CompTox Chemicals Dashboard & Bioactivity Data

Introduction: Scarlett VanDyke

CompTox Chemicals Dashboard; Bioactivity Intro: Dr. Nisha Sipes

High throughput transcriptomics (HTTr) &

High throughput phenotypic profiling (HTPP): Dr. Logan Everett

ToxCast: Madison Feshuk, MPHTM

Outline & Disclaimer

- Introduction
- High throughput transcriptomics (HTTr)
  - CCD Demo: *Bioactivity: HTTr*
- High-throughput phenotypic profiling (HTPP)
  - CCD Demo: *Bioactivity: HTPP*
- Toxicity Forecasting (ToxCast)
  - CCD Demo: *Bioactivity: ToxCast*

The views expressed in this presentation are those of the presenter and do not necessarily reflect the views or policies of the US Environmental Protection Agency.

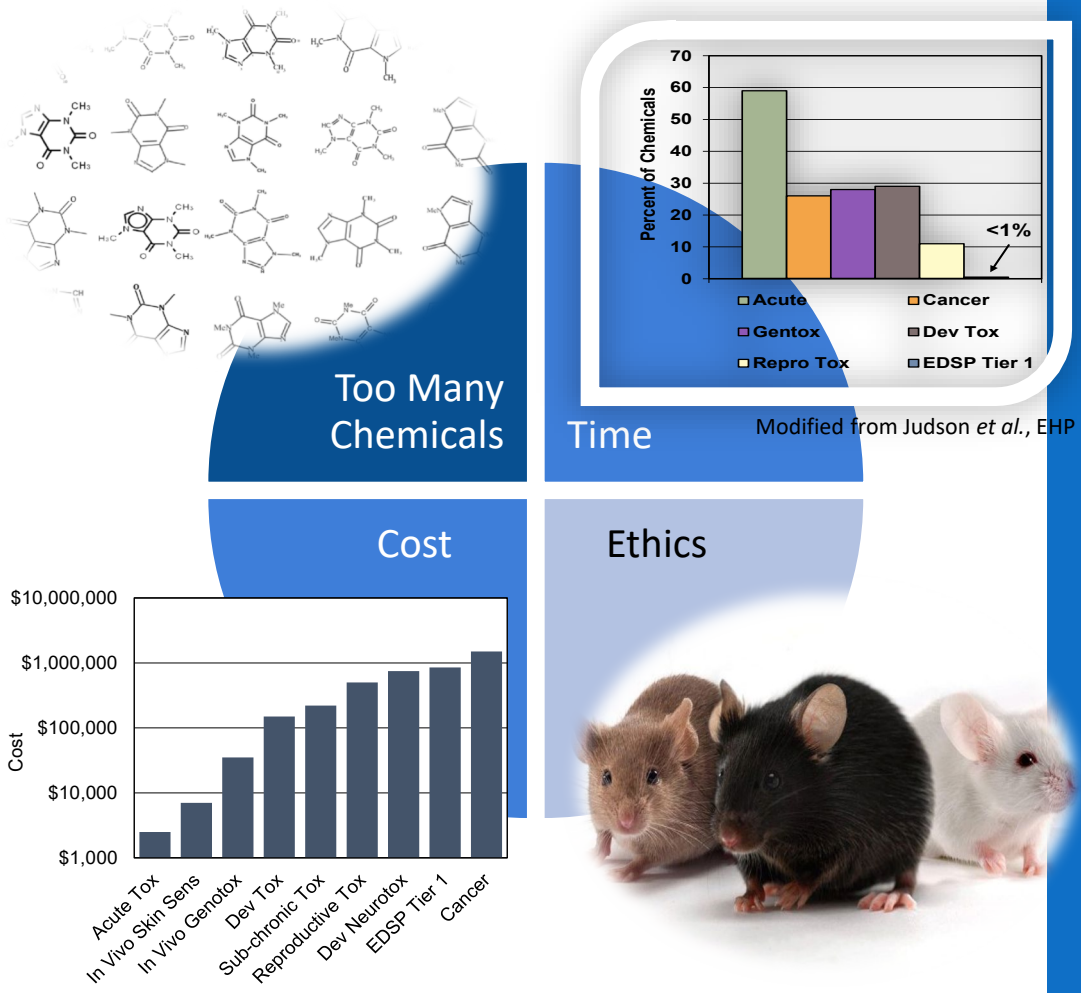
Company or product names do not constitute endorsement by US EPA.

Bioactivity Introduction

Nisha Sipes

Need for Alternative Approaches for Next-Gen Risk Assessment

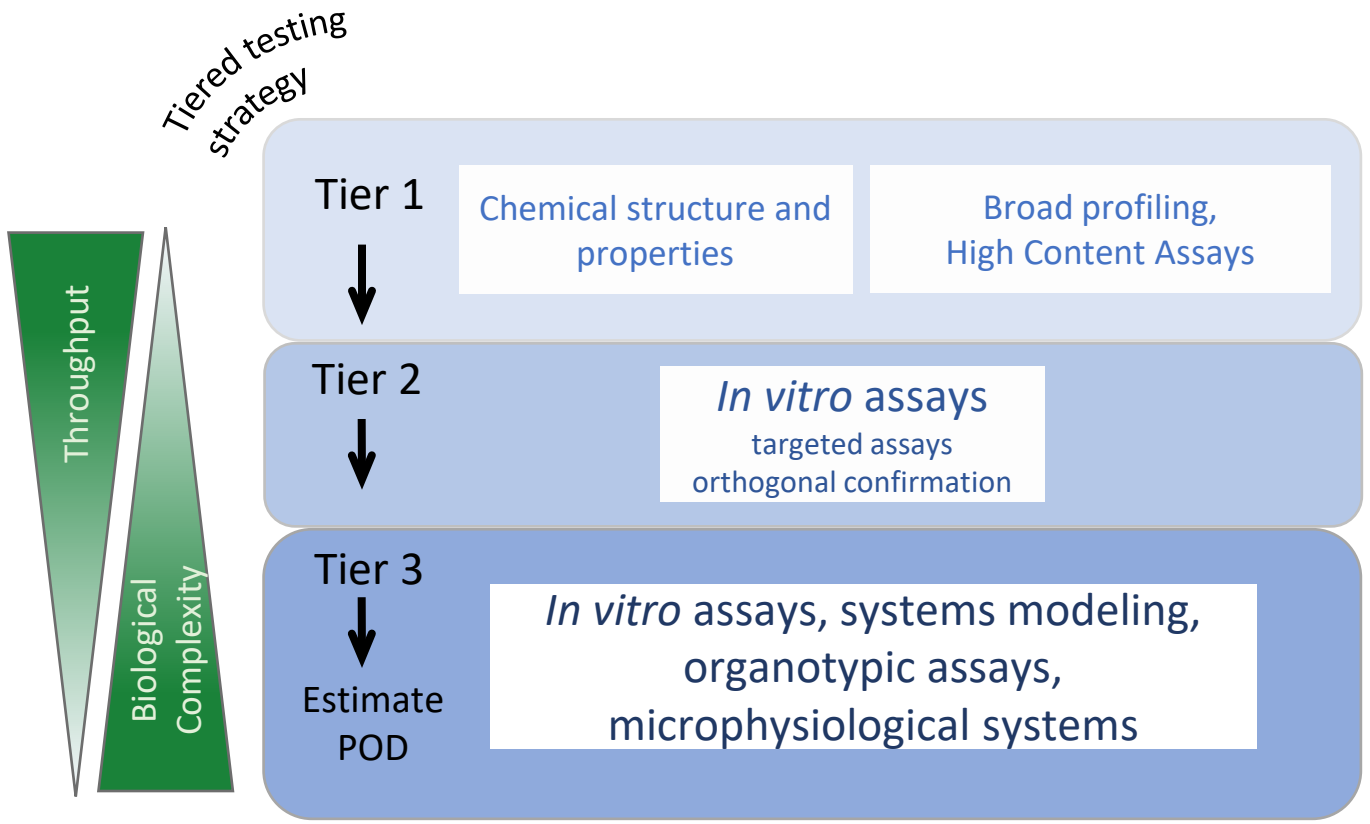
- Several limitations to traditional *in vivo* toxicology testing
- EPA needs rapid and efficient methods to prioritize, evaluate, and regulate thousands of chemicals in commerce
- New Approach Methods (NAMs) can provide information on hazard + exposure to inform research and decisions



Modified from Judson *et al.*, EHP 2010

NAMs for hazard include broad profiling and targeted approaches

Adapted from the US EPA's NexGen Blueprint of Computational Toxicology



Example assay data

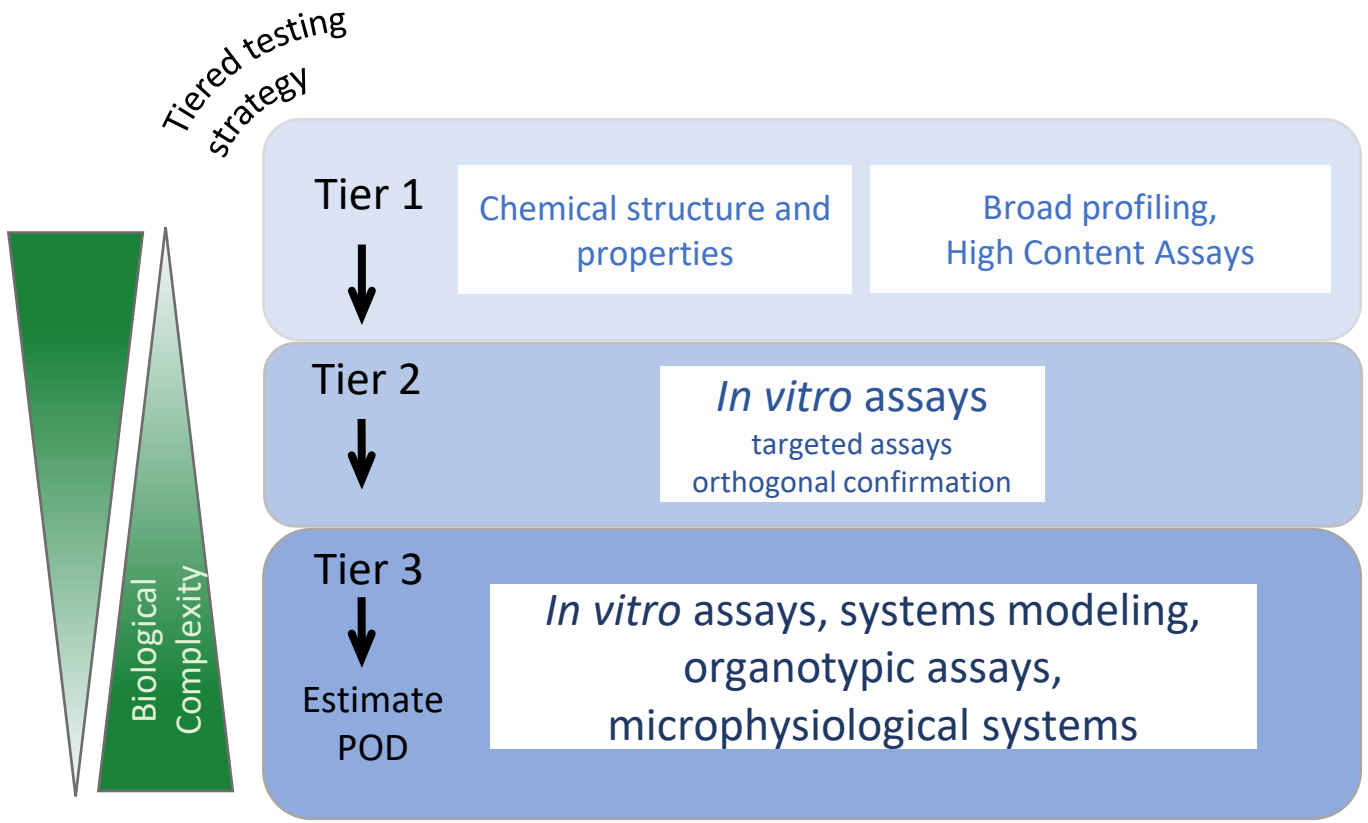
High Throughput Transcriptomics (HTTr), High Throughput Phenotypic Profiling (HTPP), *in silico* NAMs (e.g., TEST, GenRA)

ToxCast

ToxCast

NAMs for hazard include broad profiling and targeted approaches

Adapted from the US EPA's NexGen Blueprint of Computational Toxicology

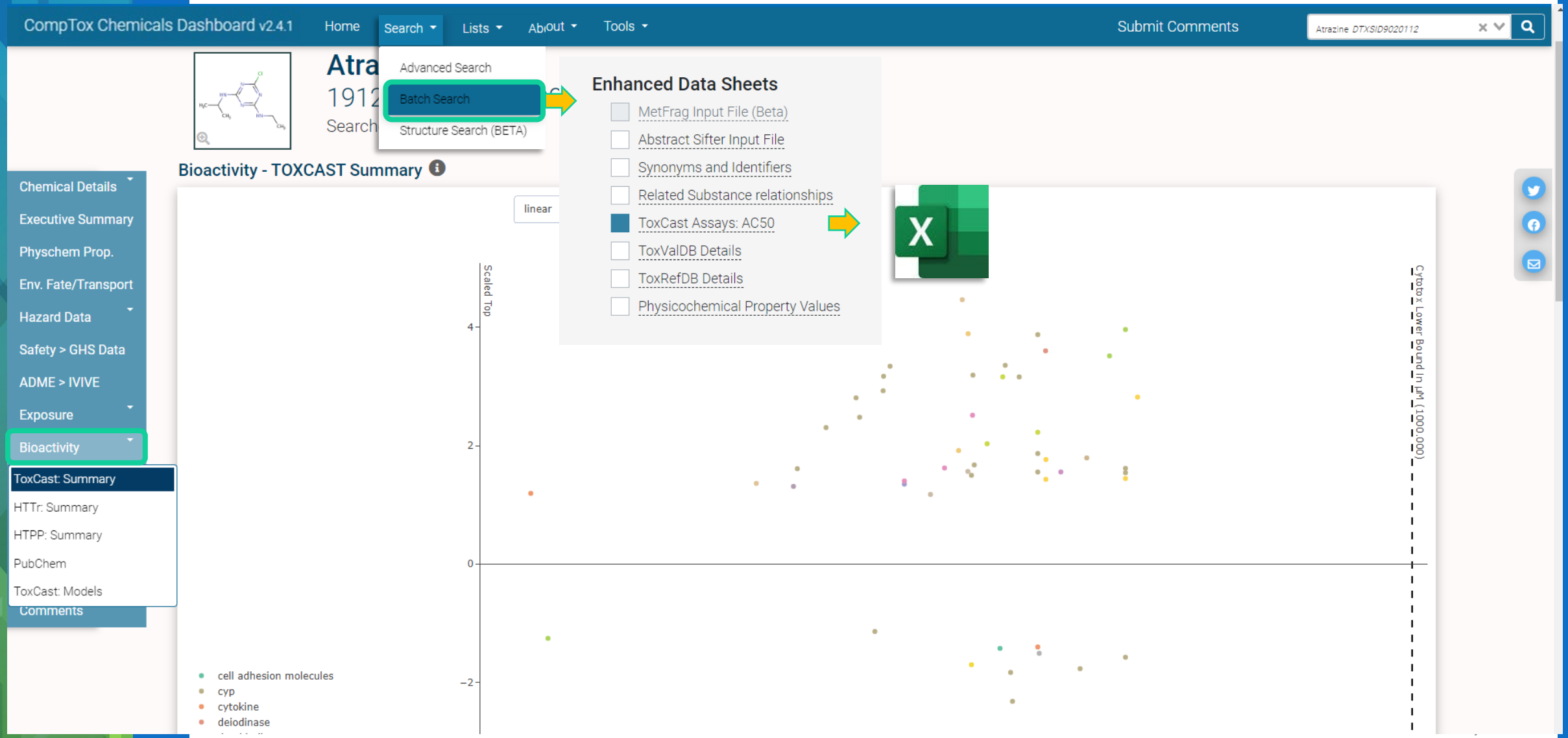


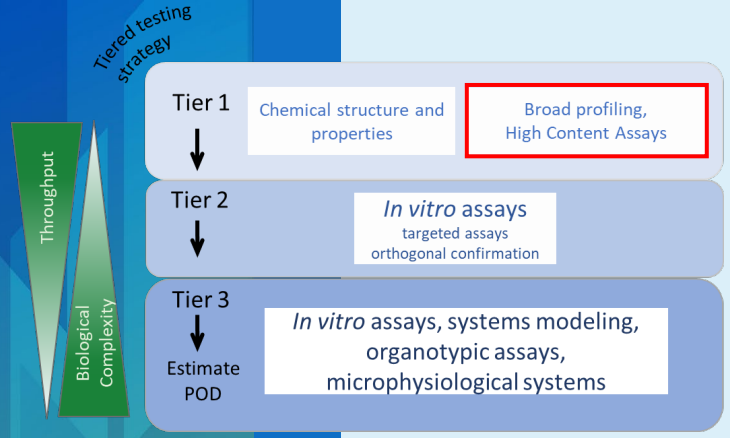
<u>How to make data comparable between tiers?</u>

A single curve-fitting approach (*tcplfit2 R package*) enables

- Comparison
- Interoperability

CompTox Chemicals Dashboard: Bioactivity Tab





High-Throughput Transcriptomics (HTTr)

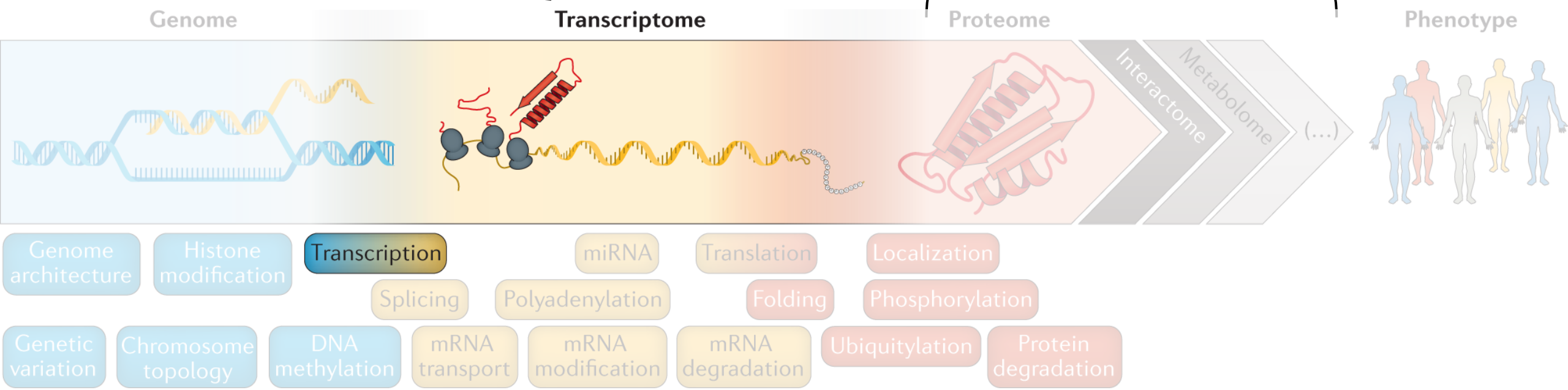
Logan Everett

(with contributions from Joshua Harrill and Richard Judson)

Why Transcriptomics?

Highly dynamic in response to environmental stimuli,
Well-established, cost-effective, high-throughput methods

Broad profiling methods are newer
or less cost-effective

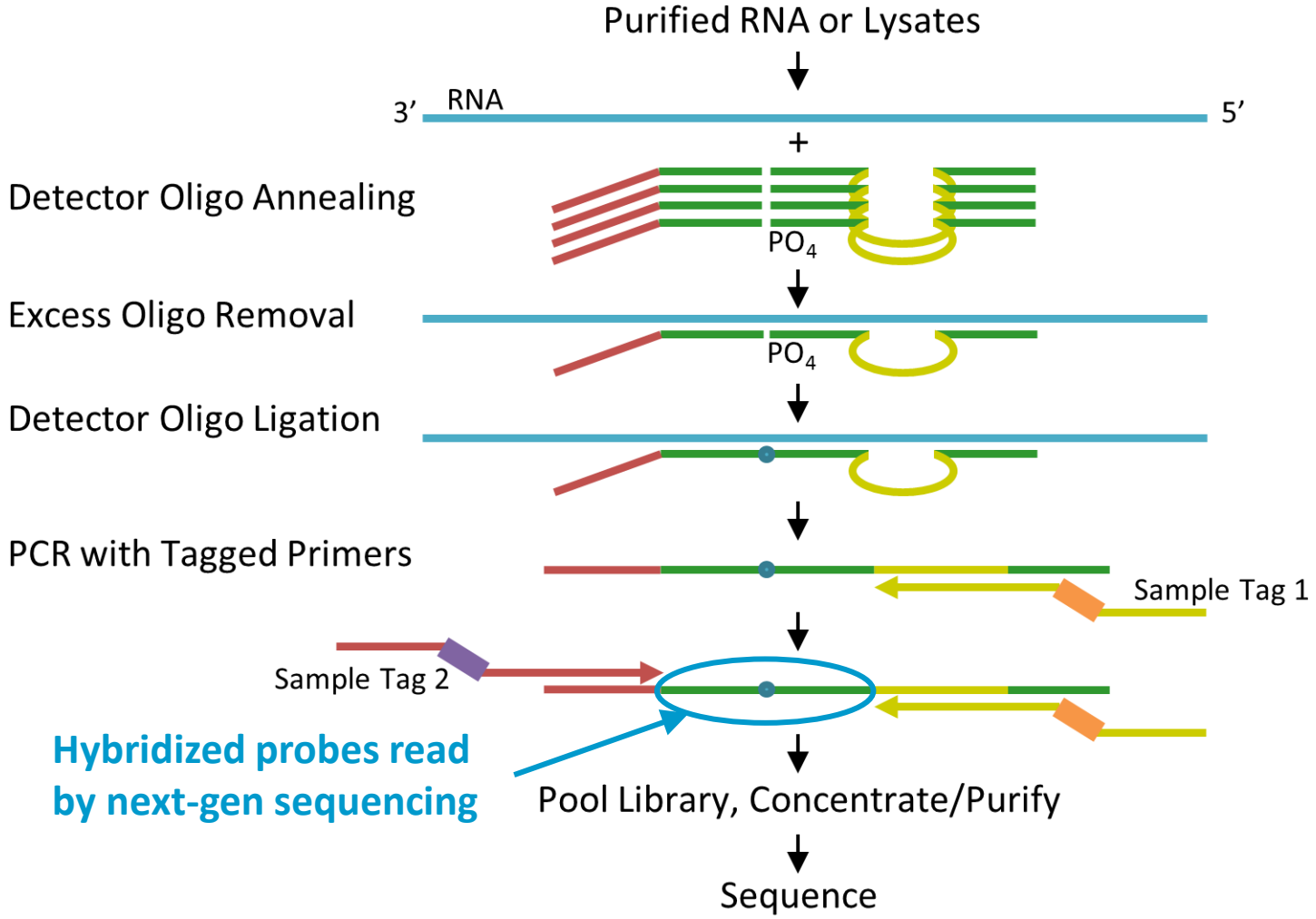


Hazardous exposures can cause perturbations
at multiple levels of gene regulation

Buccitelli and Selbach, *Nat Rev Genet* 2020
DOI: 10.1038/s41576-020-0258-4

Targeted RNA-seq Assay (TempO-seq)

- Next-Gen sequencing of targeted probes hybridized to expressed transcripts
- Whole transcriptome coverage (>20,000 genes)
- Captures gene expression at lower cost than RNA-seq or microarrays
- Compatible with raw cell lysates – ***ideal for large-scale screening***



Yeakley, *et al. PLoS ONE* (2017) DOI: 10.1371/journal.pone.0178302

Automated *in vitro* Chemical Screening Strategy

Cryopreserved Cell Stocks

Cell Expansion

Cell Line Examples:	
MCF-7	Breast cancer
U-2 OS	Bone cancer
HepaRG	Liver metabolism

Cell Plating

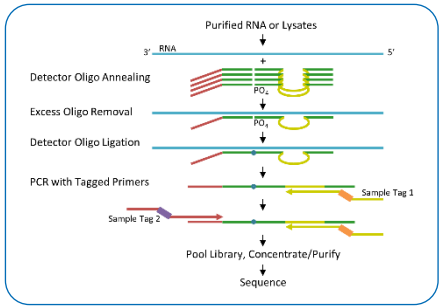
BioTek MultiFlo ™ FX

Dispensing Test Chemicals

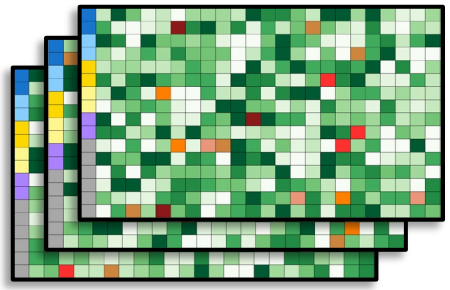
LabCyte Echo® 550 Liquid Handler

6 - 24 hr exposures

TempO-seq: Human Whole Transcriptome



Yeakley, *et al. PLoS ONE* (2017)
DOI: 10.1371/journal.pone.0178302

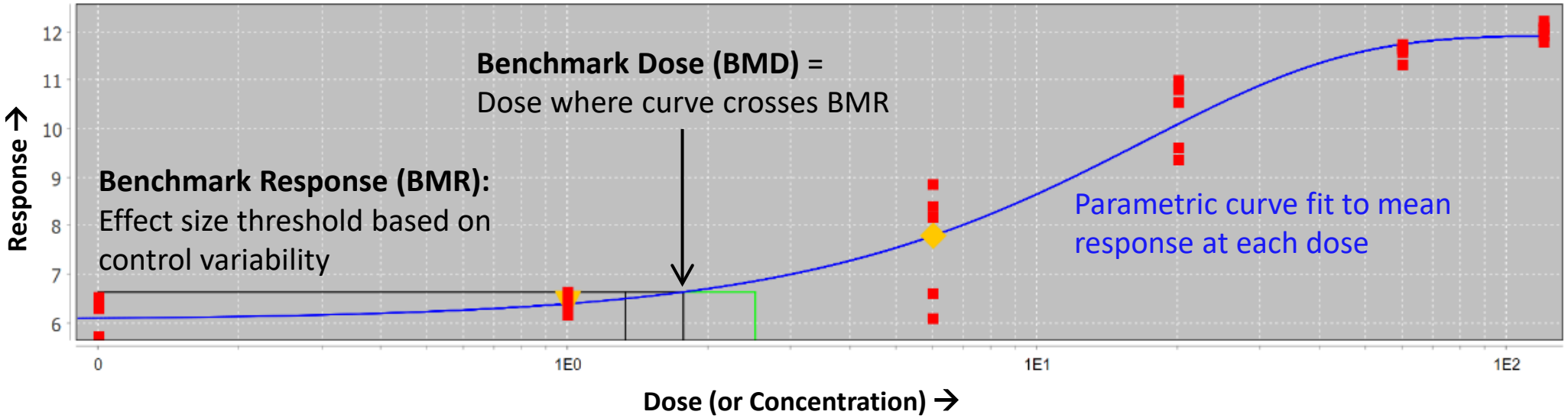


384-well test plates run in triplicate with:

- ~40 test chemicals x 8 concentrations (half-log spacing)
- Multiple vehicle controls, reference chemicals & QC samples on every plate to track assay performance
- Treatment positions randomized on each plate
- Independent culture batch on each plate

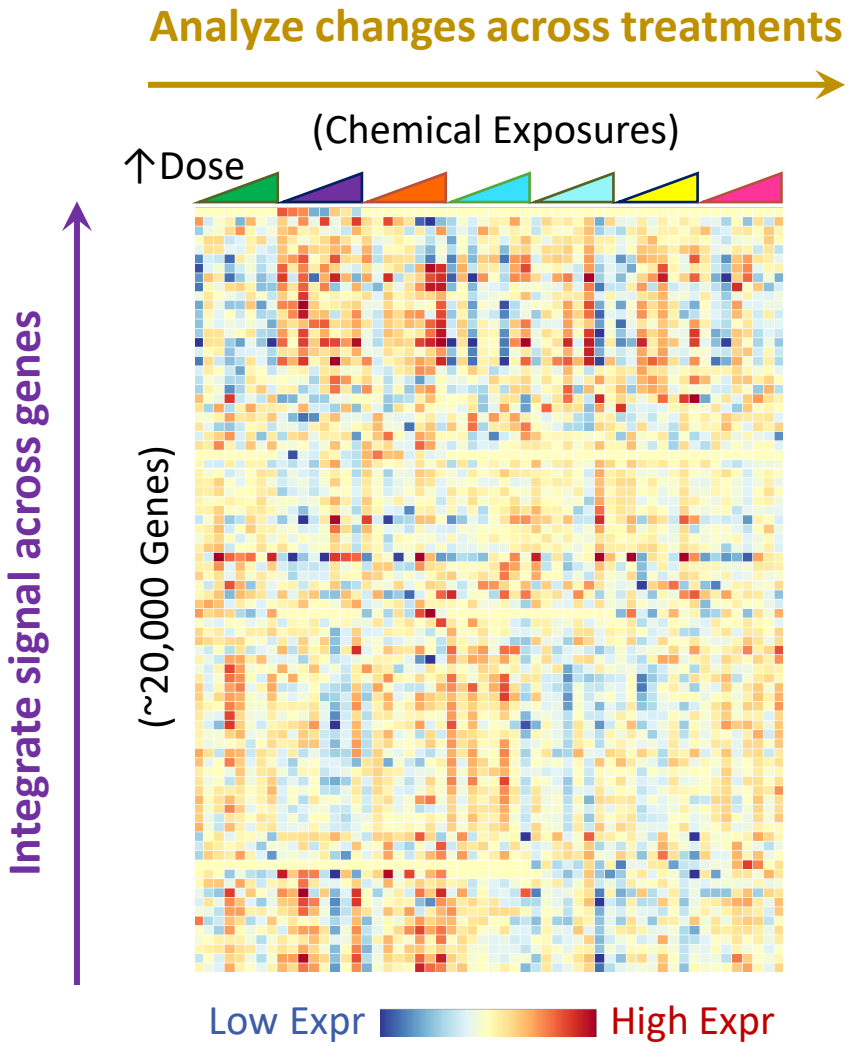
See Harrill, et al. Tox Sci 2021
DOI: 10.1093/toxsci/kfab009

Benchmark Dose



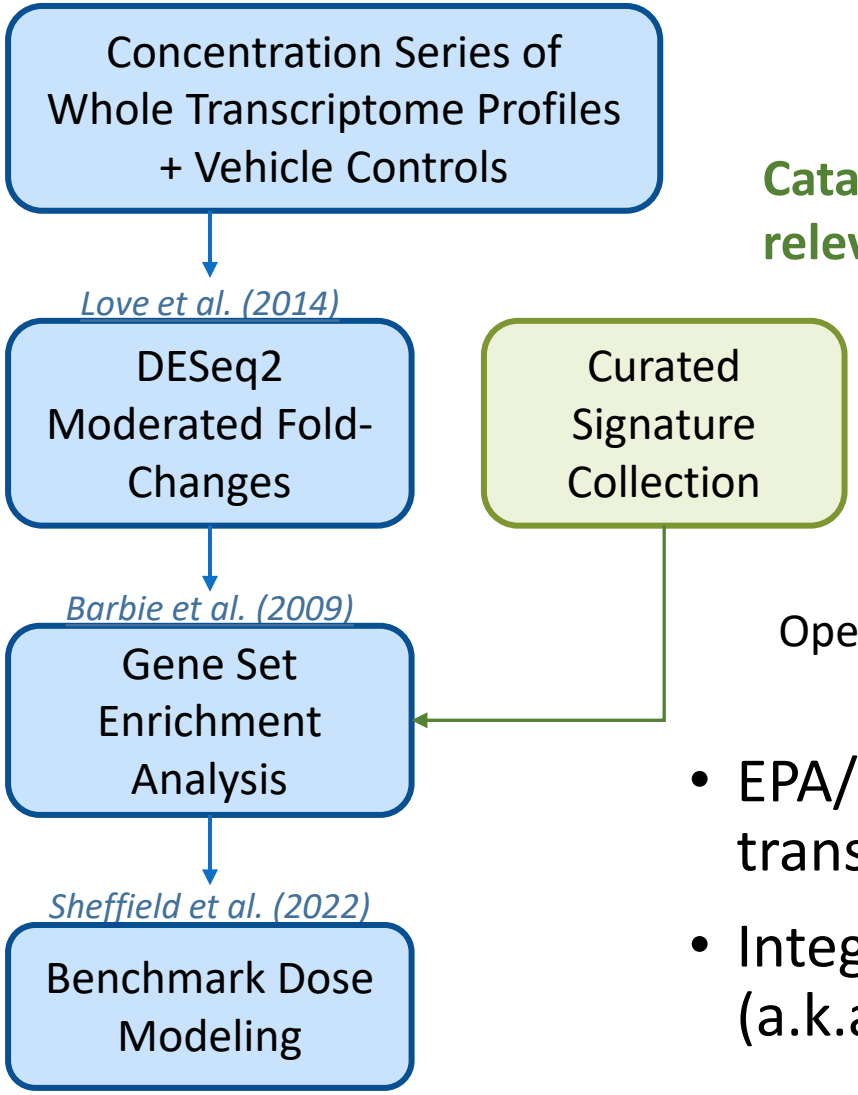
- Widely used approach for apical endpoint data
- EPA Benchmark Dose Software (BMDS): www.epa.gov/bmds

Transcriptomic Dose-Response Models



- Different genes may respond at different doses of a given exposure!
- Need to analyze both:
  - Dose-responsive trends
  - Coordinated changes in gene expression
- Gene-level data noisier in transcriptomics than targeted measurements (e.g. RT-qPCR)
- Dose-response modeling thousands of features (e.g. mRNA levels) leads to computational & statistical challenges

Dose-Response Modeling of Gene Sets/Signatures



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** *(Piñero, et al. Database 2015)*
- **MSigDB** *(Liberzon, et al. Cell Syst 2015)*

Open Source: github.com/USEPA/CompTox-httrpathway

- EPA/CCTE method for summarizing large-scale transcriptomic screening studies
- Integrates signal across known gene set (a.k.a. signature) ***before*** dose-response modeling

Dose-Response Modeling of Gene Sets/Signatures

Concentration Series of Whole Transcriptome Profiles + Vehicle Controls

Love et al. (2014)

DESeq2 Moderated Fold-Changes

Curated Signature Collection

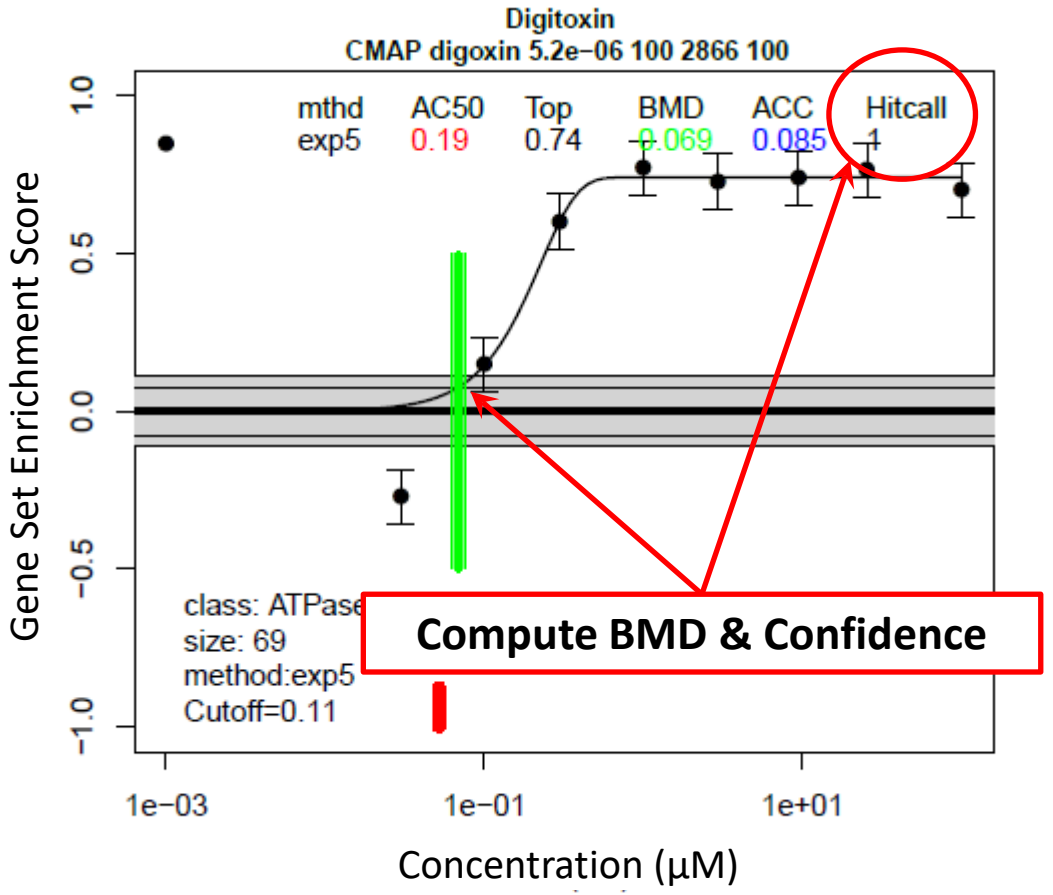
Barbie et al. (2009)

Gene Set Enrichment Analysis

Sheffield et al. (2022)

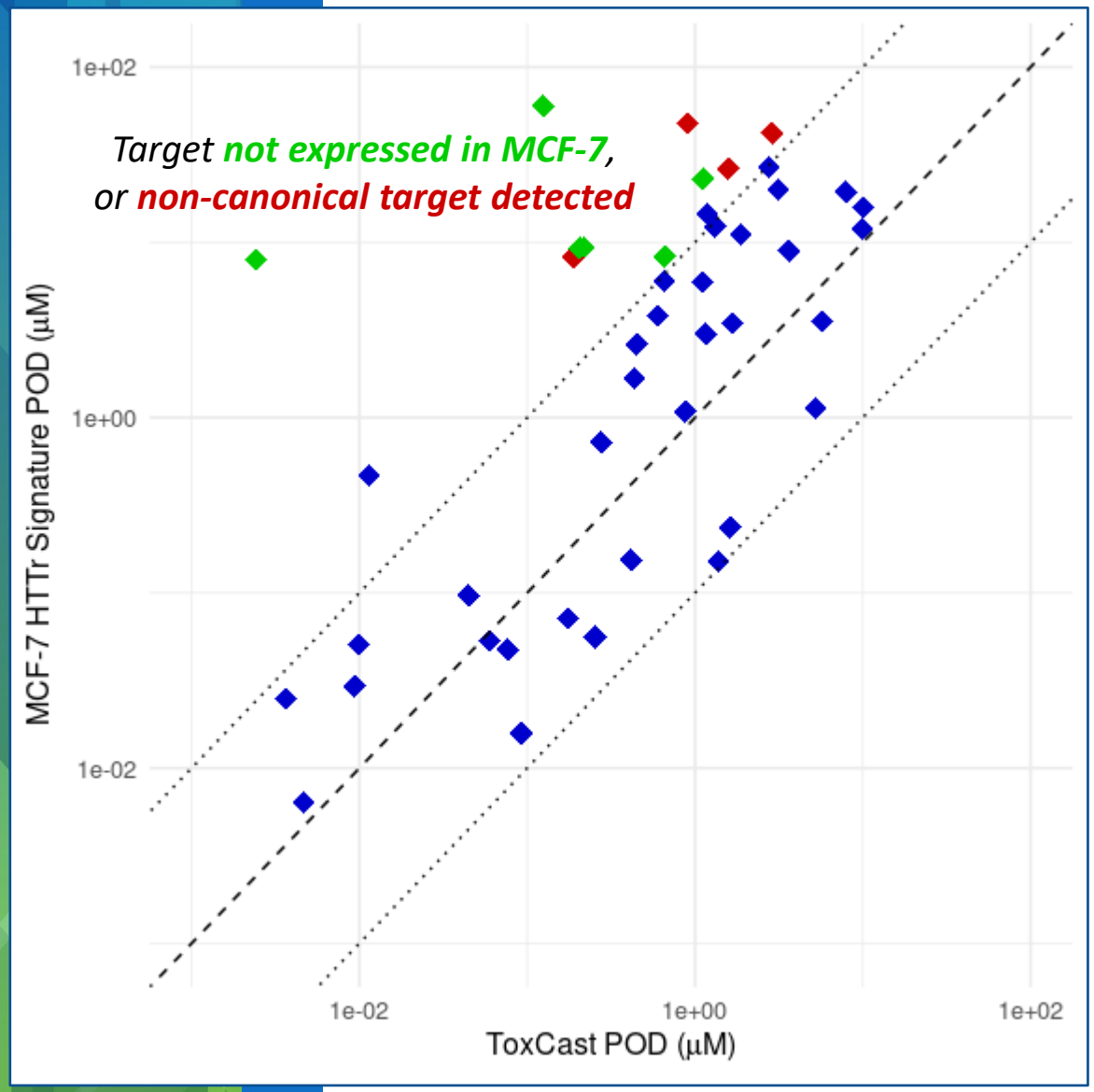
Benchmark Dose Modeling

Active Signatures have HitCall > 0.9



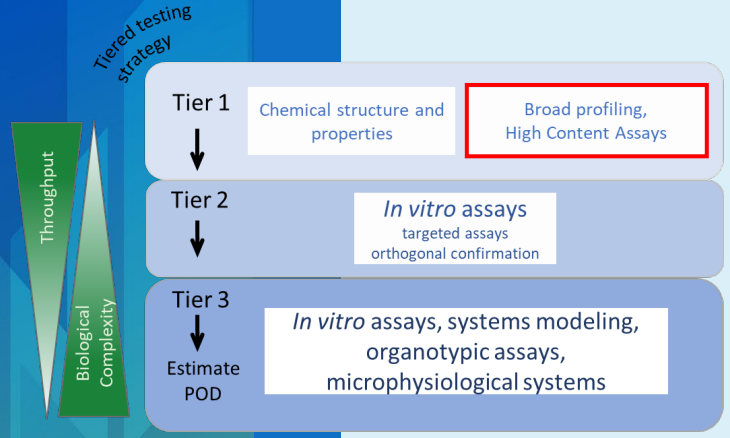
CRAN.R-project.org/package=tcplfit2

tPODs Are Concordant With ToxCast



- Computed 5th percentile of active signature/assay PODs from:
  - Pilot study of 44 well-characterized chemicals in MCF-7 cells, 6h exposure
    Harrill, et al. *Toxicol Sci* (2021)
    DOI: 10.1093/toxsci/kfab009
  - ToxCast targeted assay results *(multiple cell types, assays, and exposure lengths)*
    Paul-Friedman, et al. *Toxicol Sci* (2020)
    DOI: 10.1093/toxsci/kfz201
- Signature-based PODs are highly concordant with ToxCast results for the majority of chemicals in pilot study

Demo



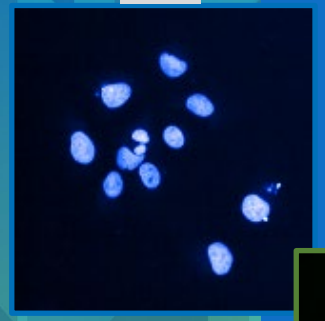
High-Throughput Phenotypic Profiling (HTPP)

Logan Everett

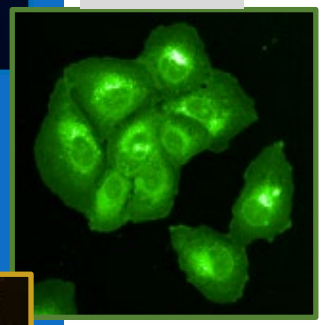
(with contributions from Joshua Harrill and Jo Nyffeler)

Cell Painting with Multiple Markers

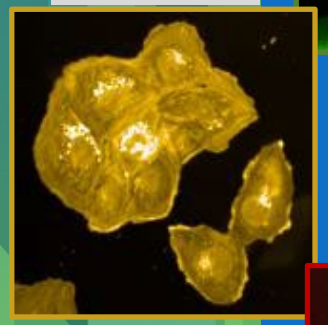
DNA



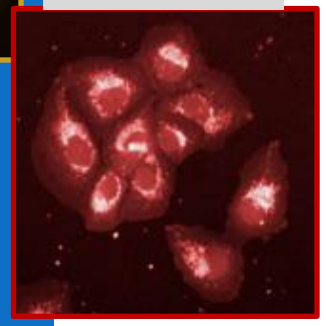
RNA + ER



Golgi + membrane + actin skeleton



Mitochondria

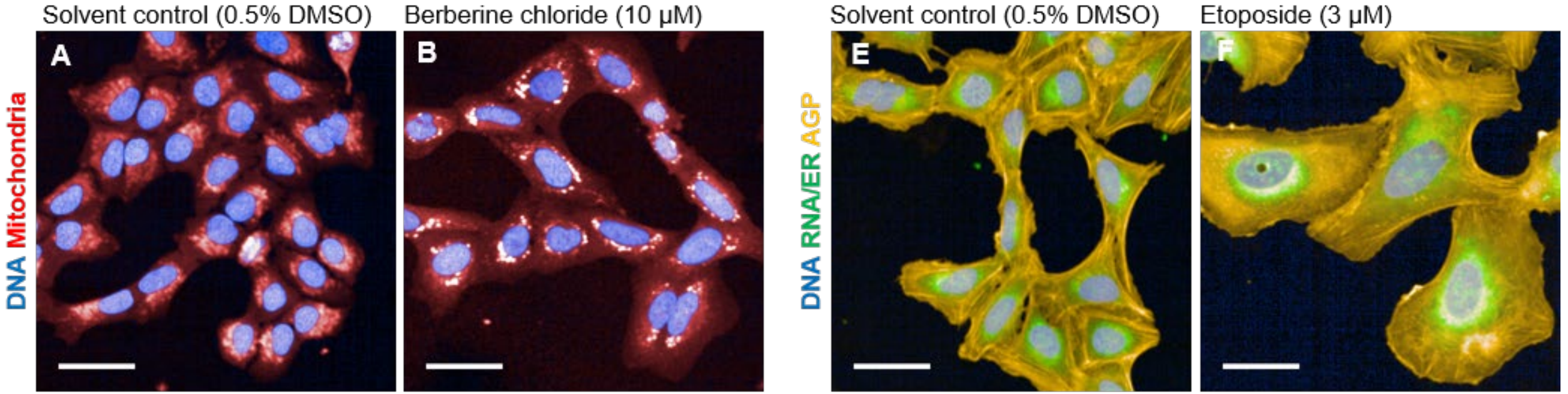


- Measures a large variety of phenotypic features in fluoroprobe labeled cells *in vitro*.
- Method based on Bray, et al. *Nat Protoc* (2016) DOI: 10.1038/nprot.2016.105

Marker	Cellular Component	Labeling Chemistry
Hoechst 33342	Nucleus	Bisbenzamide probe that binds to dsDNA
Concanavalin A – AlexaFluor 488	Endoplasmic reticulum	Lectin that selectively binds to α-mannopyranosyl and α-glucopyranosyl residues enriched in rough endoplasmic reticulum
SYTO 14 nucleic acid stain	Nucleoli	Cyanine probe that binds to ssRNA
Wheat germ agglutinin (WGA) – AlexaFluor 555	Golgi Apparatus and Plasma Membrane	Lectin that selectively binds to sialic acid and N-acetylglucosaminyl residues enriched in the trans-Golgi network and plasma membrane
Phalloidin – AlexaFluor 568	F-actin (cytoskeleton)	Phallotoxin (bicyclic heptapeptide) that binds filamentous actin
MitoTracker Deep Red	Mitochondria	Accumulates in active mitochondria

Joshua Harrill

Example Chemicals



→ Mitochondrial compactness/texture

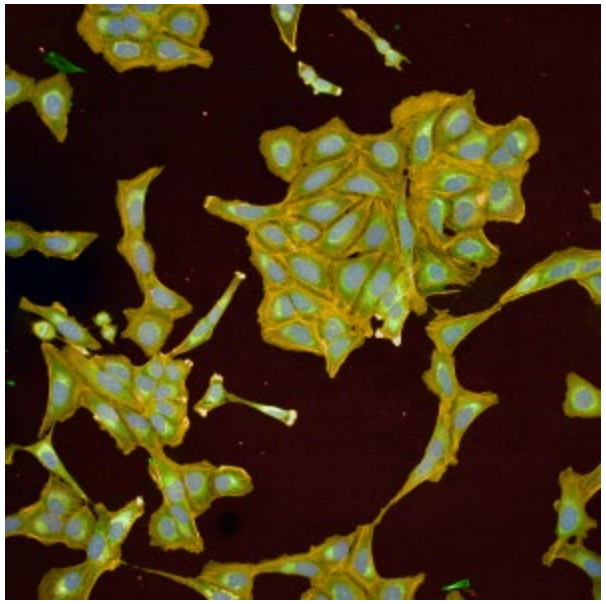
→ Cells are larger

Strong phenotypes are observable qualitatively and can be measured quantitatively using imaging processing software

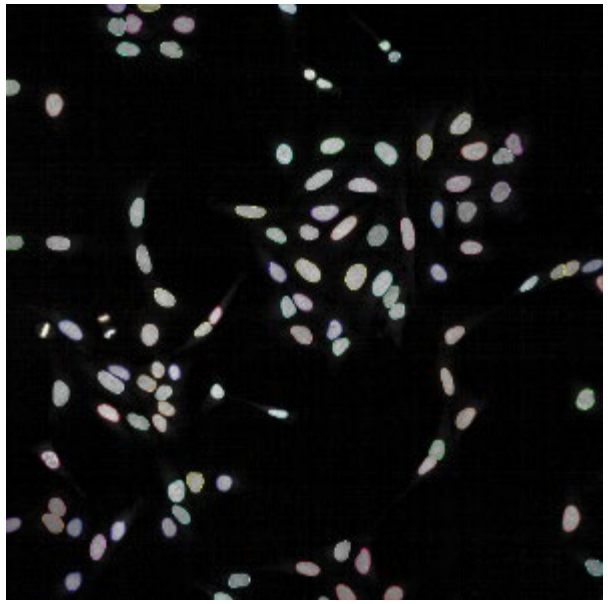
Nyffeler et al. (2020) *Tox Appl Pharm* DOI: 10.1016/j.taap.2019.114876

Joshua Harrill

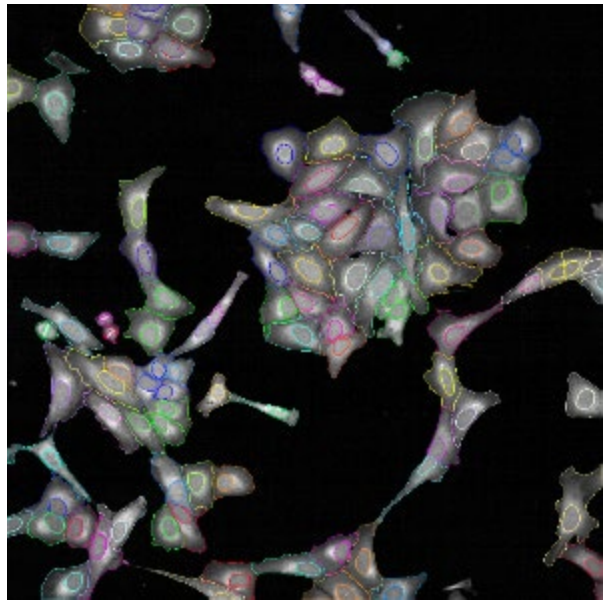
Image Segmentation



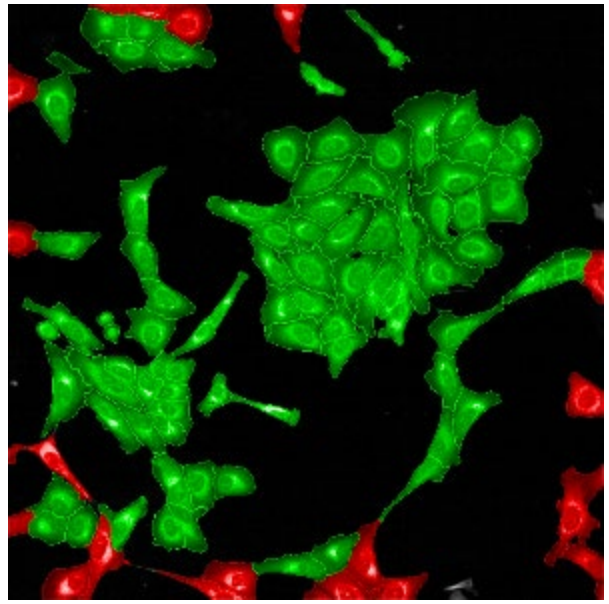
1. find nuclei



2. find cell outline



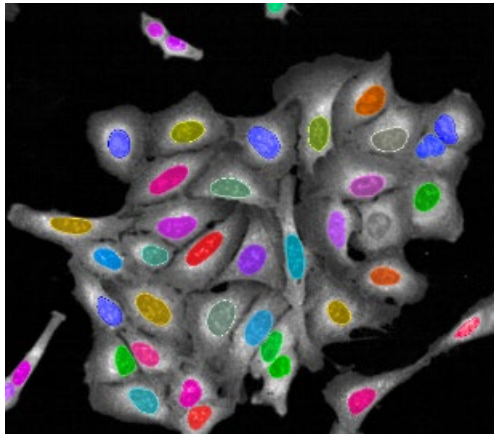
3. reject border objects



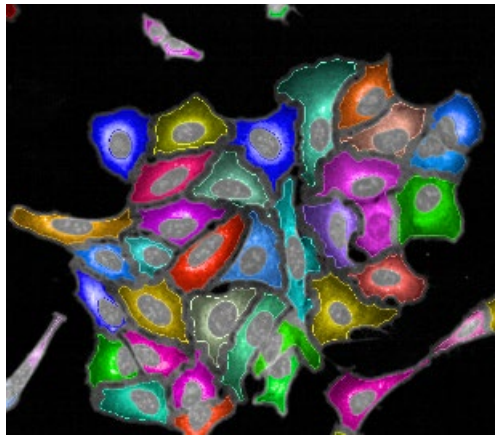
Joshua Harrill

Define Cellular Compartments

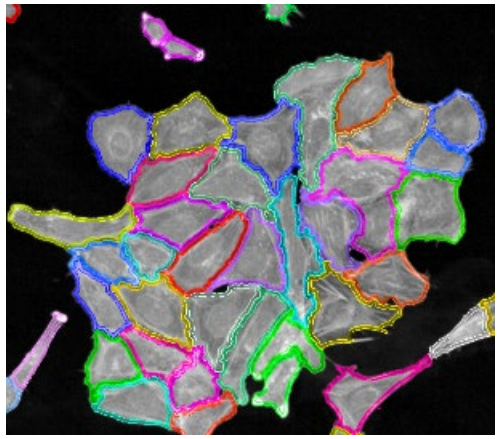
nuclei



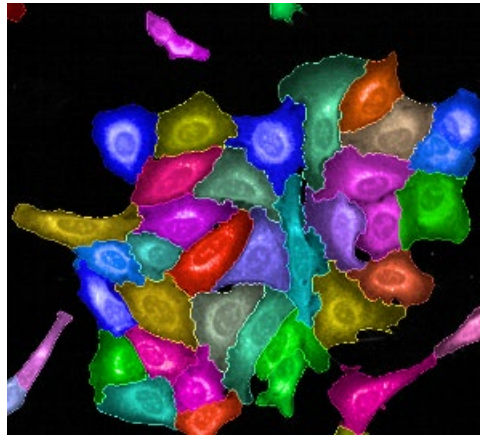
cytoplasm



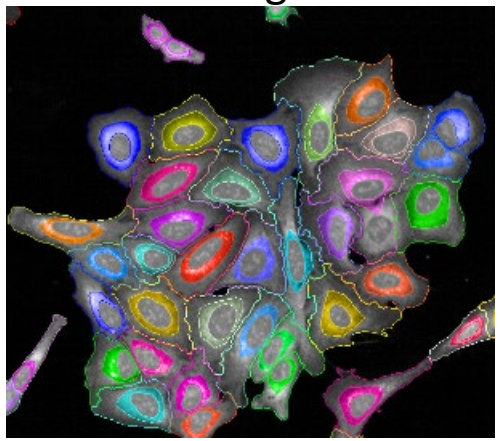
membrane



cell

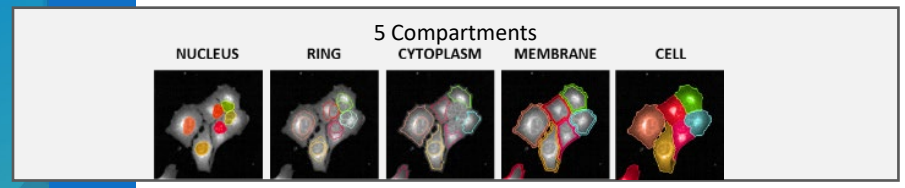


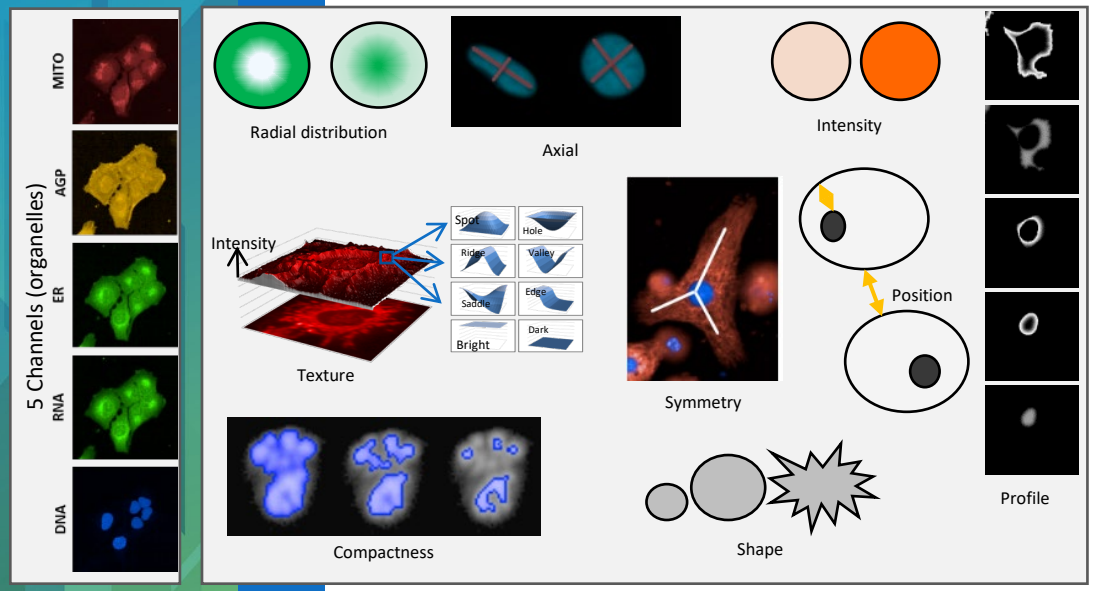
ring



Joshua Harrill

Phenotypic Feature Extraction





49 Feature Categories
(ex. MITO_Texture_Cytoplasm)

1300 features / cell

PerkinElmer Opera Phenix

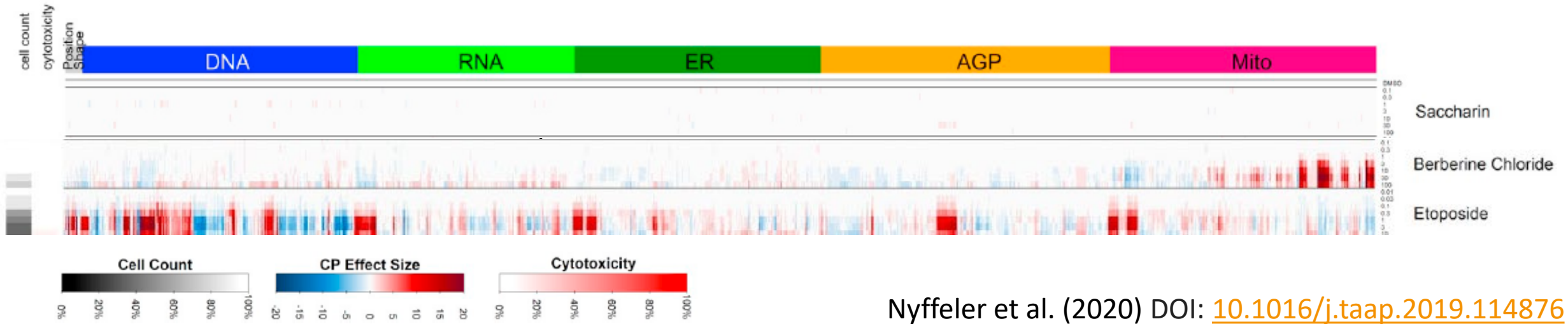
Modality: Confocal (single *z*)
Objective: 20X Water
Plate: CellCarrier-384 Ultra
Fields: 5 or 9

Harmony software performs feature extraction

Slide by Joshua Harrill
Illustrations from Perkin Elmer

		Module								
		Position [7]	Basic morph-ology [5]	SCARP morphology					Intensity [9]	Texture [14]
				Symmetry [80]	Compactness [40]	Axial [20]	Radial [28]	Profile [20-30]		
Channel	DNA			Nuclei	Nuclei	Nuclei	Nuclei Cell	Nuclei Cytoplasm	Nuclei	Nuclei
	RNA			Nuclei	Nuclei	Nuclei	Nuclei	Nuclei	Nuclei	Nuclei
	ER			Cell	Cell	Cell	Cell	Cytoplasm	Ring Cytoplasm	Ring Cytoplasm
	AGP			Cell	Cell	Cell	Cell	Nuclei Cytoplasm	Ring Cytoplasm Membrane	Ring Cytoplasm Membrane
	Mito			Cell	Cell	Cell	Cell	Nuclei Cytoplasm	Ring Cytoplasm	Ring Cytoplasm
	Not associated with a channel	Nuclei Cell	Nuclei Cell							

Quantification of Cellular Features

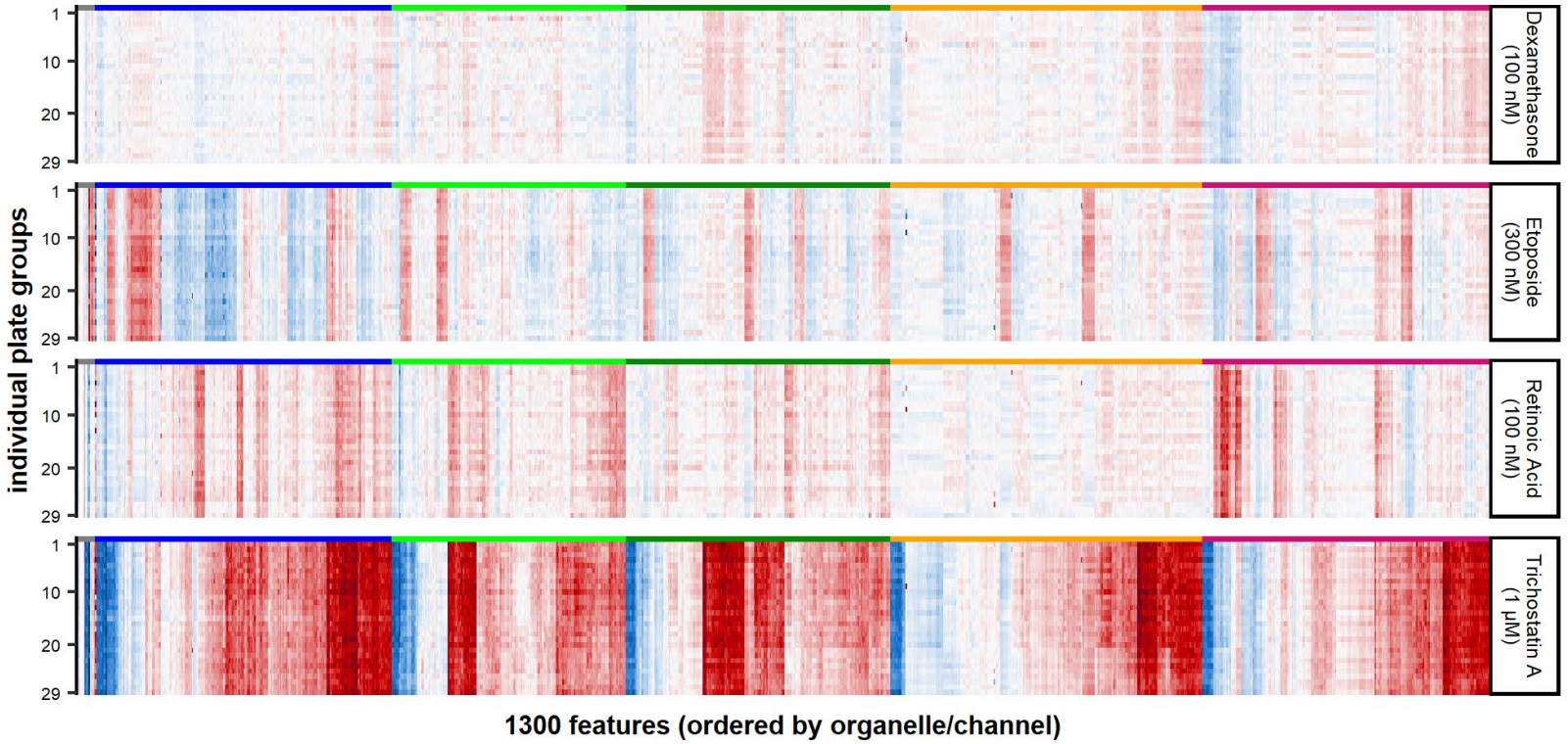


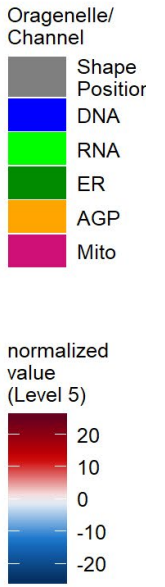
Nyffeler et al. (2020) DOI: 10.1016/j.taap.2019.114876

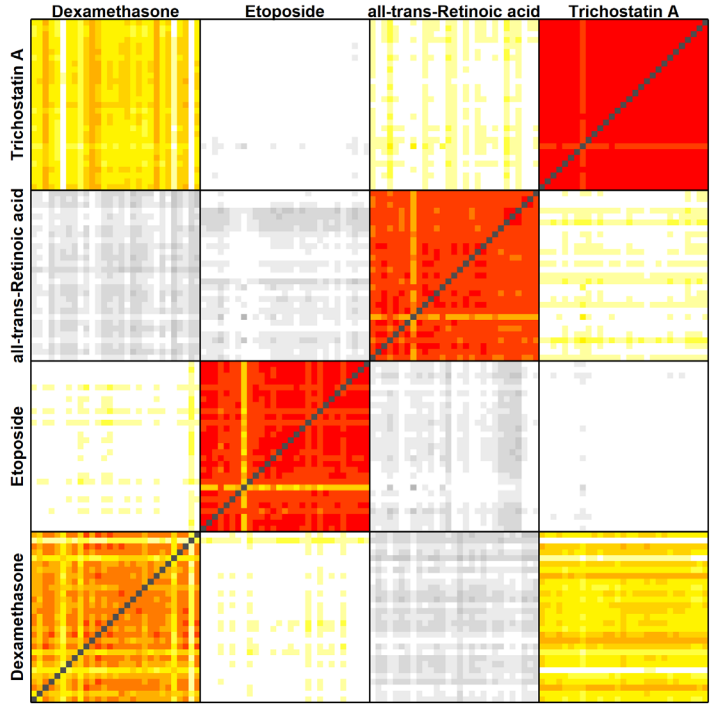
- Cell-level features are summarized per well & normalized to controls
- Different chemicals induce distinct, dose-responsive profiles

Joshua Harrill

Assay Performance / Reproducibility





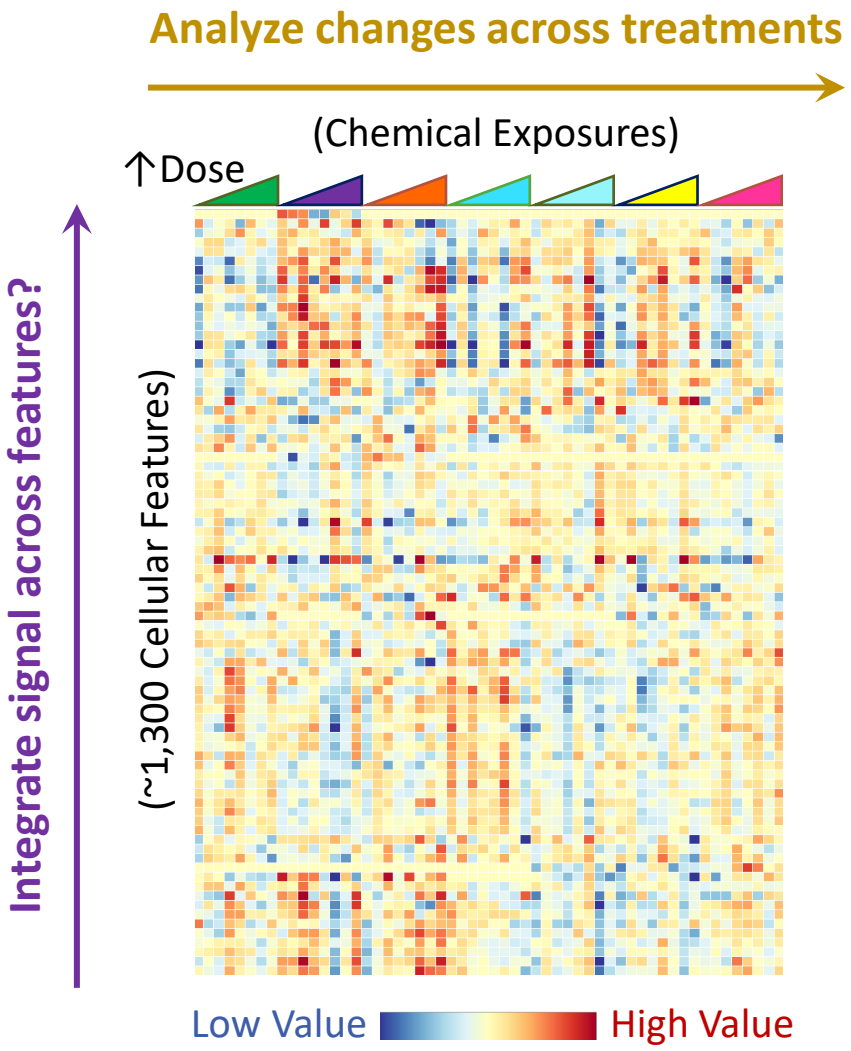


⇨ **Reference chemicals produce <u>reproducible</u> and <u>distinct</u> profiles.**

Preliminary results. Do not cite or quote.

Joshua Harrill

High-Dimensional Dose-Response Models



- Different features may respond at different doses of a given exposure!
- Need to analyze both:
  - Dose-responsive trends
  - Coordinated changes across features
- Differences from transcriptomics:
  - Lack knowledgebases linking features to biological processes
  - Certain features are highly correlated

Cell Painting Dose-Response Models

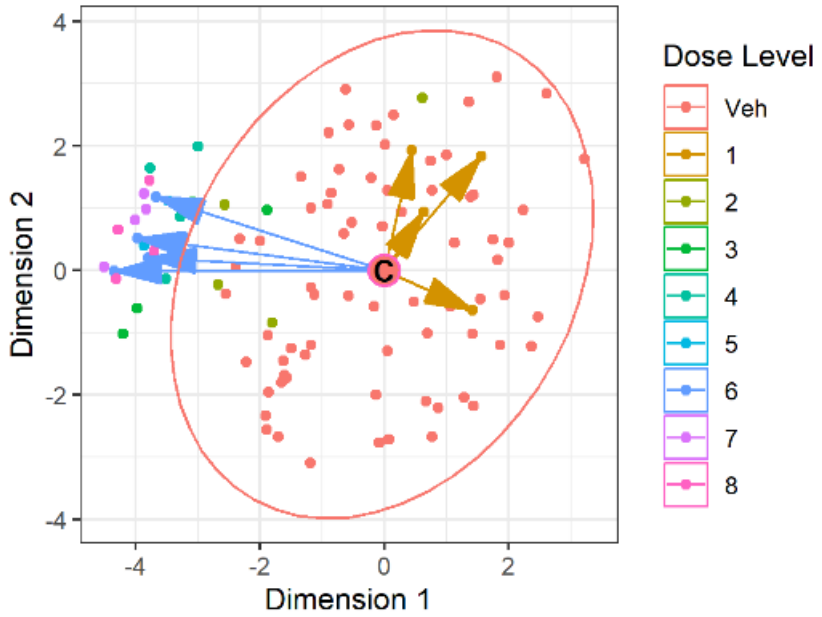
Global Mahalanobis

derive a Mahalanobis distance
(relative to control wells)

1300 features

Feature-level fitting

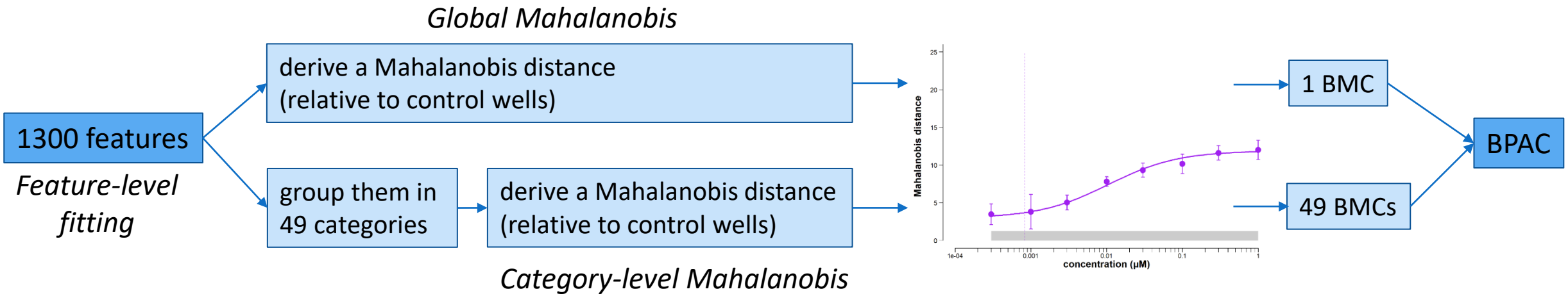
This figure is for illustrative purposes, actual calculations involve higher number of dimensions.



Mahalanobis Distance (D_M):

- A multivariate distance metric that measures the distance between a point (vector) and a distribution
- Essentially measures how much an overall cell painting profile differs from control wells
- Considers the variability of control wells in each dimension

Cell Painting Dose-Response Models



- Chemicals where a BMC can be determined using either the global or category D_M approach are considered active.
- The minimum of the global or most sensitive category BMC is the **Biological Phenotype Altering Concentration** (BPAC)

Acknowledgements

Questions?

everett.logan@epa.gov

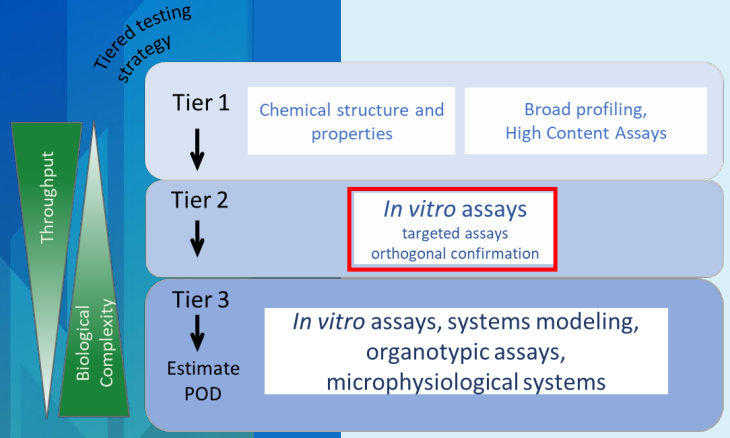
CCTE Leadership

Rusty Thomas
Sid Hunter
Katie Paul Friedman
Kimberly Slentz-Kesler

HTTr Team

Joshua Harrill
Richard Judson
Imran Shah
Derik Haggard
Joseph Bundy
Beena Vallanat
Jesse Rogers
Bryant Chambers
Laura Word
Jacob Fredenburg
Sarah Davidson-Fritz
James Johnson
Felix Harris
Khan Inan
Joshua Witten
Woody Setzer
Clinton Willis
Thomas Sheffield

Demo



Toxicity Forecasting (ToxCast)

Madison Feshuk

(with contributions from Katie Paul Friedman)

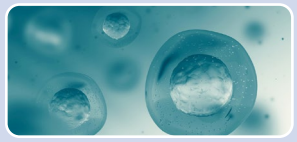
ToxCast Database Coverage

The Toxicity Forecaster (ToxCast) program curates and makes publicly available targeted bioactivity screening data. Latest database release (v4.1) includes:

- 26 Assay Sources
- 623 Unique Assays
- 1496 Unique Endpoints
- 9559 Chemicals

Including a TOX21 assay source for data generated by the TOX21 program

Diverse biology with ***over 500 mapped gene targets***, including:



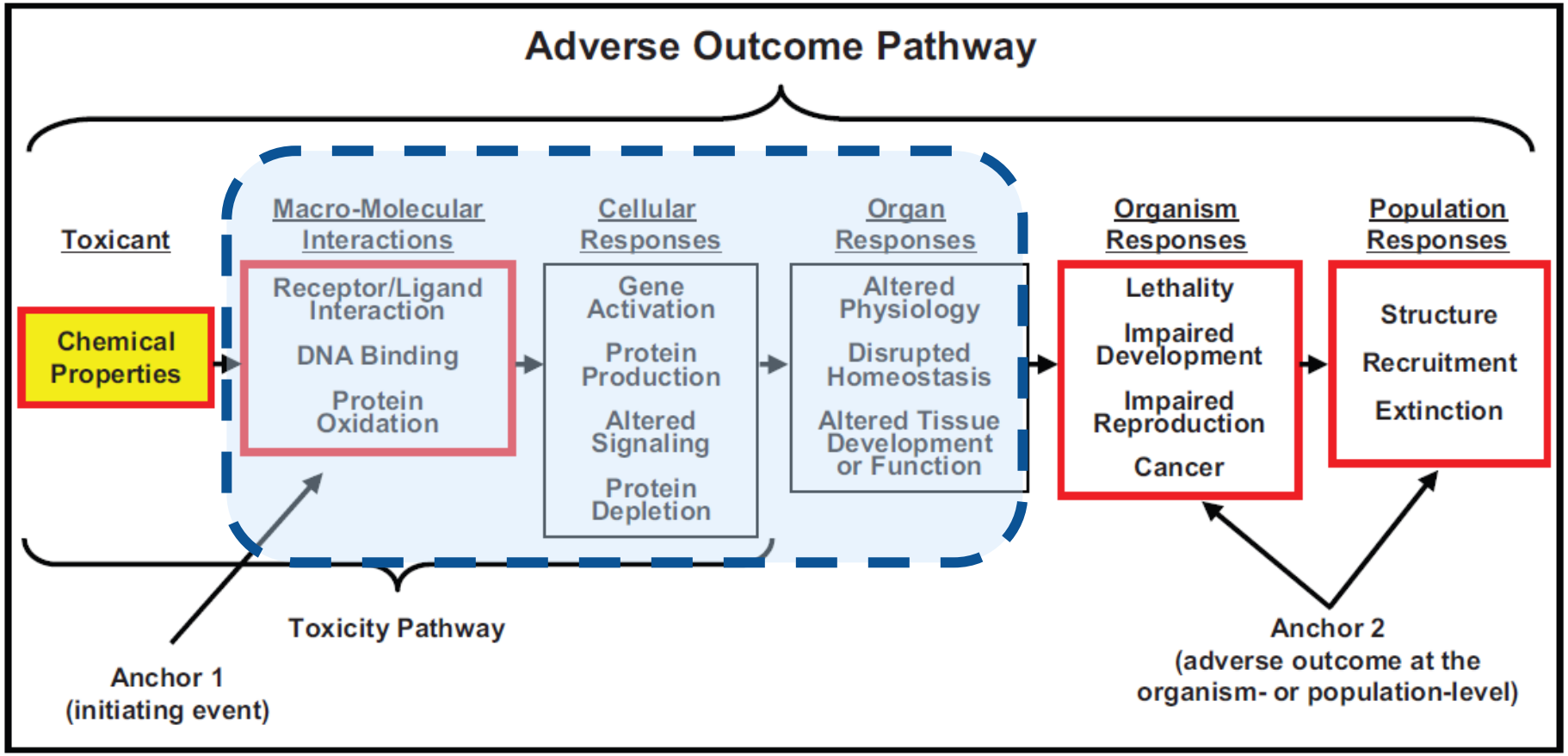
Endocrine-Related: Estrogen Receptor, Androgen Receptor, Thyroid, Steroidogenesis

Cellular Signaling Pathways: Cytotoxicity, Proliferation, Stress, Mitochondrial Toxicity

Protein Interactions: Receptors, Transporters, Ion Channels, Enzymes

Complex Responses, e.g. Immune Response, Development, Neurotoxicity, etc.

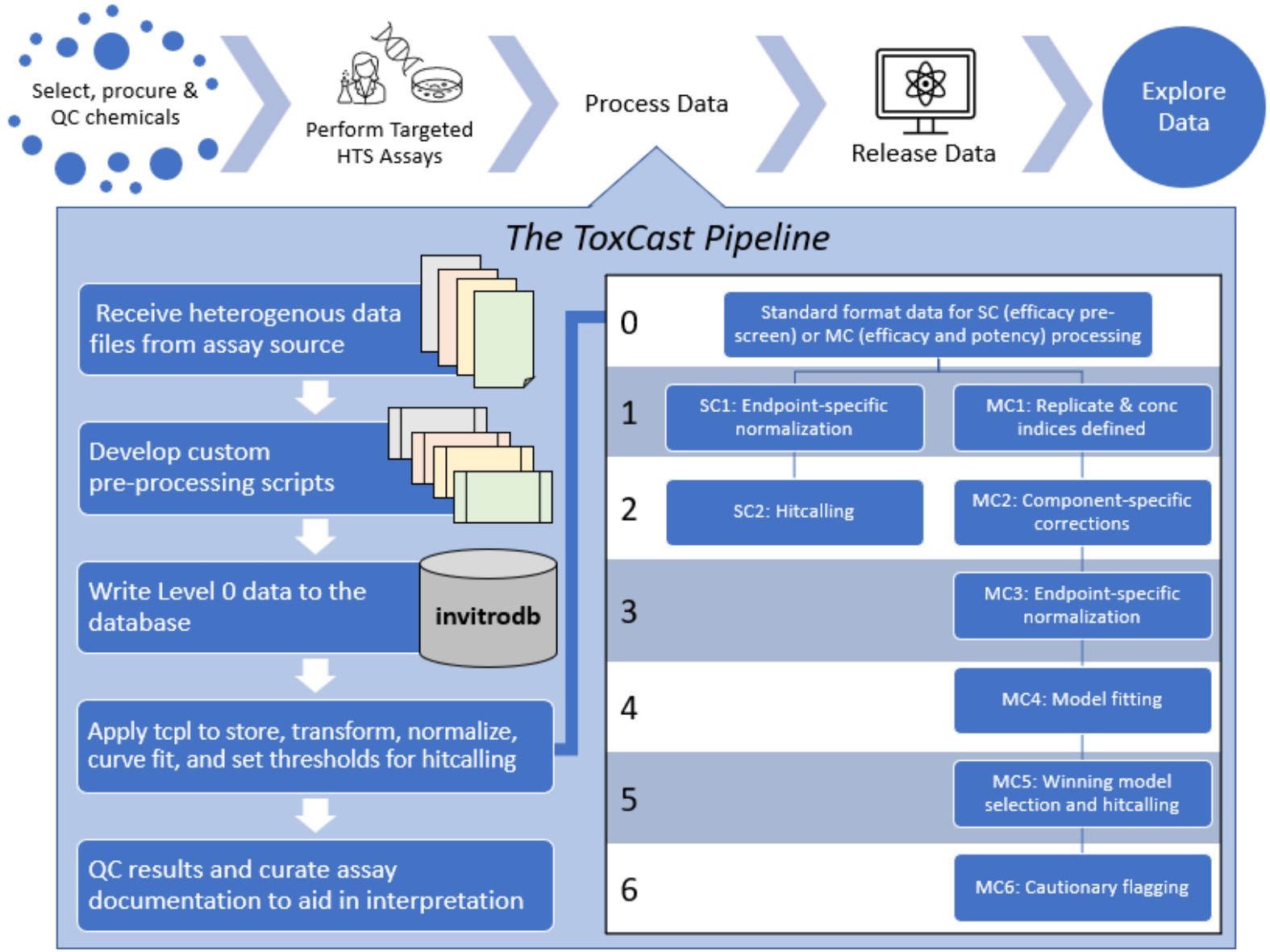
Heterogeneous targeted NAMs in ToxCast address a range of event types in the AOP framework



Ankley et al. (2010) 10.1002/etc.34

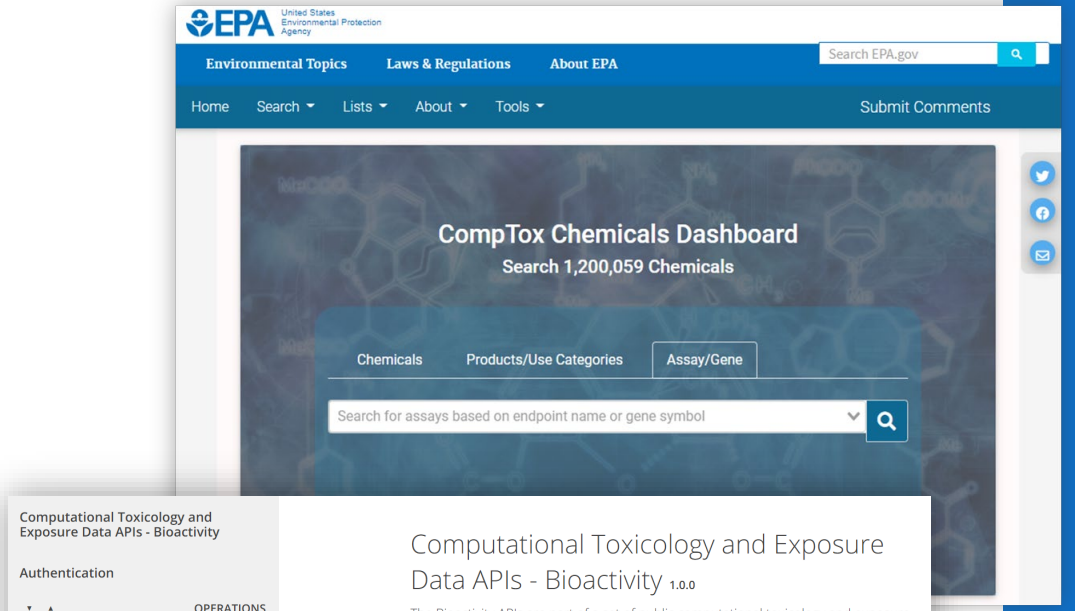
ToxCast Pipeline and Database

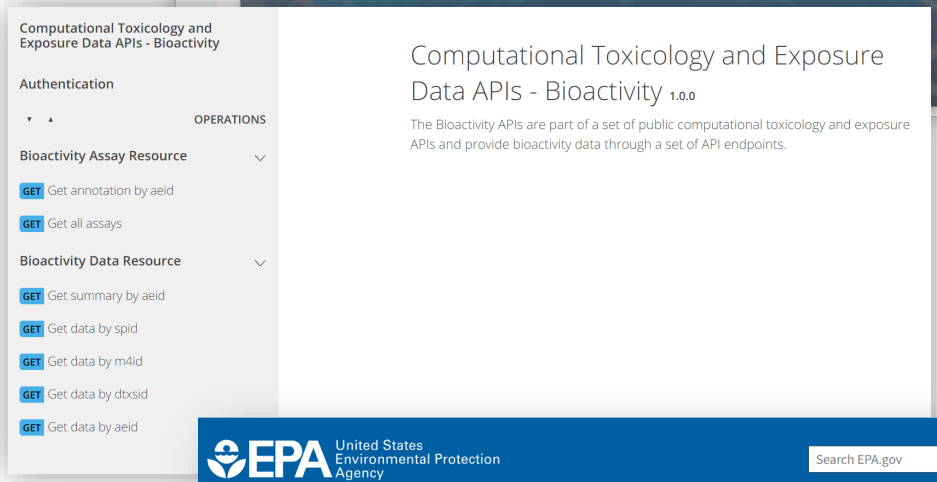
Process Overview

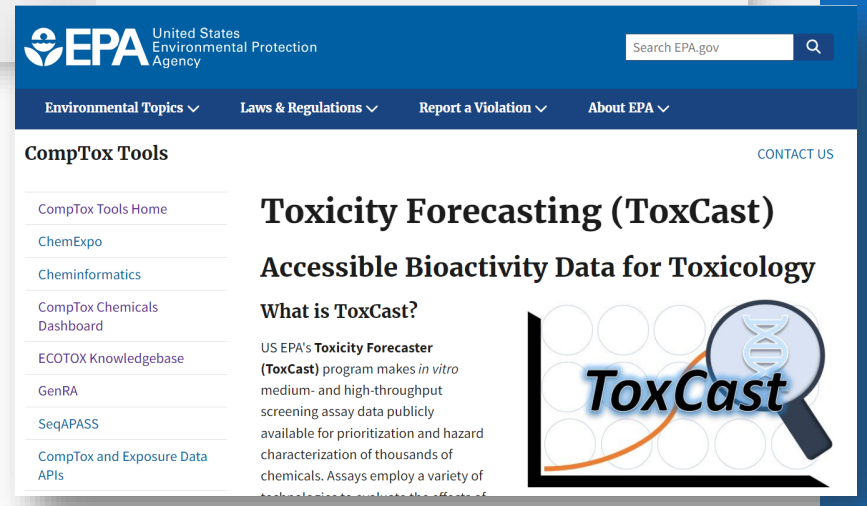


Exploring ToxCast

- Ongoing work has also focused on augmenting and diversifying how ToxCast data can be accessed for our users.
- ToxCast data is accessible via:
  - CompTox Chemicals Dashboard
  - Computational Toxicology and Exposure Bioactivity APIs
  - Downloadable Data Pages

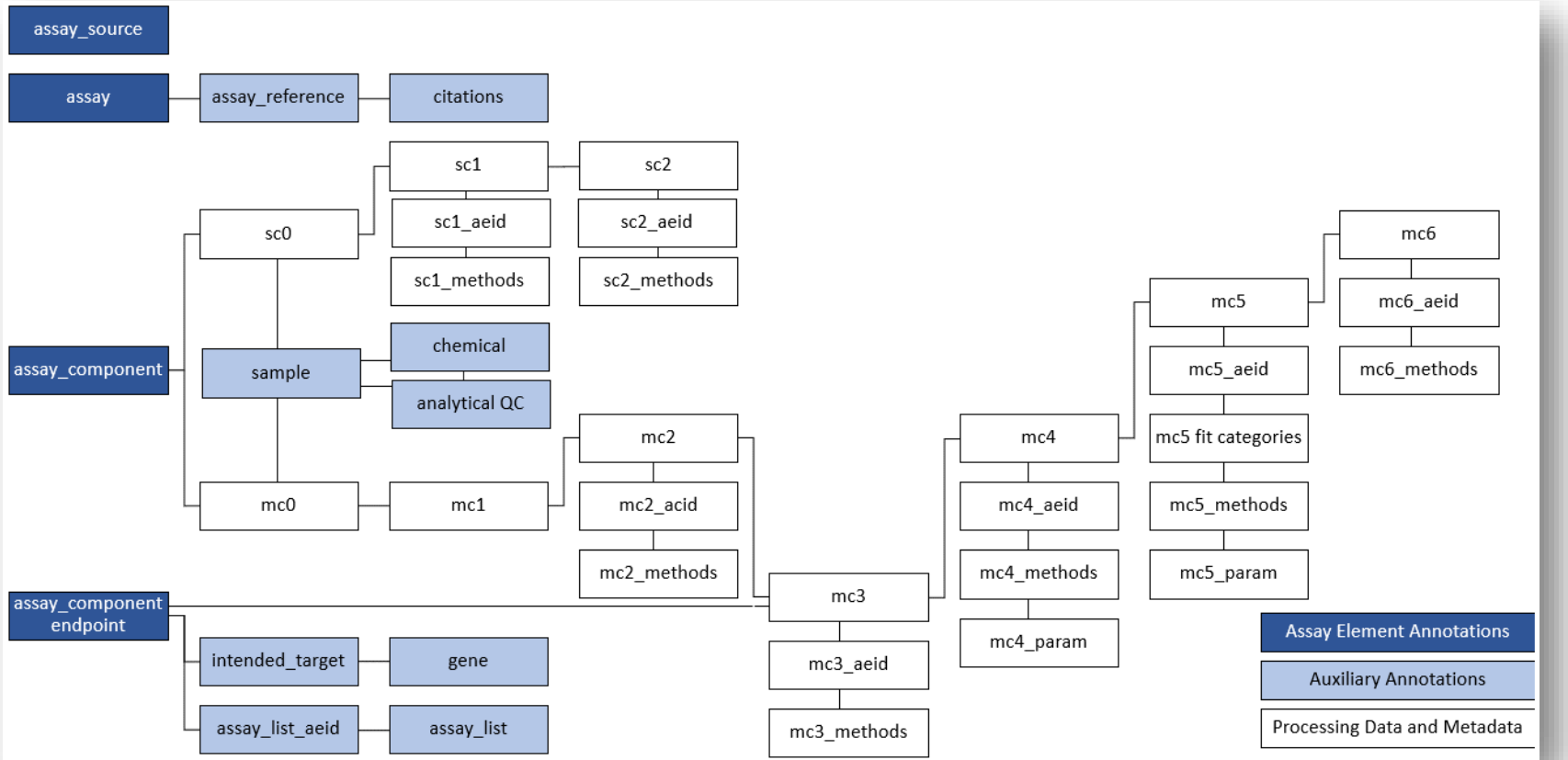






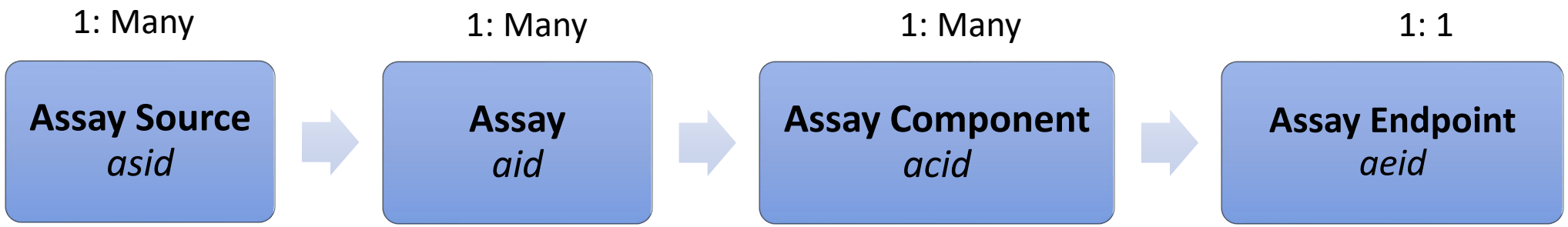
Database Structure

- ToxCast covers diverse biological space and annotations help us flexibly aggregate and differentiate processed ToxCast data for user needs
- The ToxCast database (invitrodb) captures the following types of information:
  - Assay Element Annotations
  - Auxiliary Annotations
  - Processing Data and Metadata



Assay Element Annotations

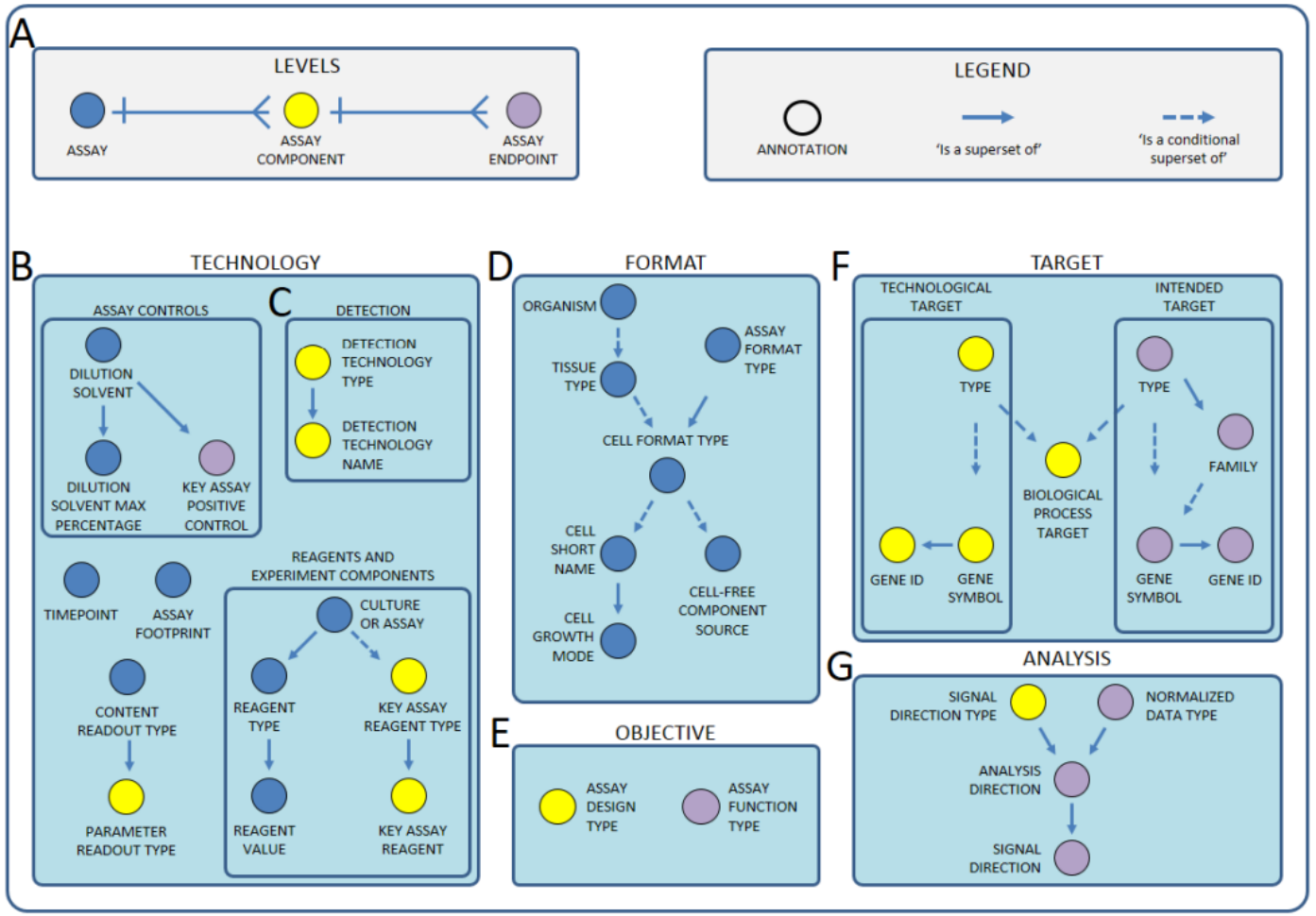
- Each annotation is assigned as a feature to an assay element level:
  - **Assay Source:** *Who* conducted the assay
  - **Assay:** *What* assay platform was used
  - **Assay Component:** "Raw" readout of *what* was measured
  - **Assay (component) Endpoint:** *How* the measurement is interpreted (i.e. normalized component data)



Note: All processing with tcpl occurs at the assay component or assay endpoint, depending on the processing type (single-concentration or multiple-concentration) and level. **No data is stored at the assay or assay source level.**

Assay Element Annotations

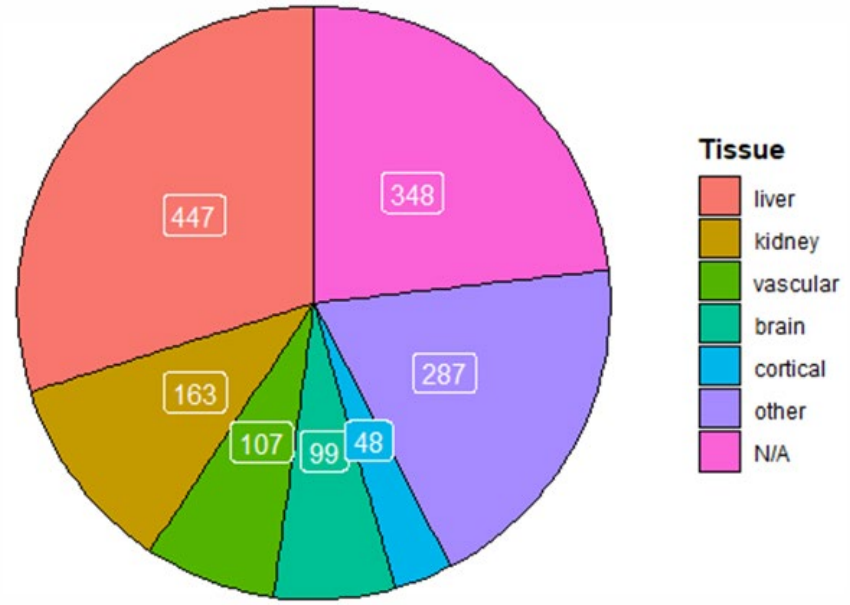
- Annotations follow Bioassay Ontology (BAO) framework capture four types of information:
  - Identification (A)
  - Design (B-E)
    - Technology
    - Format
    - Objective
  - Target (F)
    - Technological target
    - Intended target
    - Biological process
  - Analysis (G)



Assay Element Annotations

- Most annotations employ controlled vocabulary within the database
- Some annotations are hierarchical

- e.g., general ‘intended_target_family’ and more specific ‘intended_target_family_sub’

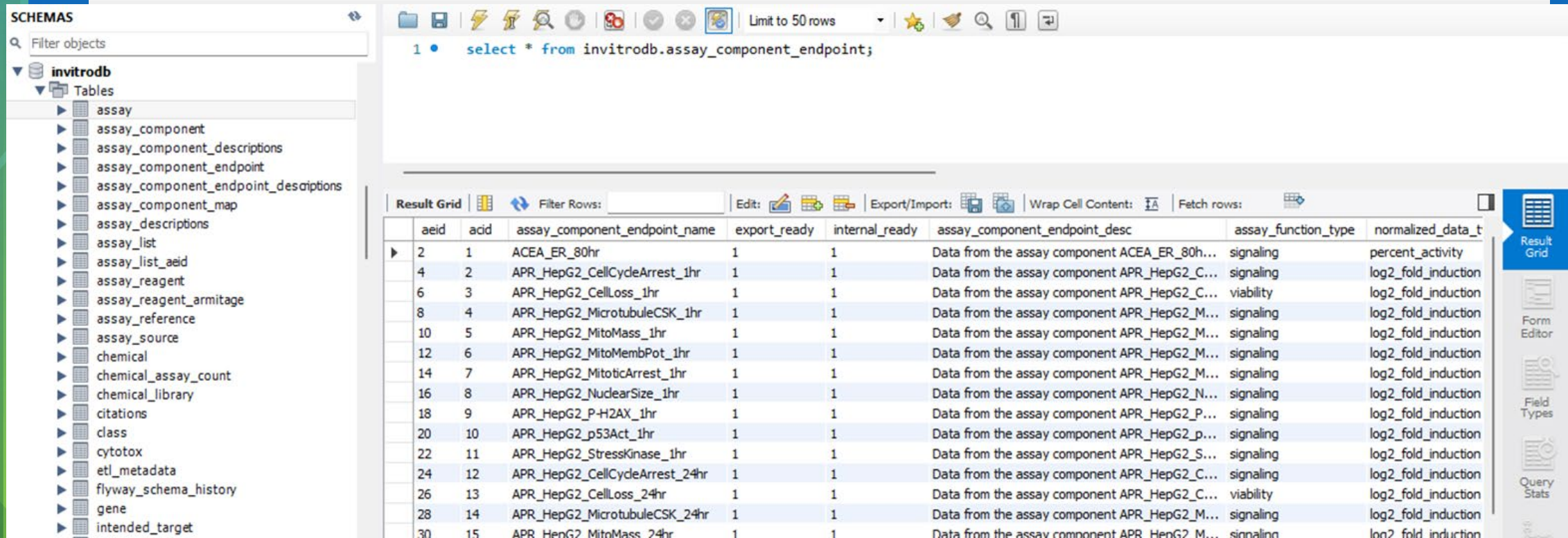


Tissue *of origin across all assays*

Intended_target_family *frequency across all endpoints*

Auxiliary Annotations

- Capture additional information, including:
  - Assay list presence
  - Relevant gene identifier(s) from National Center for Biotechnology Information (NCBI)
  - Reagents or experimental conditions
  - Publications describing assay design or results
  - Linkages to relevant Adverse Outcome Pathways (AOPs) and Key Events (KEs)



Demo

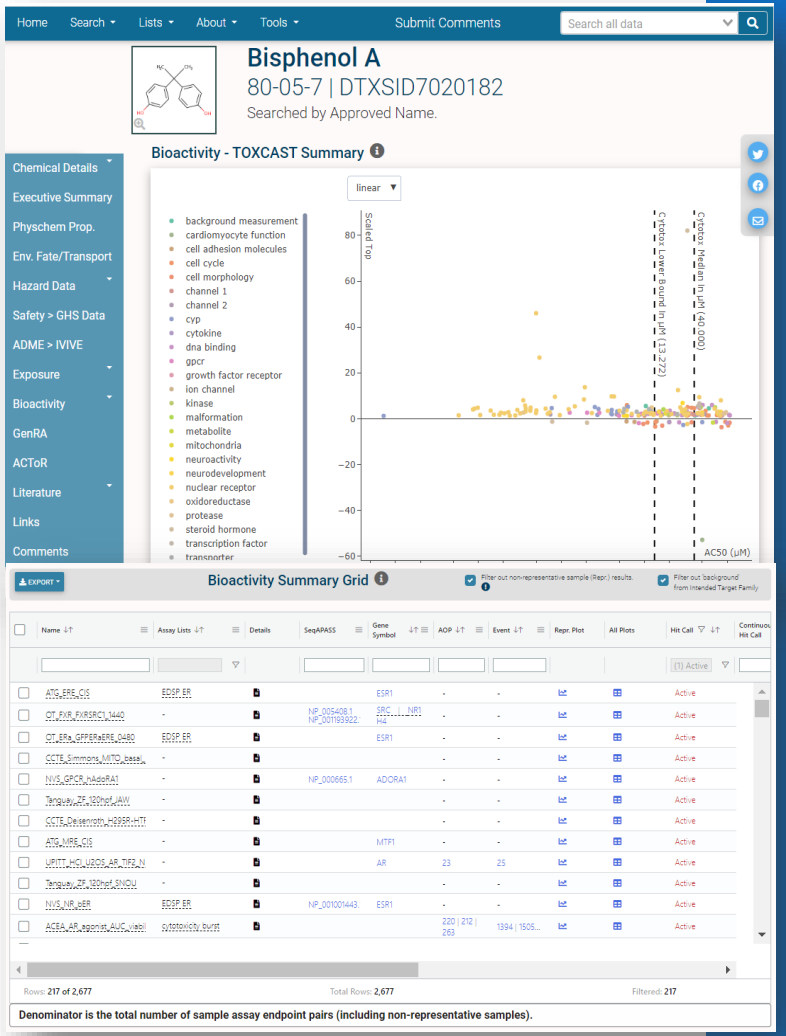
Demo

- Toxicity Forecasting (ToxCast) home page https://www.epa.gov/comptox-tools/toxicity-forecasting-toxcast
  - Exploring ToxCast Data → Download Database Package
- Tcpl CRAN: https://cran.r-project.org/web/packages/tcpl/index.html
- Tcpl GitHub: https://github.com/USEPA/CompTox-ToxCast-tcpl
- CCD: https://comptox.epa.gov/dashboard/
  - Single Chemical Search "BPA" >Navigate to ToxCast tab>
    - ToxCast Summary plot (AC50 vs Scaled Top (max modeled response/cutoff), cytotoxicity burst median and lower bounds)
    - Bioactivity grid (Adding additional fields like Annotations, Inspecting plots)
  - Search by gene "estrogen"
  - Search by assay "ACEA_ER_80hr"
  - Lists of Assay vs Chemicals > Send to Batch
  - Batch Search Export of ToxCast AC50 values
- CCTE APIs home https://api-ccte.epa.gov/docs/ (Must request API key to access)
- Bioactivity APIs https://api-ccte.epa.gov/docs/bioactivity.html
  - Overview of different request types
- ccdR for accessing APIs https://cran.r-project.org/web/packages/ccdR/index.html

CompTox Chemicals Dashboard (CCD)

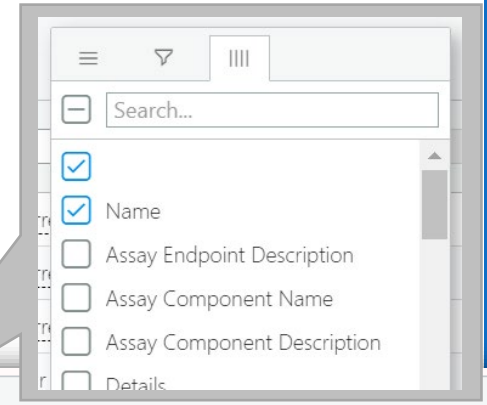
https://comptox.epa.gov/dashboard

- CCD's ToxCast bioactivity module presents a view of potency and relative efficacy metrics across ToxCast endpoints for chemicals of interest
- Users can easily sort, filter, and export ToxCast results and assay descriptions
- Notable updates in the CCD v2.3 release (December 2023) include:
  - Data was refreshed to invitrodb v4.1
  - ToxCast Summary tab is now a single tab that combines the previous ToxCast Summary and ToxCast Conc. Response tabs
  - Bioactivity Summary Grid includes v4.1 information in new columns, including benchmark dose (BMD), benchmark response (BMR), and Continuous Hitcall
- *Example on right: Bisphenol A*
  https://comptox.epa.gov/dashboard/chemical/invitrodb/DTXSID7020182



Filtering ToxCast Data on the CCD CompTox Chemicals Dashboard: https://comptox.epa.gov/dashboard

- CCD's ToxCast bioactivity module presents a view of potency and relative efficacy metrics across ToxCast endpoints for chemicals of interest
- Users can easily sort, filter, and export ToxCast results and assay descriptions
- Example: Consider BPA DTXSID7020182
  - Select ☰ in Bioactivity Summary Grid column headers to add additional annotation columns |||
  - Explore!
    - Below shows results filtered to **EDSP ER Assay List** (assays used in the ToxCast ER pathway model)



Name	Hit Call	Assay Lists	Gene Symbol	Organism	Tissue	Cell Format	Intended Target Family
		(1) EDSP ER					
ACEA_ER_80hr	Active	EDSP ER	ESR1	human	breast	cell line	nuclear receptor
ACEA_ER_AUC_viability	Inactive	cytotoxicity burst EDSP ER		human	breast	cell line	cell cycle
ATG_ERE_CIS	Active	EDSP ER	ESR1	human	liver	cell line	nuclear receptor
ATG_ERa_TRANS	Active	EDSP ER	ESR1	human	liver	cell line	nuclear receptor
NVS_NR_bER	Active	EDSP ER	ESR1	bovine	uterus	tissue-based cell-free	nuclear receptor

Filtering ToxCast Data on the CCD CompTox Chemicals Dashboard:

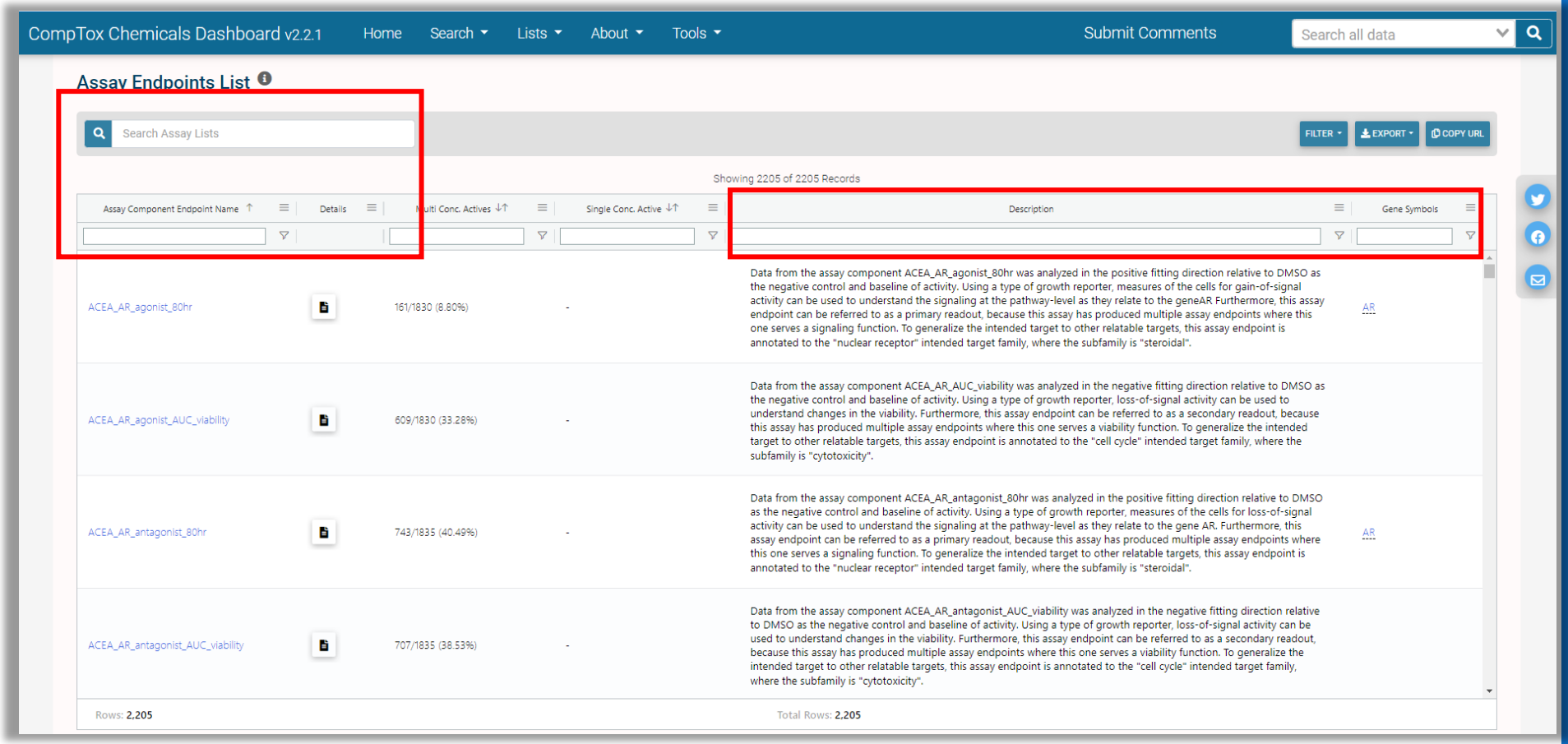
https://comptox.epa.gov/dashboard

- CCD's ToxCast bioactivity module presents a view of potency and relative efficacy metrics across ToxCast endpoints for chemicals of interest
- Users can easily sort, filter, and export ToxCast results and assay descriptions
- Example: Consider BPA DTXSID7020182
  - Select ☰ in Bioactivity Summary Grid column headers to add additional annotation columns |||
  - Explore!
    - Below shows results filtered to **Actives in human ESR1 cell-based assays**

Name	Hit Call	Assay Lists	Gene Symbol	Organism	Tissue	Cell Format	Intended Target Family
		(1) EDSP ER	ESR1	human		cell line	
ACEA_ER_80hr	Active	EDSP ER	ESR1	human	breast	cell line	nuclear receptor
ATG_ERE_CIS	Active	EDSP ER	ESR1	human	liver	cell line	nuclear receptor
ATG_ERa_TRANS	Active	EDSP ER	ESR1	human	liver	cell line	nuclear receptor
OT_ER_ERaERa_0480	Active	EDSP ER	ESR1	human	kidney	cell line	nuclear receptor
OT_ER_ERaERa_1440	Active	EDSP ER	ESR1	human	kidney	cell line	nuclear receptor
OT_ER_ERaERb_0480	Active	EDSP ER	ESR1 | ESR2	human	kidney	cell line	nuclear receptor

ToxCast data are publicly accessible from the CompTox Chemicals Dashboard

Search by gene, vendor name, etc.

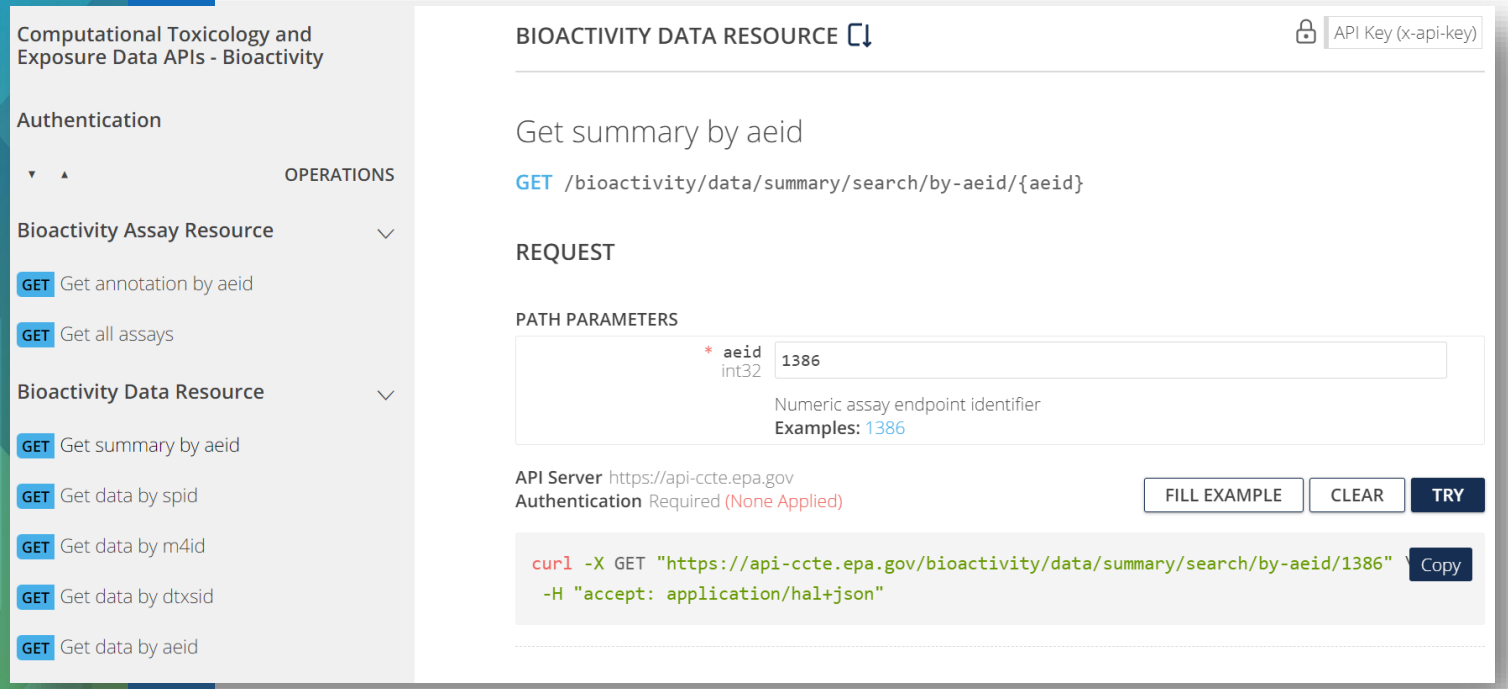


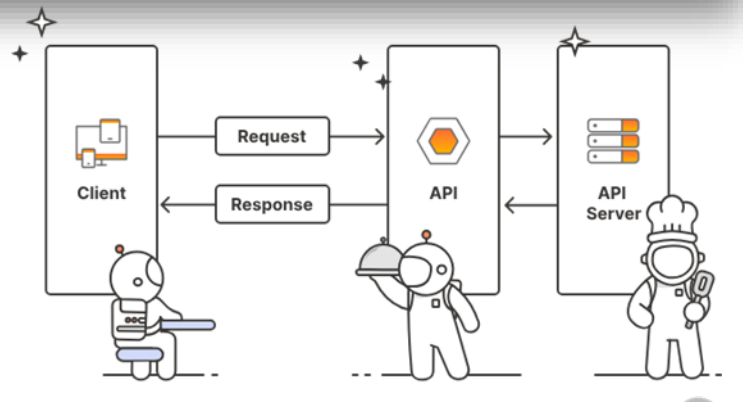
Many users are accustomed to viewing data per substance (as identified by a DSSTox identifier, or DTXSID), but you can also identify assay endpoint data by entering from Lists > Lists of Assays. These data can be exported after loading the data for the assay.

https://comptox.epa.gov/dashboard/assay-endpoints

Application Programming Interfaces (APIs)

https://api-ccte.epa.gov/docs/bioactivity.html



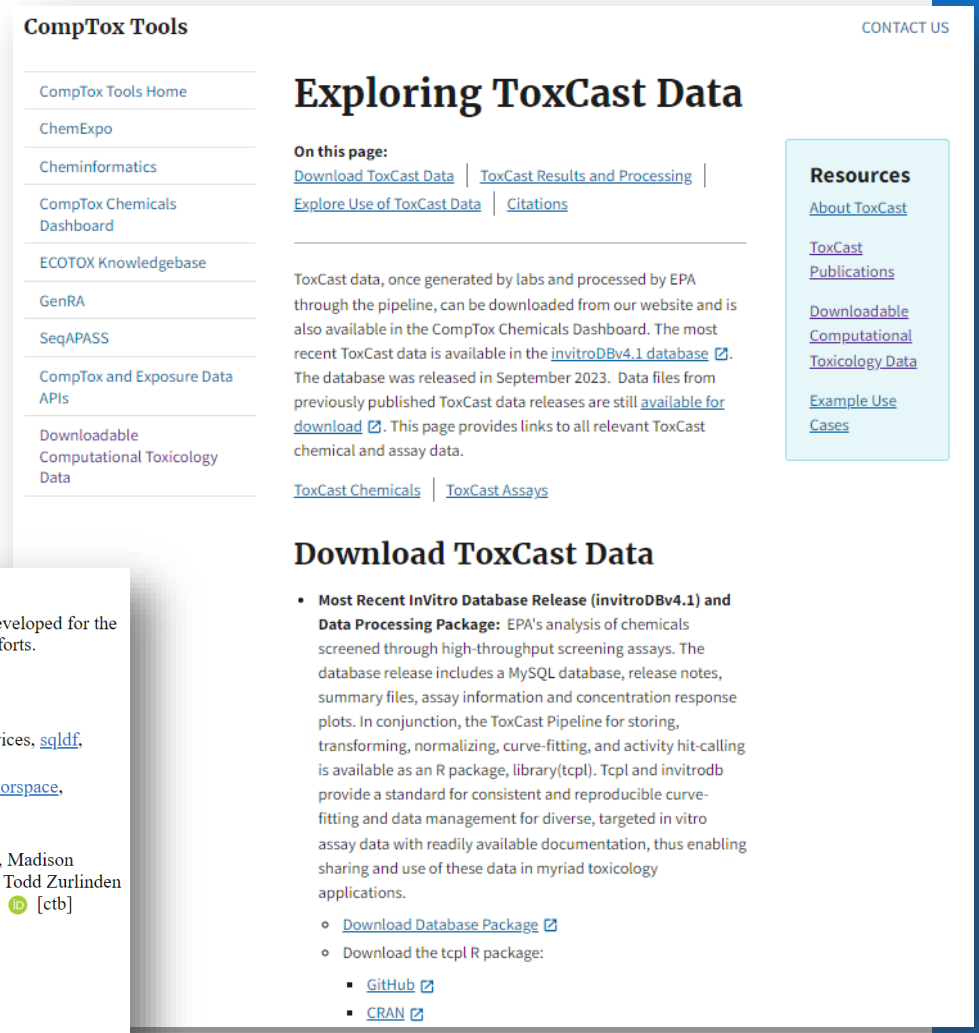


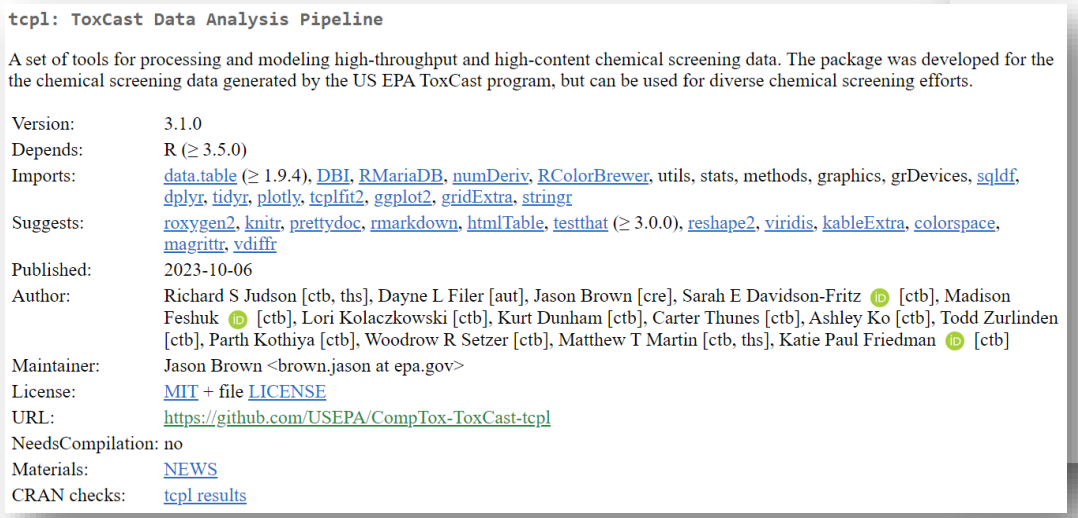
- APIs provide data for various use cases, including research and applications with user interfaces
- Users can avoid large data downloads by accessing invitrodb programmatically via an API
- This is a great read-only solution for users who require more flexibility than the CCD can provide
- For additional documentation, check out the CCTE API Home Page or ccdR R package. More integration with tcpl is coming soon

ToxCast Data Downloads

https://www.epa.gov/comptox-tools/exploring-toxcast-data

- Data downloads allow users to set up their own personal instance of the invitrodb MySQL database and interact with the data directly via the tcpl R package
- This is a preferred option for more customized or programmatic ToxCast data needs, or if users want to do their own data processing





Conclusions

- Hazard NAMs are being developed as alternatives to traditional hazard methods
- Many hazard NAM data are available in the CompTox Chemicals Dashboard, download or API
- Each assay technology may have specific limitations, which may require user discretion for more complex interpretations of the data
- Hazard NAM data may be qualitatively and quantitatively informative in different contexts