 raw HTTr profiles generated on thousands of chemicals in MCF7, U2OS and other cell types in process

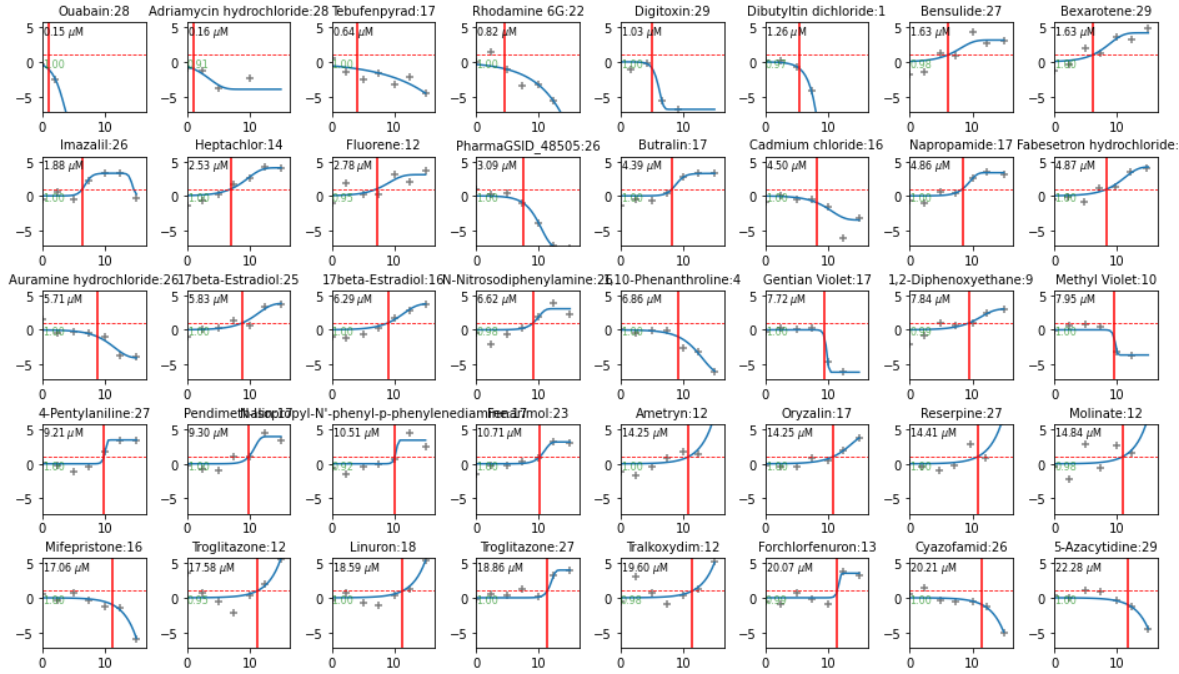


Univariate analysis of each gene is feasible …

PXR/NR1I2: CYP3A4



CAR/NR1I3: CYP2B6

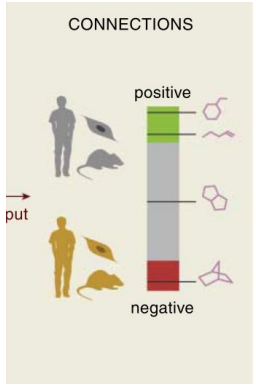


- Used HepaRG L2FC data and R/tcplFit2 (Judson *et al.*) for curve fitting to find benchmark concentrations (BMCs)
- However, this can miss subtle but coordinated regulation of multiple genes in pathways

Multivariate analysis of transcriptomic profiles

Connectivity mapping

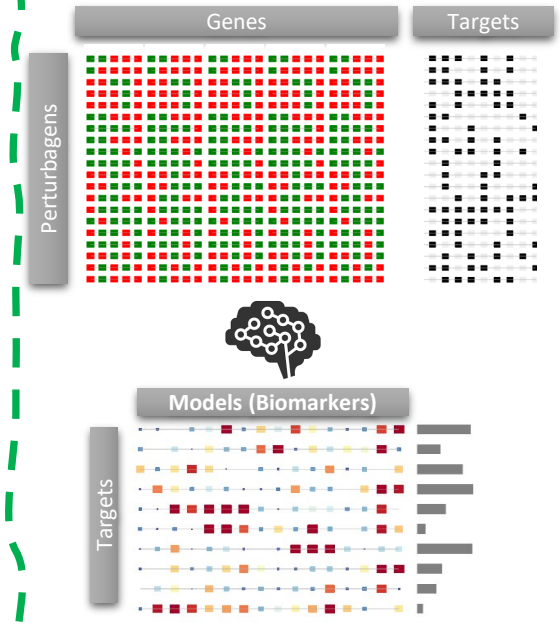
Match test profiles with gene signatures of known mechanisms



Lamb *et al.*, 2006

Machine learning

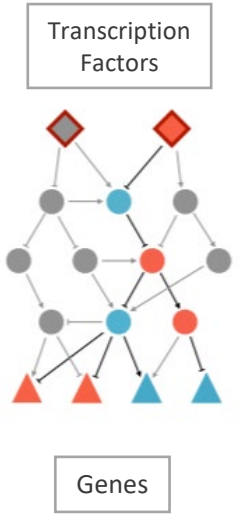
Match profiles with classifiers / biomarkers developed using machine learning



Liu *et al.*, 2017

Causal analysis

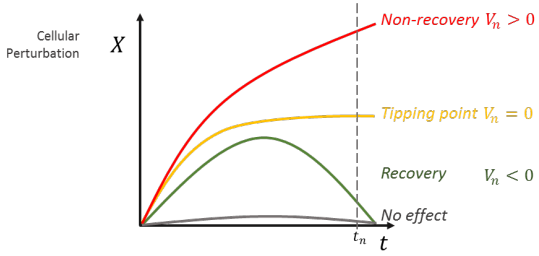
Reconstruct signaling and genetic regulatory interactions to define mechanistic networks



Liu *et al.*, 2019

Tipping Points

Analyze stress response pathway dynamics to estimate critical concentrations at homeostatic thresholds



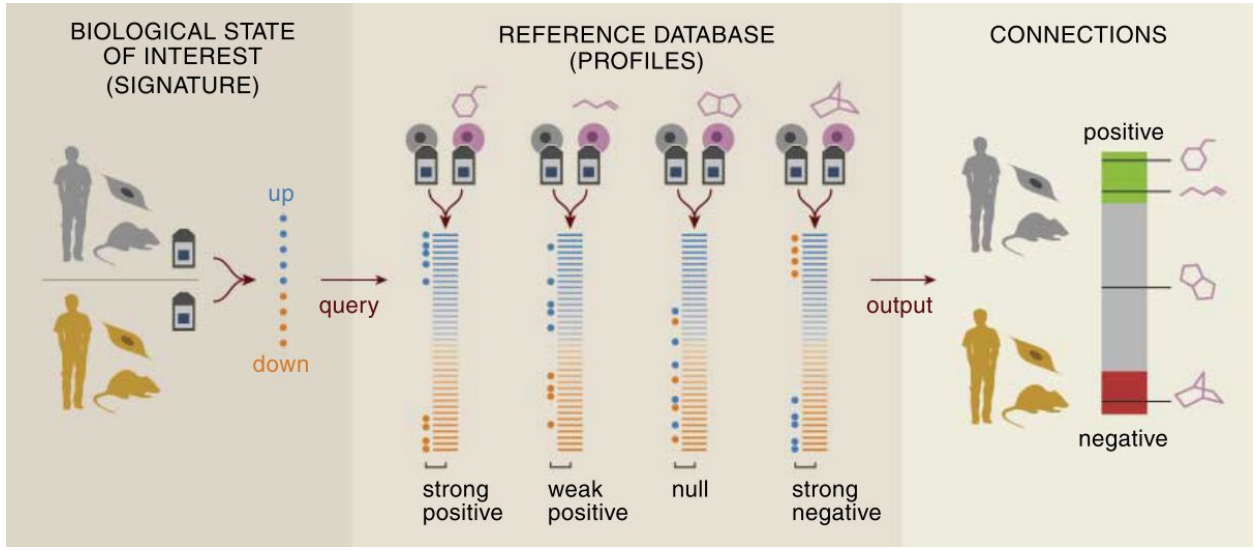
Calculate velocity $V = \frac{dX}{dt}$

Shah *et al.*, 2016

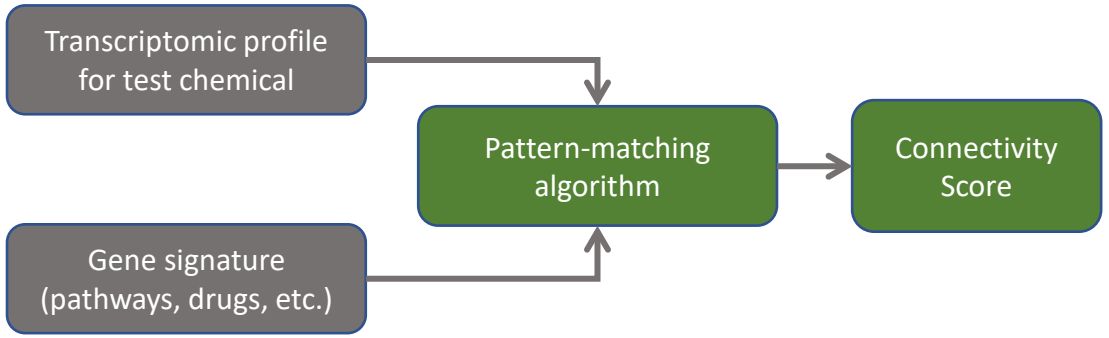
Connectivity Mapping

- "Connectivity mapping" (Lamb *et al.,* 2006): *S*imilarity between transcriptomic profiles may suggest common mechanisms
- Assumes that diverse biological states can be "fingerprinted" by the universal language of genes
- Used for linking chemicals, genes, diseases & drug-repurposing
- Increasing utility in toxicology to find putative targets (DeAbrew *et al.,* 2016; Wang *et al.,* 2016)

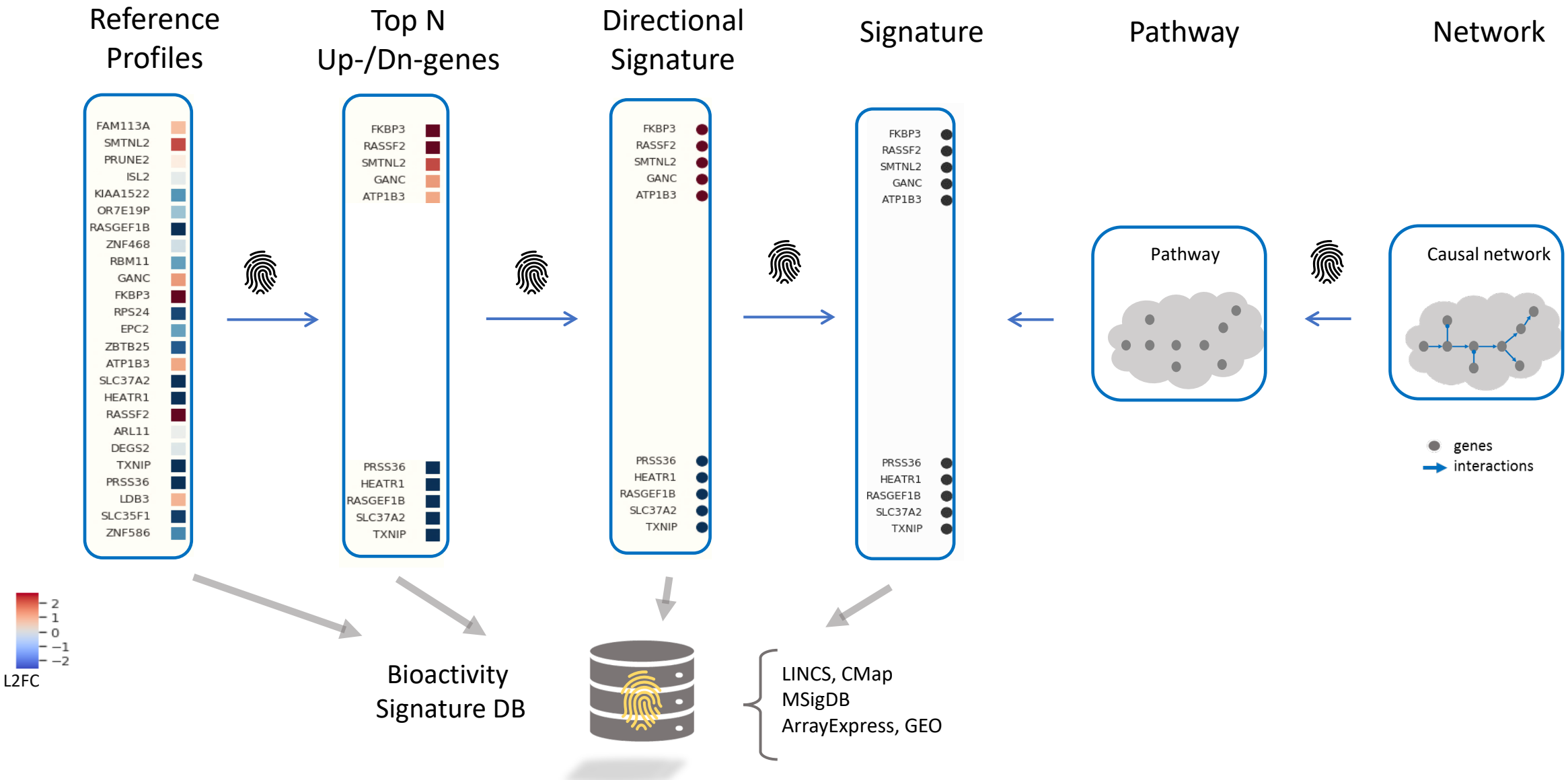
Overview



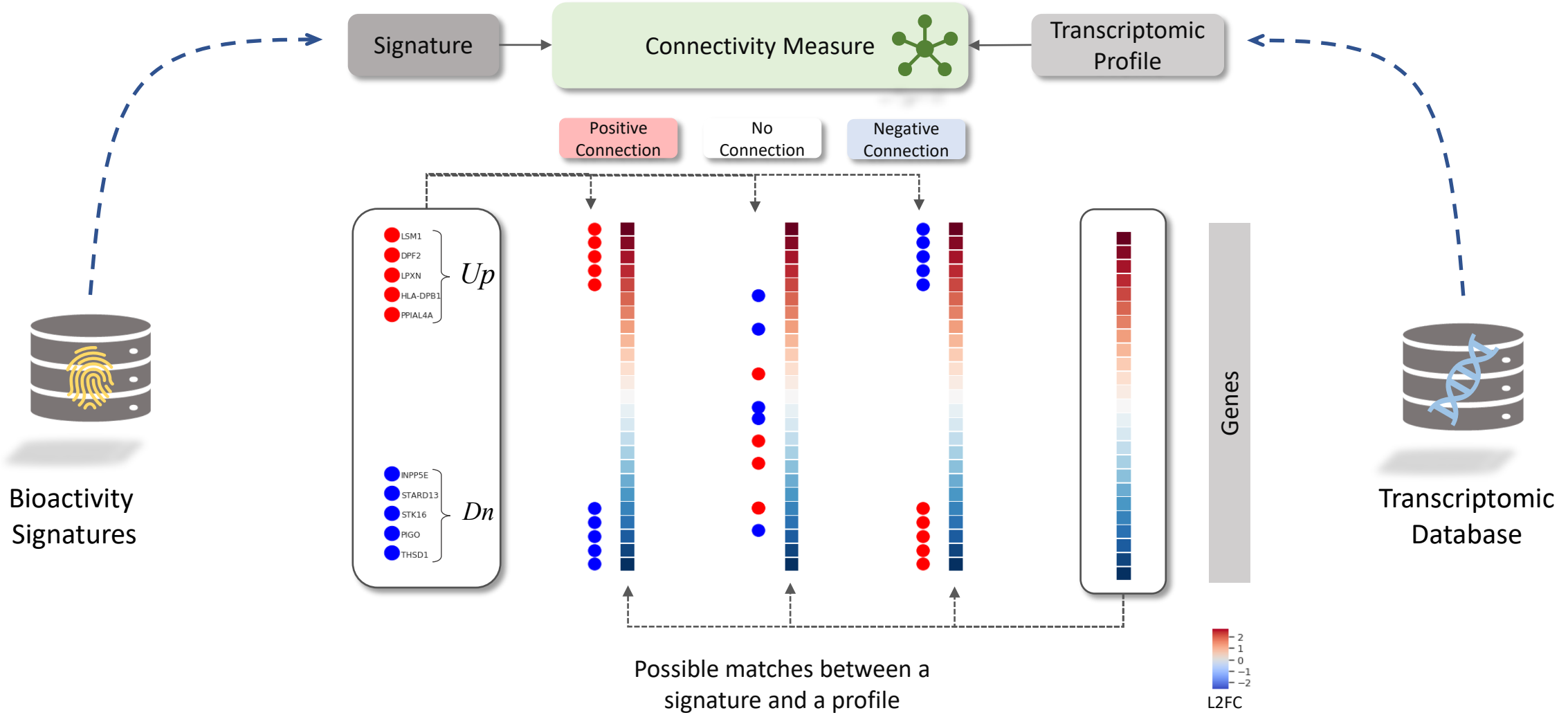
Lamb *et al.,* 2006



“Fingerprinting” bioactivity via gene signatures



Connectivity-mapping with gene signatures



Shah *et al.* (in prep)

Gene Set Connectivity Toolkit (gecco)

DB	Source	Signatures
Srp	US EPA	83
Lincs	Lincs	30,000
cmap	CMap v2	1200
msigdb	MSigDB	26860

- MongoDB for storage
- Consistent document structure
- Supports public and in-house data
- Multiple HTTr technologies

Connectivity Measures

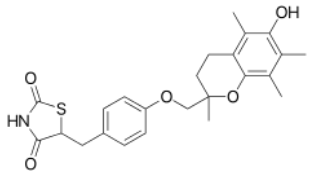
Methods	Measures
Aggregation-based	eXtreme Sum (XS), eXtreme Mean (XM), T-statistic (TT-p), Ranksum statistic (RS), Kolmogorov-Smirnov statistic (GSEA), Total enrichment score (TES)
Vector-based	Extreme Pearson correlation (XCP), Extreme Spearman Correlation (XCS), Jaccard index (JI), Signed Jaccard (SJI), Szymkiewicz–Simpson index (SI), Signed Szymkiewicz–Simpson index (SSI)

- Standardized API in Python 3
- Multiple connectivity measures
- Parallelized for speed
- Uses tcplFit2 for curve-fitting / BMD

DB	Source	Profiles
lincs	Lincs	591697
cmap	CMap v2	6100
arexp	ArrayExpress	3843
mcf7	US EPA	31352
heparg	US EPA	23102
u2os	US EPA	22980

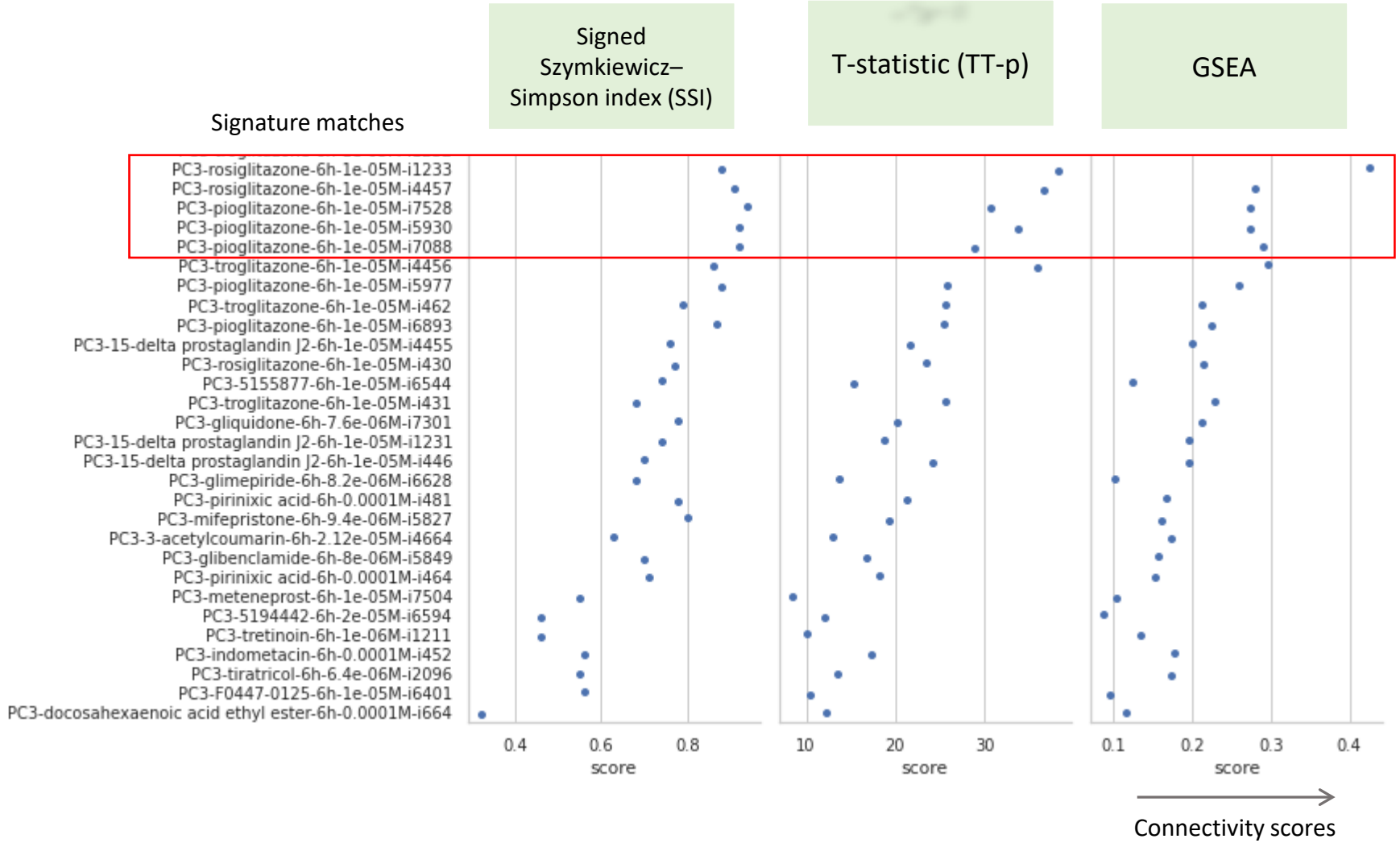
- MongoDB for storage
- Multiple HTTr technologies

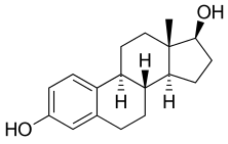
Shah *et al.* (in prep)



Matching troglitazone transcriptomic profiles with other chemical signatures

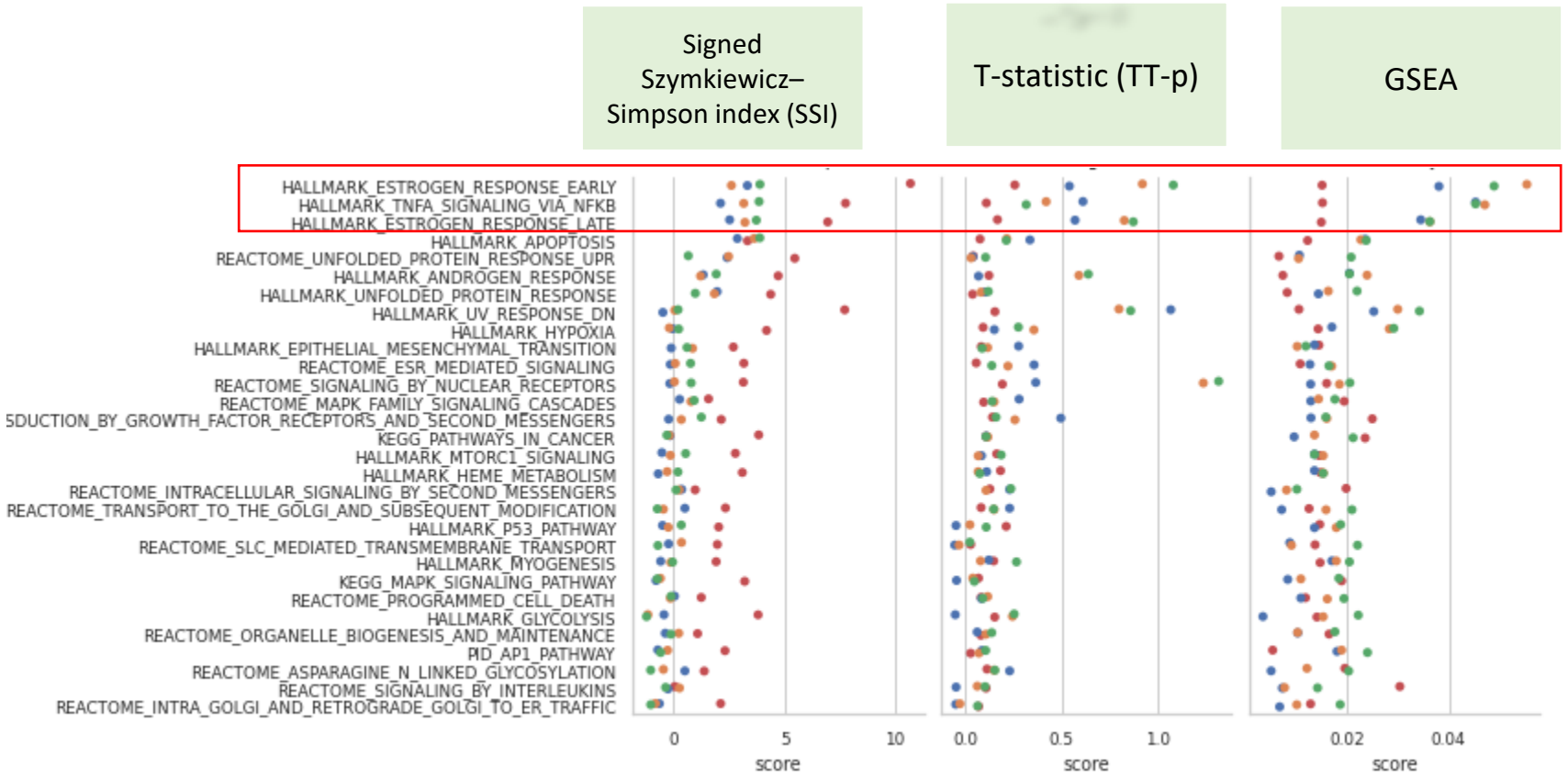
- Troglitazone is a thiazolidinedione (TZD) used as an antidiabetic and anti-inflammatory
- MIE: peroxisome proliferator activated receptor (PPARα) activator
- Use transcriptomic profile for troglitazone 10µM @ 6 h in PC3 cells
- Match 6,100 transcriptomic signatures (50-up and 50-dn genes) for 1200 chemicals in Connectivity Map v2
- Use three connectivity scores
- Best matches with other TZDs and PPARα-activators
- Can use this approach to identify putative PPARα activators





Matching estrogen transcriptomic profile with pathway signatures

- Estrogen is a female sex hormone activates the estrogen receptors (ERα/β)
- Use transcriptomic profile for troglitazone 10µM after 6 h in PC3 cells
- Match against 2,253 canonical and hallmark pathways in MSigDB v7.2
- Use different connectivity scoring methods and parameters
- Best matches are with estrogen response pathways
- Could use this approach to find putative ER-disruptors



Non-specific signatures: stress-response pathways

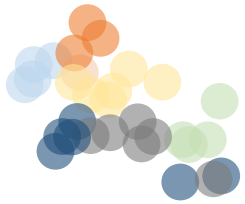
1

The Major Adaptive Stress response pathways

	Stress response pathway	Chemical inducers	TF	Activated gene promoters
OSR	Oxidative stress	Quinones, hydroperoxides, heavy metals, trivalent arsenicals	Nrf2	*HMOX1*, *NQO1*, *GST2A*
HSR	Heat shock response	Heat, Heavy Metals	HSF-1	*HSPA6*
DDR	DNA damage response	Etoposide, Methyl Methanesulfonate, N-Dimethylnitrosamine, Cyclophosphamide, UV radiation	p53	*CDKNIA*, *GADD45A*, *MDM2*, *BCL2*, *TP53I3*
HPX	Hypoxia	Hypoxia, Cobalt, Desferriozamine, Quercetin, Dimethyloxalylglycine	HIF-1	*VEGF*, *TF*, *EPO*
UPR	ER stress	Tunicamycin, Thapsigargin, Caplain, Brefeldin A	XBP-1, ATF6, ATF4	*HSP90B1*, *HSPA5*, *DNAJB9*
MTL	Metal stress	Heavy Metals	MTF-1	*MT1E*, *MT2A*
	Inflammation	Metal, PCBs, Exhaust Particles, Smoke Particles	NF-κB	*IL1A*, *TNFA*
	Osmotic stress	High salt, polyethylene glycol, mannitol	NFAT5	*AKR1B1*, *SLC6A12*, *SLC5A3*

Simmons et al., 2009

2



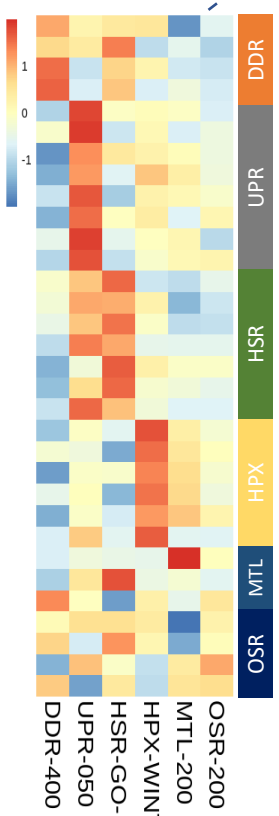
Published Signatures (MSigDB v7.2)

Overlapping Consensus Genes

Unique Consensus Signatures

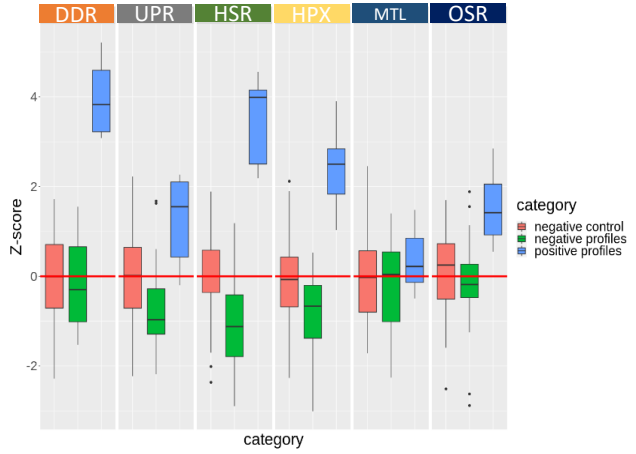
Connectivity Mapping with Transcriptomic Data for reference Chemicals (GSEA)

3



Performance

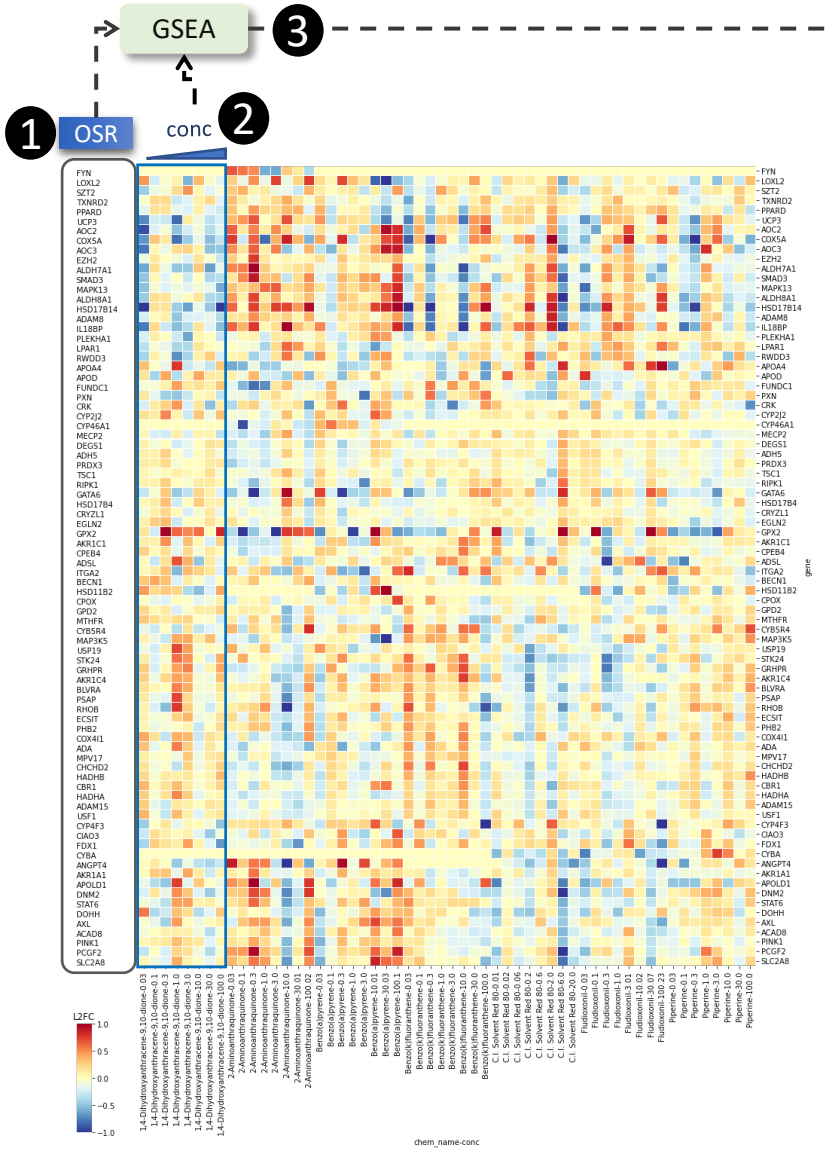
4

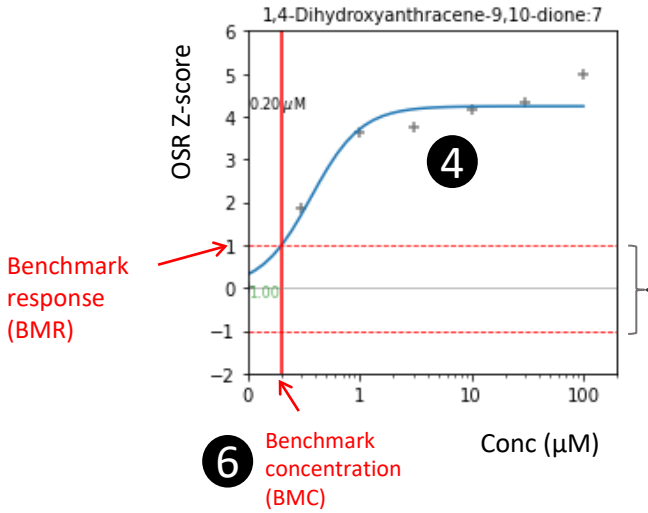


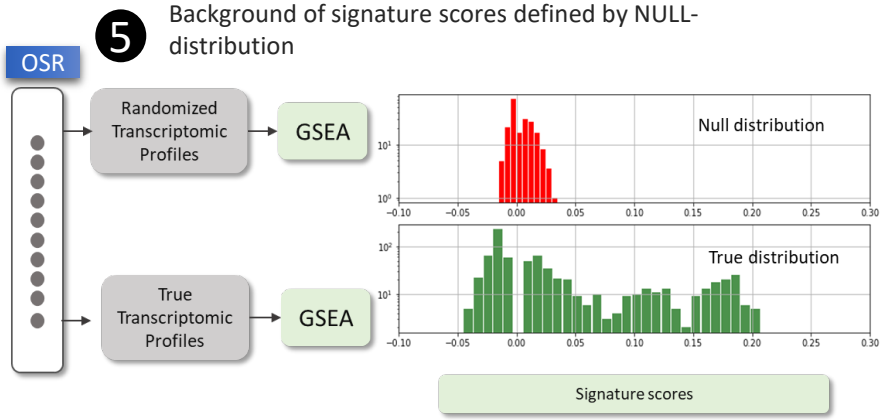
- Consensus signatures for most SRPs can identify reference perturbagens
- Can use SRP signatures to evaluate non-specific chemicals

Chambers *et al.* (*submitted*)

Using signature scores to estimate PODs

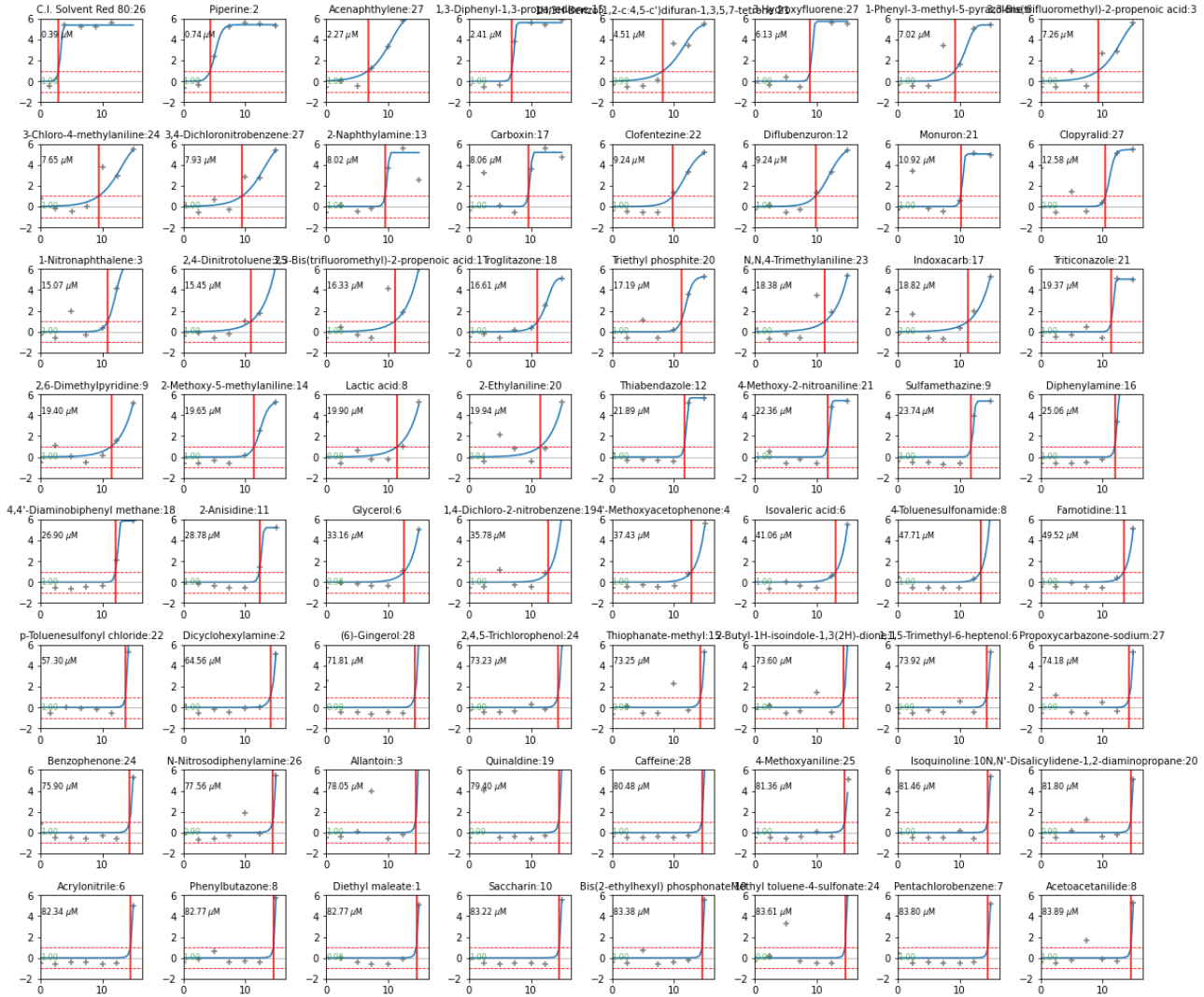


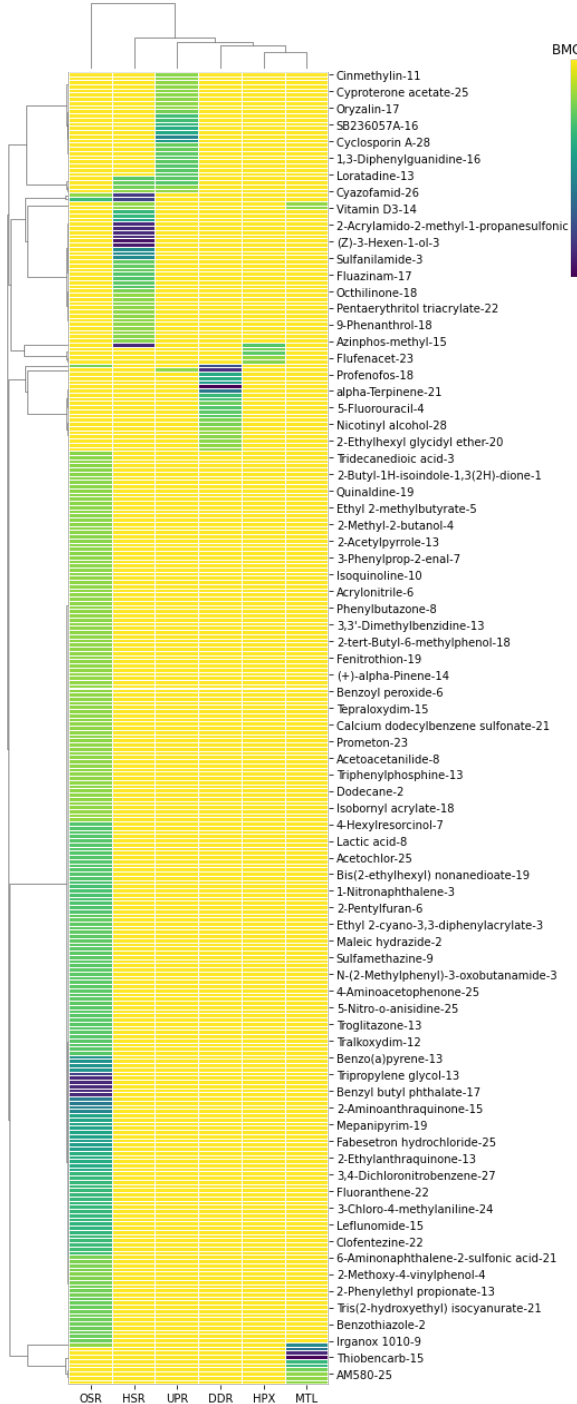




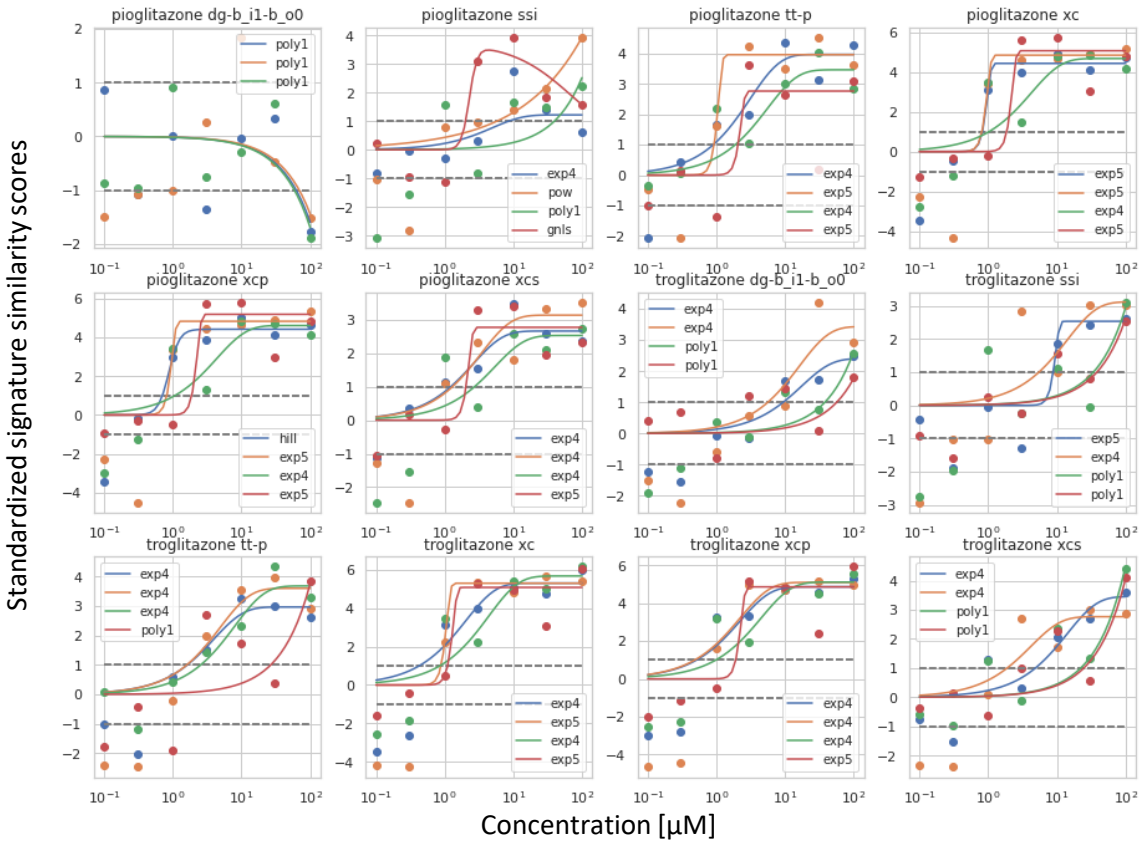
- Use
- Standardize signature scores using background distribution (Z)
- Concentration-response modeling using tcplFit2 (cnst, hill, gnls, poly1, poly2, pow, exp2, exp3, exp4, exp5)
- Estimate benchmark concentration (BMC) using benchmark response (BMR) of Z=1

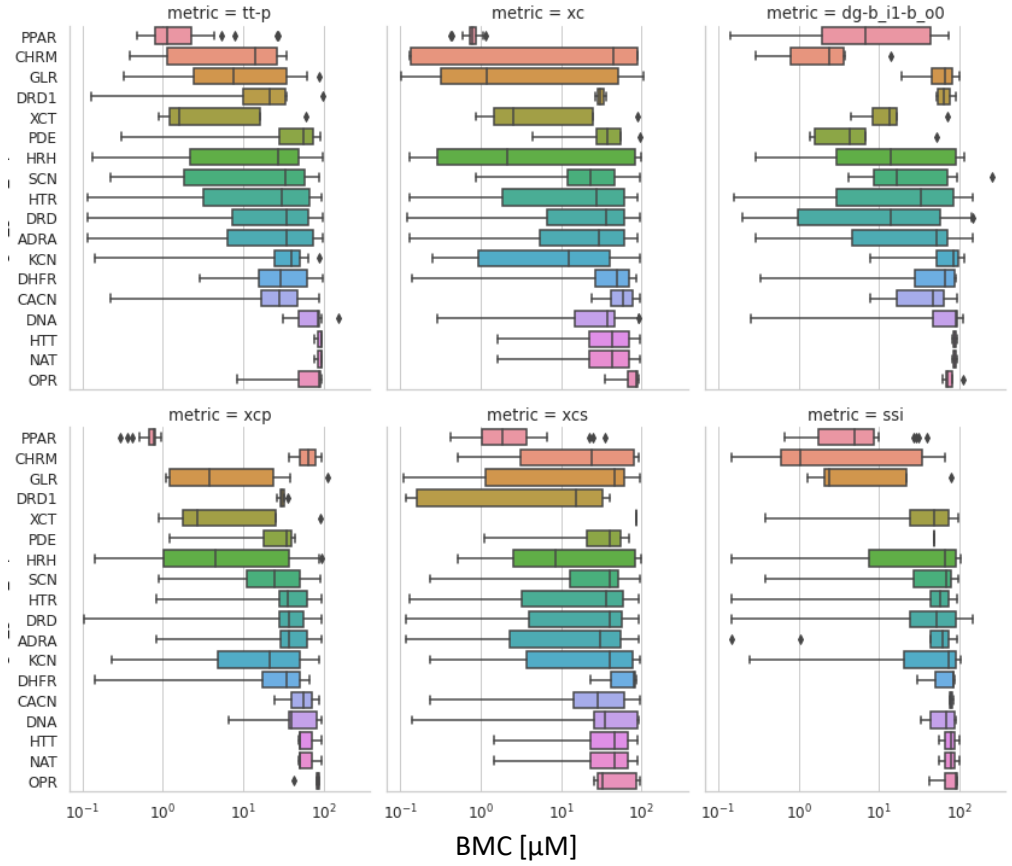
PODs for all chemicals and SRPs





Signature similarity scores to PODs: Troglitazone

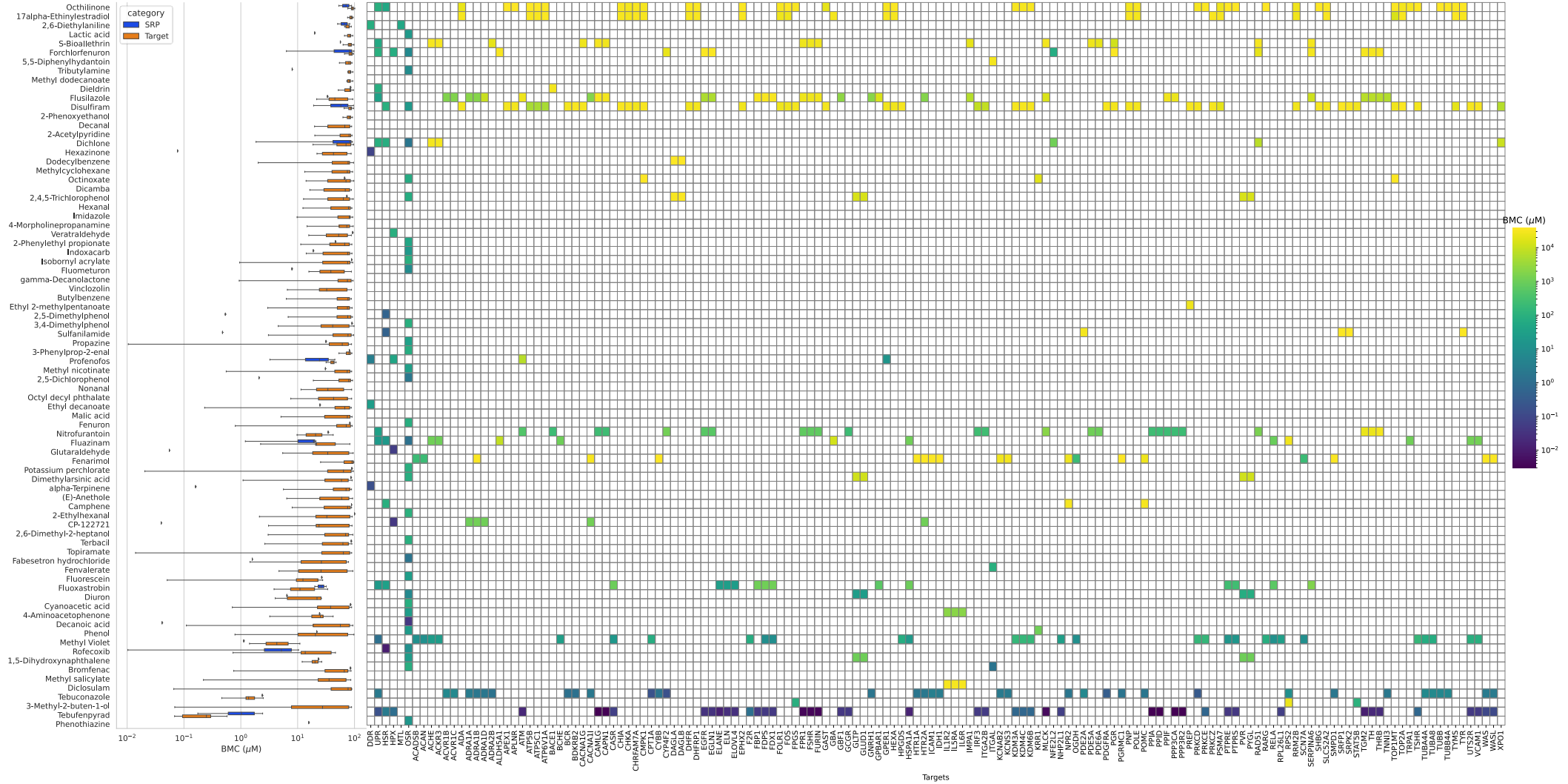




- Data: HepaRG cells treated with 1,366 chemicals 0.01-100μM
- Query: Troglitazone profiles replicated across 29 batches
- Reference: CMap v2 Affymetrix data
- Connectivity analysis using multiple signatures and metrics
- Score signatures against random profiles to estimate background

- Standardize similarity scores using background distribution (Z)
- Concentration-response modeling using tcplFit2 (cnst, hill, gnls, poly1, poly2, pow, exp2, exp3, exp4, exp5)
- Estimate benchmark concentration (BMC) using benchmark response (BMR) of Z=1

Summary of hits for HepaRG treatments



Summary

1. HTTr using TempO-Seq is a reliable & scalable tool for NAM development
2. Connectivity mapping uses *multivariate* gene signature scores for hazards and PODs
3. Additional evaluation of connectivity scoring & gene signatures is necessary
4. We are comparing PODs from transcriptomics with traditional values using qIVIVE
5. Adaptive stress response signatures is may help distinguishing specific vs. non-specific chemicals

Summary

1. HTTr using TempO-Seq is a reliable & scalable tool for NAM development

   We are using TempO-Seq technology to evaluate thousands of chemicals in multiple cell lines and have developed a high-throughput pipeline to process and analyze transcriptomic concentration-response data.

2. Connectivity mapping uses *multivariate* gene signature scores for hazards and PODs

   Environmental chemicals may not induce large changes in expression of individual genes. Gene signature-based connectivity scoring is more sensitive and can accurately identify putative targets.

3. Further evaluation of connectivity scoring & gene signatures

   We are systematically evaluating parameters of connectivity mapping and gene signature development to determine best practices for different chemical categories using gecco software.

4. Future directions

   - Systems biology of adaptive stress response pathways to distinguish specific vs non-specific chemicals
   - Investigate the molecular basis of cellular resilience and tipping points
   - Study single-cell analysis to further improve confidence in HTTr data

Acknowledgements

US EPA

Joshua Harrill

Logan Everett

Richard Judson

Bryant Chambers

Woodrow Setzer

Joe Bundy

Derik Haggard

Beena Valanat

Thomas Knudsen

Tim Shafer

Brian Chorley

John Cowden

Russell Thomas

UniLever, UK

Alistair Middleton

Joe Reynolds

Andy White

Paul Russell

Paul Carmichael

University of Cambridge, UK

Andreas Bender

Danilo Basili

Layla Hosseini-Gerami

A*Star, Singapore

Hanry Yu

Fan Lee

Yun Ting Soon

Additional information

www.linkedin.com/in/imranshah

Shah.Imran@epa.gov