Transcriptomic Connectivity for Read-Across Inference of Chemical Bioactivity

QSAR 2021
June 2-6, 2021

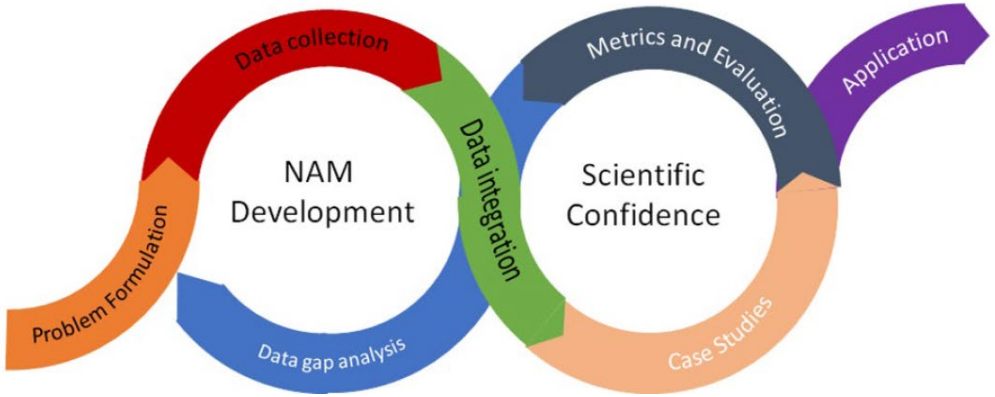
Imran Shah
Center for Computational Toxicology & Exposure

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.

Reducing use of animals in chemical testing

- On September 10, 2019 EPA Administrator Andrew Wheeler signed a directive that prioritizes efforts to reduce animal testing. The memorandum calls for the agency to:
  - reduce its requests for, and funding of, mammal studies by 30 percent by 2025, and
  - eliminate all mammal study requests and funding by 2035.
- This will be achieved via new approach methodologies (NAMs): any technology, methodology, approach, or combination of methods that can provide information about chemical hazard and point of departure (POD) without using whole animals

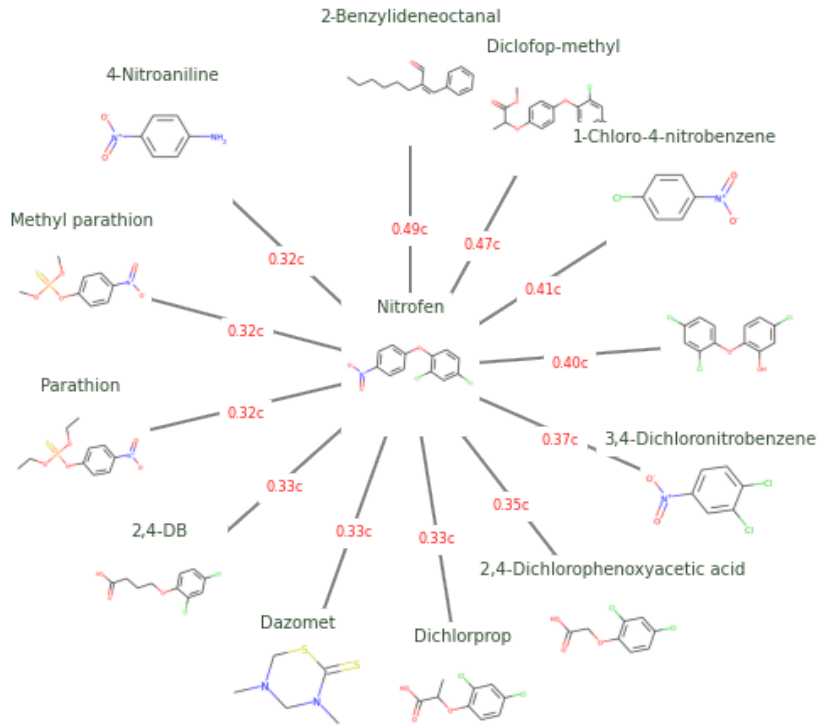
New Approach Methods Work Plan



EPA 2020

Read-across (RAX)

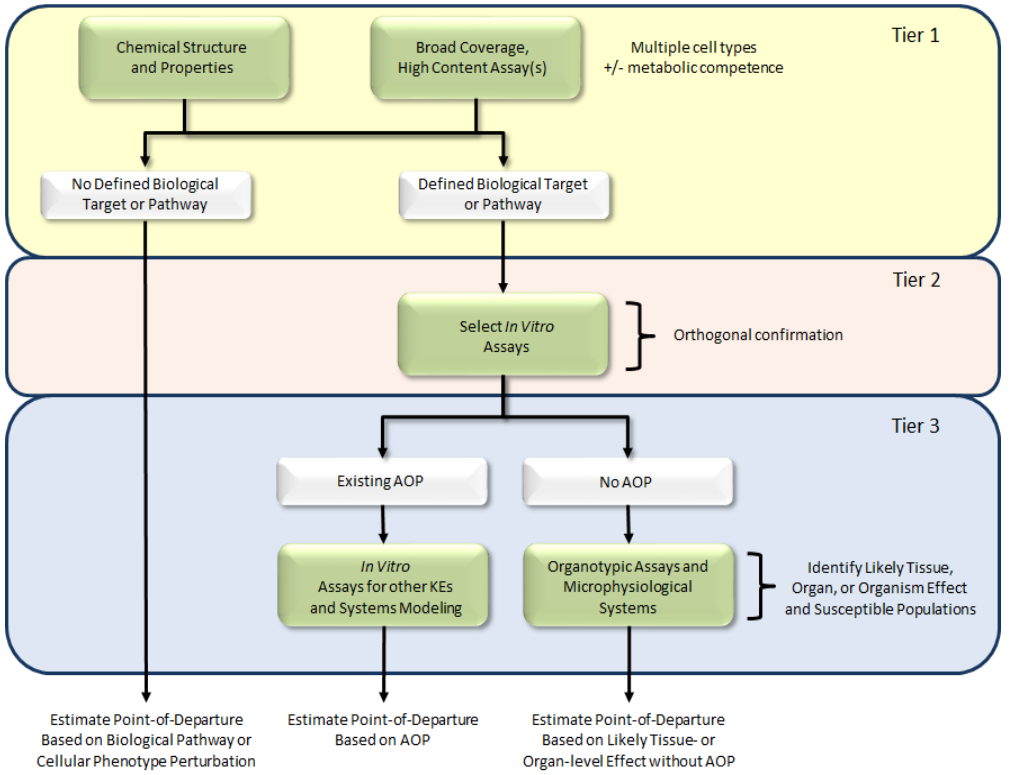
- “Read-across” (RAX) techniques are often used to fill data gaps by inference from a ‘similar’ substance or substances (OECD, 2017):
  - Identify analogues using between structure / physico-chemical similarity
  - Assign hazard and POD value based on analogue(s)
- Problem: Many chemicals of interest do not have any structural analogues with any bioactivity data
- We have developed generalized read-across (GenRA) to automate RAX using physico-chemical, bioactivity and metabolic contexts of similarity
- How can we use NAMs to define new contexts of similarity and aid RAX?



Generalized Read-across (GenRA)
Shah *et al.* 2016, 2021

Regulatory Context: Tiered Hazard Evaluation

- The “CompTox Blueprint” lays out a tiered approach for evaluating untested chemicals with NAMs
- Tier 1 NAMs based on high-throughput profiling (HTP) assays are flexible, portable and cost-efficient platforms to comprehensively evaluate the potential effects of thousands of chemicals
  - Identify hazards - putative targets and pathways
  - Estimate POD *in vitro* associated with hazards
- Two types of HTP assays:
  - High-throughput transcriptomics (HTTr)
  - High-throughput phenotypic profiling (HTPP)



Thomas *et al.*, 2018

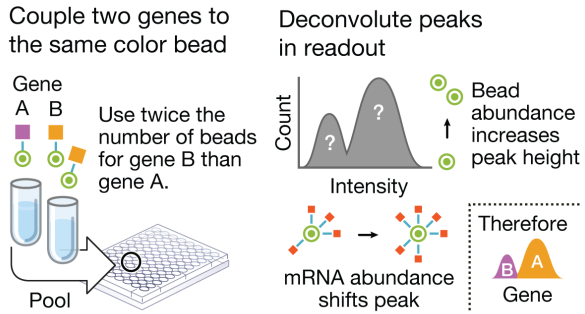
Some HTTr Technologies …

GeneChip

Affymetrix.com

- Established technology with vast amount of legacy gene expression data
- Multiple resources on chemical bioactivity including Connectivity Map v2, Open TG-GATES, and others

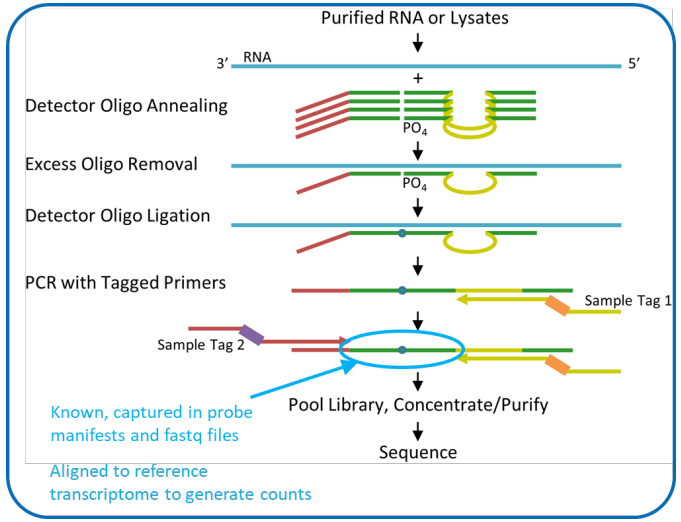
“Landmark 1000” (L1000)



- Bead-based assay to measure expression of 998 “landmark” genes used to infer expression of ~13,000 genes
- Used by the Broad Institute LINCS project
- Used to evaluate bioactivity of thousands of chemicals

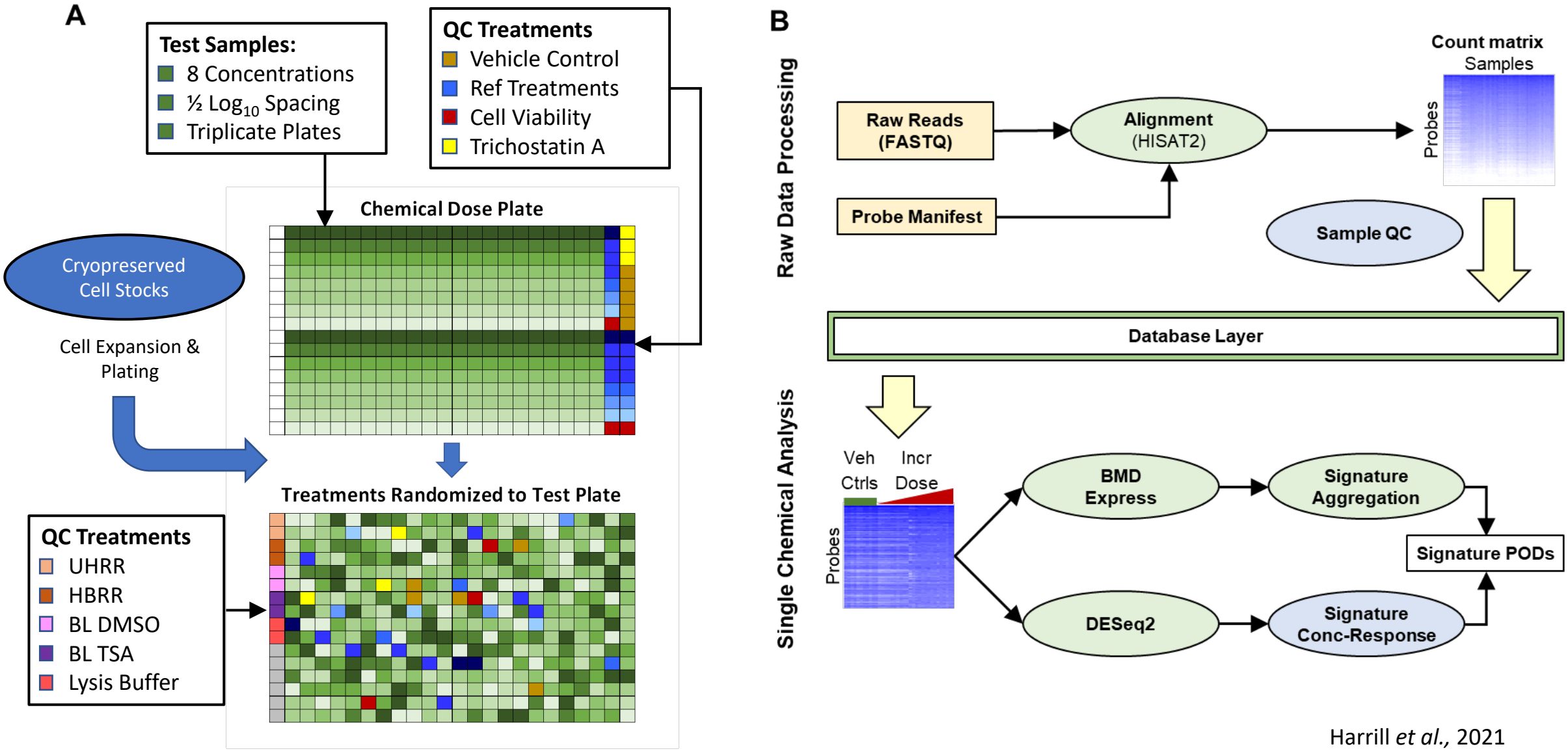
“Templated Oligo with Sequencing Readout” (TempO-Seq)

Yeakley, et al. PLoS ONE 2017



- Targeted RNA-Seq uses unique 50-mer oligos for mRNA detection
- Measures 21,000 unique mRNA
- Read space focused on known genes
- Compatible with whole cell lysates
- Being used by the US EPA for screening environmental chemicals

EPA HTTr Experimental Design and Bioinformatics Workflow for TempO-Seq data

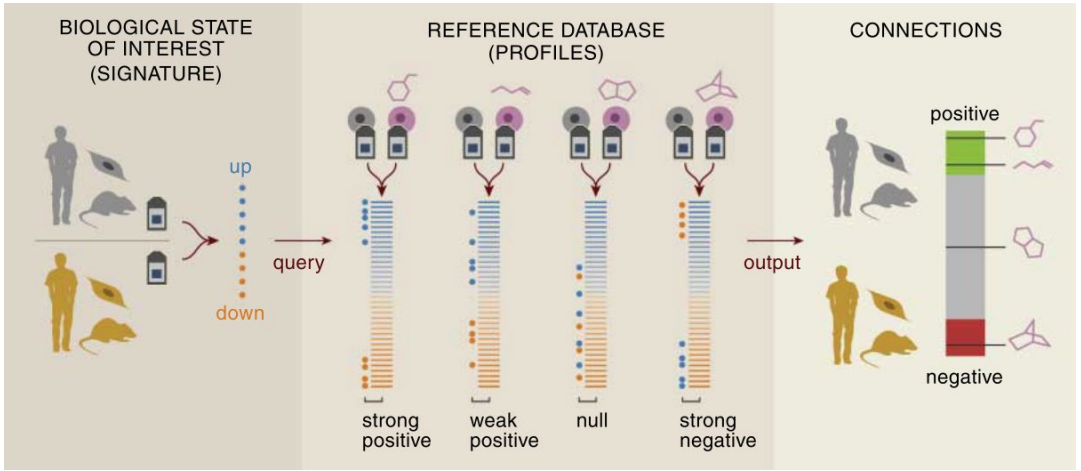


Harrill *et al.*, 2021

Connectivity Mapping and Read-Across

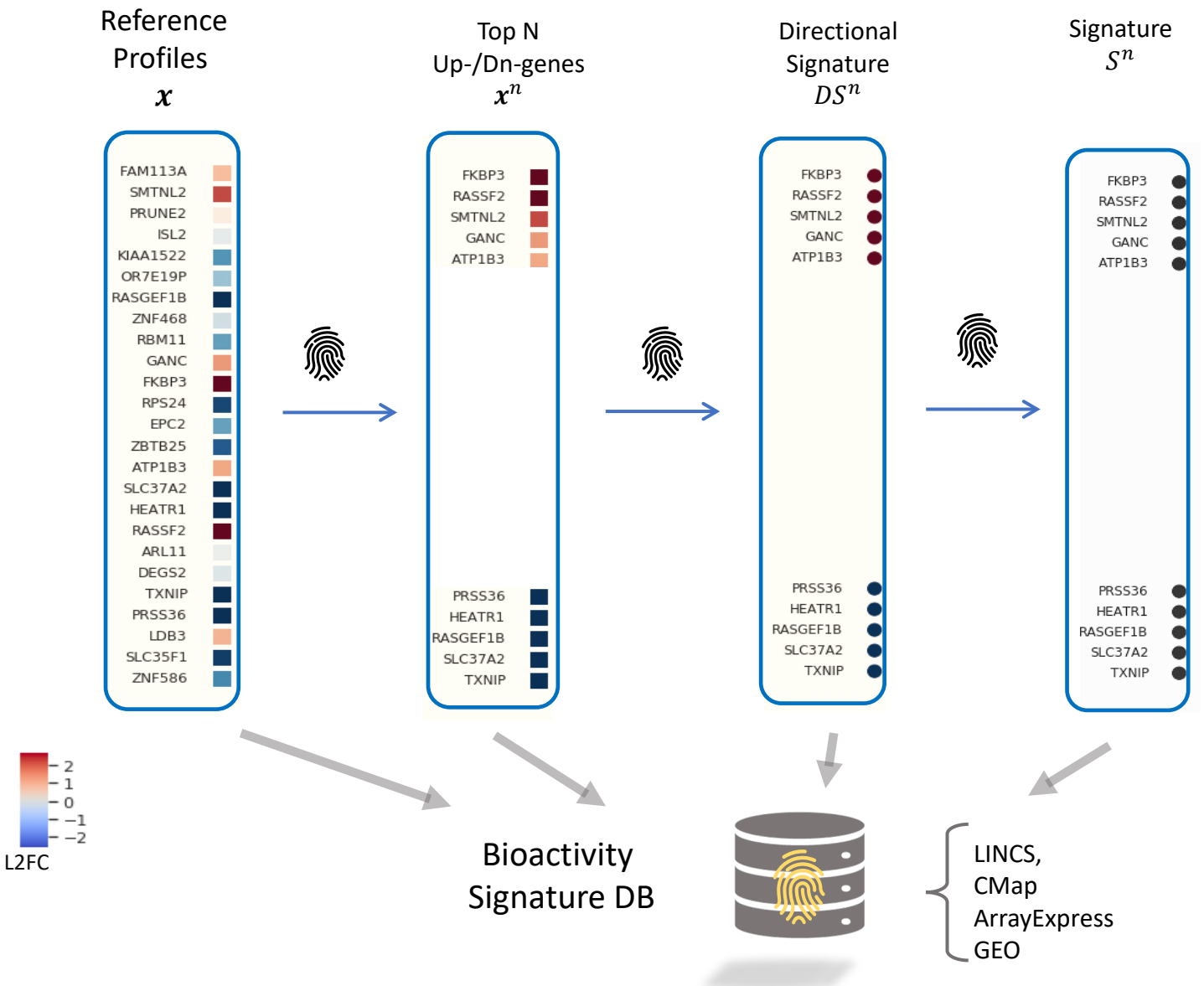
- Connectivity mapping (Lamb *et al*., 2016) developed to interpret gene sets using similarity with reference HTTr profiles.
- Assumptions:
  - Biological state of samples represented by transcriptomic descriptors
  - Similarity between transcriptomic representation implies common mechanisms
- Transcriptomic connectivity provides a new context of similarity for evaluating untested chemicals by read-across
- Key questions:
  - How do we represent transcriptomic data?
  - How can we measure similarity

Connectivity Mapping



Lamb *et al.,* 2006

Gene sets as transcriptomic “fingerprints”



$$O \in \{x, x^n, DS^n, S^n\}$$

Generalising Connectivity Analysis

- Can generalize connectivity analysis as:

$$s = SM(O_q, O_r)$$

s = similarity / connectivity score

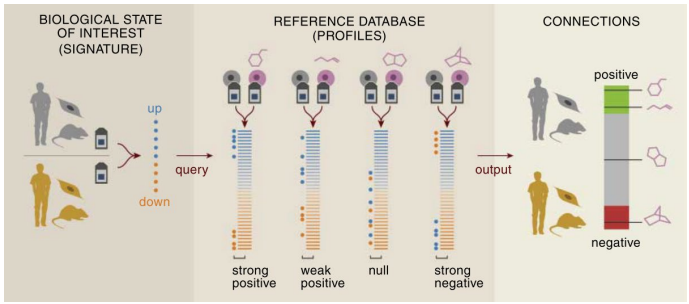
SM = similarity metric

O = gene set "Object"
$O \in \{\boldsymbol{x}, \boldsymbol{x}^n, DS^n, S^n\}$

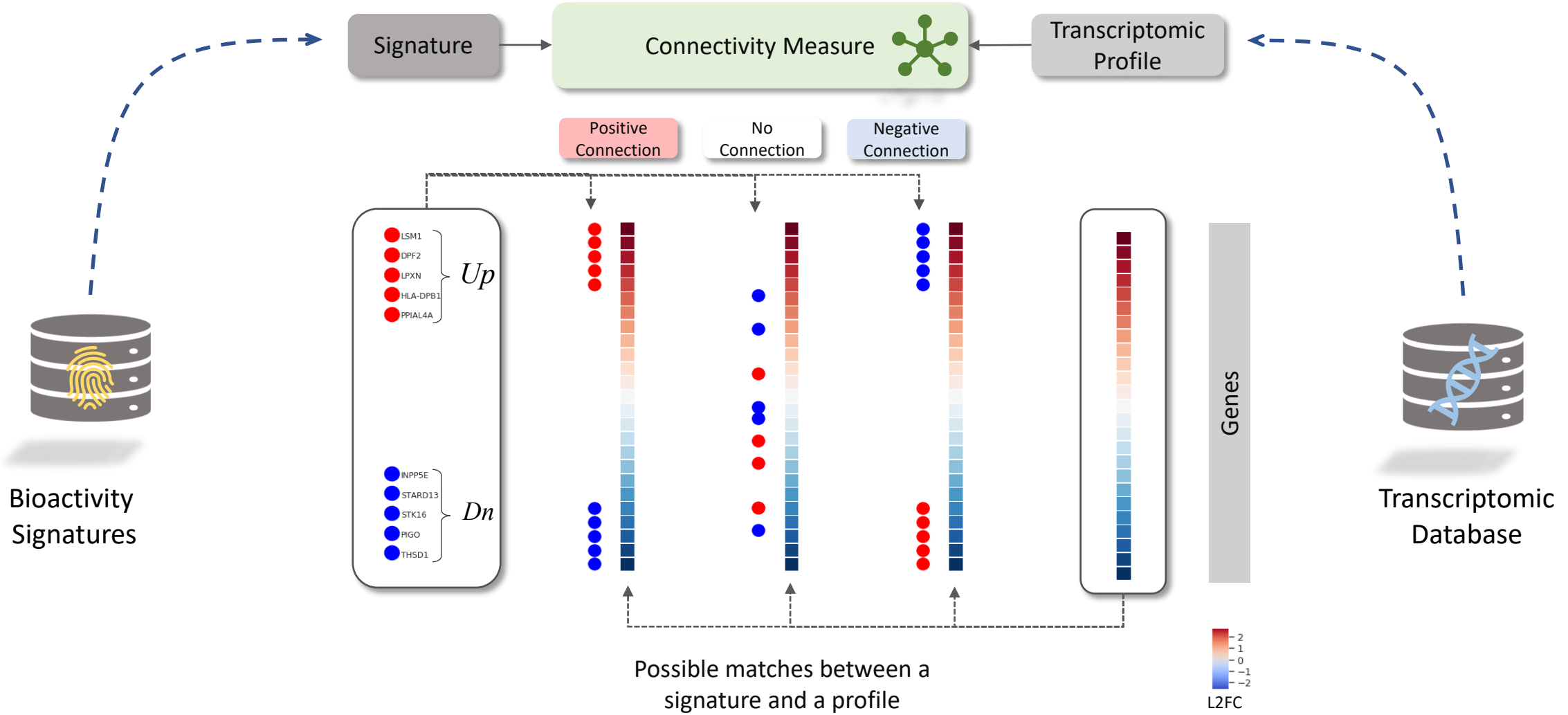
Connectivity mapping (Lamb *et al.,* 2006)

Query (O_q): directional signature ($\boldsymbol{DS}^n$)
Reference (O_r): transcriptomic profiles in CMap v2 ($\boldsymbol{x}$)
Similarity measure (SM): Gene set enrichment Analysis ($GSEA_b$)
(Subramanian *et al.* 2005)

$$SM(DS^n, x)$$



Connectivity-mapping with gene signatures



Shah *et al.* (in prep)

Generalised Connectivity Toolkit (gecco)

DB	Source	Signatures
Srp	US EPA	83
Lincs	Lincs	30,000
cmap	CMap v2	1200
msigdb	MSigDB	26860
dorothea	Dorothea	1333

$$O \in \{\boldsymbol{x}^n, DS^n, S^n\}$$

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

Similarity Measures

$$s = SM(O_q, O_r)$$

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)

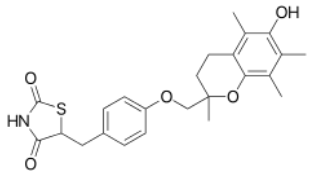
- Standardized API in Python 3
- Multiple connectivity measures
- Parallelized for speed
- Uses tcplFit2 for curve-fitting / BMD

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

$$O \in \{\boldsymbol{x}, \boldsymbol{x}^n\}$$

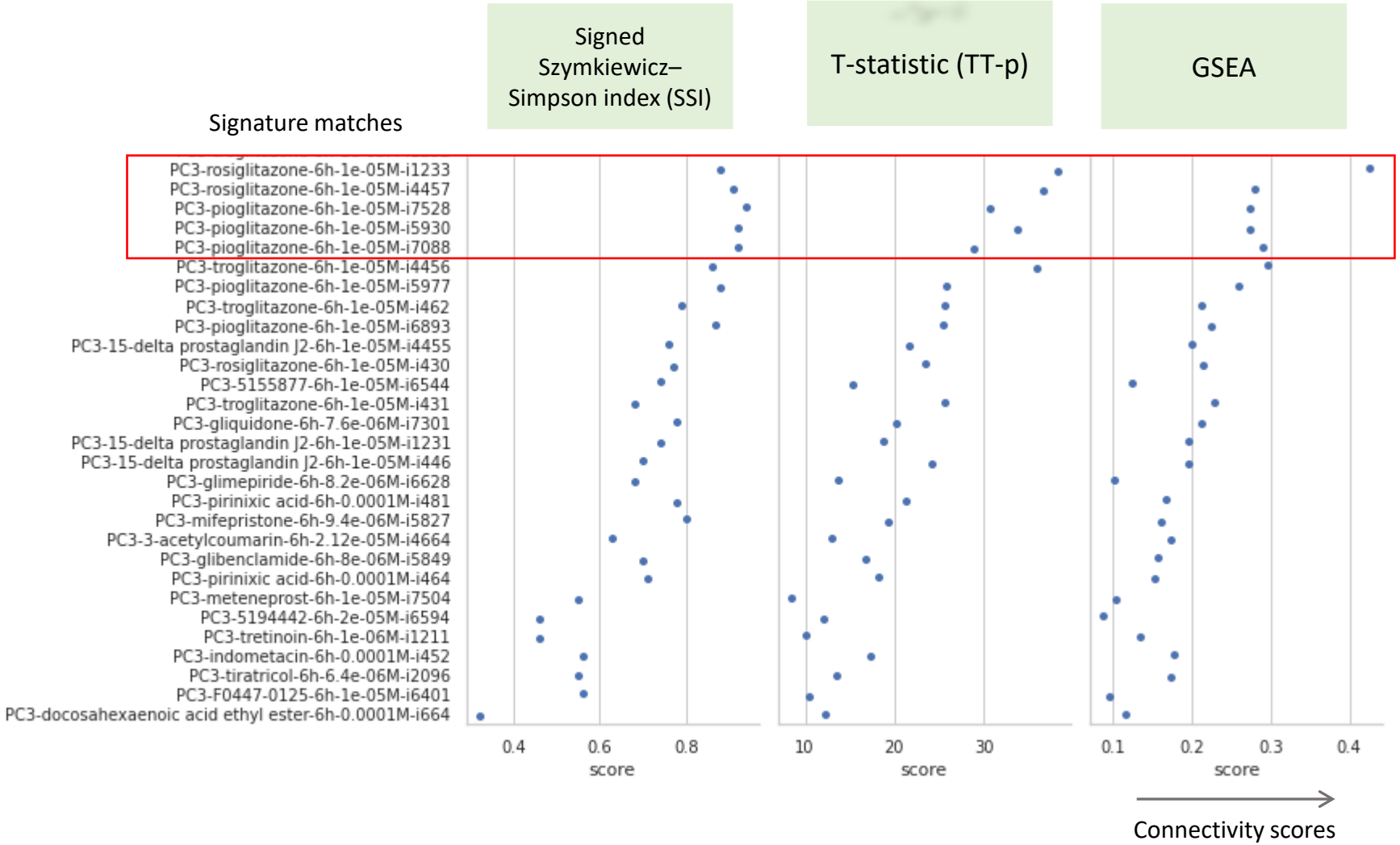
- MongoDB for storage
- Multiple HTTr technologies

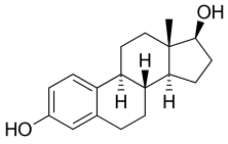
Shah *et al.* (in prep)



Matching troglitazone transcriptomic profiles with other chemical signatures

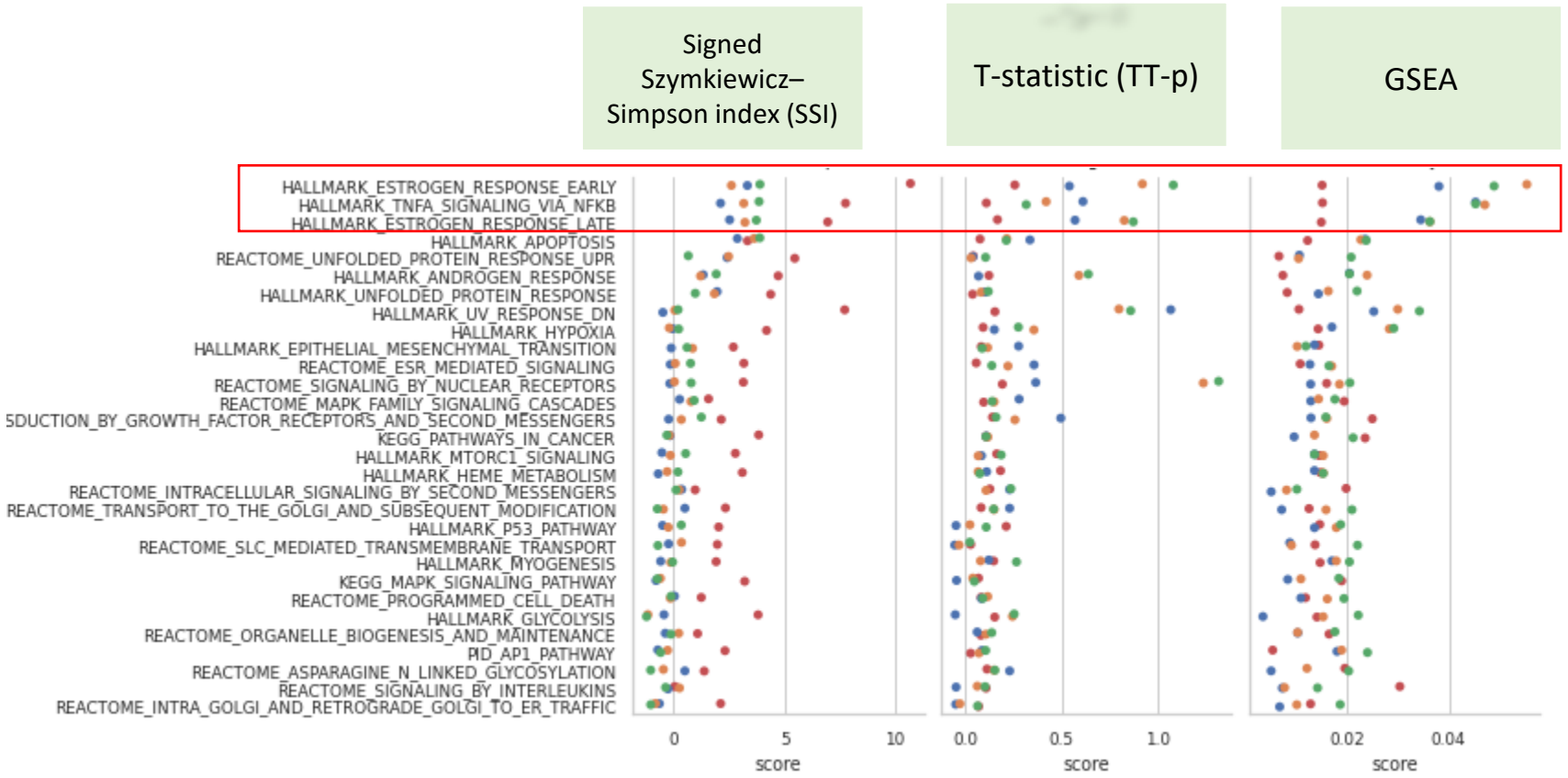
- Troglitazone is a thiazolidinedione (TZD) used as an antidiabetic and anti-inflammatory
- MIE: peroxisome proliferator activated receptor (PPARα) activator
- Use transcriptomic profile (**x**) for troglitazone 10µM @ 6 h in PC3 cells
- Match 6,100 transcriptomic signatures DS^{100} for 1200 chemicals 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





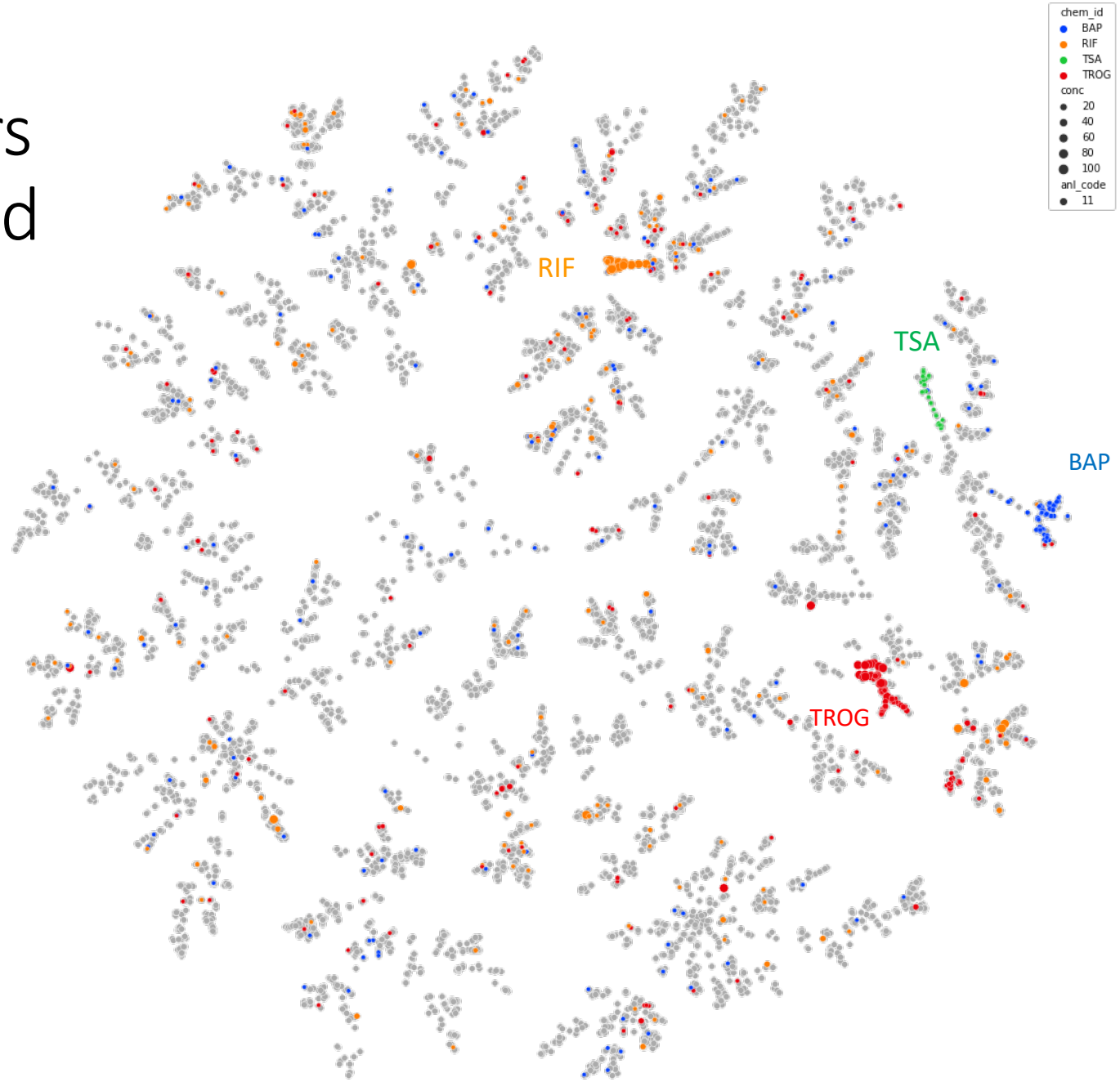
Matching estrogen transcriptomic profile with pathway signatures

- Estrogen is a female sex hormone activates the estrogen receptors (ERα/β)
- Use transcriptomic profile for estrogen 14µM after 6 h in MCF7 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



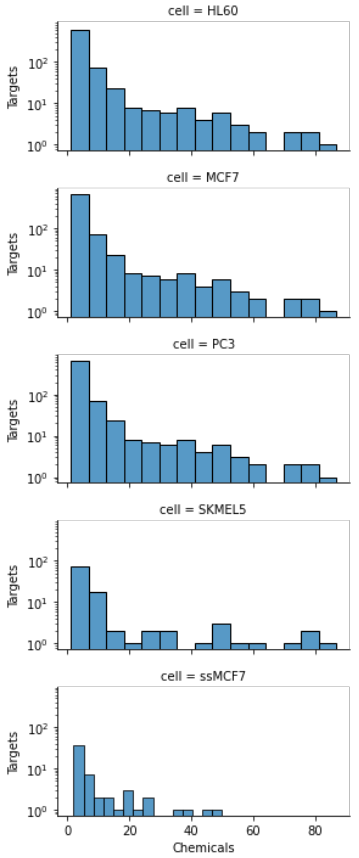
Transcriptomic neighbours are mechanistically related

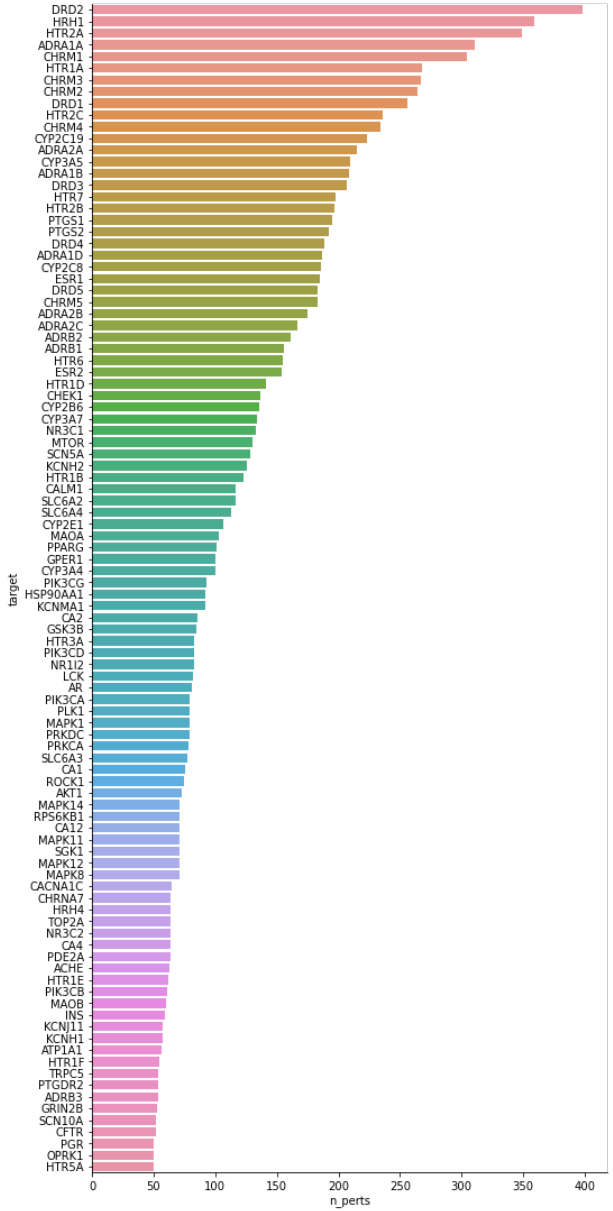
- HepaRG cells treated with 1,366 chemicals 8 concentrations (0.01-100µM) for 24h
- Data processed by EPA HTTr pipeline to produce 11,551 L2FC profiles x ~12,250 genes
- **Clustering of all 11,551 transcriptomic fingerprints DS^{100} using Jaccard Index**
- Four reference chemicals: benzo(a)pyrene (BAP●), rifampicin (RIF●), trichostatin A (TSA●) and troglitazone (TROG●)



Signatures of targets using CMap

- Obtain target annotations from CMap
- 833 targets
- Create "consensus" target signatures for cell type
  - Approach 1
    - For each DS^n for target in cell
      - Create consensus signature from the the n most frequent up/dn genes
  - Approach 2
    - For each $\boldsymbol{x}$ for target in cell
      - Find the consistently up/dn regulated genes (e.g. based on median L2FC or otherwise)
      - Create consensus signature as DS^n and $\boldsymbol{x^n}$





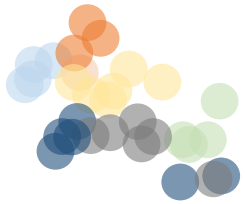
Signatures of stress-response pathways: non-specific chemicals

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



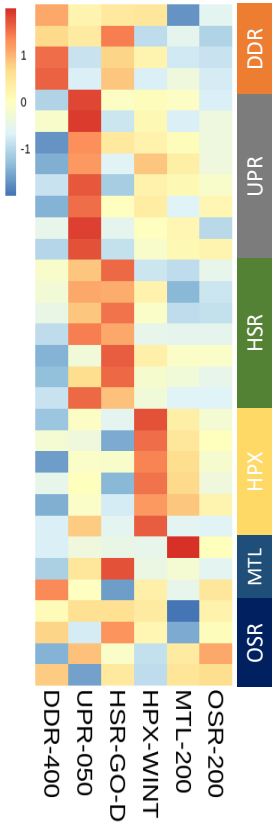
2

Published Signatures (MSigDB v7.2) → Overlapping Consensus Genes → Unique Consensus Signatures

$$O \in \{\boldsymbol{x}, \boldsymbol{x}^n, DS^n, S^n\}$$

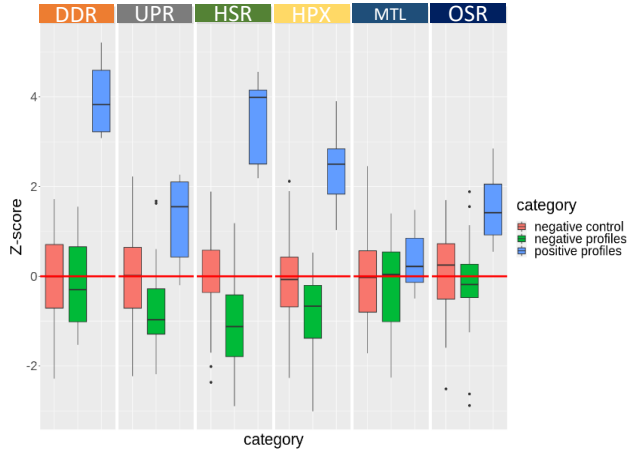
Connectivity Mapping with Transcriptomic Data for reference Chemicals (GSEA)

3



Performance

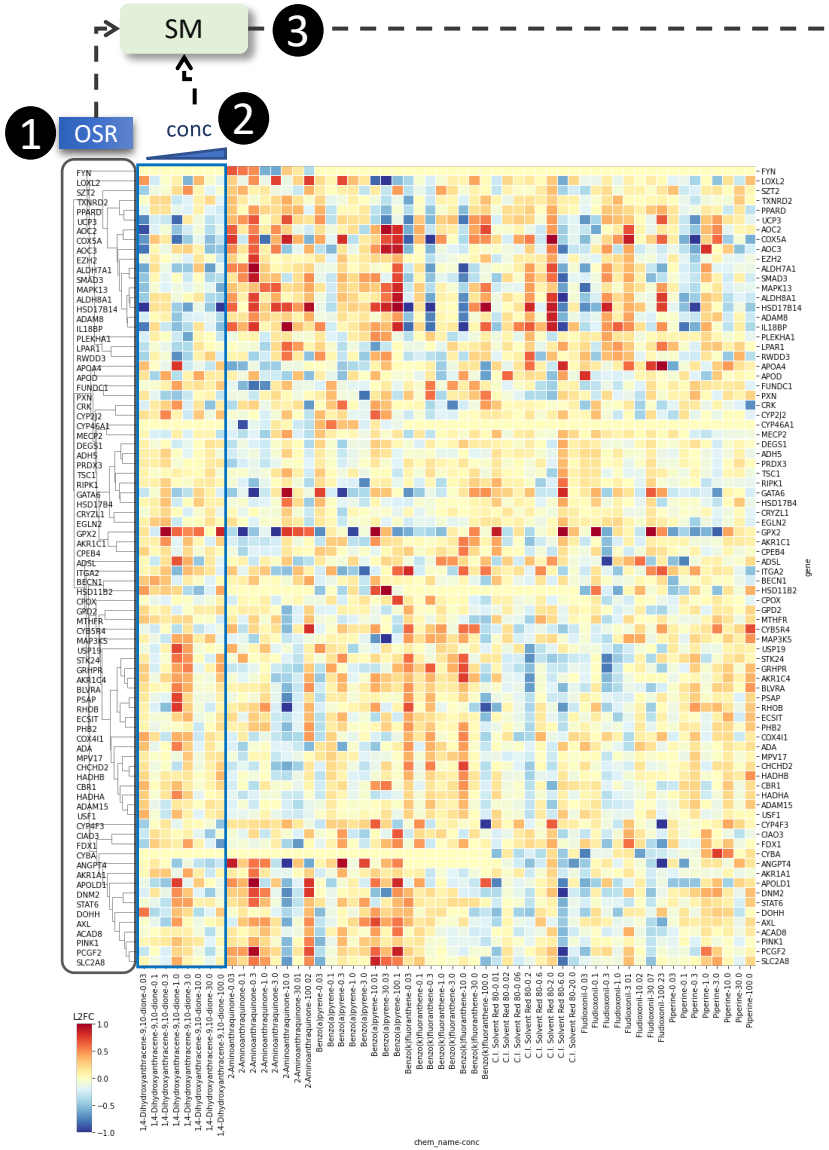
4

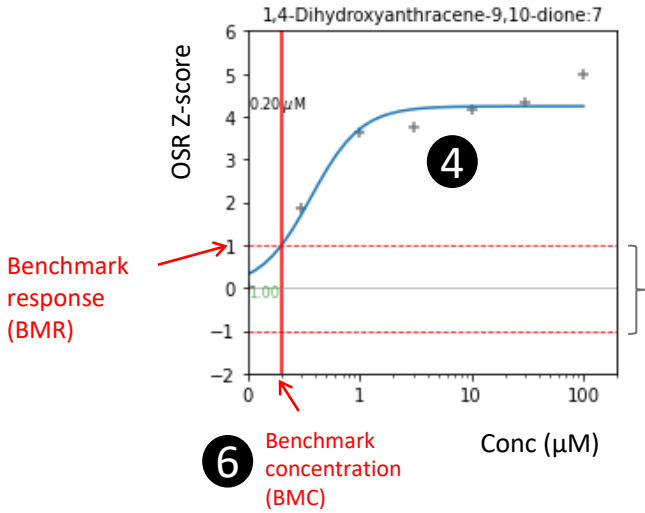


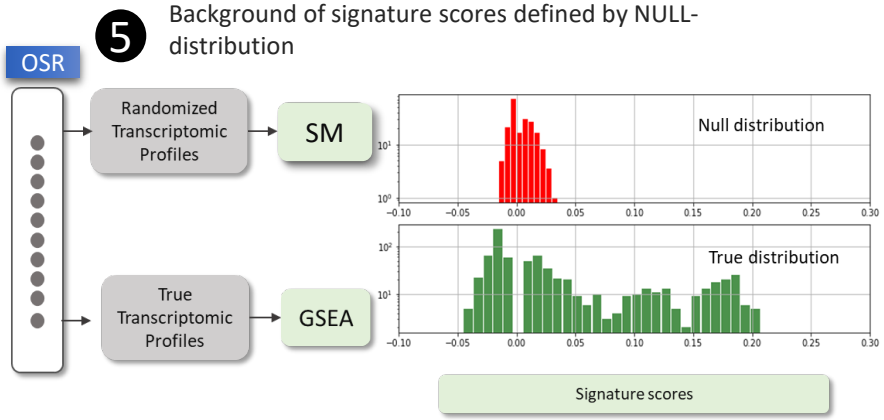
- Consensus signatures for most SRPs can identify reference perturbagens
- Can use SRP signatures to evaluate non-specific chemicals

Chambers *et al.* (in prep)

Using signature scores to estimate PODs

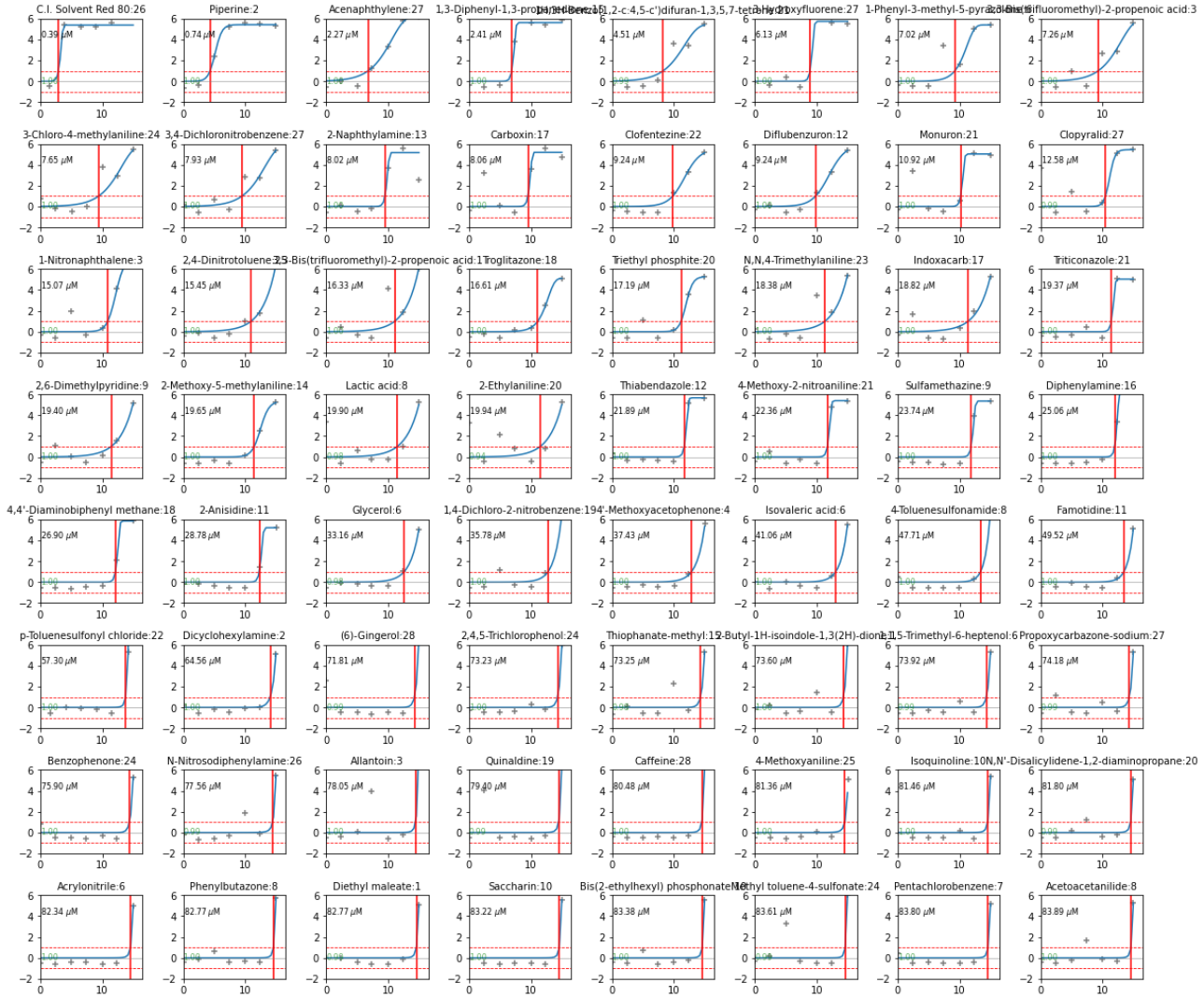


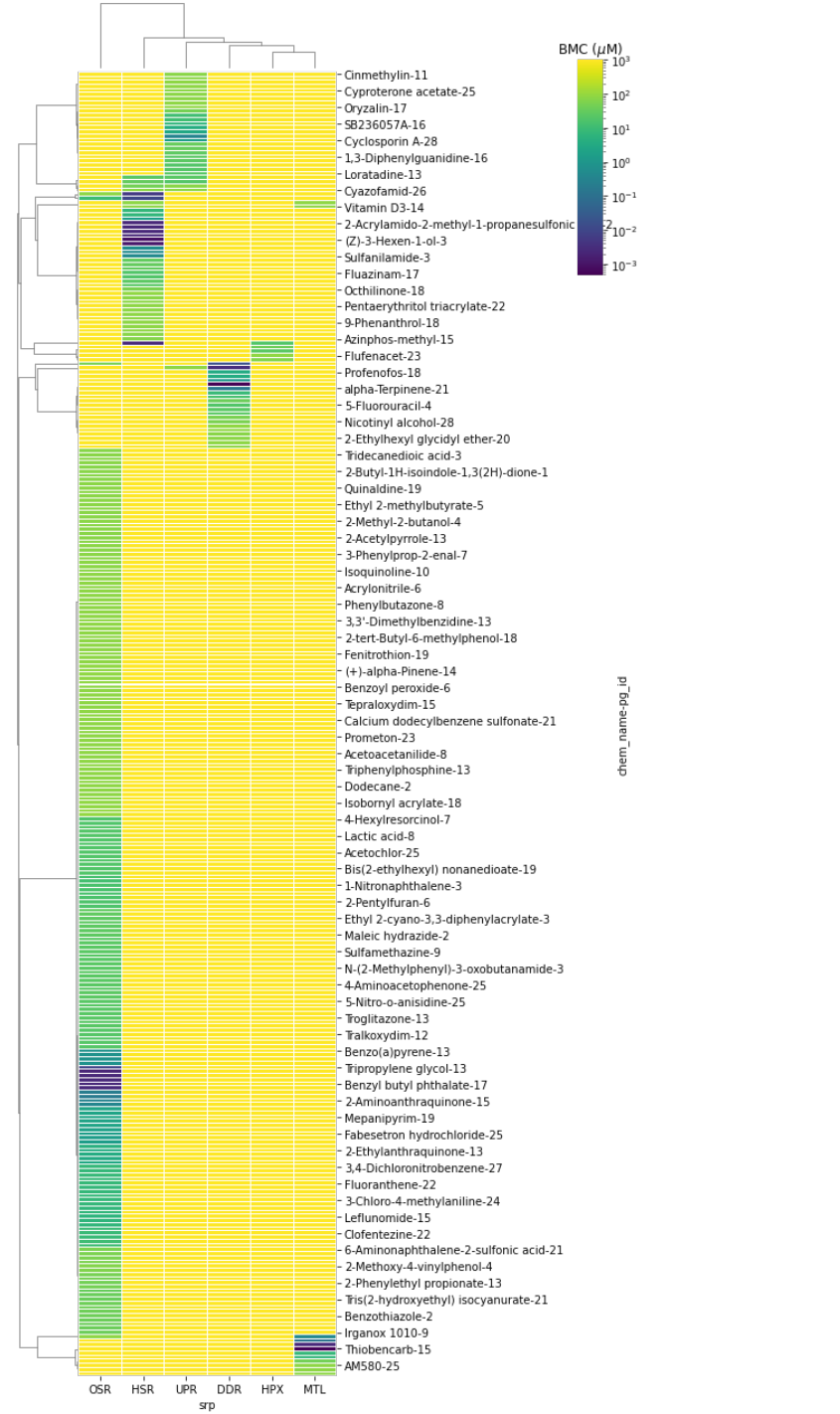


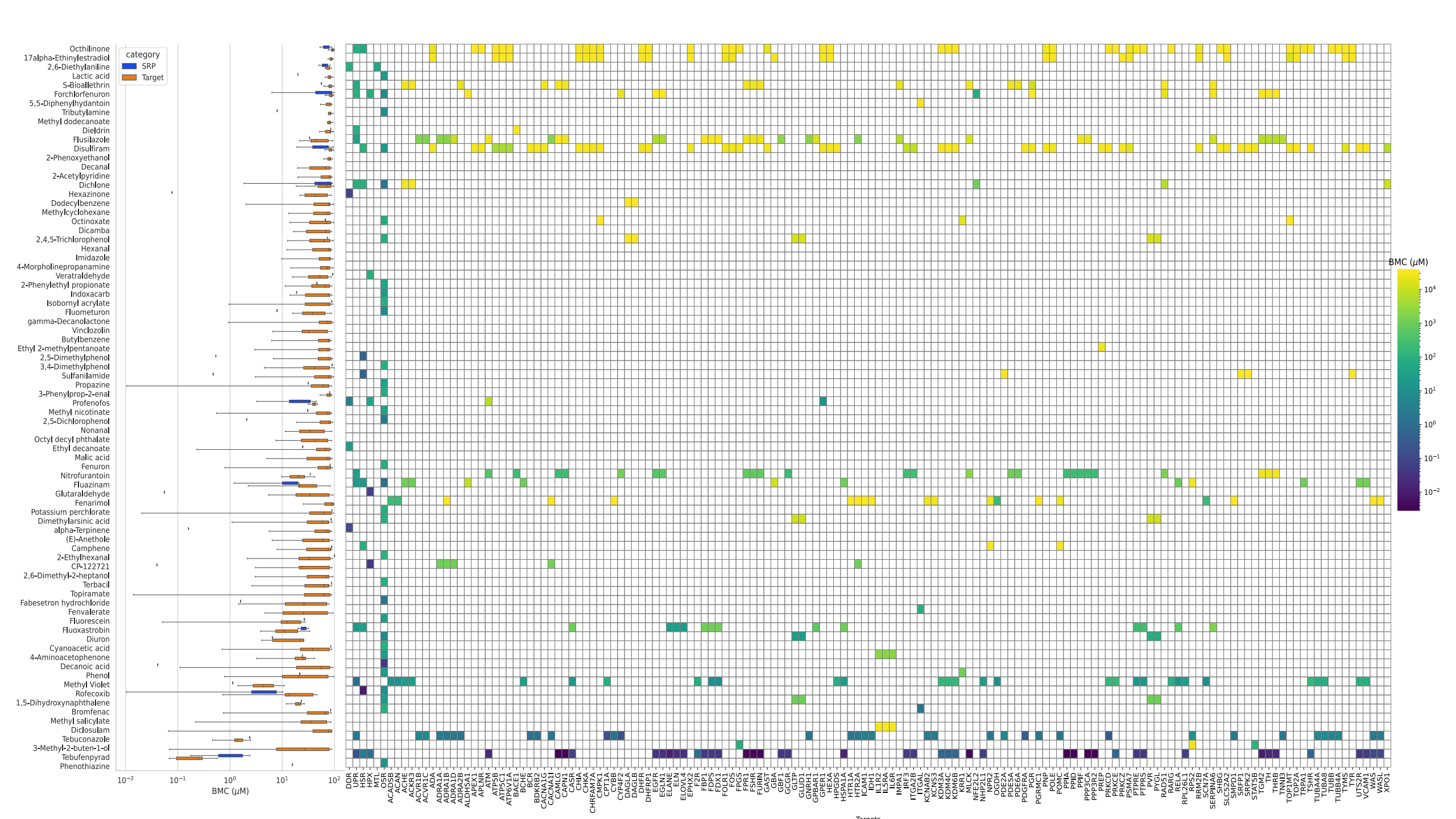


- Standardize signature scores using background distribution (Z)
- Concentration-response modeling using tcplFit2 (cnst, hill, gnls, poly1, poly2, pow, exp2, exp3, exp4, exp5)
- Estimate benchmark concentration (BMC) using benchmark response (BMR) of Z=1

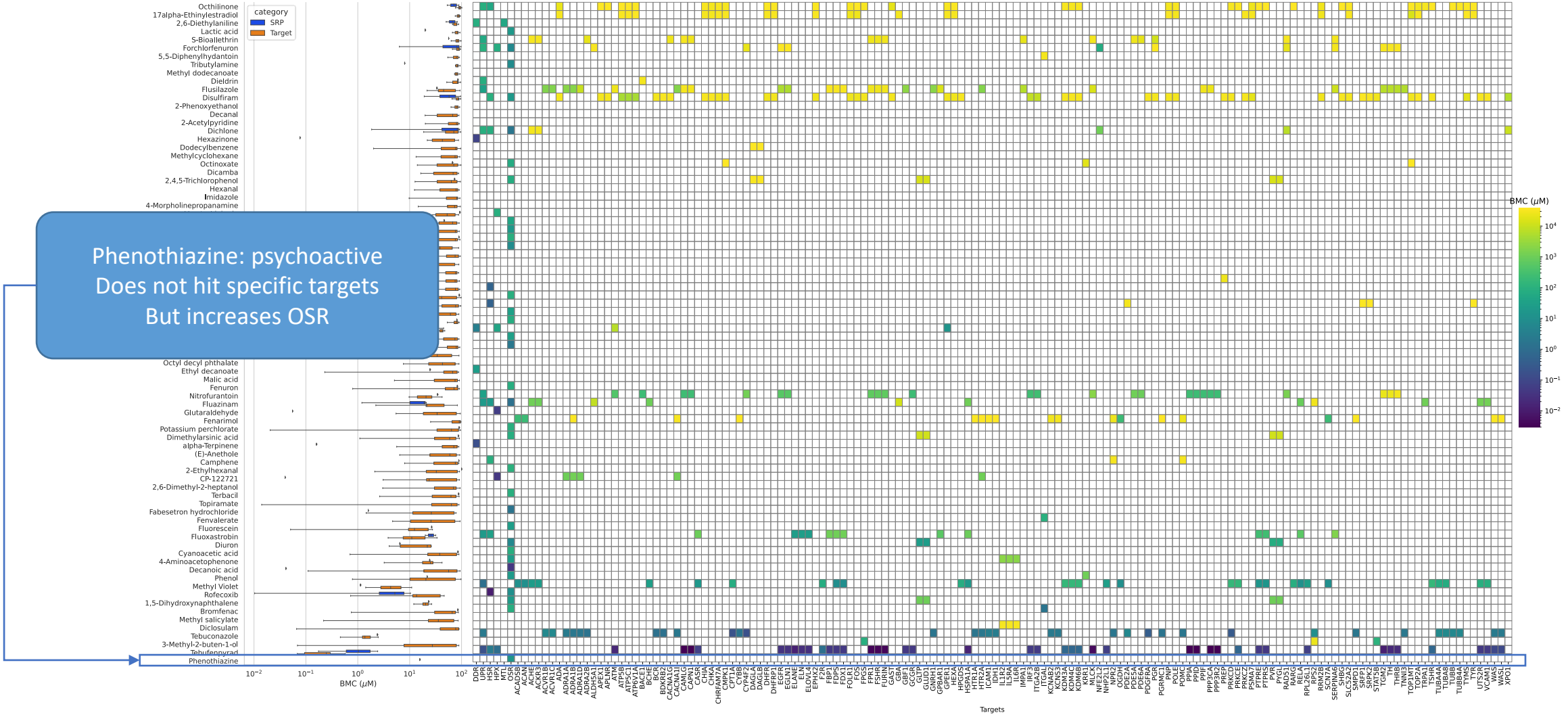
PODs for all chemicals and SRPs



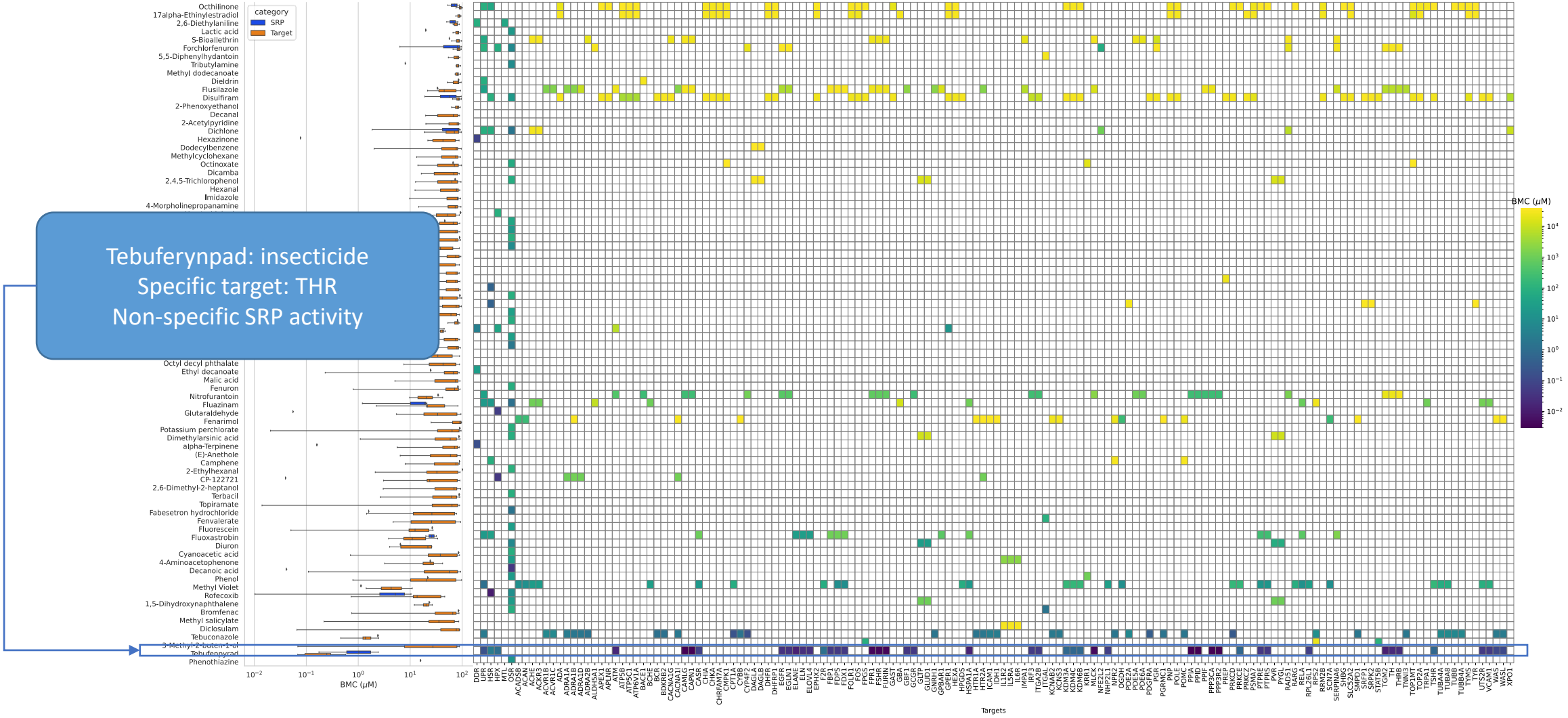



category
SRP
Target
Octhilinone
17alpha-Ethinylestradiol
2,6-Diethylaniline
Lactic acid
S-Bioallethrin
Forchlorfenuron
5,5-Diphenylhydantoin
Tributylamine
Methyl dodecanoate
Dieldrin
Flusilazole
Disulfiram
2-Phenoxyethanol
Decanal
2-Acetylpyridine
Dichlone
Hexazinone
Dodecylbenzene
Methylcyclohexane
Octinoxate
Dicamba
2,4,5-Trichlorophenol
Hexanal
Imidazole
4-Morpholinepropanamine
Veratraldehyde
2-Phenylethyl propionate
Indoxacarb
Isobornyl acrylate
Fluometuron
gamma-Decanolactone
Vinclozolin
Butylbenzene
Ethyl 2-methylpentanoate
2,5-Dimethylphenol
3,4-Dimethylphenol
Sulfanilamide
Propazine
3-Phenylprop-2-enal
Profenofos
Methyl nicotinate
2,5-Dichlorophenol
Nonanal
Octyl decyl phthalate
Ethyl decanoate
Malic acid
Fenuron
Nitrofurantoin
Fluazinam
Glutaraldehyde
Fenarimol
Potassium perchlorate
Dimethylarsinic acid
alpha-Terpinene
(E)-Anethole
Camphene
2-Ethylhexanal
CP-122721
2,6-Dimethyl-2-heptanol
Terbacil
Topiramate
Fabesetron hydrochloride
Fenvalerate
Fluorescein
Fluoxastrobin
Diuron
Cyanoacetic acid
4-Aminoacetophenone
Decanoic acid
Phenol
Methyl Violet
Rofecoxib
1,5-Dihydroxynaphthalene
Bromfenac
Methyl salicylate
Diclosulam
Tebuconazole
3-Methyl-2-buten-1-ol
Tebufenpyrad
Phenothiazine
10^{-2}
10^{-1}
10^{0}
10^{1}
10^{2}
BMC (μM)
DDR UPR HSR HPX MTL OSR ACADSB ADCY1 ACHE ACKR3 ACVR1B ACVR1C ADA ADRA1A ADRA1B ADRA2D ADRA2B ALDH5A1 APEX1 APLNR ATM ATP5B ATP5C1 ATP5F1A BACE1 BCHE BCR BDKRB2 CACNA1G CACNA1I CAMLG CAPN1 CASR CHIA CHKA CHRFAM7A CMPK1 CPT1A CPT2 CYP27B1 DAGLA DAGLB DHFR DHFRP1 EGFR EGLN1 ELANE ELN ELOVL4 EPHX2 F2R FBP1 FDP5 FDX1 FOLR1 FPGS FPR1 FSHR FURIN GAST GBA GBE1 GCGR GLTP GLUD1 GNRH1 GPBAR1 GPER1 HEXA HPGDS HSPA1A HTR1A HTR2A ICAM1 IDH1 IL1R2 IL5RA IL6R IMPA1 IRF3 ITGA2B ITGAL KCNAB2 KCNS3 KDM3A KDM4C KDM4B KRR1 MLCK NFE2L2 NHP2L1 NPR2 OGDH PDE2A PDE3A PDE5A PDE6A PDGFRA PGR PGRMC1 PNP POLE POMC PPM1A PPID PPIF PPP3CA PPP3R2 PREP PRKCD PRKCE PRKCZ PSMA7 PTPRE PTPRS PVR PYGL RAD51 RANGAP1 RLA1 RPL26L1 RPS2 RRM2B SCN7A SERPINA6 SHBG SLC22A2 SLCO2B1 SMPD1 SRFP1 SRPK2 STAT5B TGM2 TH THRB TNNT3 TOP1MT TOP2A TRPA1 TSHR TUBA4A TUBA8 TUBB TUBB4A TYMS TYR UTS2R VCAM1 WAS WASL XPO1
Targets
BMC (μM)
10^{4}
10^{3}
10^{2}
10^{1}
10^{0}
10^{-1}
10^{-2}

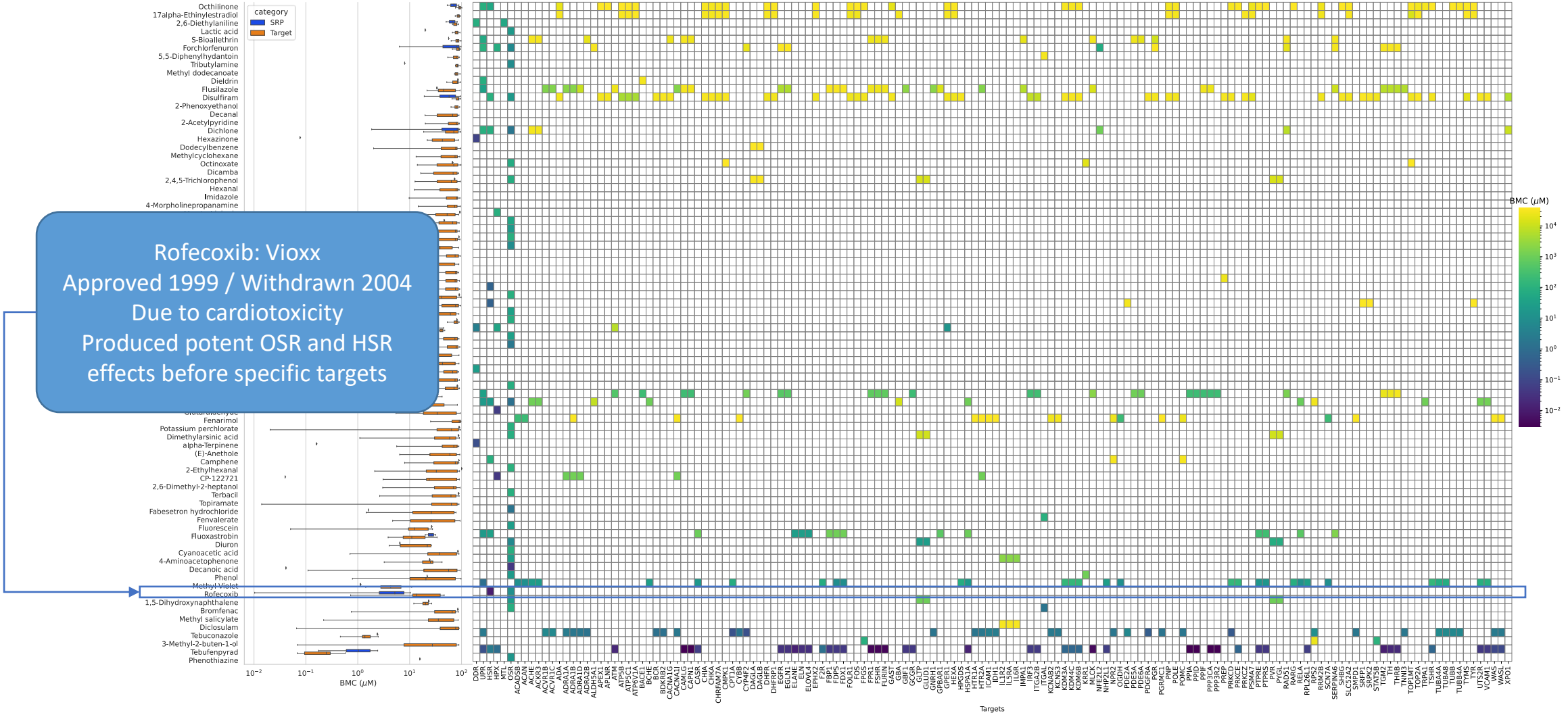
Summary of hits for HepaRG treatments



Summary of hits for HepaRG treatments



Summary of hits for HepaRG treatments



Summary

1. High-throughput transcriptomics is promising for NAM development

   We are using TempO-Seq technology (targeted RNA-Seq) 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. Feasible to identify hazard and estimate POD using gene signature “similarity”

   We are systematically evaluating gene signature-based connectivity mapping and other approaches for identifying putative targets, AOPs and *in vitro* POD values. Gene signature-based approaches are more sensitive than single gene-based techniques.

3. Connectivity mapping, read-across and risk assessment

   Transcriptomic nearest-neighbor techniques are conceptually like expert read-across approaches, which are widely used to fill data gaps for untested chemicals. Could be easier “sell” than more sophisticated network inference and AI/ML/DL.

4. Future directions

   Systems biology of adaptive stress response pathways using transcriptomics to investigate the molecular basis of cellular resilience and tipping points; streamline the development of NAMs for evaluating untested chemicals based on adaptive stress responses.

Acknowledgements

US EPA

Bryant Chambers

Joe Bundy

Derik Haggard

Joshua Harrill

Logan Everett

Richard Judson

Woodrow Setzer

Beena Valanat

Thomas Knudsen

Tim Shafer

Brian Chorley

John Cowden

Maureen Gwinn

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

References

Thomas, R.S., T. Bahadori, T.J. Buckley, J. Cowden, C. Deisenroth, K.L. Dionisio, J.B. Frithsen, et al. 2019. "The next Generation Blueprint of Computational Toxicology at the U.S. Environmental Protection Agency." *Toxicological Sciences* 169 (2). https://doi.org/10.1093/toxsci/kfz058.

Yeakley, Joanne M., Peter J. Shepard, Diana E. Goyena, Harper C. VanSteenhouse, Joel D. McComb, and Bruce E. Seligmann. 2017. "A Trichostatin A Expression Signature Identified by TempO-Seq Targeted Whole Transcriptome Profiling." Edited by Yan Xu. *PLOS ONE* 12 (5): e0178302. https://doi.org/10.1371/journal.pone.0178302.

Harrill, Joshua A, Logan J Everett, Derik E Haggard, Thomas Sheffield, Joseph Bundy, Clinton M Willis, Russell S Thomas, Imran Shah, and Richard S Judson. 2021. "High-Throughput Transcriptomics Platform for Screening Environmental Chemicals." *Toxicological Sciences*, February. https://doi.org/10.1093/toxsci/kfab009.

Lamb, Justin, Emily D Crawford, David Peck, Joshua W Modell, Irene C Blat, Matthew J Wrobel, Jim Lerner, et al. 2006. "The Connectivity Map: Using Gene-Expression Signatures to Connect Small Molecules, Genes, and Disease." *Science (New York, N.Y.)* 313 (5795): 1929–35. https://doi.org/10.1126/science.1132939.

Liu, Jie, Grace Patlewicz, Antony J. Williams, Russell S. Thomas, and Imran Shah. 2017. "Predicting Organ Toxicity Using in Vitro Bioactivity Data and Chemical Structure." *Chemical Research in Toxicology* 30 (11): 2046–59. https://doi.org/10.1021/acs.chemrestox.7b00084.

Liu, Anika, Panuwat Trairatphisan, Enio Gjerga, Athanasios Didangelos, Jonathan Barratt, and Julio Saez-Rodriguez. 2019. "From Expression Footprints to Causal Pathways: Contextualizing Large Signaling Networks with CARNIVAL." *Npj Systems Biology and Applications* 5 (1): 1–10. https://doi.org/10.1038/s41540-019-0118-z.

Shah, Imran, R Woodrow Setzer, John Jack, Keith A Houck, Richard S Judson, Thomas B Knudsen, Jie Liu, et al. 2016. "Using ToxCast™ Data to Reconstruct Dynamic Cell State Trajectories and Estimate Toxicological Points of Departure." *Environmental Health Perspectives* 124 (7): 910–19. https://doi.org/10.1289/ehp.1409029.

Abrew, K. Nadira De, Raghunandan M. Kainkaryam, Yuqing K. Shan, Gary J. Overmann, Raja S. Settivari, Xiaohong Wang, Jun Xu, et al. 2016. "Grouping 34 Chemicals Based on Mode of Action Using Connectivity Mapping." *Toxicological Sciences* 151 (2): 447–61. https://doi.org/10.1093/toxsci/kfw058.

Wang, Rong-Lin, Adam D Biales, Natalia Garcia-Reyero, Edward J Perkins, Daniel L Villeneuve, Gerald T Ankley, and David C Bencic. 2016. "Fish Connectivity Mapping: Linking Chemical Stressors by Their Mechanisms of Action-Driven Transcriptomic Profiles." *BMC Genomics* 17 (January): 84. https://doi.org/10.1186/s12864-016-2406-y.

Chambers, B, and I. Shah. n.d. "Elucidating Stress Response Pathway Activity Using Transcriptomics." (*submitted*)

Contact

www.linkedin.com/in/imranshah

Shah.Imran@epa.gov

github.com/i-shah