EPA
United States
Environmental Protection
Agency

High-throughput Transcriptomic Profiling of Chemicals for Risk Assessment Applications

Richard Judson

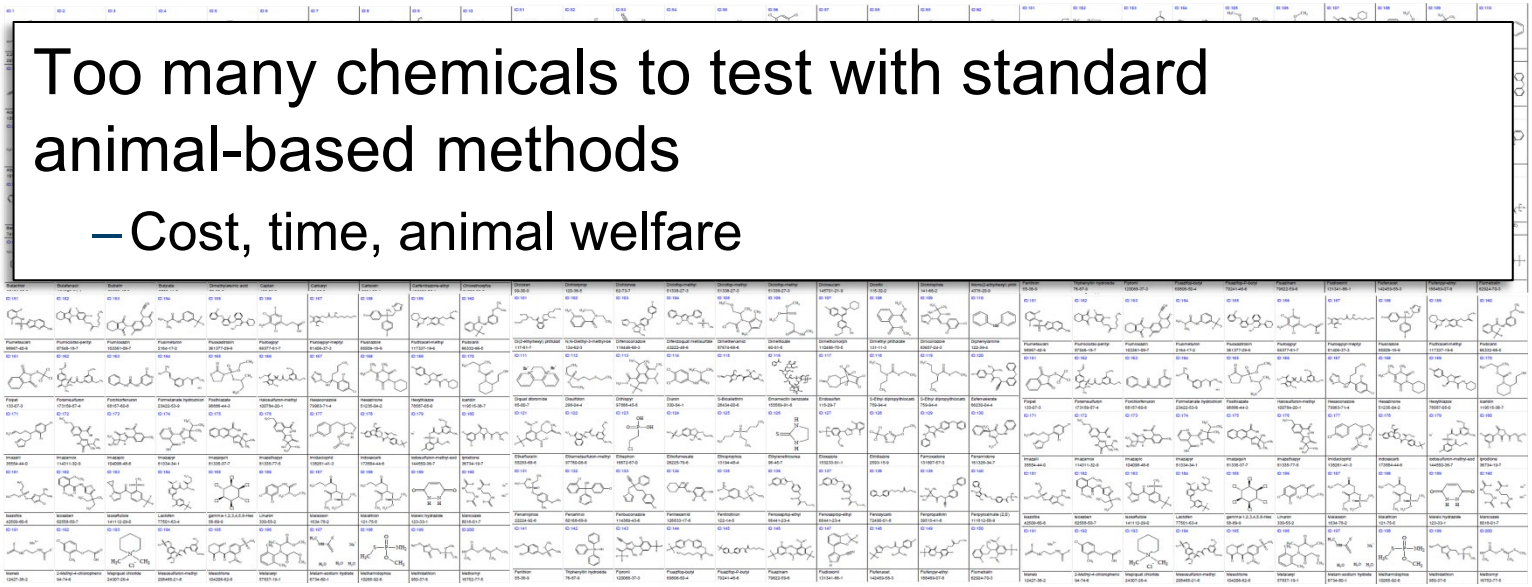
MAQC Society
April 29, 2021

Office of Research and Development
Center for Computational Toxicology and Exposure

Problem Statement

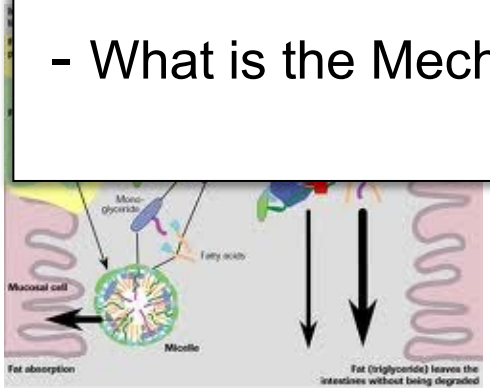
Too many chemicals to test with standard animal-based methods

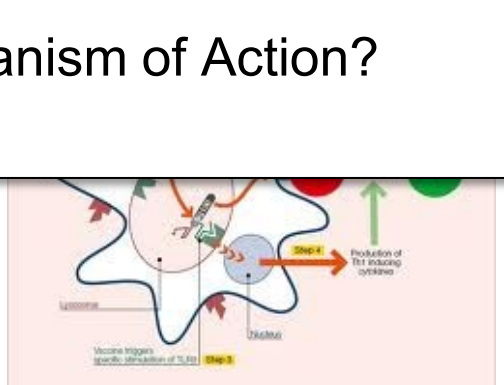
- Cost, time, animal welfare

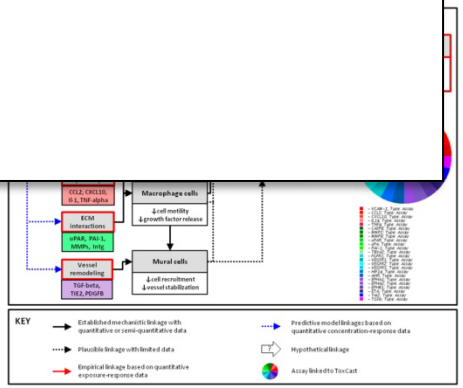


Need for better mechanistic data

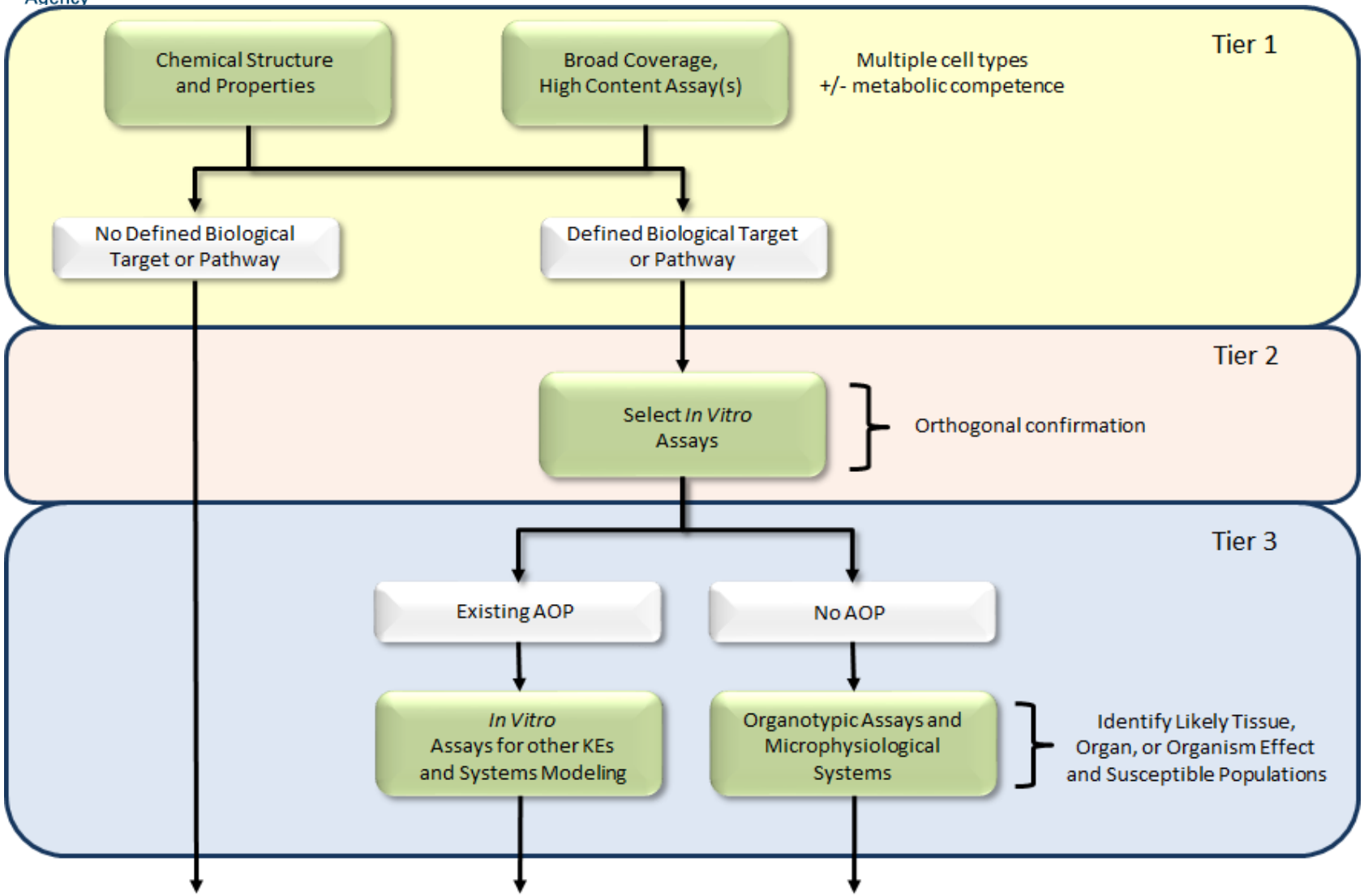
- Determine human relevance
- What is the Mechanism of Action?







Tiered Hazard Evaluation Approach



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

Experimental Workflow

United States Environmental Protection Agency

Cryopreserved Cell Stocks

Cell Expansion

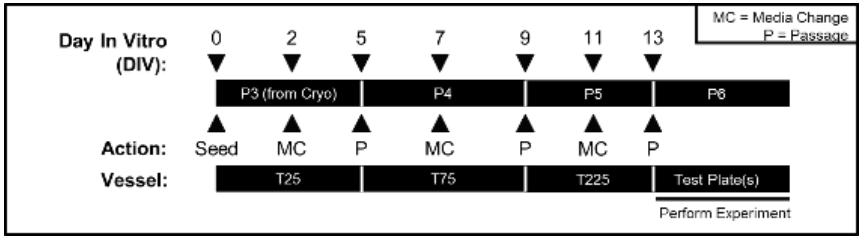
Cell Plating

BioTek MultiFlo ™ FX

Dispensing Test Chemicals

LabCyte Echo® 550 Liquid Handler

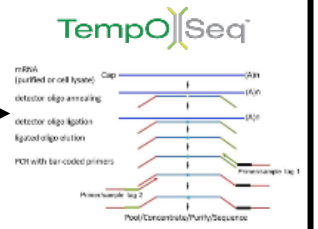
Standardized Expansion Protocol



Track 1: Targeted RNA-Seq

Generating Cell Lysates

TempO-Seq WT



Track 2: Apoptosis / Cell Viability

Reagent Dispensing

High Content Imaging

Perkin Elmer Opera Phenix™ High Content Screening System

HTTr Datasets

Dataset	MCF7 Pilot	MCF7 Screen	HepaRG Screen	U2OS Screen
Tissue	Breast	Breast	Liver	Bone
Chemicals	44	1593 [3]	1323	1324
Samples [1]	350	12959	10825	10766
Genes [2]	10149	9137	12116	11815

Notes:

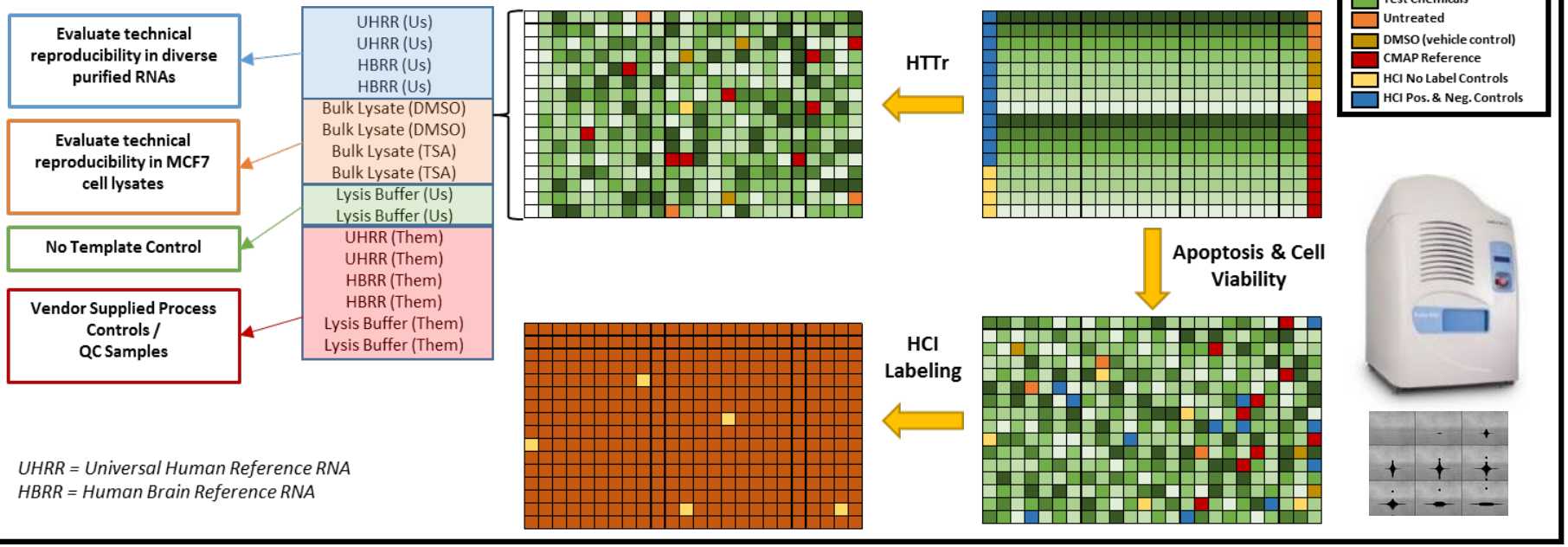
[1] Includes 8 concentrations / chemical and replicates, but not reference chemicals

[2] There may be more than one probe per gene. At least 95% of samples must have at least 5 counts for probe to be included

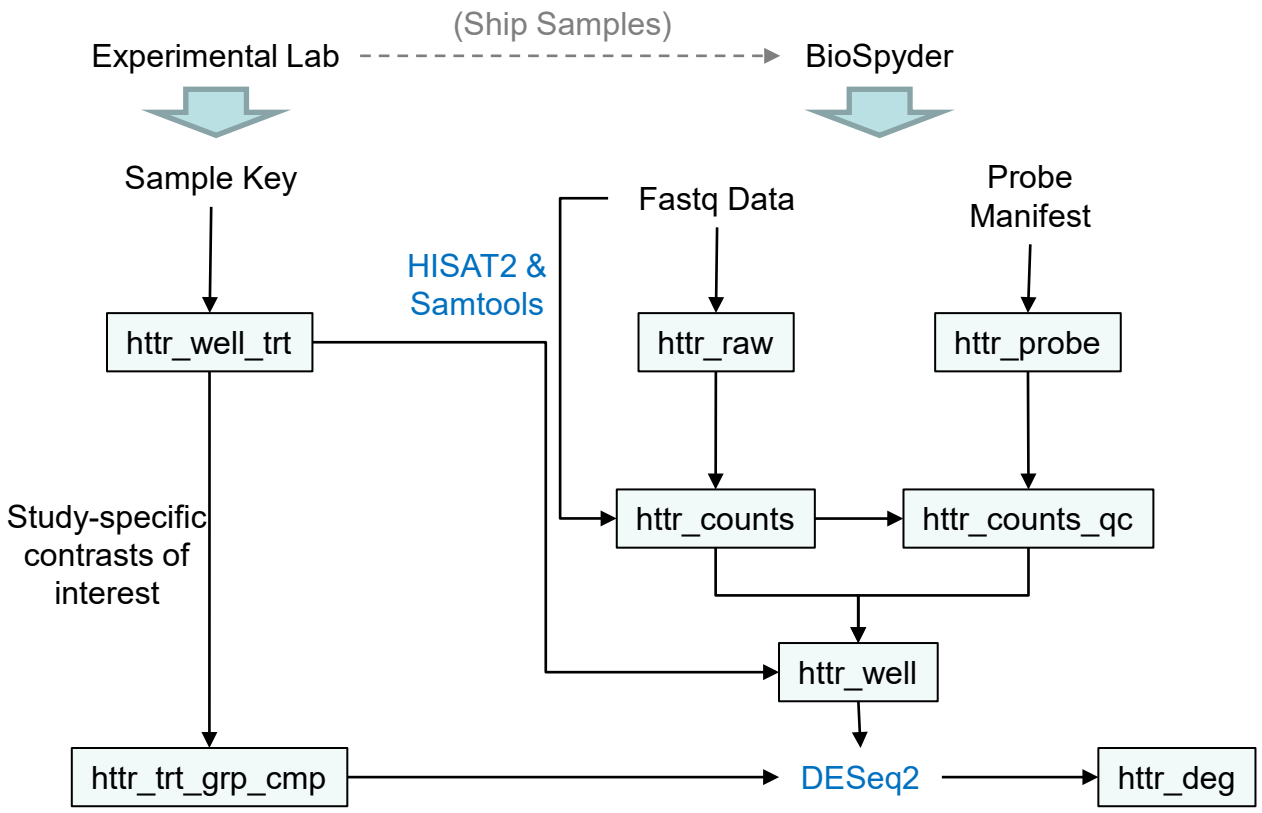
[3] After samples from bad plate groups were removed

Treatment Randomization & Quality Control Samples

Treatment Randomization: *Each test plate uniquely randomized with respect to treatment.*
QC Samples: *Quality Control samples included on each plate*



HTTr Data Management



Scheduled backups
Recovery plan
Rapid export
Open-source tech

L. Everett

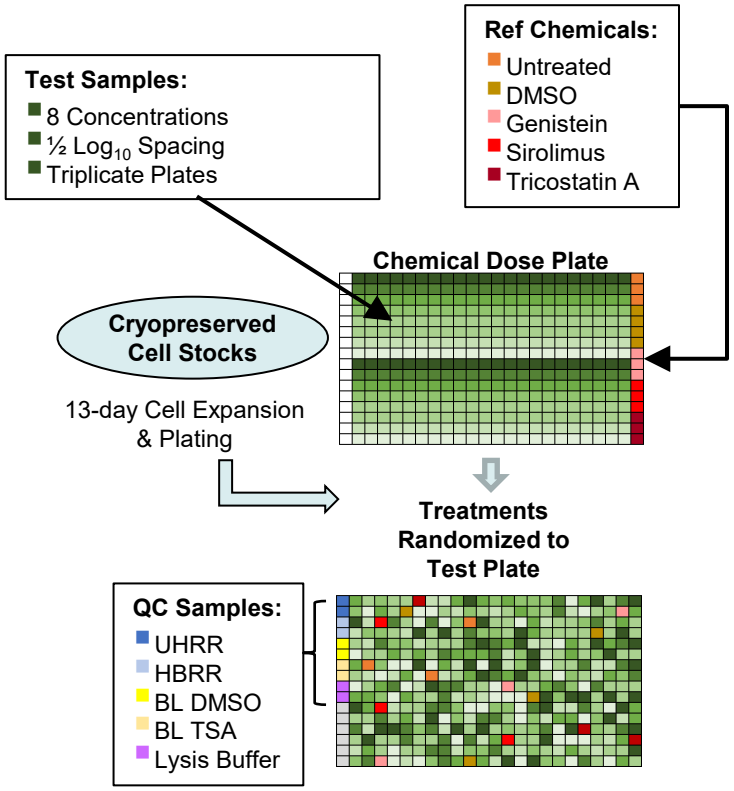
Raw Processing Options

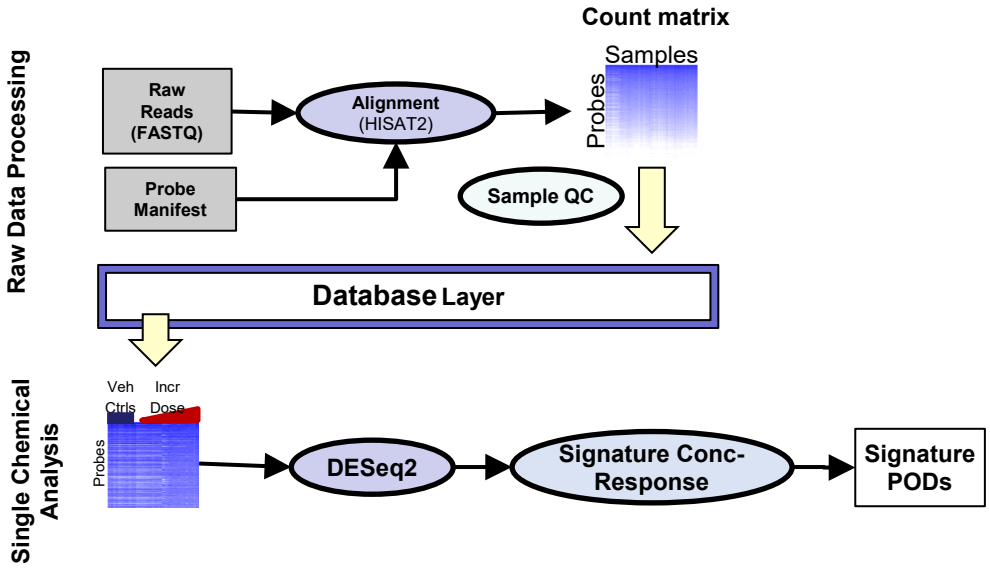
- Alignment Pipeline – using HISAT2, comparable to STAR
  - Now trims 51bp reads prior to alignment
  - Allowed soft-clipping with per base penalty

- Probe Homology can be an issue
  - Mapped homology within probe manifest (some probes have 49bp overlap)
  - >95% of reads map uniquely to one probe with current parameters
  - HISAT2 was better at resolving unique matches for homologous probes
  - Multi-mapping probes discarded for final counts

L. Everett

HTTr Overall Process





Differential Gene Expression Analysis

- Most recent version of DESeq2 (v1.24.0)
  - Evaluated questions about choice of plate effect and shrinkage using reference chemicals
  - Newer shrinkage methods (Ashr, Apeglm) results less reliable
- Analyze one chemical at a time with matched DMSO controls
- DEG analysis by four DESeq2 options:-
  1. Plate effect - , Shrinkage -
  2. Plate effect - , Shrinkage +
  3. Plate effect + , Shrinkage -
  4. Plate effect + , Shrinkage + (Recommended)

EPA United States Environmental Protection Agency

Gene Sets: “Signatures”

- Understanding the results of changes in expression of 10,000-20,000 genes is hard
- Group genes into gene sets (“Signatures”)
- Examples of signature types
  - Genes that are perturbed in diseased tissue vs. health tissue
  - Genes perturbed in individuals with congenital diseases vs. those without
  - Genes perturbed by drugs or other chemicals
  - Genes perturbed by gene knockdowns / knockouts
- Example use
  - If a chemical perturbs the genes upregulated in a cancer type, the chemical is a candidate carcinogen (or candidate anti-cancer drug)

- Each signature has a hand-annotated “super target” class to help with annotation
- ~10,000 signatures
- ~1000 super targets

Signature Scoring

- Start with matrix of samples x genes with l2fc from DESeq2
- For each concentration of each sample, calculate score for each signature using
  - GSEA (ssGSEA)
  - **FC (mean(l2fc|in signature) – mean(l2fc|out of signature))**
- Distribution of signature scores are zero centered
- For bidirectional signatures collapse score to that of parent
  - Score(chemical, concentration, parent)=score(up) – score(down)
  - Retains directionality
- For unidirectional signatures, parent score=signature score

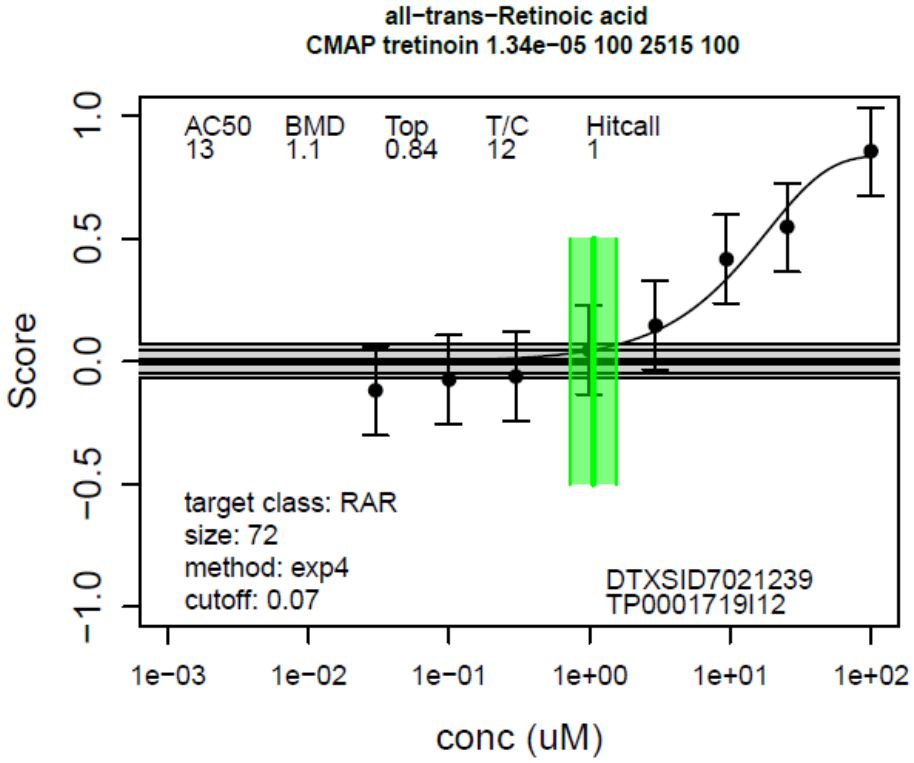
Predicting Potency

- At what concentration does the chemical cause an effect?
- "Point of Departure"
  - AC50: concentration at 50% of effect
  - Benchmark Dose/Concentration: concentration where signal exceeds noise
- Measure this in vitro
- Can also predict in vivo dose where effect happens using toxicokinetics

Concentration-response modeling

- Use variant of ToxCast tcpl concentration-response fitting method
- Expanded to include all models used in BMDExpress
  - cnst, hill, gnls, poly1, poly2, pow, exp2, exp3, exp4, exp5
  - Fitting in both up and down directions
  - Model with lowest AIC is selected
- Produces a continuous hit call value
- Implemented in R package tcplFit2
  - https://github.com/USEPA/CompTox-ToxCast-tcplFit2
- Create null distribution of 1000 randomly select “chemicals” created by permuting columns of sample x gene matrix
- Real chemical response has to exceed 95% CI of the null distribution

Example Signature Concentration-Response plot

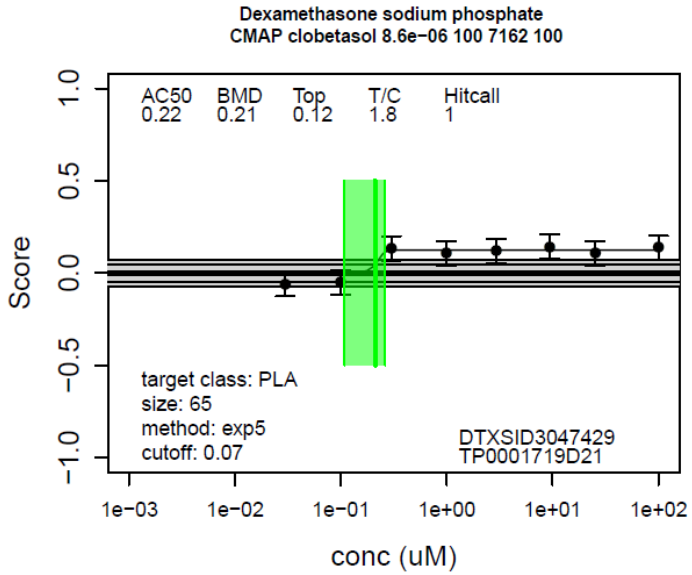


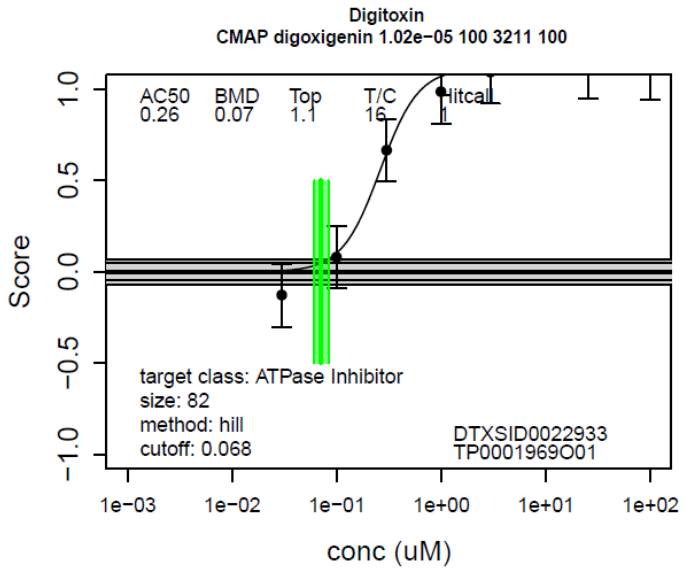
Confidence Interval (CI) around points from the fitting error term

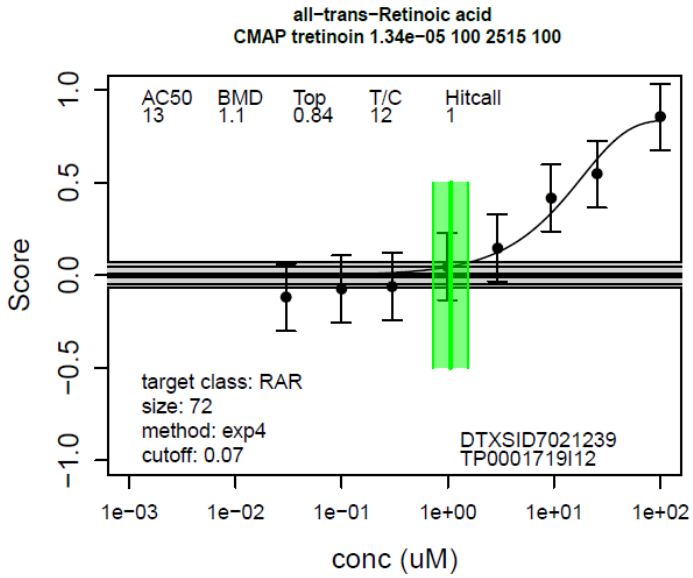
Outer gray band is 95% CI of null dist. Inner lines are benchmark response

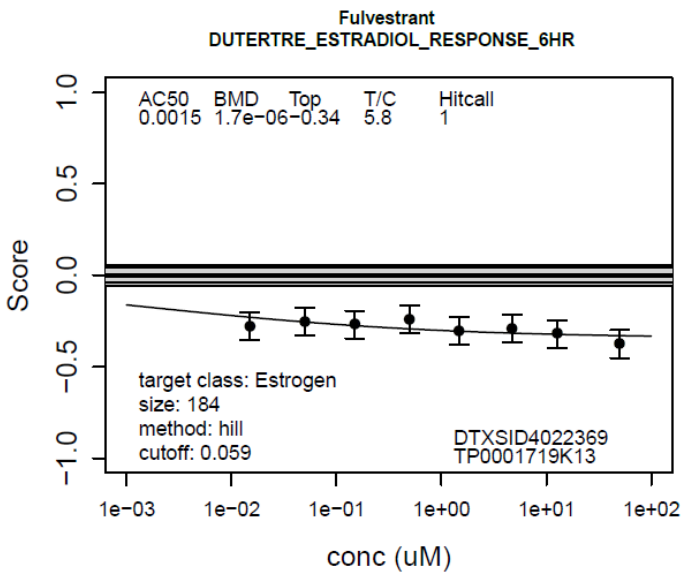
Green vertical band is BMD and 95% CI

More activity that just Estrogen Receptor

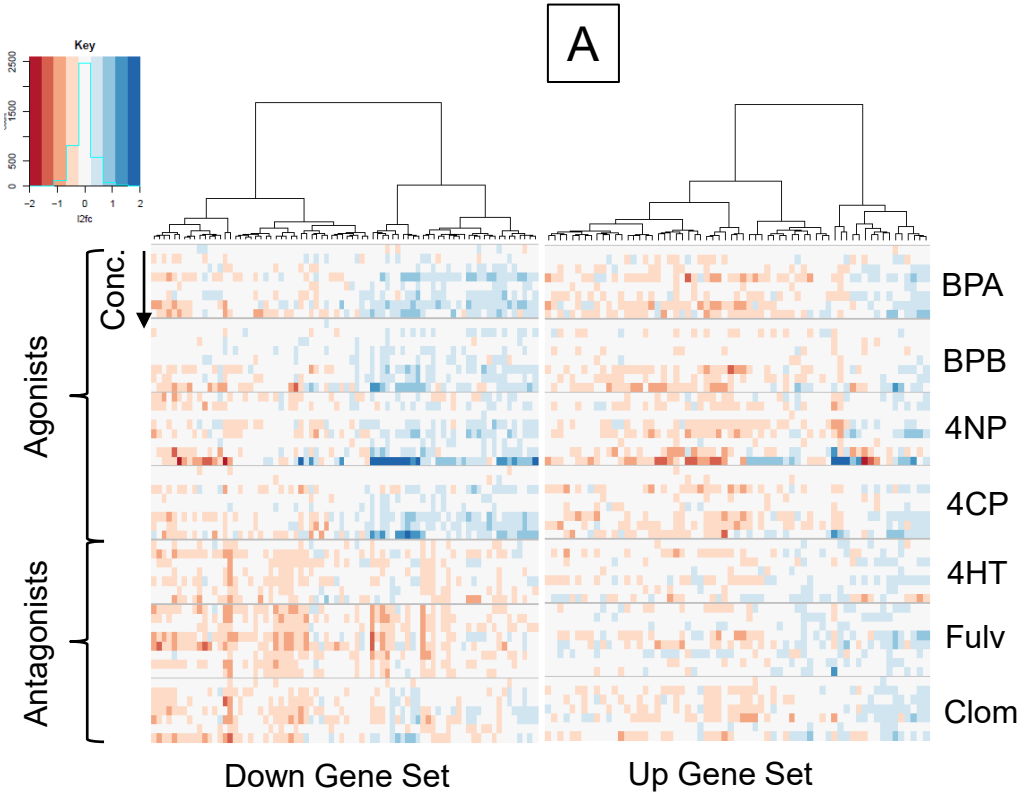


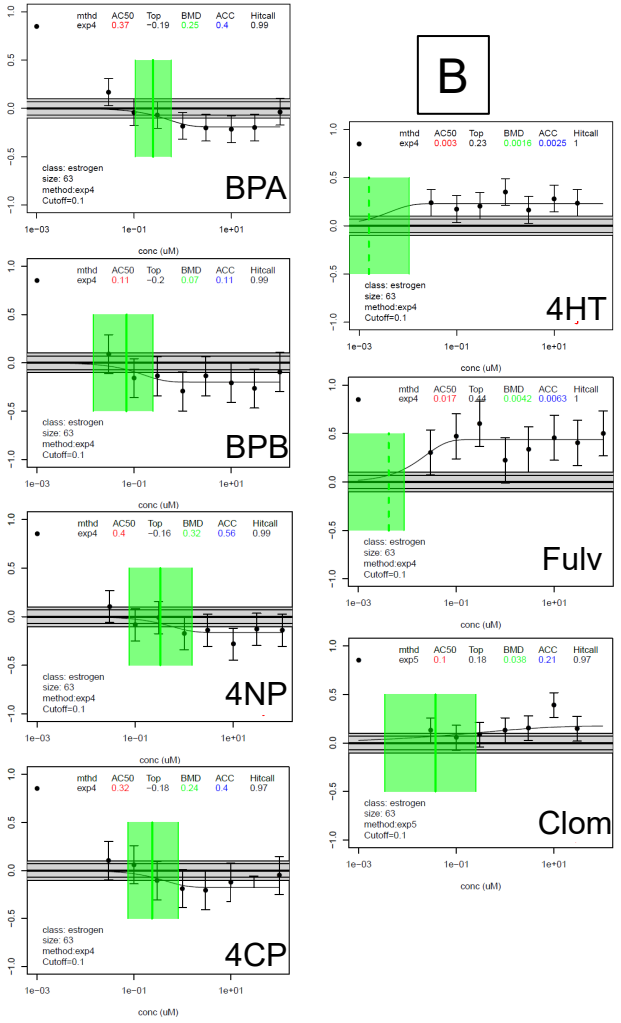




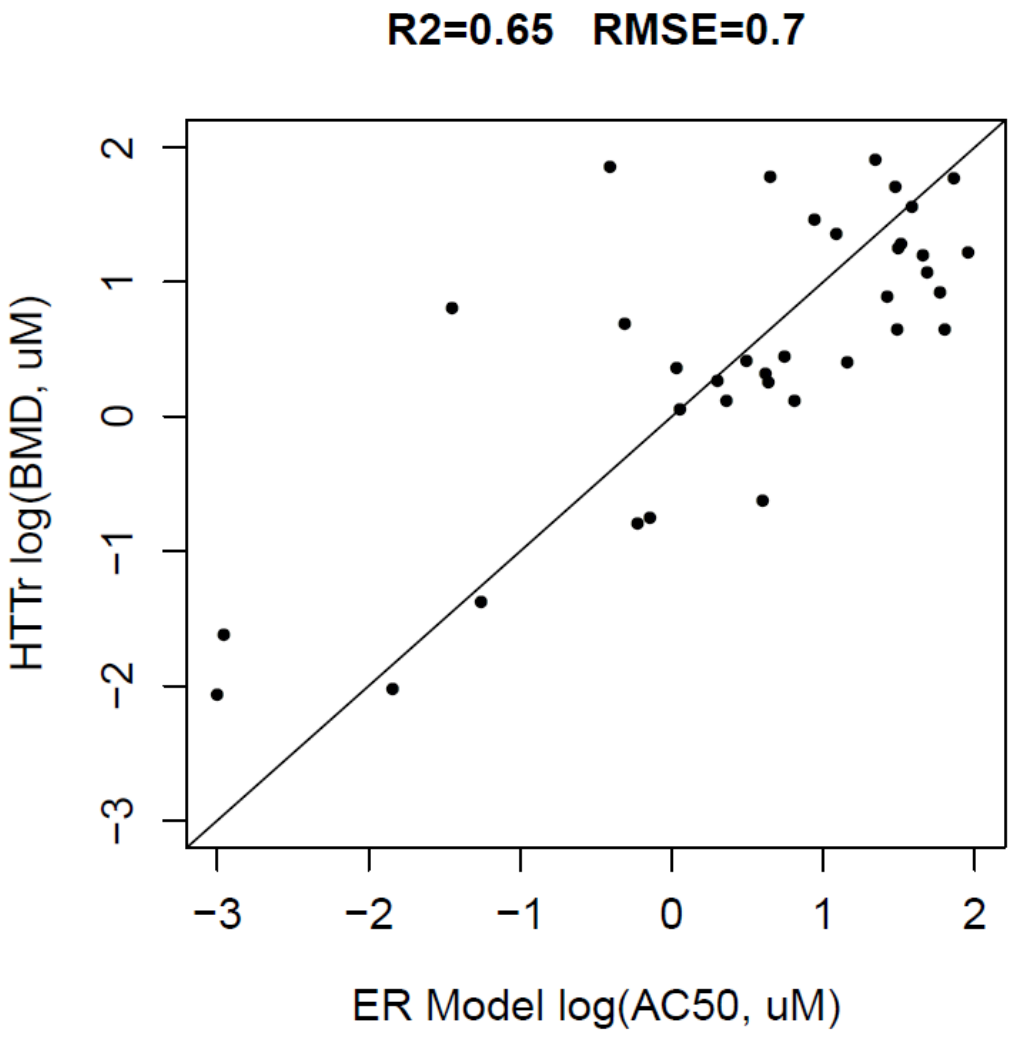


Gene-level to signature score





How do potencies compare with other in vitro assays?



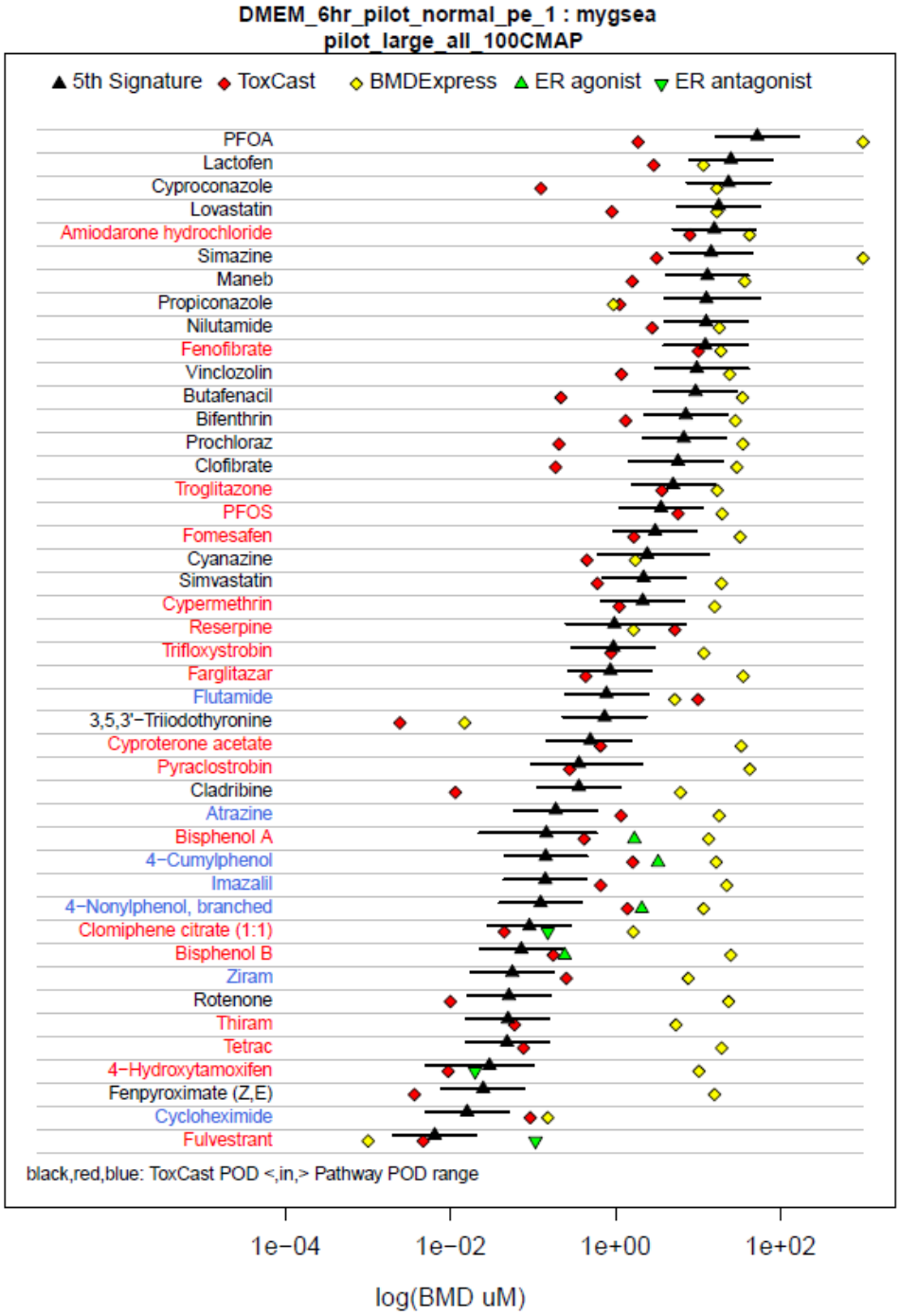
Compare potency with estimates from 18 in vitro agonist and antagonist high-throughput screening assays.

United States Environmental Protection Agency

Ranking Chemicals by potency
And
Comparing Technologies

Black: lowest 5%-ile signature
Red: ToxCast 5% POD
Yellow: BMD Express
Green: ToxCast ER Model

Data from MCF7 Pilot



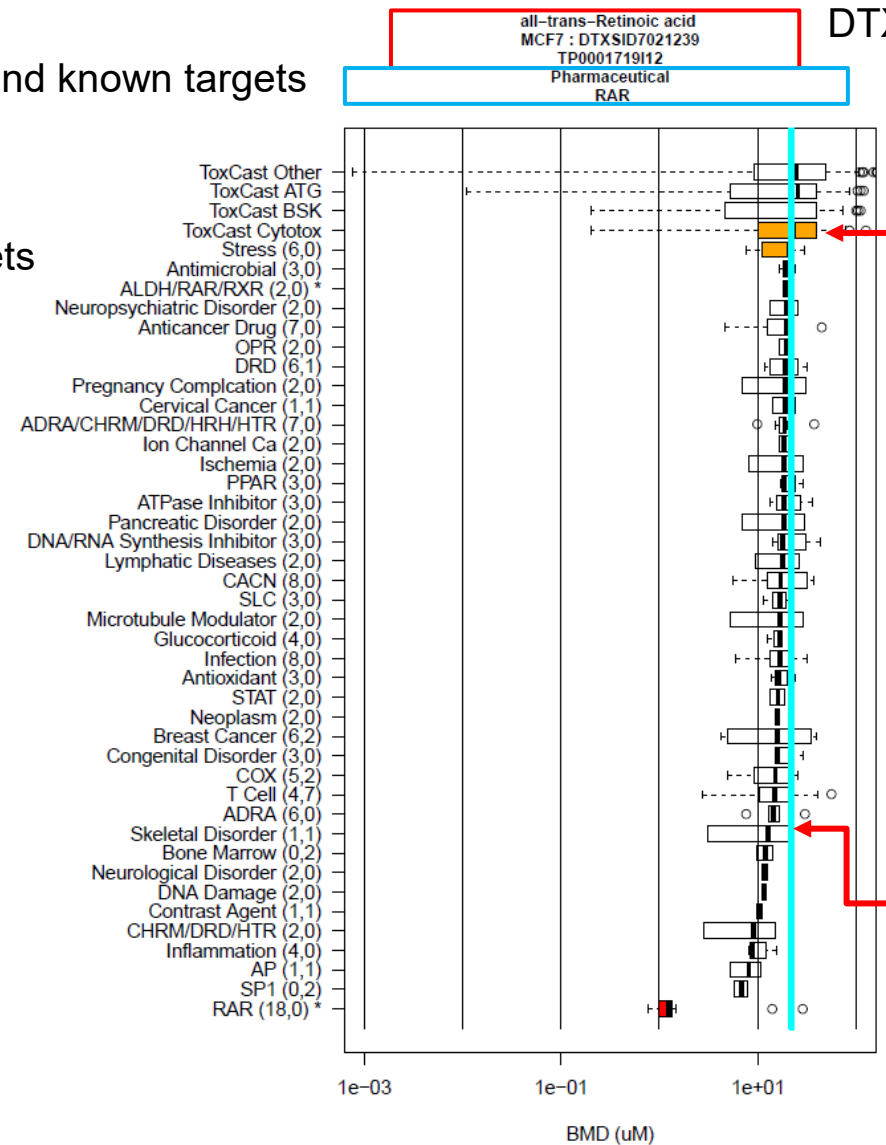
Predicting Effect

- What signatures or pathways are activated?
- Are they target-specific?
- Are they related to generalized cell stress?

Super Target Summary Plot

Chemical name, DTXSID and sample ID

Use class and known targets



Super targets

Stress signatures

Median of all super target BMD medians

Boxplot shows range of BMD values for signatures for the super targets

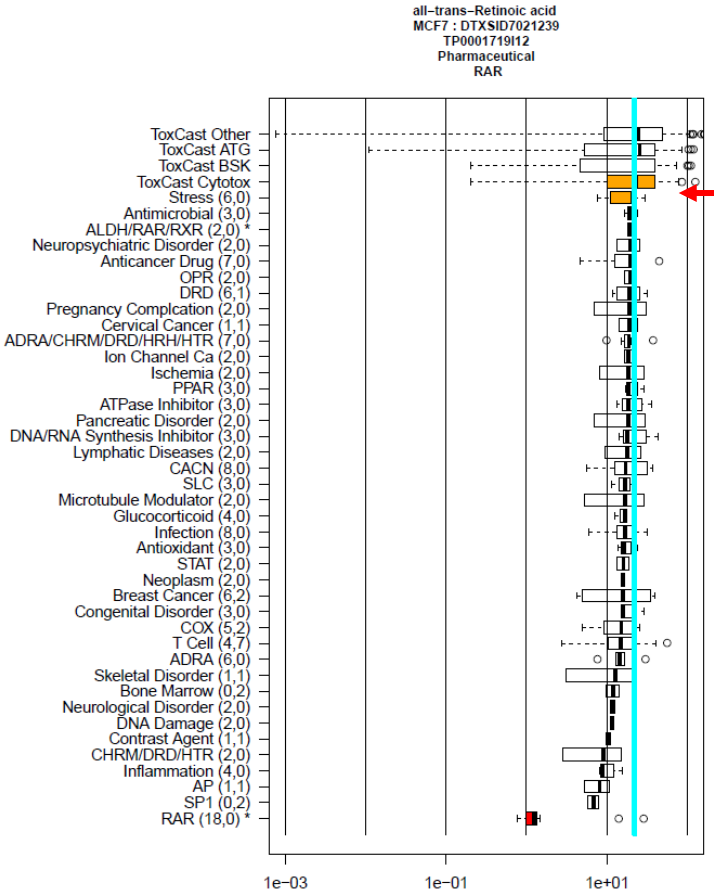
Red indicates that the super target is a target of the chemical

Why Cell Stress is Important

- Activity can be specific or non-specific

- Specific
  - Chemical interacts with a target that causes genes to be up or down-regulated
  - Examples are nuclear receptors (ER, AR, RAR)
- Non-specific
  - Chemical causes some kind of general stress
  - Disrupts cell membranes, oxidative stress, apoptosis
  - Cell responds by turning on generalized stress response pathways
  - Large number of genes are mis-regulated
  - “Burst” of activity across the genome

Specific vs. Non-specific



Stress signatures

Non-Specific – tens to hundreds of signatures

Specific
1 signature

EPA United States Environmental Protection Agency

Summary

- Need to screen thousands of chemicals for potency and mechanism of action
- We can now do this with HTPP, HTTr and HTS
- Application areas in current use
  - Prioritizing chemicals for further investigation
  - Clustering chemicals by activity profile
  - Identifying areas of concern for emerging contaminants
  - Estimating safe exposure levels for chemicals
  - Animal-free evaluation of chemical safety for cosmetics ingredients (with Unilever)

Acknowledgements

- Josh Harrill
- Logan Everett
- Imran Shah
- Rusty Thomas
- Laura Taylor
- Derik Haggard
- Joseph Bundy
- Beena Vallanat
- Bryant Chambers
- Woody Setzer
- Thomas Sheffield
- Clinton Willis
- Richard Brockway
- Johanna Nyffeler
- Megan Culbreth
- Dan Hallinger
- Terri Fairley