High-throughput Transcriptomics Concentration-Response Screen of 1593 Chemicals in MCF7 Cells

Richard Judson, Logan Everett, Derik Haggard, Joseph Bundy, Bryant Chambers, Laura Taylor, Beena Vallanat, Imran Shah, Joshua Harrill

US EPA, Office of Research and Development

Richard Judson | email: judson.richard@epa.gov | 919-449-7514
ORCiD: orcid.org/0000-0002-2348-9633

Introduction

Goal: Use whole-genome transcriptomics methods to define points of departure for chemicals and determine mechanism of action

Key Topics

- Concentration-response transcriptomics data can now be generated on hundreds to thousands of compounds
- We have twin goals of using this data for hazard identification (what pathways / targets will a chemical activate) and estimation of points of departure (POD)
- Data can be analyzed at the gene or gene-set / signature level
- Here we present preliminary work on using signature modeling methods to analyze data from a 1593-chemical screen run by EPA

Computational Details

Experimental Details (detailed in [1]):

- 1593 unique chemicals
- 8-point concentration-response in MCF7 cells, 3 biological replicates
- 33 pairs of replicate chemicals
- All cell culture and chemical exposures performed at EPA
- Cells were lysed 6 hours after exposure
- Transcriptomics performed by BioSpyder using the Temp-O-Seq platform, human probe set version 1

Computational Pre-processing

- Raw count is input to CCTE pipeline
- This is converted to log2 fold-change (L2FC) using DESeq2 - https://bioconductor.org/packages/release/bioc/html/DESeq2.html
- L2FC data (one value per chemical sample / probe / time / concentration) is the input to signature modeling
- A "signature" is a collection of genes
- Signatures are taken from (total of 10456)
  - MSigDB (http://software.broadinstitute.org/gsea/msigdb)
  - NCATS BioPlanet
  - DisGenNET – disease gene sets
- Signatures are labeled with a "target", which can be a molecular target, or a process at the molecular, cellular, tissue or organism level, e.g. Estrogen, PPAR, Cancer. There are multiple signatures per target
- Molecular and process targets for tested chemicals have also been annotated, using the same terminology as for the signatures

References

[1] Harrill et al. "High-Throughput Transcriptomics Platform for Screening Environmental Chemicals", Tox Sci 2021
[2] Judson et al. "Integrated Model of Chemical Perturbations of a Biological Pathway Using 18 *In Vitro* High-Throughput Screening Assays for the Estrogen Receptor", Tox Sci 2015
[3] Judson et al. "Analysis of the Effects of Cell Stress and Cytotoxicity on *In Vitro* Assay Activity Across a Diverse Chemical and Assay Space", Tox Sci 2016

Concentration-Response Approach

Method:

- We use the "Fold-Change" (FC) method
- Score (one sample, one concentration) = median(L2FC, genes in the signature) – median(L2FC, genes out of the signature)
- Background level of activity is calculated using a null distribution
  - Permute the entire data set 1000 times (creating 1000 concentration-response series), drawing from the underlying distribution for each genes
  - Correlation between genes is broken
  - Scores calculated for each signature
  - Cutoff = 95% CI of the null distribution
- Curve fitting uses package *tcplfit2*, tries multiple methods and selects the one with the lowest AIC
- Benchmark dose (BMD) is calculated as the concentration where the selected curve exceeds the benchmark response level (BMR=1.349 SD of the null)
- A continuous hitcall (range 0 to 1) is calculated
- Activity is defined here where hitcall>0.95 and top / cutoff>1.5

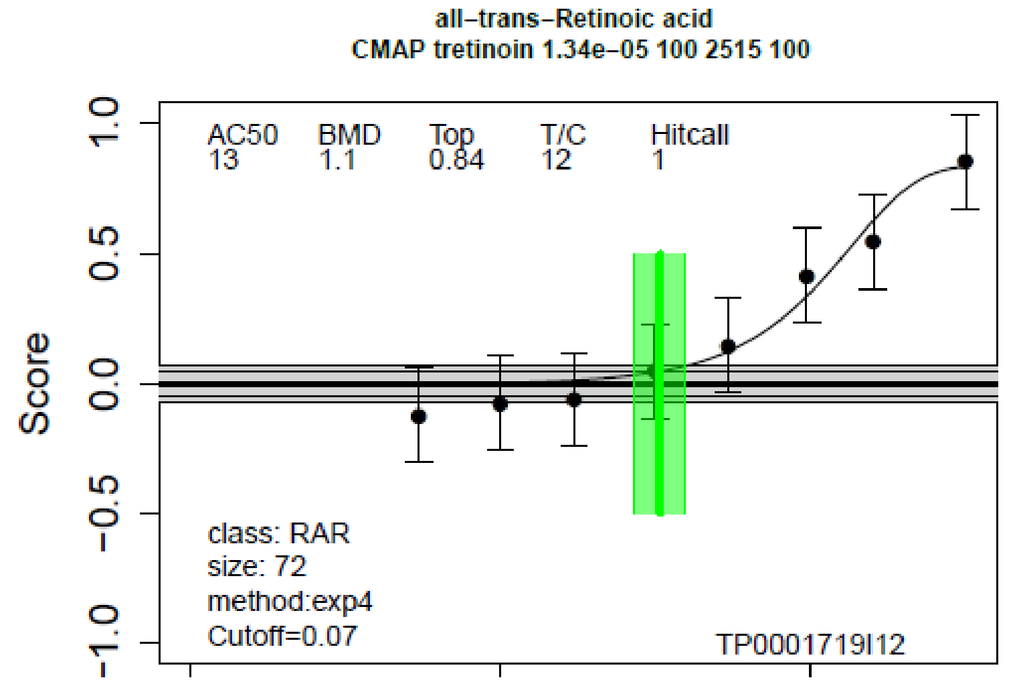

Figure 1: Example of a concentration-response plot. The header shows the chemical and the specific signature. The gray band is the null background, and the inner black horizontal line is the BMR. The green band shows the BMD and its upper and lower 95% CI. At the bottom left of the graph is the class or target, the number of genes in the signature, the winning curve-fit method and the cutoff value. (RAR=Retinoic Acid Receptor)

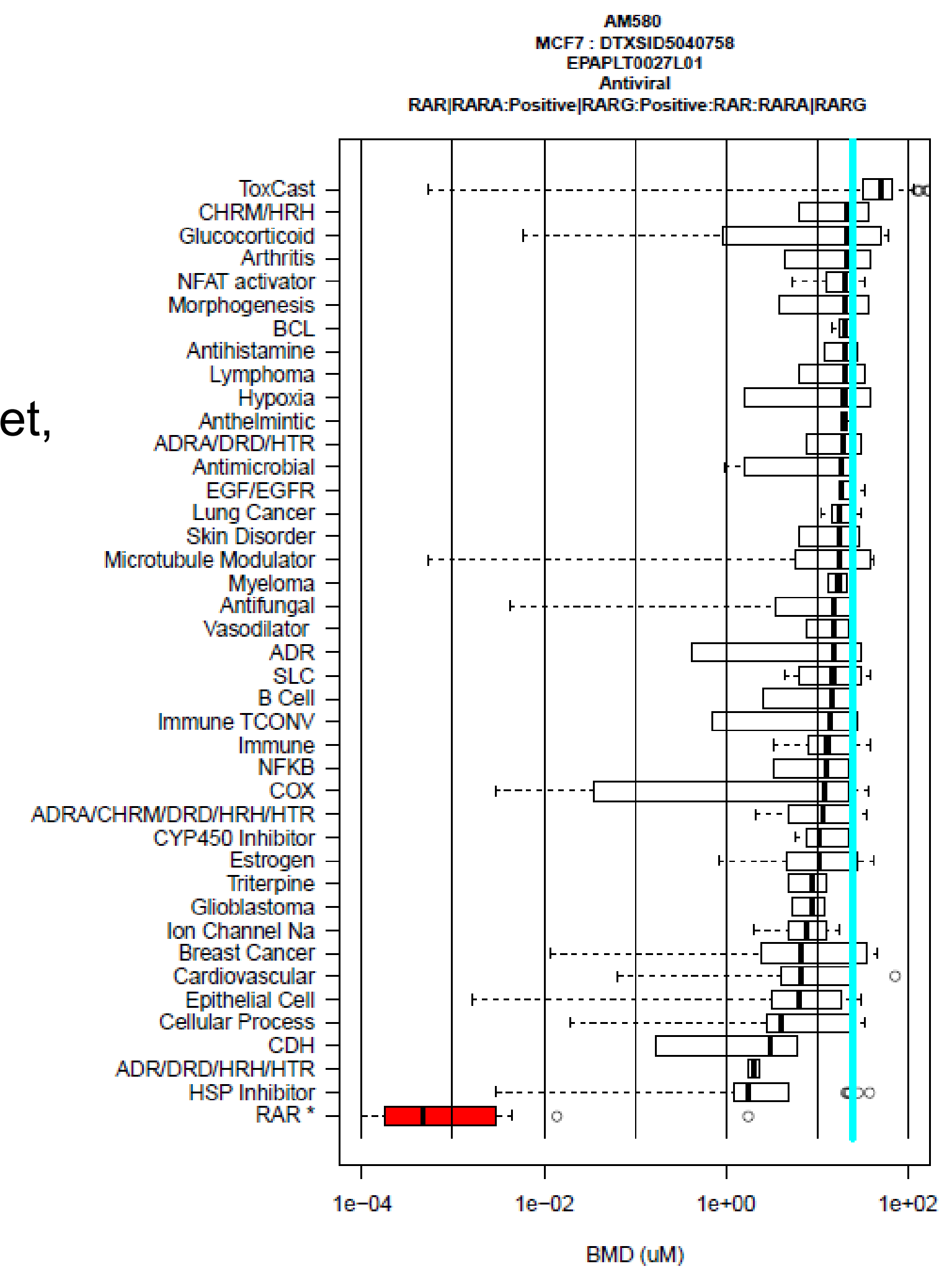

Figure 2: Summary of activity across all targets for a single chemical. For each target, the box shows the BMDs for all signatures corresponding to the target. Where the molecular and the chemical targets match, the box is colored red.
The blue vertical line indicates the median value of all active targets. Only the most potent 40 targets are shown. At the top of the boxes is the distribution of potency values from the in vitro assays in the ToxCast program for this chemical.
The "Target-based POD" is the median BMD across signatures for the most potent target, here RAR.

Potency Trends

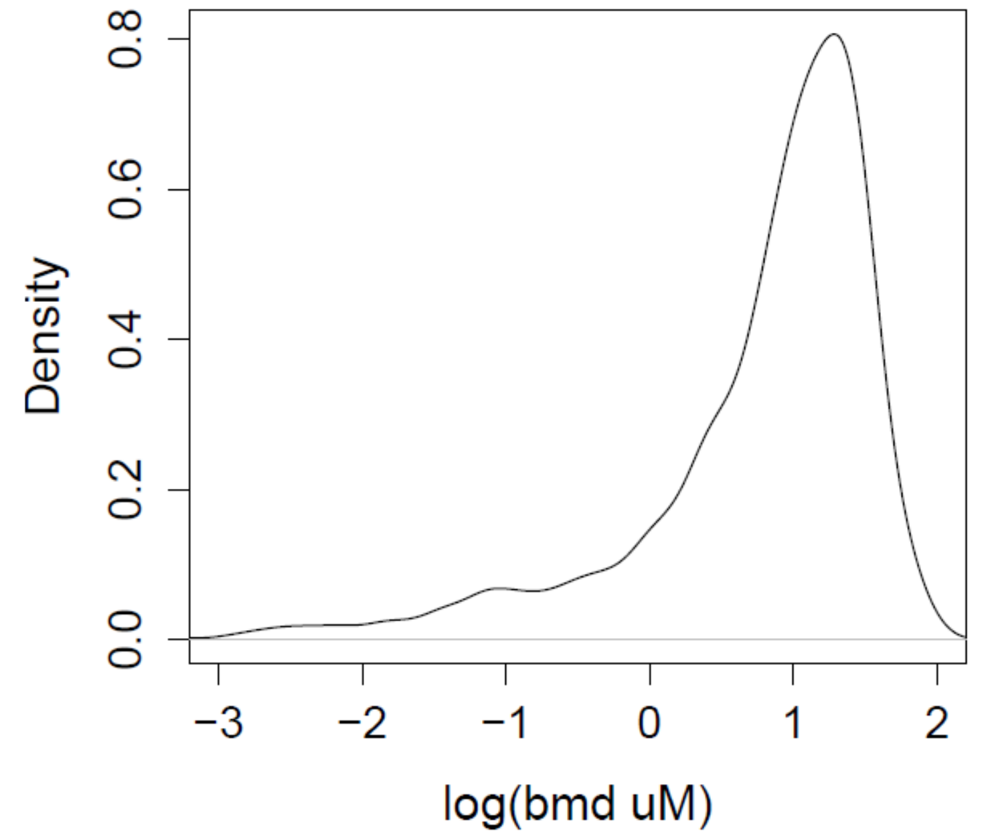

Figure 3: Density plot of chemical-level BMD values for all active chemicals. The metric used is the BMD of the most potent target that has at least 2 active signatures. Note that the majority of activity is > ~ 1 uM

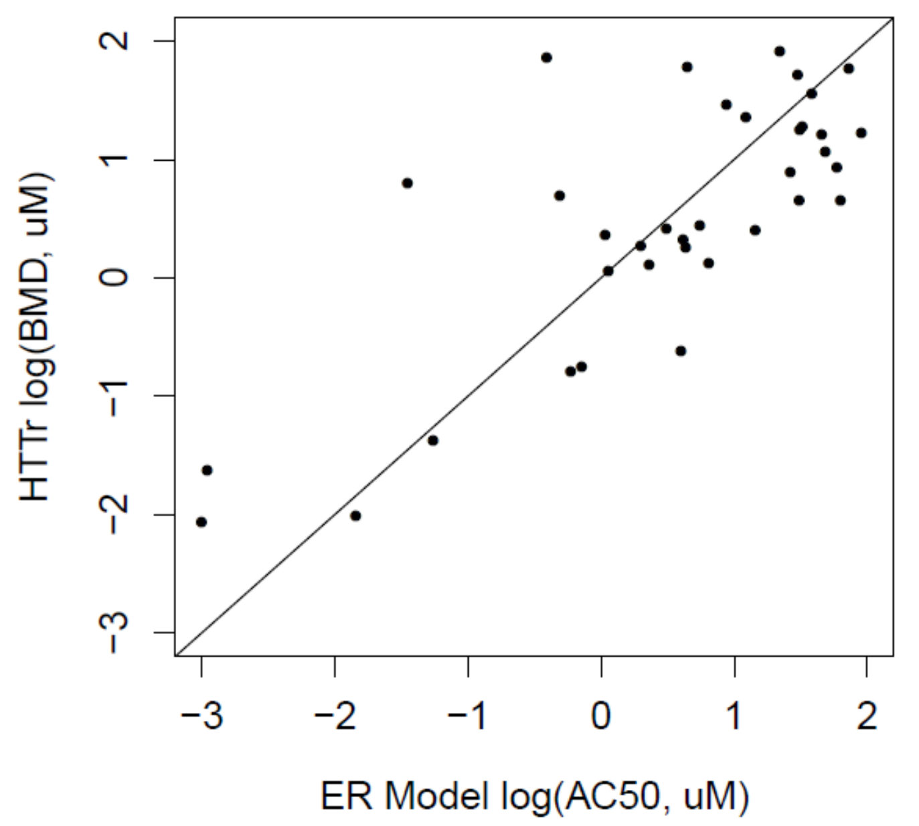

Figure 4: Comparison of Estrogen target PODs from transcriptomics against POD values for estrogen activity from 18 in vitro assays[2]. The R^2 value is 0.65 and RMSE=0.7 in log units.

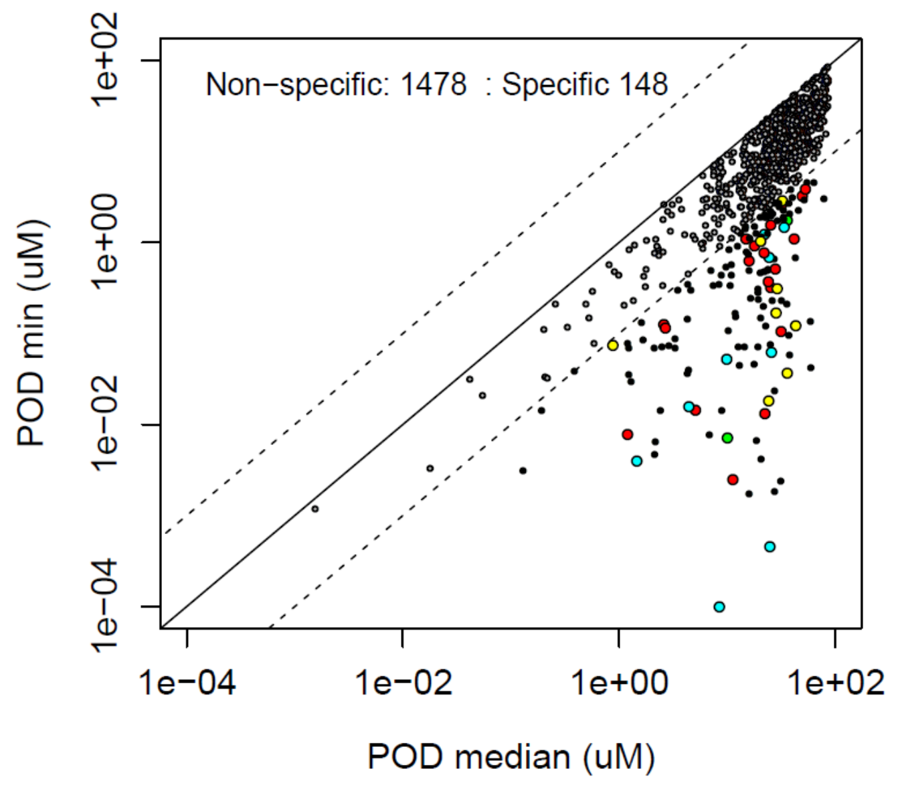

Figure 5: Specificity analysis. Each point is a single chemical sample. The x-axis is the target-based POD and the y-axis is the median BMD across all targets and corresponds to the high-concentration cell stress burst [3]. Chemicals with no activity below ~burst/10 are defined to be non-specific (90% of chemicals). The 10% specific chemicals are most active against a small set of targets (Estrogen: red, RAR: cyan, Glucocorticoid: green, Androgen: yellow)

Replication Analysis

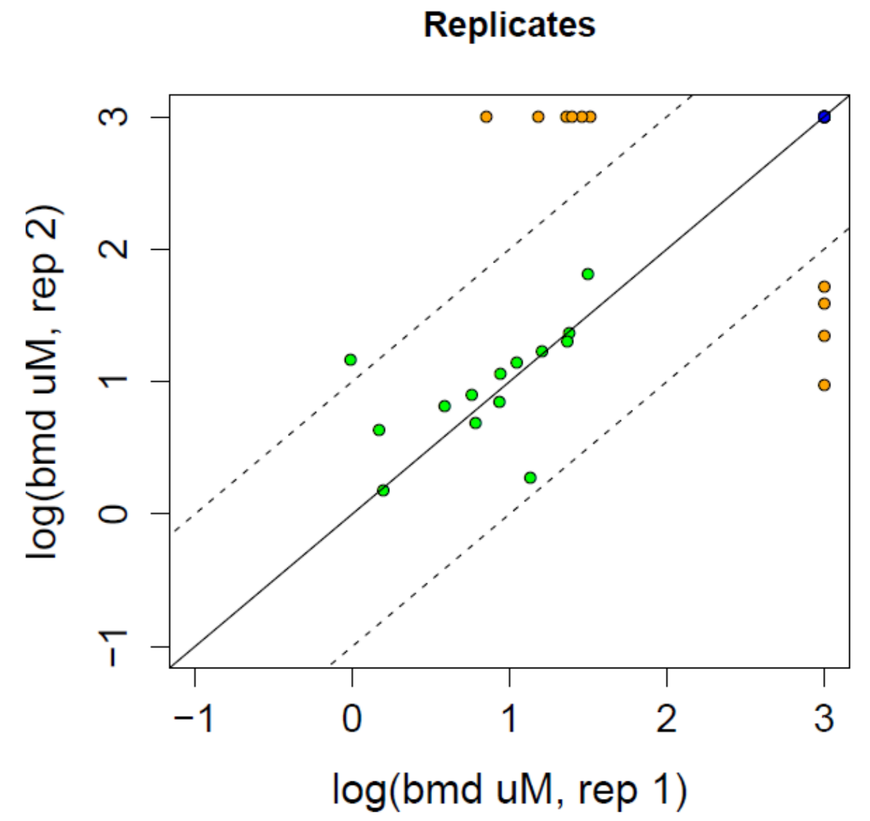

Figure 6: Replication Analysis. The x and y axes are the target-based POD for the two replicates for the 33 chemicals with 2 samples. Green points are chemicals where both samples are active (14/33). Orange points are chemicals where one sample is active and one is inactive (10/33, most where activity is >10 uM, close to the upper limit of testing of 100 uM), Blue points (position (3,3)) are inactive in both samples (9/33).

Reference Chemical Analysis

Goals:

- Use data from public sources to annotate chemicals with a use category and molecular and other biological process targets
- Key question is "What (types of) targets are transcriptionally accessible?". Nuclear receptor targets should be, but enzymes, transporters and other receptors may not be.

Method

- Create 2x2 matrix for each target with
  - TP=# of chemical annotated to be active in the target that are active
  - FP=# of chemical not annotated to be active in the target that are active
  - FN=# of chemical annotated to be active in the target that are not active
  - TP=# of chemical not annotated to be active in the target that are not active

Super Target	TP	FP	FN	TN	Sens	Spec
Cardiac Glycoside Drug	4	248	0	1016	1.00	0.80
Cholinergic Muscarinic Receptor	4	476	0	788	1.00	0.62
ATPase Inhibitor	4	572	0	692	1.00	0.55
Estrogen Receptor	51	713	2	502	0.96	0.41
Serotonin Receptor	9	526	1	732	0.90	0.58
Glucocorticoid Receptor	14	459	2	793	0.88	0.63
Retinoic Acid Receptor	6	309	1	952	0.86	0.75
Dopamine Receptor	6	524	1	737	0.86	0.58
Mitochondria	9	516	2	741	0.82	0.59
DNA Synthesis Inhibitor	6	355	2	905	0.75	0.72
Androgen Receptor	25	417	10	816	0.71	0.66
DNA	4	296	2	966	0.67	0.77
GABA Receptor	13	482	7	766	0.65	0.61
Progesterone Receptor	9	370	5	884	0.64	0.70
CYP450 Inhibitor	16	404	9	839	0.64	0.67
Antimicrobial	16	361	14	877	0.53	0.71

Table 1: Statistics for reference chemical comparison for all targets with more than 3 reference chemicals and sensitivity > 0.5. The major classes of targets are nuclear receptors, GPCRs and DNA. However, other targets (CYP450 and ATPase) also yield reasonable sensitivity.

"Off-Target" Activity:

- There are a large number of chemicals active in each of the targets in Table 1 (False Positives). Some of these may be unannotated true positives, but we hypothesize that much of this activity is high-concentration non-specific activity due to cell stress corresponding to the well-known stress "burst" [3]

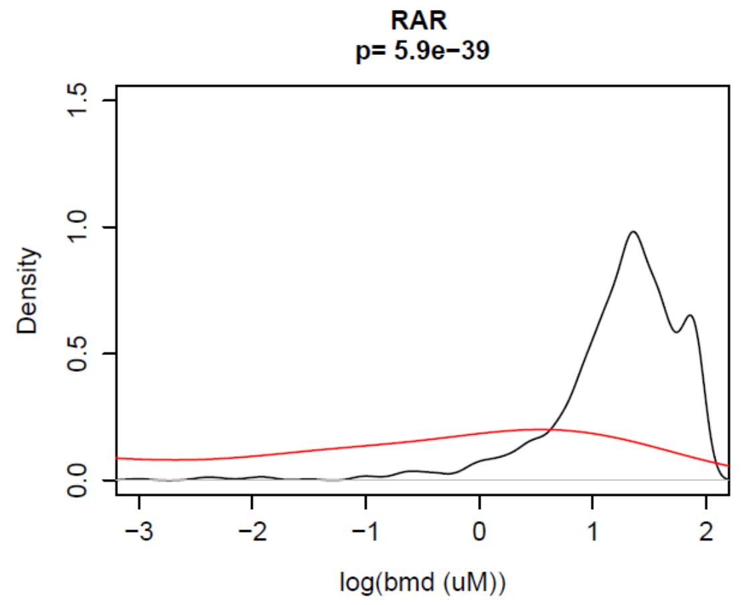

Figure 7: Comparison of the on-target (red, TP) activity for one target (Retinoic Acid Receptor) vs. the off-target (black, FP)

Conclusions

- We have analyzed 1593 environmental chemicals in concentration-response transcriptomics. Pipelines to process this data are now fully automated
- We have shown that multiple targets and target types are transcriptionally accessible
- Samples do not replicate at 100% level, but most instances of non-replication occur at high concentrations
- This data set allows us to determine transcriptional PODs and targets for many chemicals

Innovative Research for a Sustainable Future

This poster does not necessarily reflect U.S. EPA policy