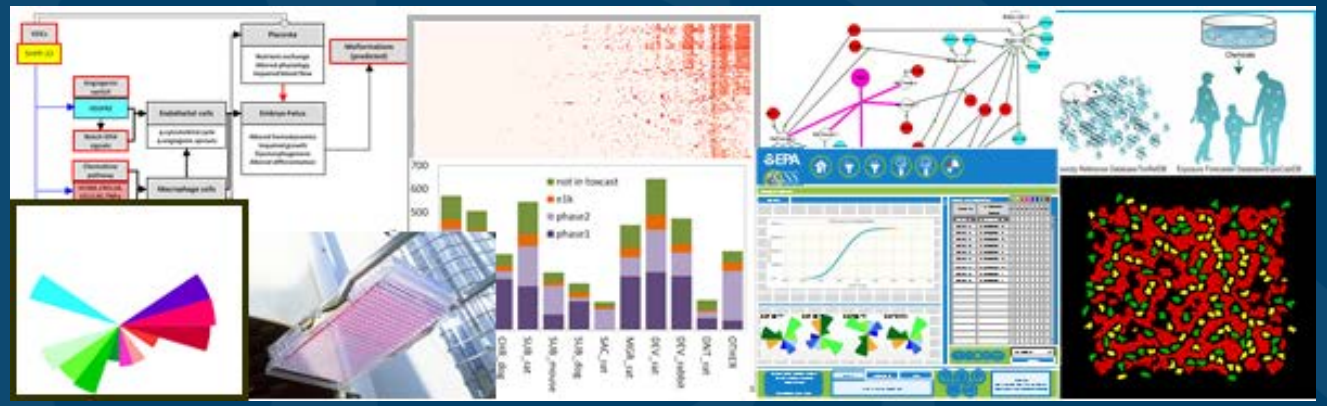
EPA United States Environmental Protection Agency

Update on NCCT Transcriptomic and Metabolic Retrofit Projects



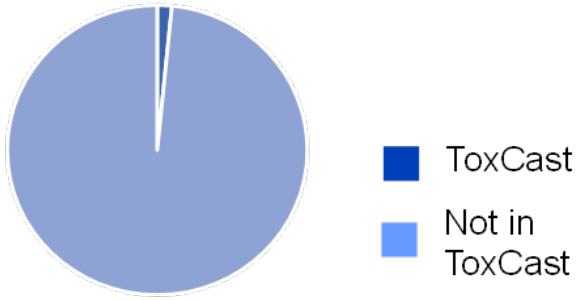
Endocrine Policy Forum
10 October 2107

Kevin M. Crofton, PhD
Deputy Director
National Center for Computational Toxicology

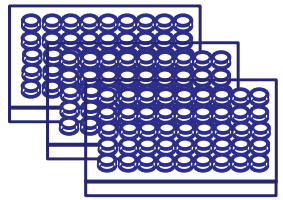
The views expressed in this presentation are those of the presenter and do not necessarily reflect the views or policies of the U.S. EPA

Major Limitations in *In Vitro* Test Systems

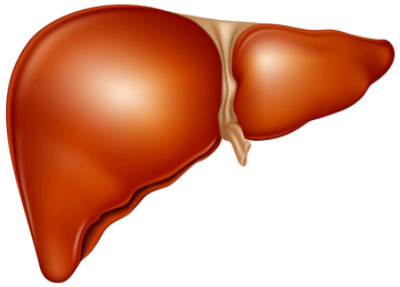
Biological Coverage (Gene Basis)



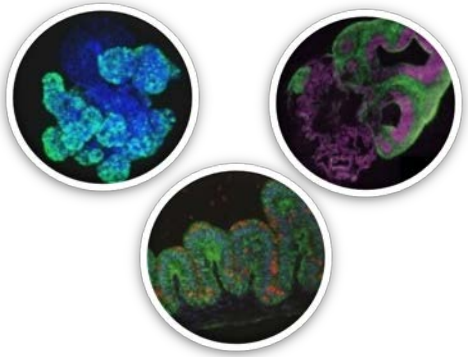
Chemical Coverage and Specific Chemical Types (e.g., VOCs)



Metabolic Competence



Organ and Tissue Responses



Four ongoing Project to Systematically Address Limitations in Alternative Test Systems

United States Environmental Protection Agency

High-Throughput Transcriptomic Screening

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
A	MAQC-A (Us)	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	non-treated
B	MAQC-A (Us)	30	30	30	30	30	30	30	30	30	30	30	30	30	30	30	30	30	30	30	30	30	30	non-treated
C	MAQC-B (Us)	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	non-treated
D	MAQC-B (Us)	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	DMSO
E	Bulk Lysate (DMSO)	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	DMSO
F	Bulk Lysate (DMSO)	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	DMSO
G	Bulk Lysate (Trichostatin)	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	DMSO [No Label]
H	Bulk Lysate (Trichostatin)	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	Trichostatin (1 µM)
I	Lysis Buffer (Us)	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	Trichostatin (1 µM)
J	Lysis Buffer (Us)	30	30	30	30	30	30	30	30	30	30	30	30	30	30	30	30	30	30	30	30	30	30	Trichostatin (1 µM)
K	MAQC-A (Them)	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	Genistein (10 µM)
L	MAQC-A (Them)	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	Genistein (10 µM)
M	MAQC-B (Them)	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	Genistein (10 µM)
N	MAQC-B (Them)	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	Sirolimus (0.1 µM)
O	Lysis Buffer (Them)	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	Sirolimus (0.1 µM)
P	Lysis Buffer (Them)	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	Sirolimus (0.1 µM)

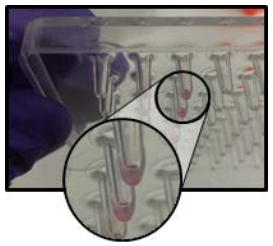
- Whole transcriptome
- Low cost
- 384-well
- Lysate
- Automatable

- Performance controls on each plate
- Replicate correlation ~0.75
- Replicate CVs ~20%

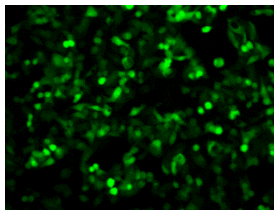
Collaboration with Unilever

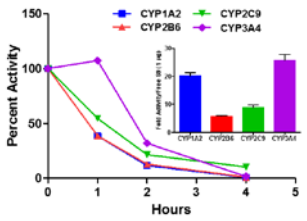
Assay Retrofit for Metabolism

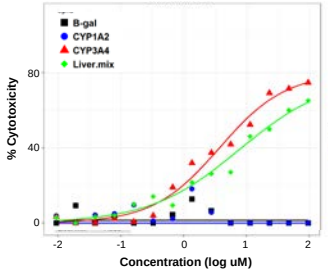
Extracellular



Intracellular

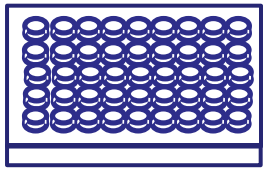






Collaboration with Unilever

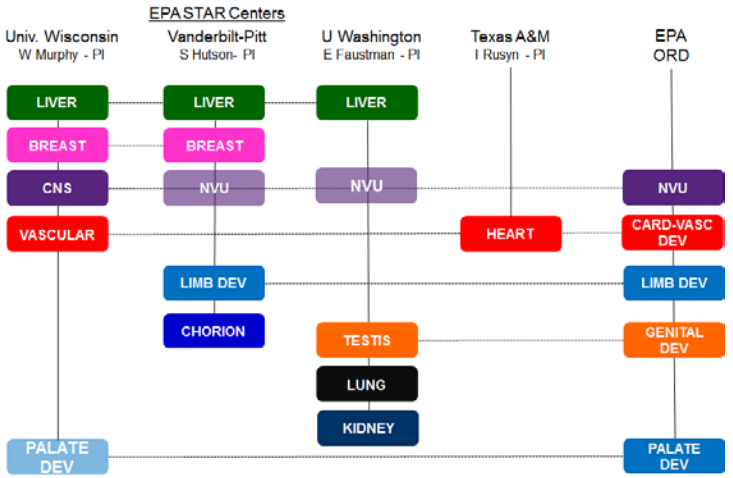
VOC *In Vitro* Exposure System and Water Library



Initial test library of ~70 chemicals

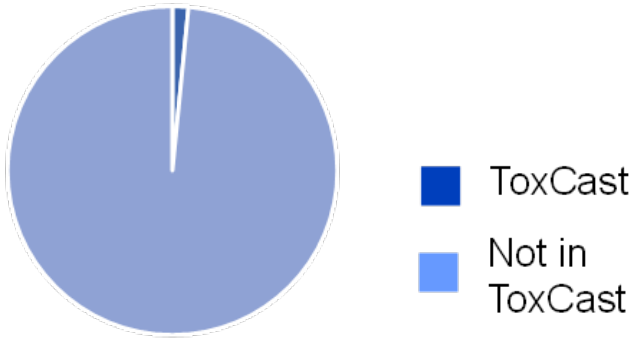
M. Higuchi (EPA-NHEERL)

Organotypic Model Development and Virtual Tissue Modeling

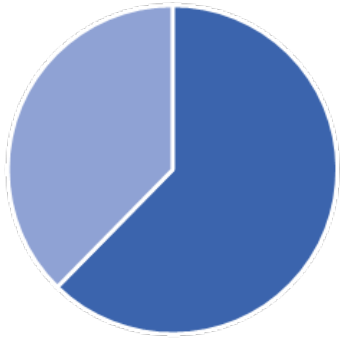


Biological Space Coverage of ToxCast

Gene Coverage



Pathway Coverage*



*At least one gene from pathway represented

- ToxCast assays cover
  - about 320 genes
- Pathway coverage is higher – but still leaves large gaps
- Recent technological advances in genomics are very promising for rapid and cost-effective whole genome screening

Progress on Development of High-Throughput Transcriptomics (HTTr)

2016

- Compared three HTTr platforms
  - Selected TempO-Seq (BioSpyder, Inc)
- Developed 384 well format for MCF7 cells
- Designed and tested custom attenuated whole transcriptome assay

2017

- Studies to determine Performance Metrics
- Workflow Pilot to determine optimal experimental conditions
- Develop data analysis pipeline
- Finalized chemical list & protocols for screening of 2200 chemicals
- Began large-scale screening

HTTr MCF7 Workflow Pilot 2017

Pilot study to validate workflow and refine experimental design

- Cell type: MCF7
- 44 chemicals, 8 conc
- Three time points: 6 , 12, 24 h
- Two types of media:
  - PRF- / PRF+ (DMEM +10% HI-FBS)
- Data: 6,804 samples x 21,111 transcripts
- Goal: Determine “optimal” time point and media

Large-scale screen (Summer 2017):

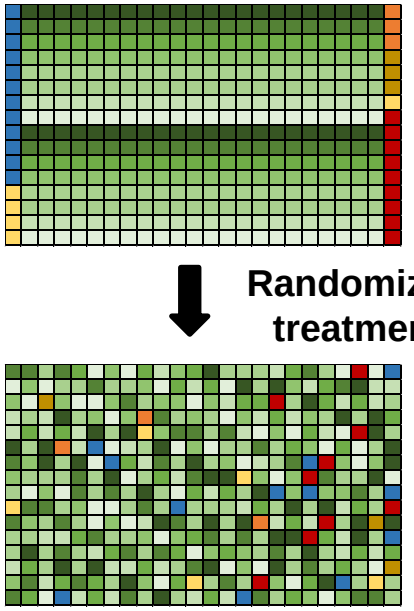
- Cell type: MCF7
- Compounds: 1,000 (ToxCast Phase I/II)
- Time Point: Single
- Concentration Response: 8
- HTTr ~25,000 x 21,111 transcript

Assay Design for Rigorous QC and Performance Validation

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
A	MAQC-A (Us)	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	non-treated
B	MAQC-A (Us)	30	30	30	30	30	30	30	30	30	30	30	30	30	30	30	30	30	30	30	30	30	30	non-treated
C	MAQC-B (Us)	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	non-treated
D	MAQC-B (Us)	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	DMSO
E	Bulk Lysate (DMSO)	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	DMSO
F	Bulk Lysate (DMSO)	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	DMSO
G	Bulk Lysate (Trichostatin)	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	DMSO [No Label]
H	Bulk Lysate (Trichostatin)	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	Trichostatin (1 µM)
I	Lysis Buffer (Us)	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	Trichostatin (1 µM)
J	Lysis Buffer (Us)	30	30	30	30	30	30	30	30	30	30	30	30	30	30	30	30	30	30	30	30	30	30	Trichostatin (1 µM)
K	MAQC-A (Them)	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	Genistein (10 µM)
L	MAQC-A (Them)	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	Genistein (10 µM)
M	MAQC-B (Them)	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	Genistein (10 µM)
N	MAQC-B (Them)	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	Sirolimus (0.1 µM)
O	Lysis Buffer (Them)	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	Sirolimus (0.1 µM)
P	Lysis Buffer (Them)	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	Sirolimus (0.1 µM)
	No Cells																							

Screen Design

- Large bank of cytogenetically and functionally characterized cells
- Multiple controls
  - MAQCs, lysate, DMSO, trichostatin, genistein, sirolimus
- 8 point concentration response
- Plate randomization technology
- Parallel HCI screen



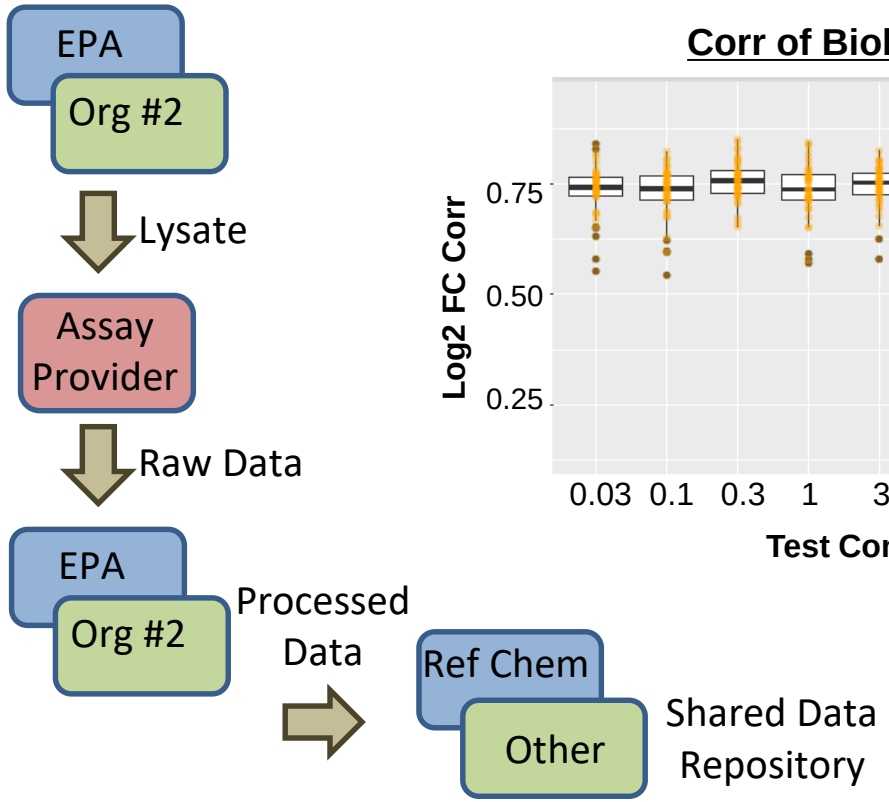
Josh Harrill, Unpublished

Workflow Pilot Performance Metrics

High-Throughput Transcriptomic Screen

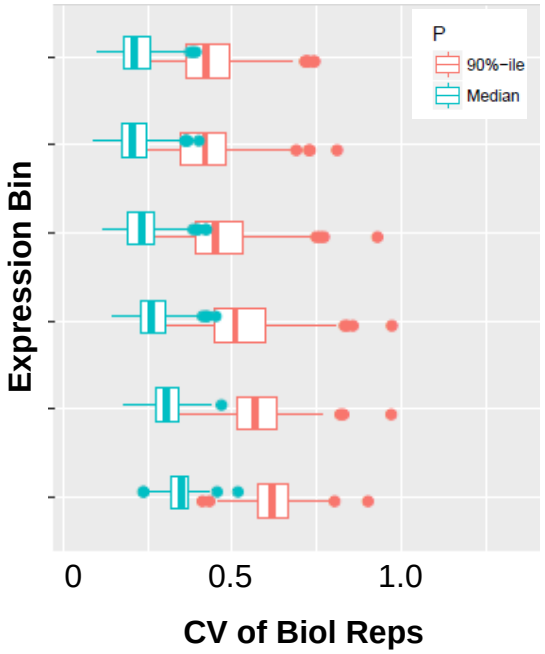
- TempO-Seq whole transcriptome assay
- Low cost
- 384-well, *cell lysate compatible*
- Automatable

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
A	MAQC-A (Us)	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	non-treated
B	MAQC-A (Us)	30	30	30	30	30	30	30	30	30	30	30	30	30	30	30	30	30	30	30	30	30	30	non-treated
C	MAQC-B (Us)	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	non-treated
D	MAQC-B (Us)	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	DMSO
E	Bulk Lysate (DMSO)	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	DMSO
F	Bulk Lysate (DMSO)	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	DMSO
G	Bulk Lysate (Trichostatin)	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	DMSO [No Label]
H	Bulk Lysate (Trichostatin)	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	Trichostatin (1 µM)
I	Lysis Buffer (Us)	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	Trichostatin (1 µM)
J	Lysis Buffer (Us)	30	30	30	30	30	30	30	30	30	30	30	30	30	30	30	30	30	30	30	30	30	30	Trichostatin (1 µM)
K	MAQC-A (Them)	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	Genistein (10 µM)
L	MAQC-A (Them)	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	Genistein (10 µM)
M	MAQC-B (Them)	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	Genistein (10 µM)
N	MAQC-B (Them)	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	Sirolimus (0.1 µM)
O	Lysis Buffer (Them)	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	Sirolimus (0.1 µM)
P	Lysis Buffer (Them)	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	Sirolimus (0.1 µM)



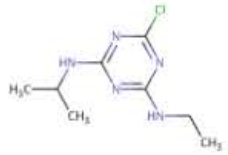
Corr of Biol Reps

CVs Biol Reps



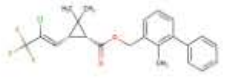
Collaboration with Unilever

Example Workflow Pilot Chemicals with known MIE/MoA



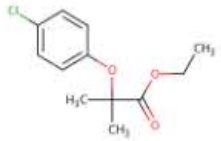
Atrazine
1912-24-9

cAMP inducer



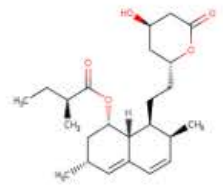
Bifenthrin
82657-04-3

Na+ channel



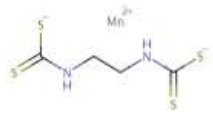
Clofibrate
637-07-0

Hypolipidemic



Lovastatin
75330-75-5

HMGCR



Maneb
12427-38-2

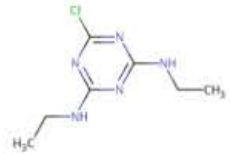
SH Reactive

Reserpine
50-55-5

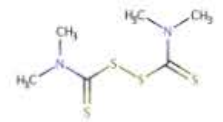
VMAT

Rotenone
83-79-4

Mito Complex I



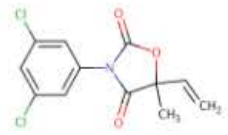
Simazine
122-34-9



Thiram
137-26-8

SH Reactive

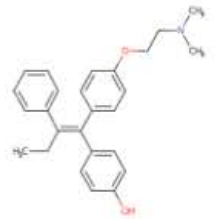
Ziram
137-30-4



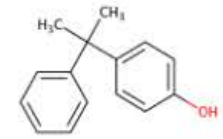
Vinclozolin
50471-44-8

Antiandrogen

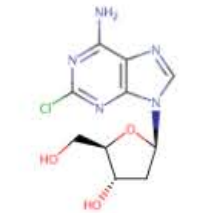
Fulvestrant
129453-61-8



(Z)-4-Hydroxytamoxifen
68047-06-3

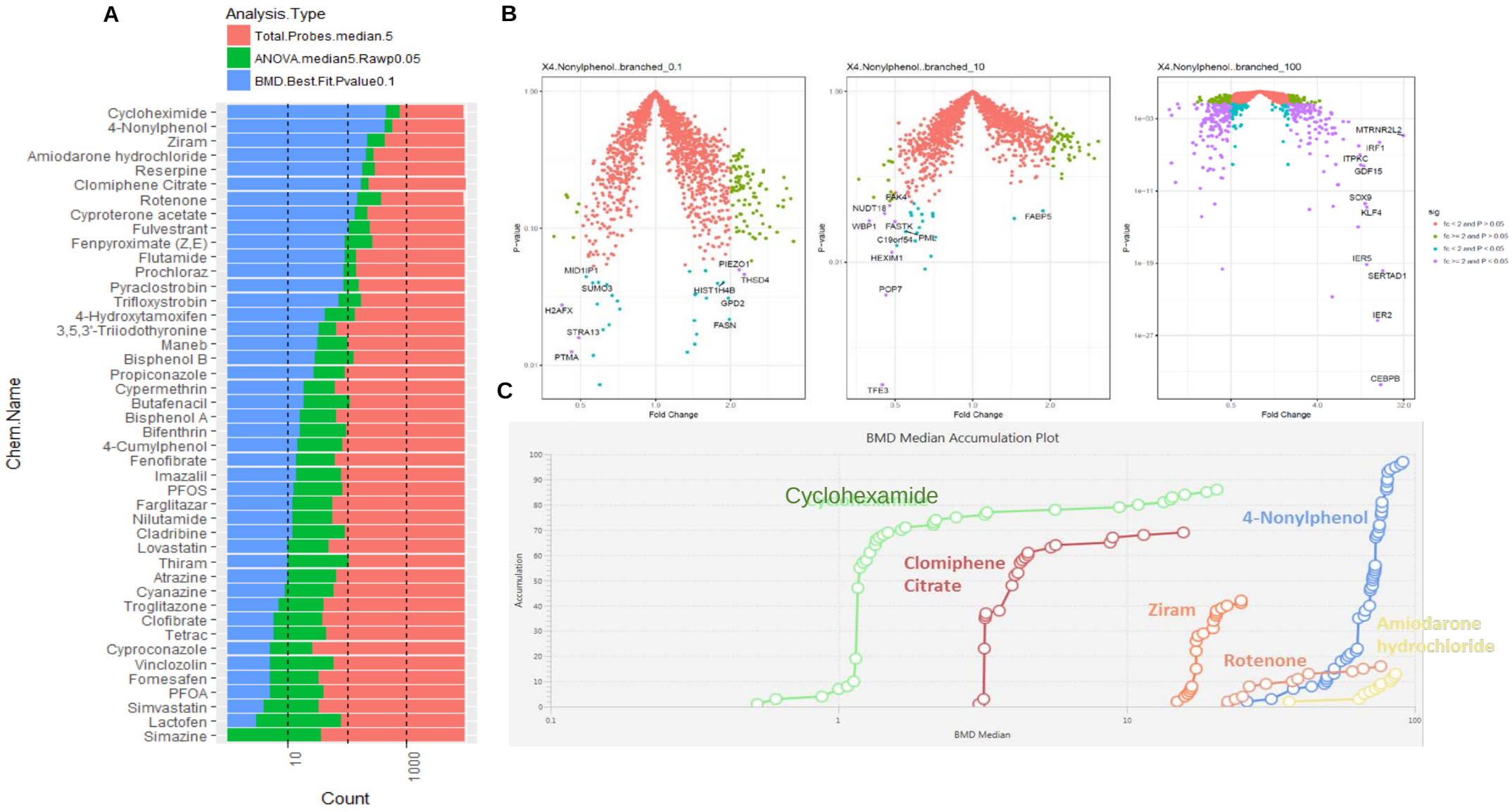


4-Cumylphenol
599-64-4



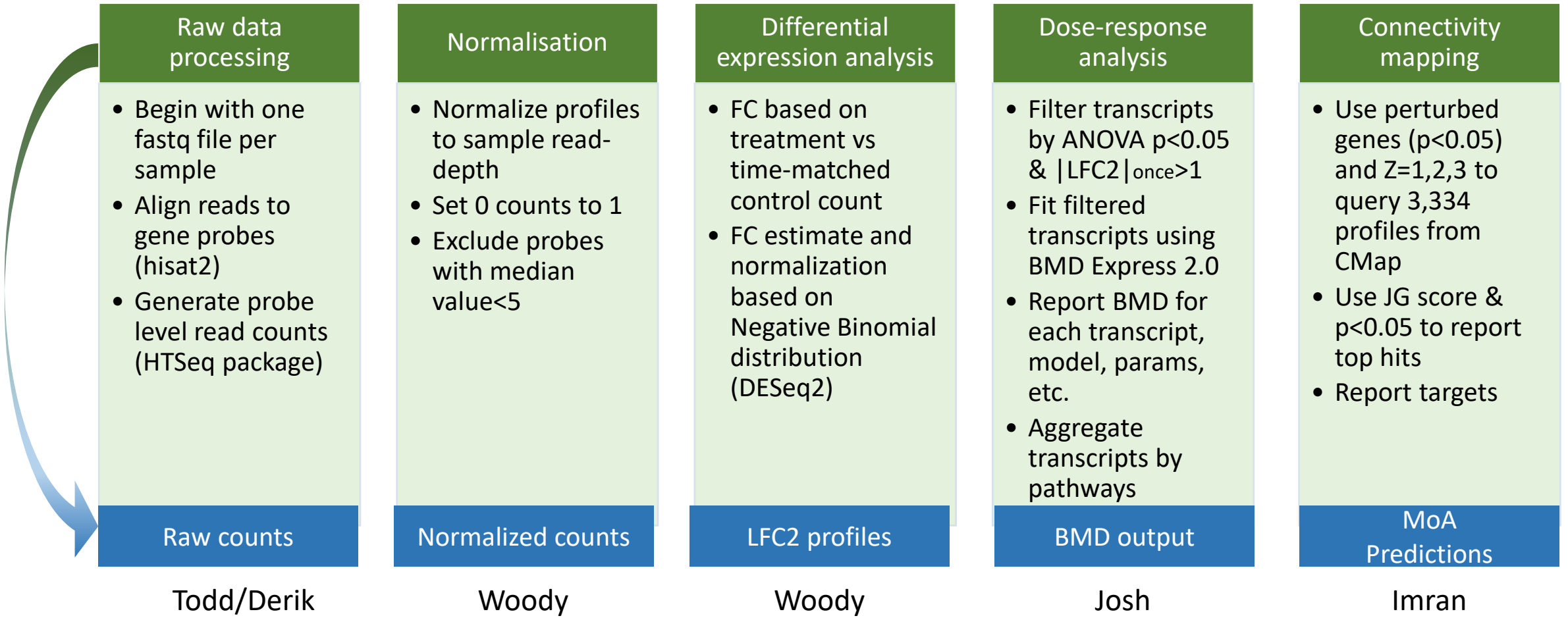
Cladribine
4291-63-8

Workflow Pilot - Preliminary Concentration-Response Modeling



- BMDExpress 2.0 to identify transcriptional point-of-departures (i.e. BMDs)

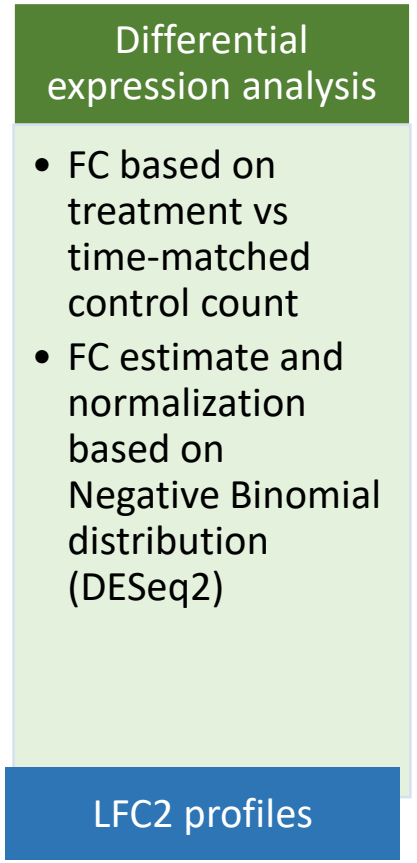
HTTr Analysis Pipeline Development



LFC2 = log2(fold change)

Analyzing count data presents some issues ..

- Due to …
  - Variance in total # of transcripts between samples
  - Heterogeneous transcript distributions
- Simple normalization schemes may call transcripts expressed the same level differentially expressed (DE)
- Use raw data to estimate significant differences in read counts
  - E.g. Negative Binomial (NB) distribution which is implemented in DESeq2 R package

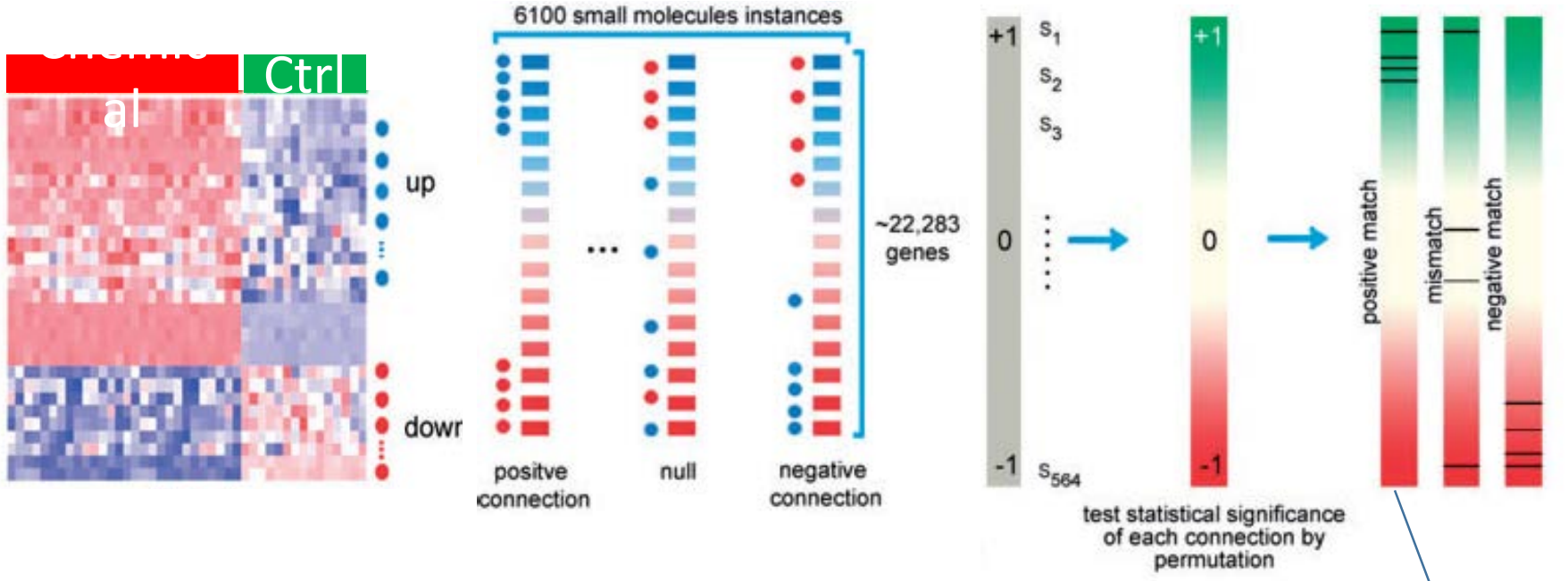


Connectivity Mapping

DE LFC2 profiles

Query CMap2.0 or LINCS

Find best positive matches



Connectivity mapping

- Use perturbed genes to query CMap (linked to targets)
- Identify molecular targets / MIE / MoA

Putative target / Selectivity

Infer

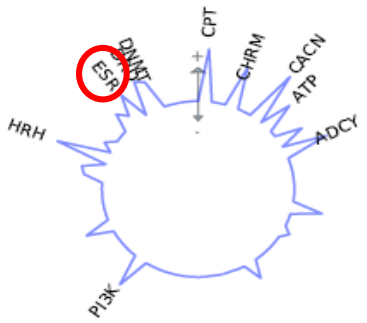
MIE/Target by best

Lamb *et al* (2006)
Musa *el al* (2017)

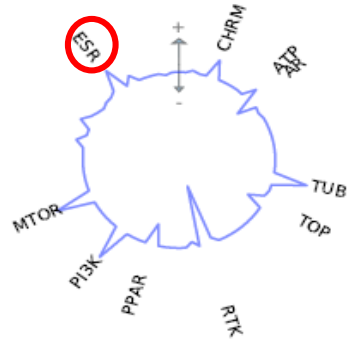
Z=1

Preliminary Analyses - Data Visualization – Spider Plots

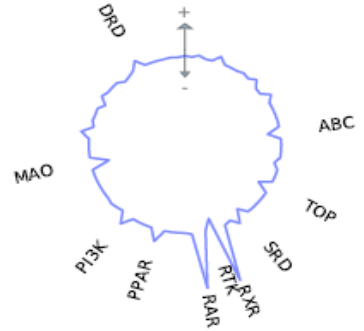
Genistein DMEM 06h



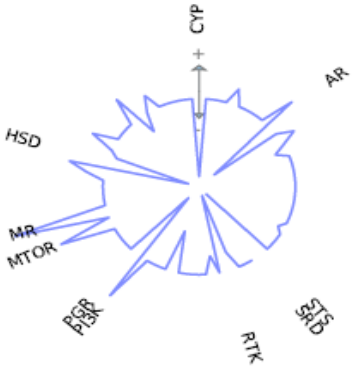
Genistein DMEM 12h



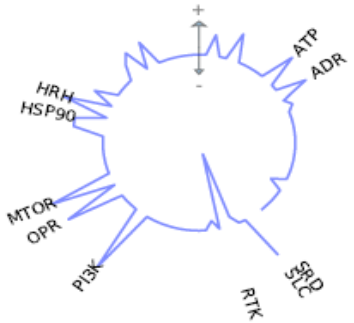
Genistein DMEM 24h



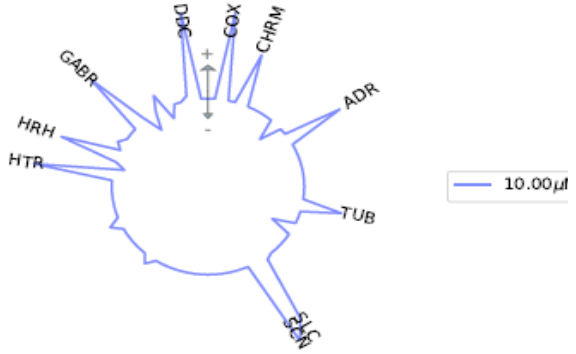
Genistein PRF-DMEM 06h



Genistein PRF-DMEM 12h



Genistein PRF-DMEM 24h



Ongoing HTTr Activities.

- Large Scale Chemical Screening
  - Block 1 (of 4) chemical testing blocks complete.
  - Anticipated completion date of Blocks 2 through 4 by Dec ‘17.

- Data analyses
  - HTTr data analysis pipeline refinement
  - Collaboration with NTP on a command line version of BMDExpress 2.0
  - tclp also being considered for concentration-response modeling

- Refinement / development of MIE prediction methodologies.
  - CMAP work ongoing
  - Development of unsupervised machine learning

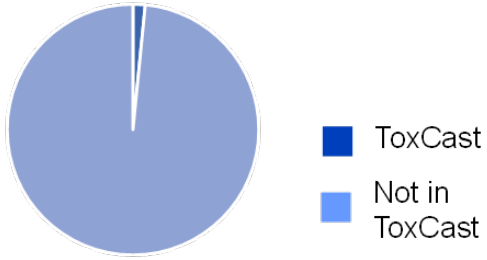
- IT Infrastructure
  - Investing in expanded storage (lots of terabytes of data)

HTTr Summary

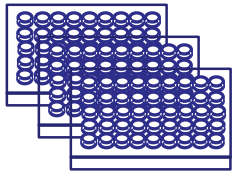
- HTTr will provide an unprecedented amount of bioactivity data on chemicals
- We have a pipeline to process, analyze, store and interpret this information
- QA/QC is being built into the system
- Just beginning to address MoA prediction

Addressing Limitations in *In Vitro* Test Systems

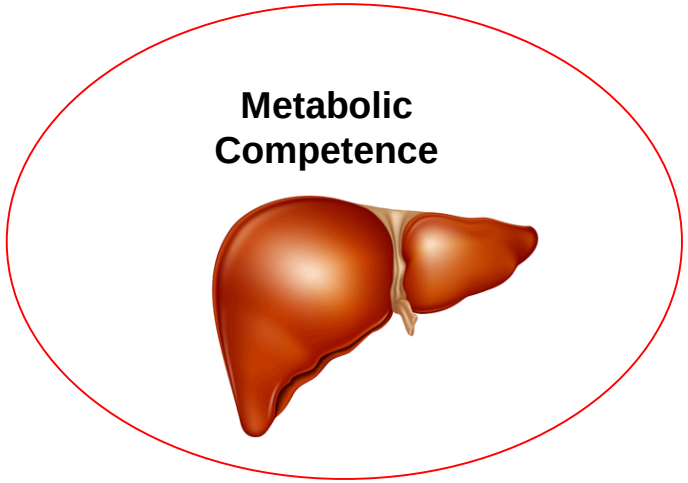
Biological Coverage (Gene Basis)



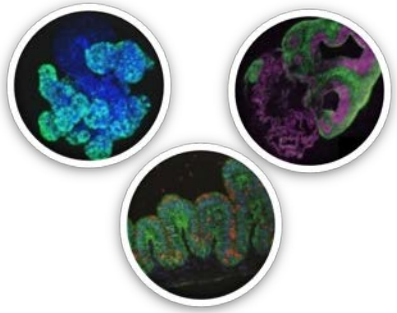
Chemical Coverage and Specific Chemical Types (e.g., VOCs)



Metabolic Competence



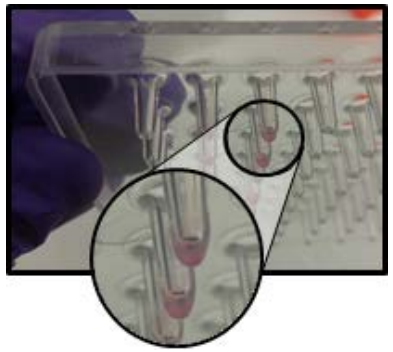
Organ and Tissue Responses

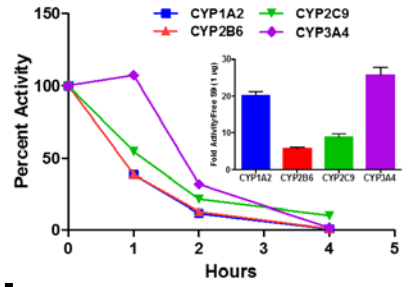


Developing Methods to Address Metabolic Competence – Two Approaches

"Extracellular" Approach

Chemicals metabolism in the media or buffer of cell-based and cell-free assays

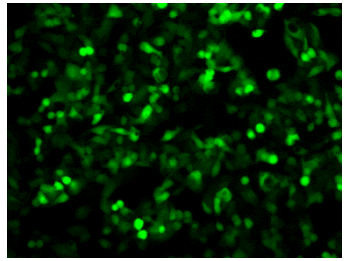


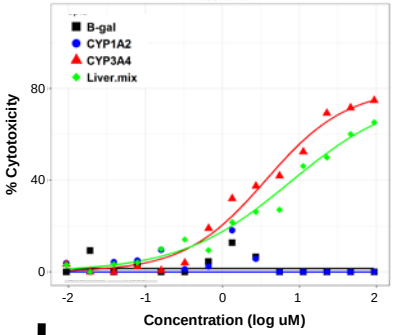


More closely models effects of hepatic metabolism and generation of circulating metabolites

"Intracellular" Approach

Capable of metabolizing chemicals inside the cell in cell-based assays





More closely models effects of target tissue metabolism

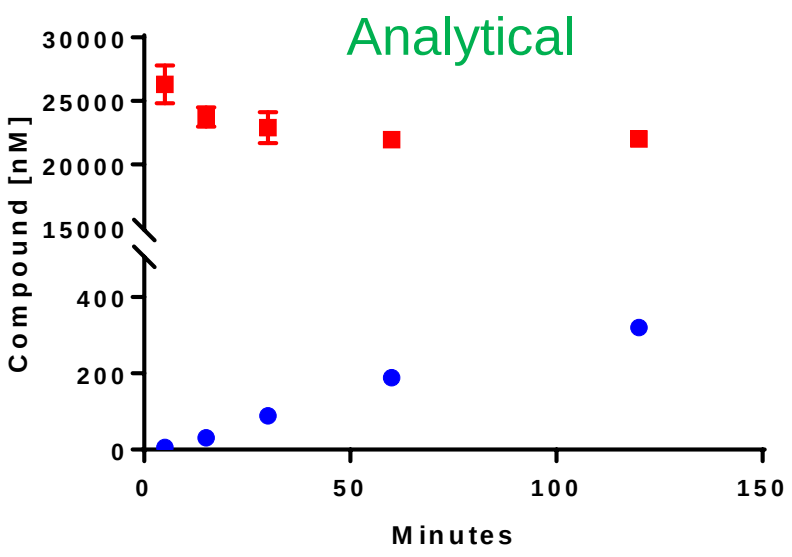
Integrated approach to model *in vivo* metabolic bioactivation and detoxification

Collaboration with Unilever

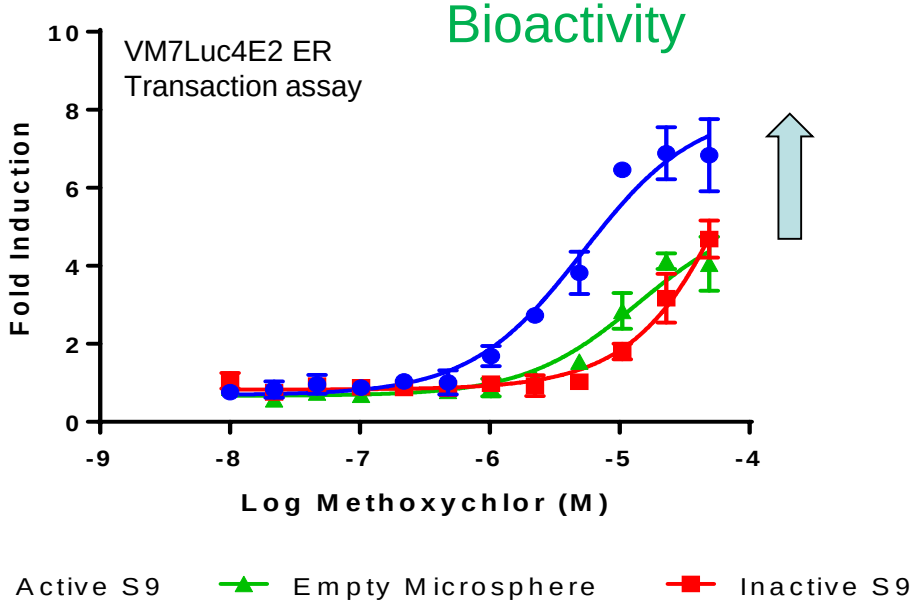
Extracellular Metabolism

Alginate Immobilization of Metabolic Enzymes (AIME)

Scalable to 96- and 384-well microplates



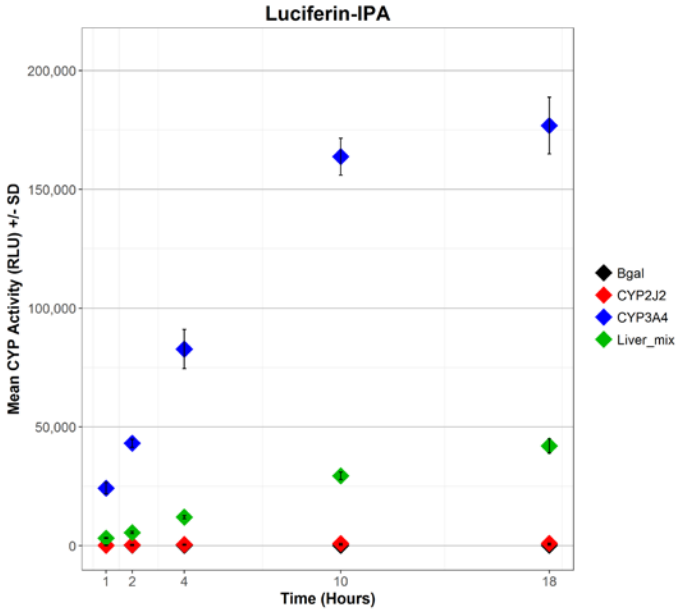
AIME forms predicted metabolites from cytochrome P450 probe substrates

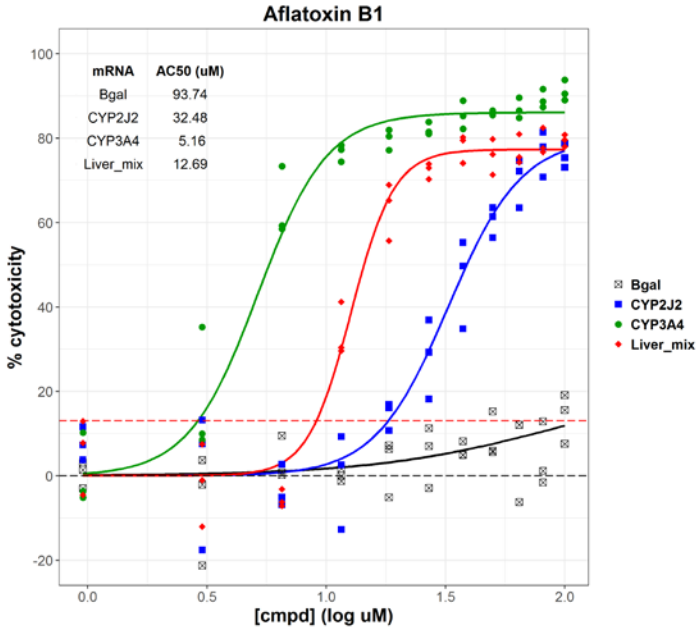


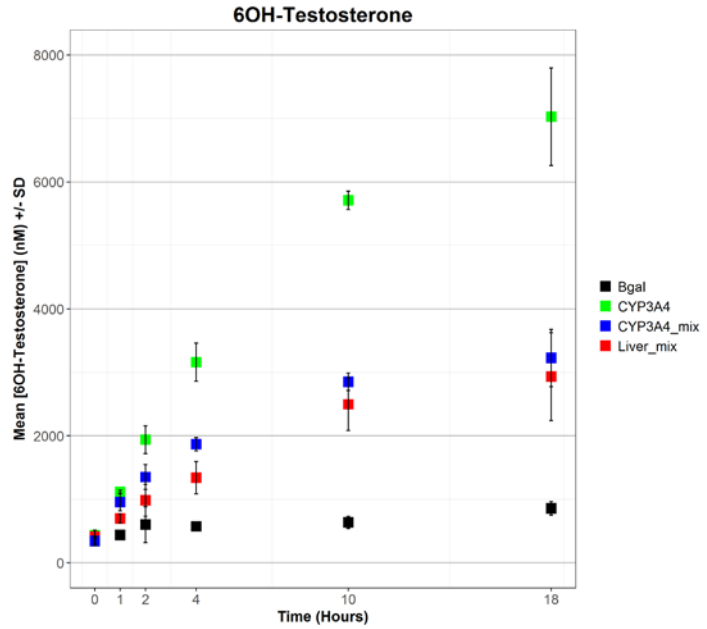
AIME produces the expected increase in methoxychlor ER activity with induced rat hepatic S9 (384-well format)

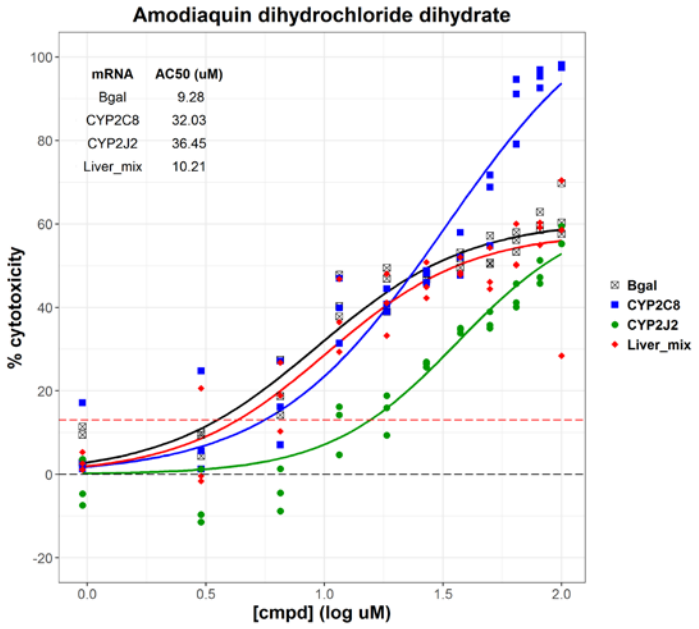
Intracellular Metabolism

Analytical



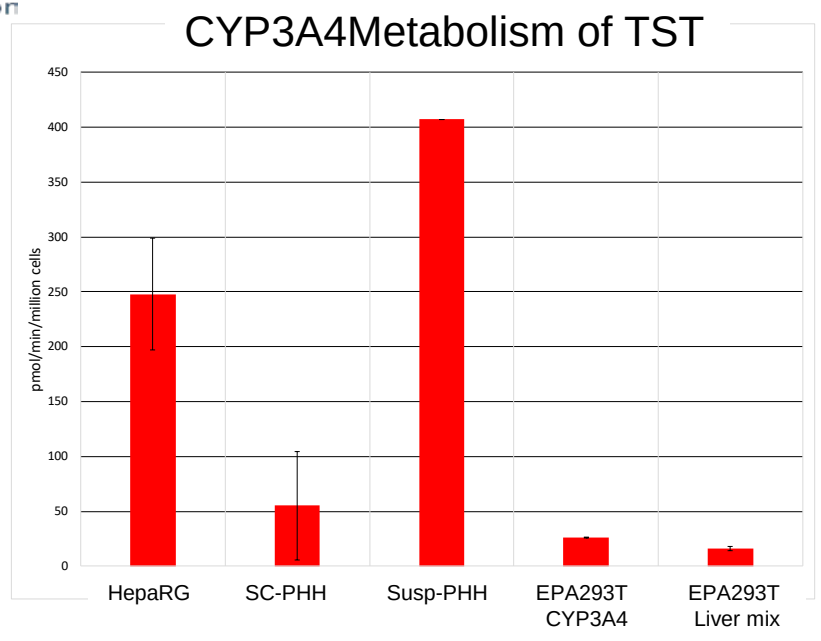


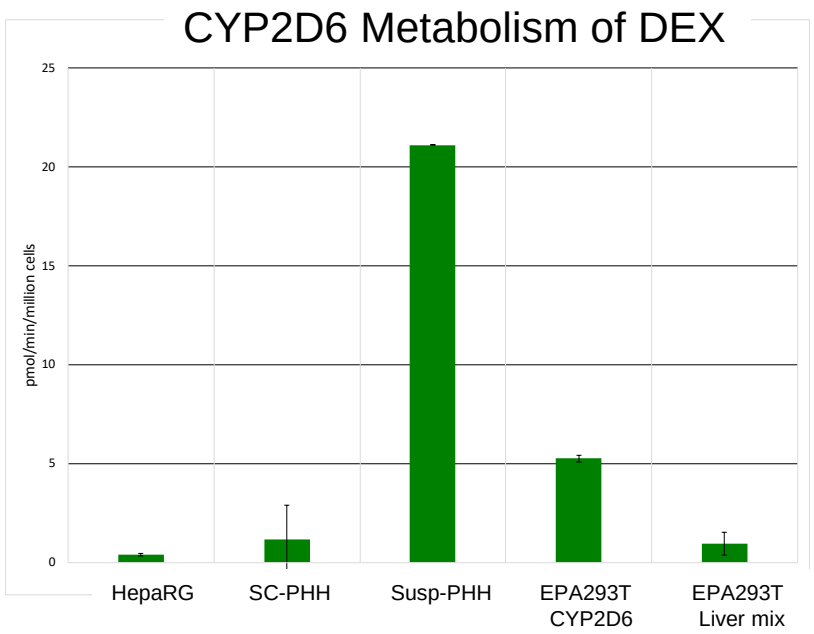


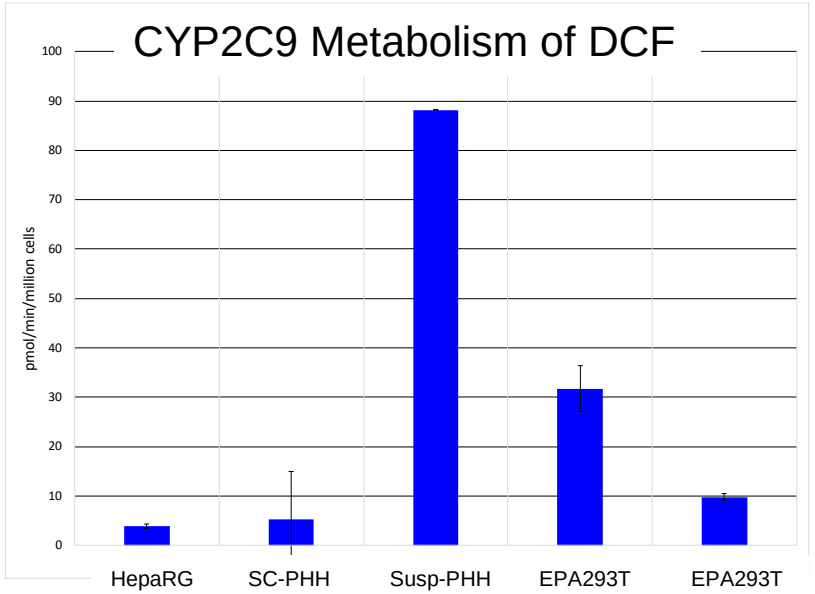


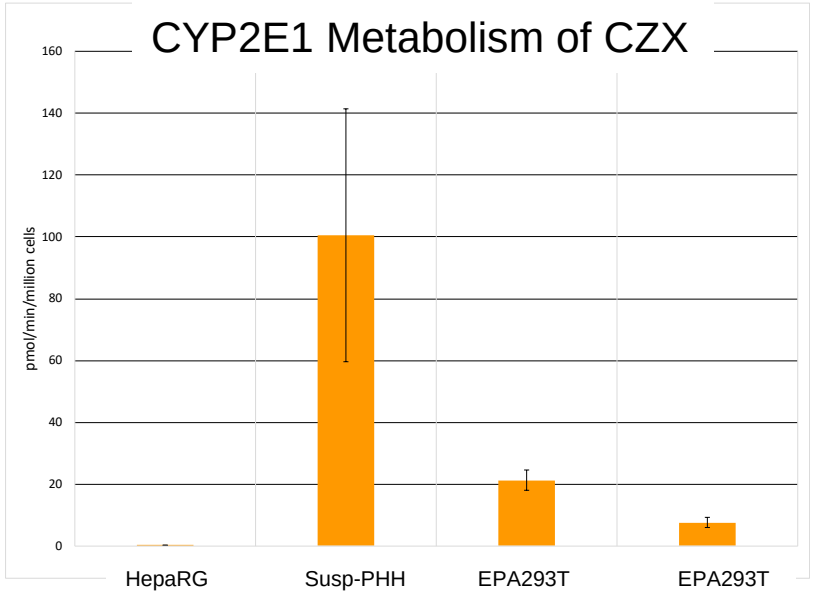
Bioactivity

Comparison to Other Models









Summary - Metabolic Competence

Extracellular AIME

- Established 96 and 384 well plate systems
- Works well with rat S9
  - Metabolite formation
  - Increased bioactivity in ER assay
- Large scale screening project to begin soon

Intracellular

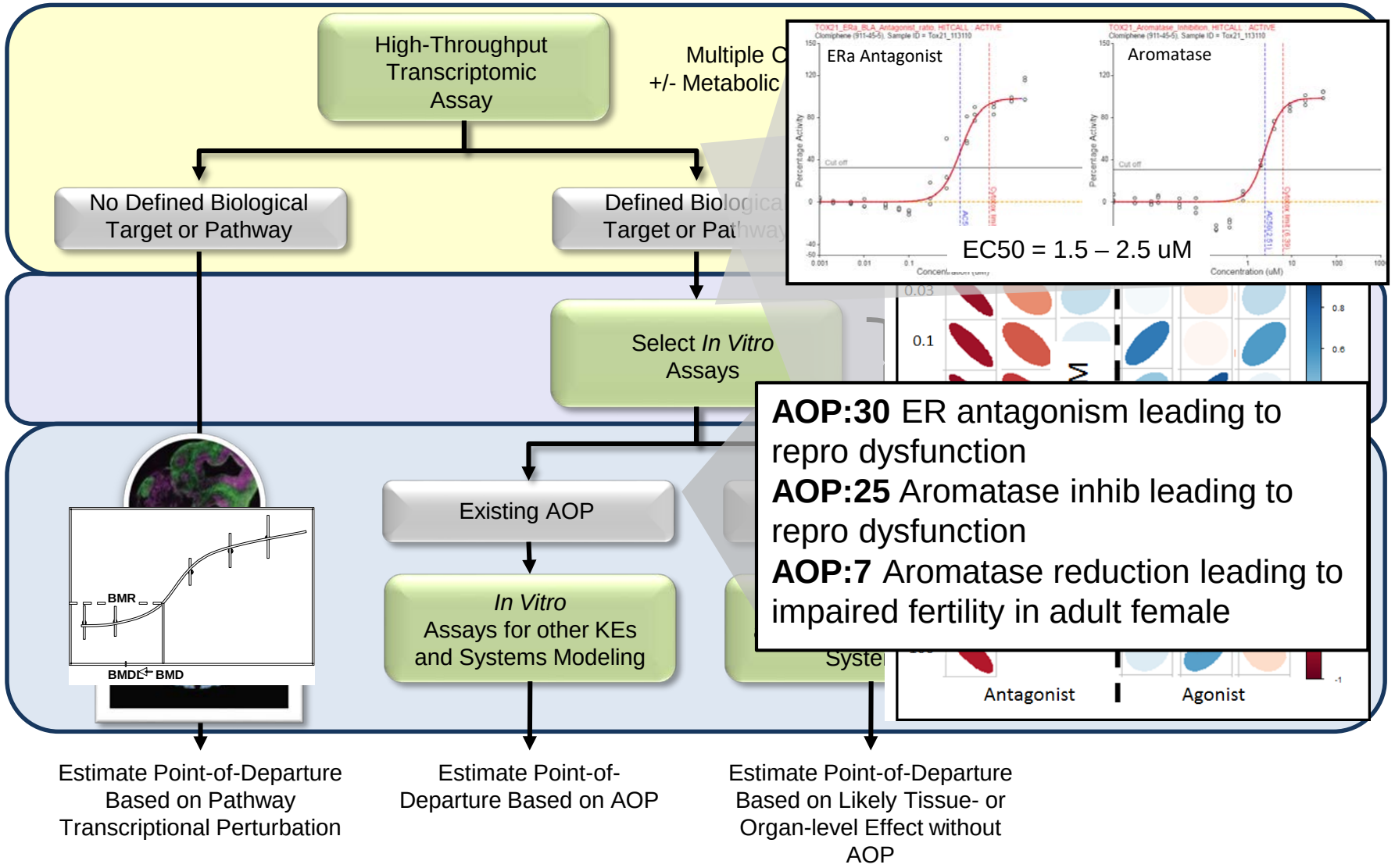
- Established 384 well plate system
- Using HEK293T cells transfected with human CYP mRNAs
  - Increased metabolism
  - Bioactivation and detoxification
- DeGroot et al (submitted)

EPA United States Environmental Protection Agency

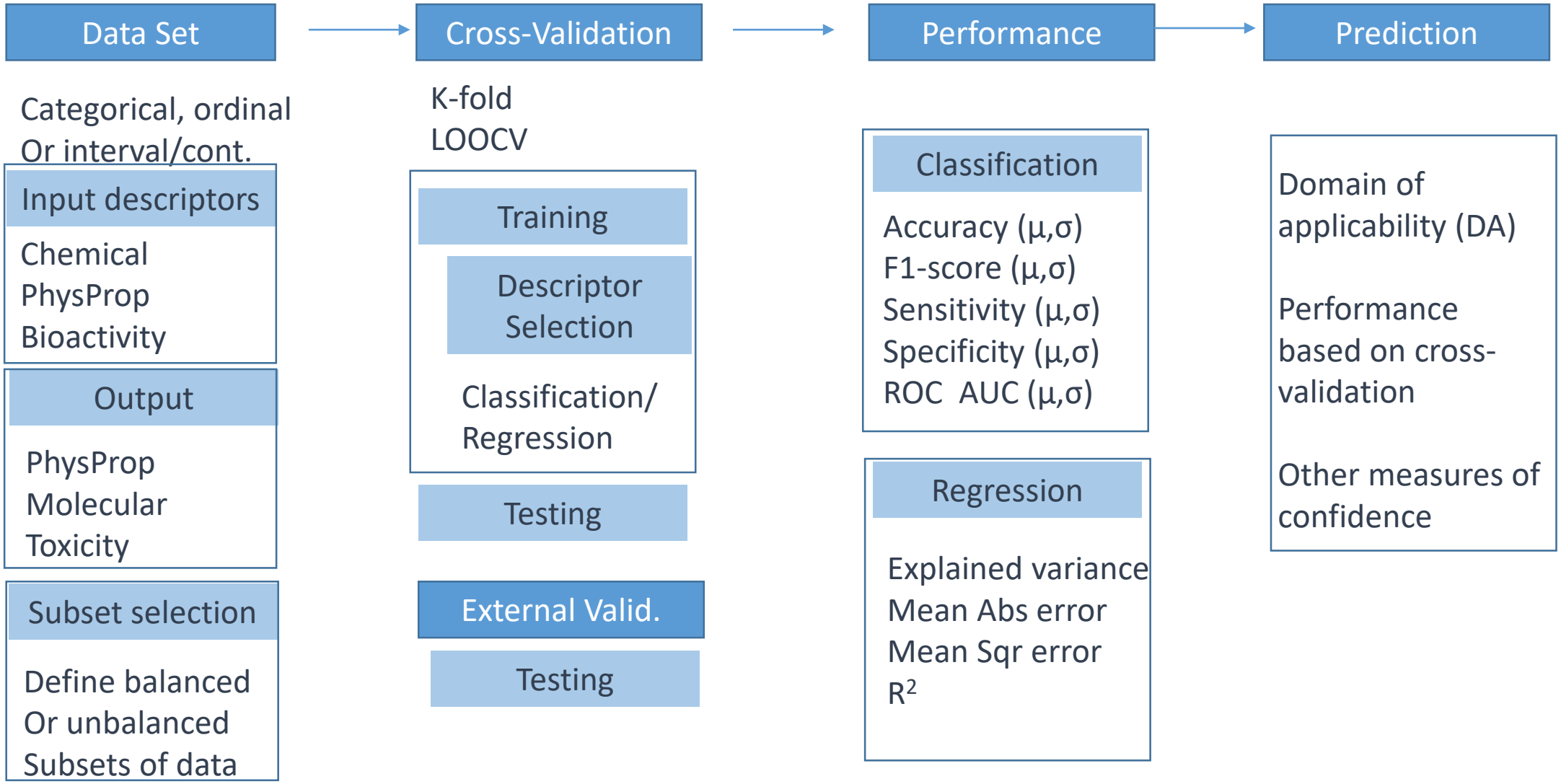
Questions?

EPA's National Center for Computational Toxicology

Integrating Components Into a Tiered Testing and Assessment Strategy



Machine Learning to Predict MoA



Transparency/QA/QC

01

Working with QA/QC team to document the complete HTTr workflow

02

The entire HTTr sequence of chemical - results is being documented:

- chemical -> biological sample: NCCT internal lab workflow steps (Josh)
- biological sample -> raw fastq (transcriptomics): sequencing steps (BSP)
- raw fastq -> analysis results: NCCT internal computational workflow steps (Imran)

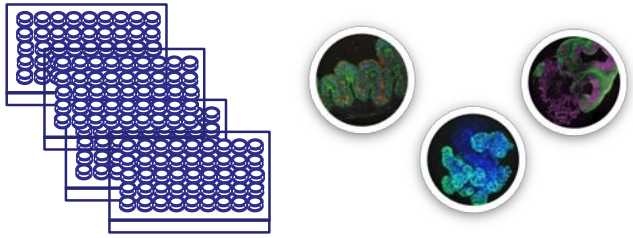
03

The workflow will be stored in database. Currently in spreadsheets and Jupyter notebook

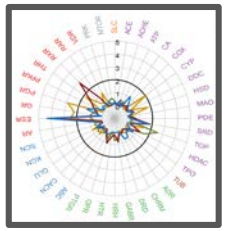
04

Generate complete “traceback” of all lab/comp steps for any result

Developing a Portfolio of TGX Experimental and Analytical Tools



High-Throughput Toxicogenomics Screen
- Multiple cell types
- Thousands of chemicals
- Whole transcriptome (EPA)
- S1500+ (NTP)



Karmaus, Unpublished

Mode of Action/MIE
- Refined CMAP tool
- Curating CMAP database for MIE and directional response
- >60 MIEs and growing

Dose Response Analysis
- BMDExpress
- Tcpl

- Large scale screen of 2,200 chemicals (ToxCast I/II) in single cell type this summer
- Additional screens across multiple cell types/lines
- Additional reference chemicals and genetic perturbations (RNAi/CRISPR/cDNA)

BMD Analysis

- BMDS is standard approach for concentration-response analysis
- BMD 2.0 Java GUI-based interactive wrapper for BMDS, currently maintained by Scott Auerbach@NIH
- Scott facilitating the implementation of BMD 2.0 command-line version (currently alpha)
  - BMD2 –input data –config bmd2.json –out bmd.json
- All model fits and BMD values will be exported for storage into httrdb
- Available Nov 2017 and expect integration into pipeline ~ Dec 2017