

# High Throughput Transcriptomics (HTTr) Concentration-Response Screening in MCF7 Cells

Joshua A. Harrill, Ph.D.



#### **Conflict of Interest Statement**



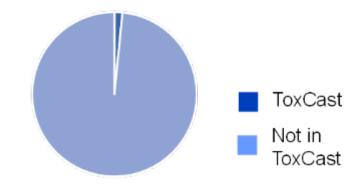
- No conflict of interest declared.
- Disclaimers:
  - The views expressed in this presentation are those of the author and do not necessarily reflect the view or policies of the USEPA.
  - This presentation does not necessarily reflect USEPA policy. Mention of trade names or commercial products does not constitute an endorsement or recommendation for use by USEPA.
  - Data in this presentation is the result of preliminary analyses.

#### **Outline**

- Background & Objectives
- HTTr Pilot Experiment
  - Optimization Steps
  - Attenuation
  - Experimental Layout
- Results
  - Assay Performance Metrics
  - Concentration-Response Modeling
- Current Activities & Future Directions

#### **Background**

#### Gene Coverage



#### Pathway Coverage\*



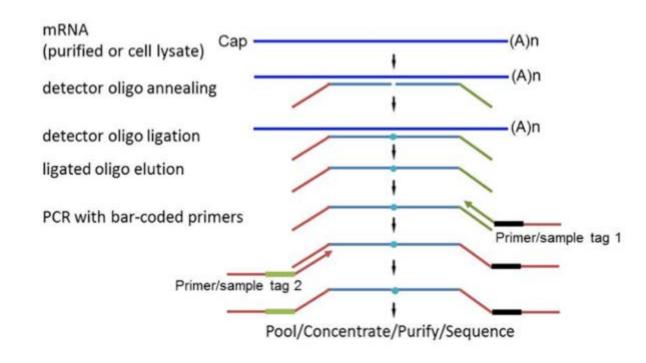
\*At least one gene from pathway represented

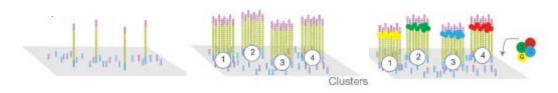
- ToxCast assays cover about 320 genes.
- Pathway coverage is higher but still leaves large gaps
- Recent technological advances in transcriptomics are very promising for rapid and cost-effective whole transcriptome screening.
- Increase biological coverage by using high throughput transcriptomics (HTTr) as broad-based Tier 0 bioactivity screen.

#### **BioSpyder TempO-Seq**

- Targeted RNA-Seq technology
- Whole transcriptome assay provides output on > 20,000 transcripts.
- Requires very low input (< 10 pg total RNA).
- Performed on "standard" PCR and Next Gen Sequencers.
- Compatible with purified RNA or cell lysates.







## **Objectives**

- Optimize culture and assay conditions for HTTr screening in MCF7 cells using the TempO-Seq human whole transcriptome assay.
- Perform a pilot experiment with a limited number of chemicals (n=44) in order to:
  - 1) Evaluate TempO-Seq assay performance.
  - Determine the ability of the TempO-Seq assay to detect known biological signatures following chemical perbations
  - 3) Guide experimental design of larger screening studies.

#### **HTTr Pilot: Experimental Design**

Parameter	Multiplier	Notes
Cell Type(s)	1	MCF7
Culture Condition	2	DMEM + 10% HI-FBS PRF-DMEM + 10% CS-HI-FBS
Chemicals	44	see subsequent slides
Time Points:	3	6, 12, 24 hours
Assay Formats:	3	TempO-Seq HCI-Apoptosis HCI-Cytotoxicity
Concentrations:	8	3.5 log <sub>10</sub> units; ½ log <sub>10</sub> spacing
Biological Replicates:	4	3 TempO-Seq; 1 Reserve

<sup>&</sup>lt;sup>a</sup> MCF7 cells cultured in DMEM + 10% HI-FBS was selected as the test system to facilitate comparability to the Broad Institute Connectivity Map (CMAP) database (<a href="http://portals.broadinstitute.org/cmap/">http://portals.broadinstitute.org/cmap/</a>).

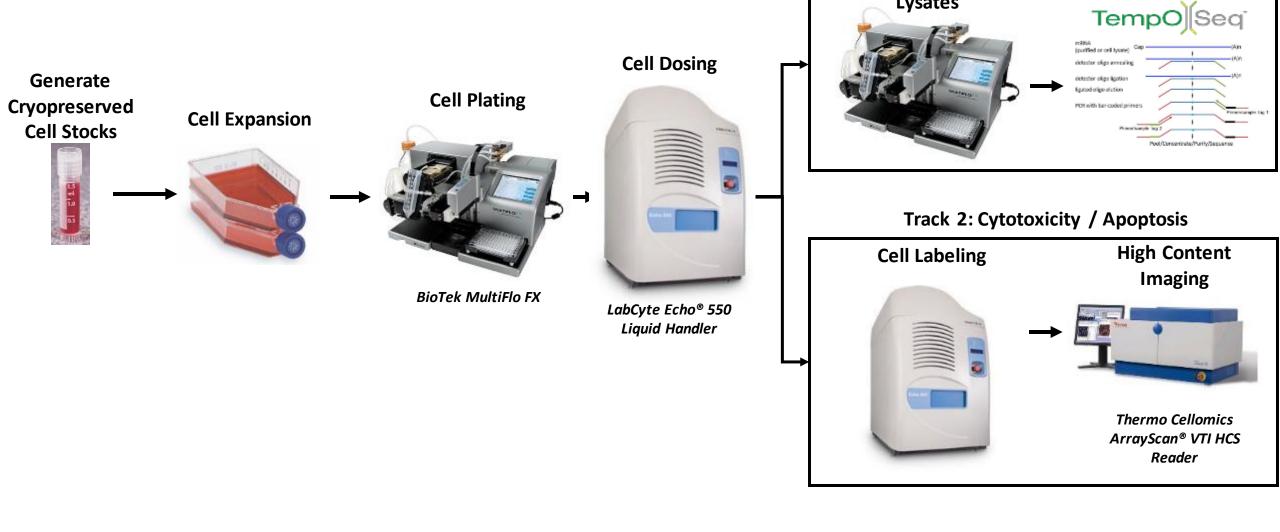
#### **HTTr Pilot: Workflow**

Track 1: Targeted RNA-Seq

TempO-Seq WT

**Generating Cell** 

Lysates



# **Assay Optimization**

#### MCF7 Cell Culture

- Authentication
- Expansion Protocol
- Media Formulation
- Seeding Density

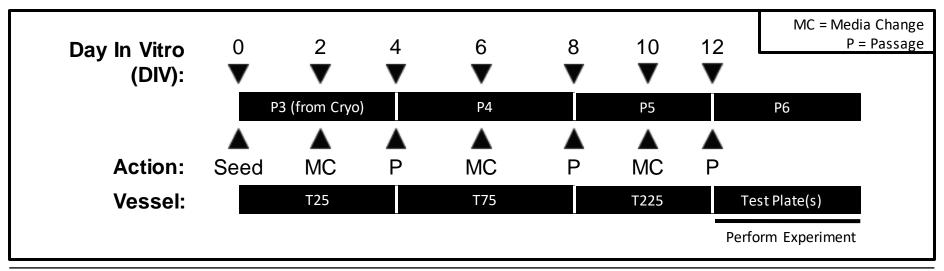
#### TempO-Seq Assay

- Lysis Conditions
- Attenuation of Highly Expressed Genes

#### Chemical Treatments

- Concentration Range
- Plate Map Design
- Exposure Duration

# **MCF7 Expansion Protocol**



Stage	Culture Vessel	Average Cell Yield <sup>a</sup>	Number of Treatment Wells <sup>b</sup>	Number of Test Plates <sup>c</sup>
Initial Seeding	NA	1.28x10 <sup>7</sup>	182	0.47
P (3→4)	T25	$2.43x10^7$	346	0.90
P (4→5)	T75	5.86x10 <sup>7</sup>	837	2.18
P(5 <del>→</del> 6)	T225	1.47x10 <sup>8</sup>	2100	5.47

<sup>&</sup>lt;sup>a</sup> Median values from c2017-08-14, c2017-08-15, c2017-08-19, c2017-08-20

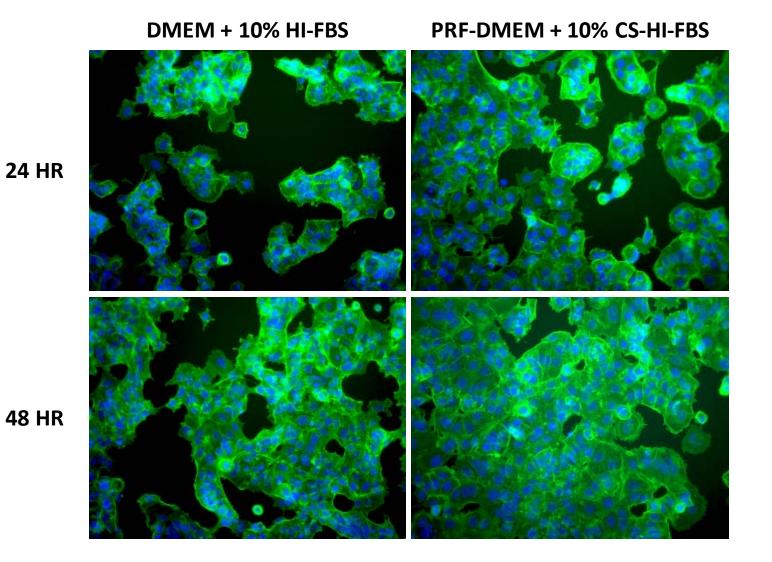
MCF7 Cells authenticated by STR Profiling and karyotyping prior to use in screening studies.

<sup>&</sup>lt;sup>b</sup> Assumes 384 well plate, 10,000 cells / well.

<sup>&</sup>lt;sup>c</sup> For experimental needs > 5 plates / experiment, expand multiple cryopreserved MCF7 cell aliquots in parallel. Pool at each passaging stage.

#### **Media Effects on MCF7 Growth**

- DMEM + 10% HI-FBS contains phenol red and an unknown compliment of serum factors which may stimulate ER activation.
- Phenol red-free media with charcoal-stripped FBS reduces endogenous estrogen receptor activation.

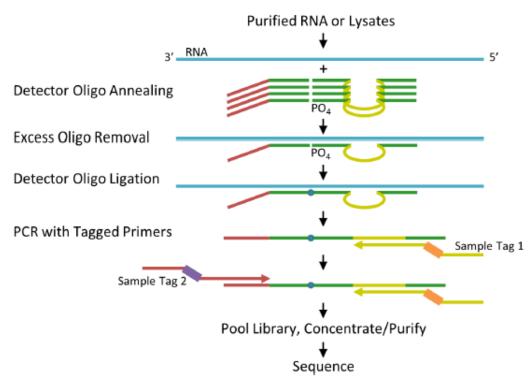


#### **Qualitative Observations**

- More cell attachment and cell spreading with PRF-DMEM + 10% CS-HI-FBS.
- Greater increase in cell confluency over time in PRF-DMEM + 10% CS-HI-FBS.
- More proliferation over time in DMEM + 10% HI-FBS.

#### **Attenuation**

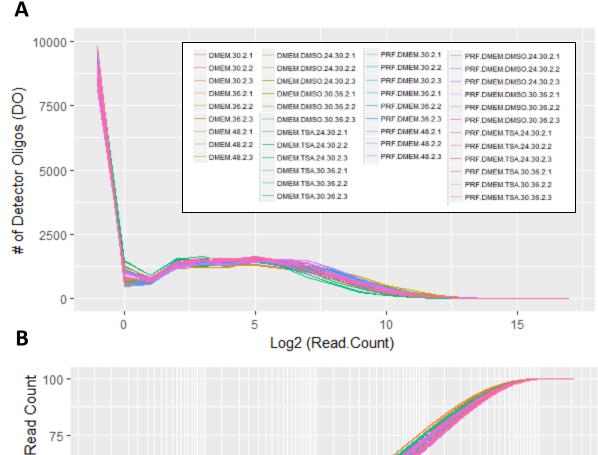
- A method used with BioSpyder TempO-Seq assay to prevent highly expressed genes from occupying a disproportionate amount of available read space and increase the ability to quantify low abundance transcripts.
- Attenuation is accomplished by adding "cold probes" which compete with matching DOs for hybridization sites on target RNAs.
- The attenuation probe will bind to the same site as the detector oligos, thus decreasing the amount of the target RNA species available for PCR amplification.
- For attenuation, the end user must define:
  - The set of genes to be attenuated, and...
  - What degree of attenuation is appropriate
- Question(s):
  - Is additional attenuation needed in the MCF7 cell model?
  - If so, how is the attenuation set defined?



**Fig 1. TempO-Seq biochemical scheme.** RNAs are targeted by annealing to DOs that contain target-specific sequences (green) as well as primer landing sites (red and yellow) that are shared across all DOs. Excess oligos are removed by a 3' exonuclease, then the hybridized oligos are ligated and amplified using primers that contain sample tag (index) sequences (orange and purple bars), and adaptors required for sequencing. The amplified assay products are pooled for a library, purified/concentrated and sequenced.

https://doi.org/10.1371/journal.pone.0178302.g001

#### **Distribution of Read Counts**



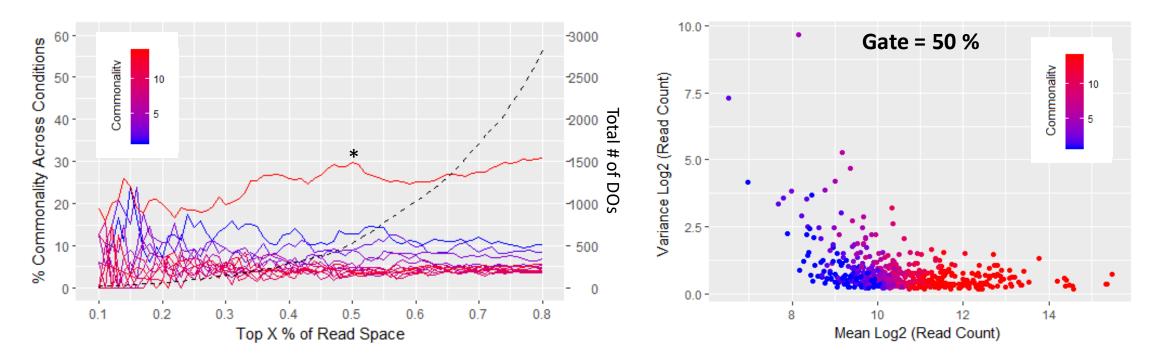
В		Ó	5 Log2 (Rea	10 ad.Count)	15
Cumulative % of Total Read Count	100 - 75 -				
% of Total	50 ~				
Sumulative	25 -				
	0 -	1 ' ' '10		1000 Detector Oligos	10000

	Table 1. Number of DOs Accounting for 50% of Total Read Space, Per Sample Basis													
	Media	Treatment	Treatment	Sample Time,	Replicate Number									
	Туре	Туре	Time, h	h	1	2	3							
	DMEM			30	242	246	186							
rse	DMEM			36	273	220	208							
Course	DMEM			48	238	249	239							
Je (	PRF.DMEM			30	276	288	289							
Time	PRF.DMEM			36	268	248	244							
	PRF.DMEM			48	240	240	262							
7	DMEM	DMSO	24	30	308	259	269							
C.Resp.1	DMEM	TSA, 1 μM	24	30	231	248	253							
.Re	PRF.DMEM	DMSO	24	30	307	303	322							
O	PRF.DMEM	TSA, 1 μM	24	30	273	278	303							
2	DMEM	DMSO	30	36	242	233	249							
C.Resp.2	DMEM	TSA, 1 μM	30	36	192	222	208							
.Re	PRF.DMEM	DMSO	30	36	245	242	232							
O	PRF.DMEM	TSA, 1 μM	30	36	220	273	263							
			Range of D	O Counts:	1	186 - 322								

#### **Results**

- Read count distributions similar across samples.
- Broad range of read counts within each sample (0 ~32K).
- Within each sample, ~50-60% of DOs with non-zero read counts.
- Between 186 322 DOs account for 50% of the available read space (varies with sample).

#### **Evaluating Commonality of Highly Expressed Genes Across Test Conditions**



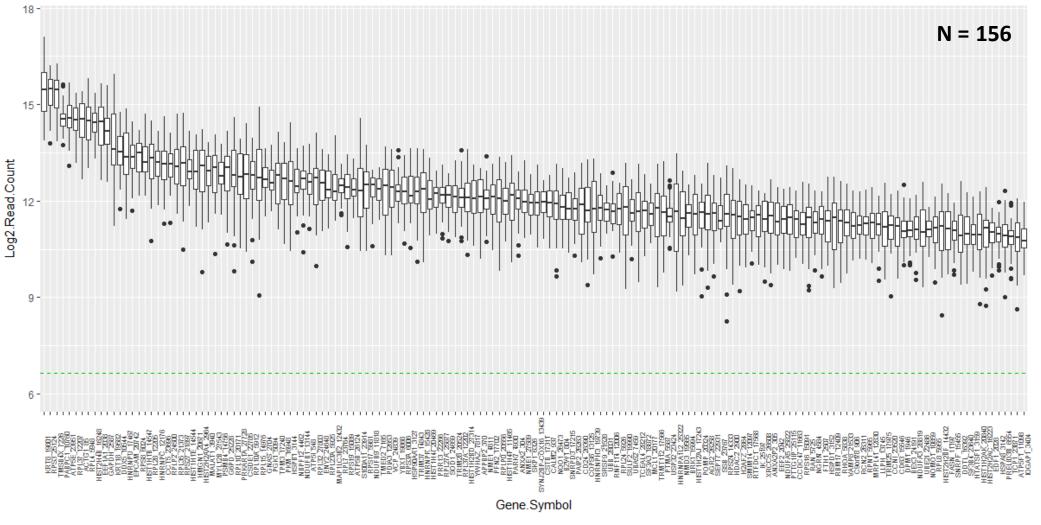
#### Using a Gate of 50 % of the total read space (\*):

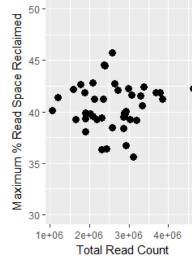
- Commonality Score = 14: ~ 30% of the DOs are identified as "highly-expressed" in all 14 test conditions (red).
- Commonality Score = 1: ~12.5% are identified as "highly-expressed" in only 1 test condition (blue).
- Commonality Score = 2 13: Varying number of DOs (< 10%) identified as "highly-expressed" in 2 to 13 test conditions.
- Variance: Tended to increase in DOs with lower commonality scores.

#### **Conclusions:**

- At Gate = 50 %, DOs with Commonality Scores of 14 are consistently identified as "highly-expressed" across all test conditions and have relatively lower variance and higher read counts across all test conditions.
- N = 156 DOs identified as candidates for attenuation.

# **Candidate "Highly Expressed Genes" for Attenuation**





- Rank ordered on x-axis by average read count across all test conditions.
- Green line → Raw read count = 100.
- The most highly expressed genes in the attenuation set are "housekeeping" genes.

#### **HTTr Pilot: Chemical Test Set**

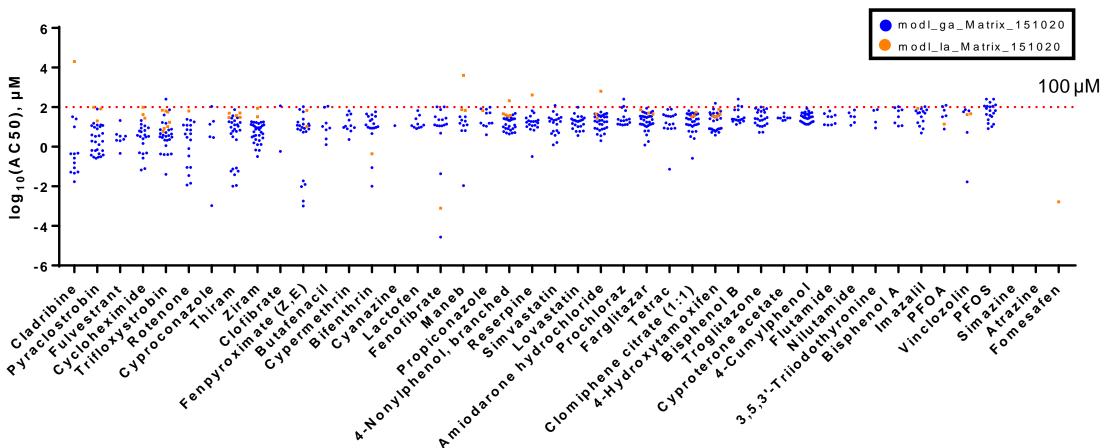
Chemical Name	MIE Family	Chemical Name	MIE Family
Flutamide		Rotenone	MITOCHONDRIA
Nilutamide	ANTIANDROOFN	Fenpyroximate (Z,E)	(COMPLEXI)
Cyproterone acetate	ANTIANDROGEN	Trifloxystrobin	MITOCHONDRIA
Vinclozolin		Pyraclostrobin	(COMPLEXII)
4-Hydroxytamoxifen		PFOS	
Clomiphene citrate (1:1)	ANTIESTROGEN	PFOA	PPAR
Fulvestrant		Troglitazone	FFAIX
Atrazine	cAMP INDUCERS/	Farglitazar	
Cyanazine	PDE INHIBITORS	Lactofen	PPO INHIBITOR / PPAR
Simazine	1 DE INTIBITORO		TTO INTIBITOR/TT/AIC
Cladribine	CYTOTOXICANTS	Fomesafen	PPO INHIBITOR
Cycloheximide	01101070070110	Butafenacil	1101111111
Bisphenol A		Maneb	
Bisphenol B	ESTROGENS	Thiram	SH REACTIVE
4-Nonylphenol, branched	ECTROCEINO	Ziram	
4-Cumylphenol		Imazalil	
Clofibrate	FIBRATES	Prochloraz	STEROIDOGENESIS
Fenofibrate	TIBIOTIES	Cyproconazole	OTENOIDOCENEDIO
Lovastatin	LIMOCD	Propiconazole	
Simvastatin	HMGCR	Tetrac	THR
Bifenthrin		3,5,3'-Triiodothyronine	1111
Cynormothrin	NA+ CHANNEL	Reserpine	VMAT
Cypermethrin		Amiodarone hydrochloride	V IVIA I

• Chemical set covers broad range of mechanistic diversity with redundancy within mechanistic class.

#### **Dose Range Selection**

#### Cytotoxicity-Related Assays

Judson et al. (2016)
\*\*Data from INVITRODB\_V2\_SUMMARY\*\*

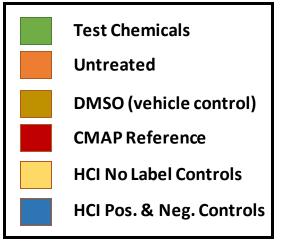


- Upper bound in testing range set at 100 μM based on upper limit of cytotoxicity range for most chemicals.
- Final dose range: 0.03, 0.1, 0.3, 1, 3, 10, 30, 100 μM

#### **Dosing Plate Layout**

										DOSI	NG PI	LATE N	/IAP												
	_	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
1	Α	Ionomycin (30 μM)	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	non-treated
2	В	Ionomycin (30 μM)	30	30	30	30	30	30	30	30	30	30	30	30	30	30	30	30	30	30	30	30	30	30	non-treated
3	С	Ionomycin (30 μM)	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	non-treated
4	D	Staurosporine (1 μM)	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	DMSO
5	E	Staurosporine (1 μM)	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	DMSO
6	F	Staurosporine (1 μM)	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
7	G	Saccharin (100 μM)	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]
8	н	Saccharin (100 μM)	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)
9	1	Saccharin (100 μM)	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)
10	J	Sorbitol (100 μM)	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)
11	К	Sorbitol (100 μM)	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)
12	L	Sorbitol (100 μM)	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)
13	М	Ionomycin (30 μM) [No Label]	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)
14	N	Staurosporine (1 µM) [No Label]	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)
15	0	Saccharin (100 μM) [No Label]	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)
16	Р	Sorbitol (100 μM) [No Label]	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)

- 44 chemicals in 8-point concentration-response → all on one plate
- Non-treated (n=3) and DMSO (n=3) control wells.
- Three "CMAP" Reference Compounds, single point, in triplicate
- First column reserved for addition of RNA QC samples by NCCT (pre-shipment) and BioSpyder (post-shipment).



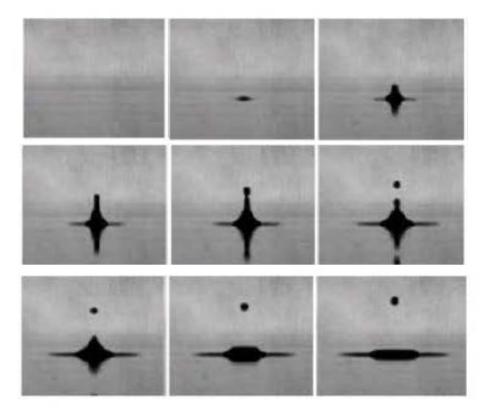
#### **Dose Randomization using Echo 550**

#### **Acoustic dispensing technology:**

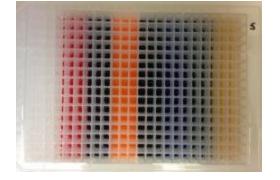
- Uses soundwaves to precisely transfer small quantities of liquid (nL) from source plate to test plate.
- Allows for randomization of test wells → mitigate potential edge effects without "losing real estate."



LabCyte Echo® 550 Liquid Handler



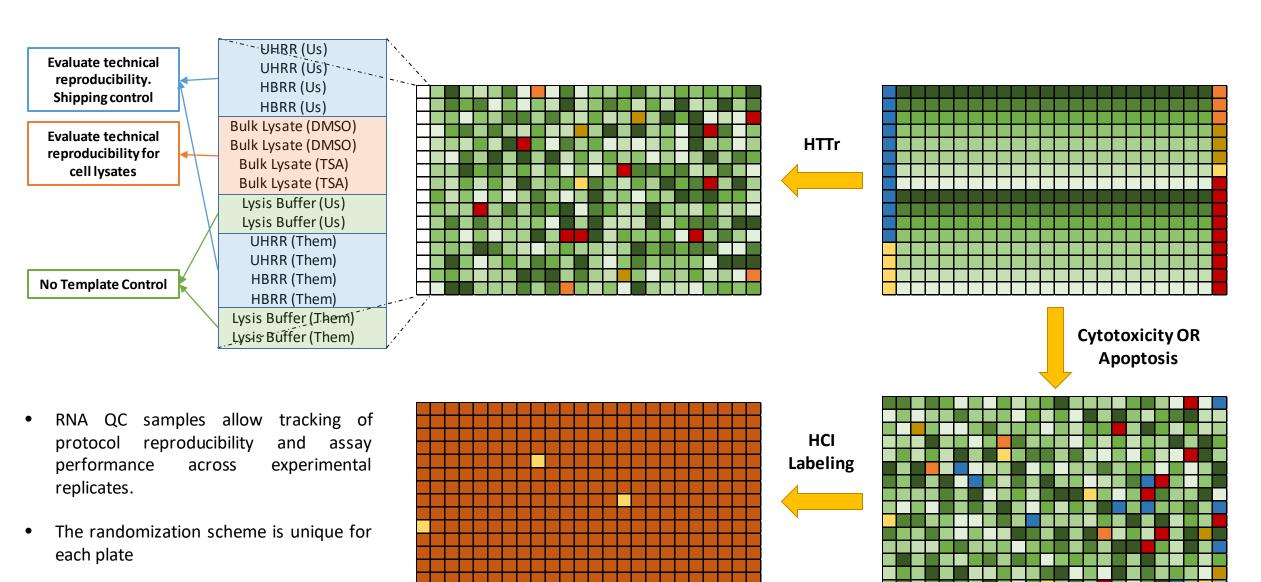
#### **Source Plate**



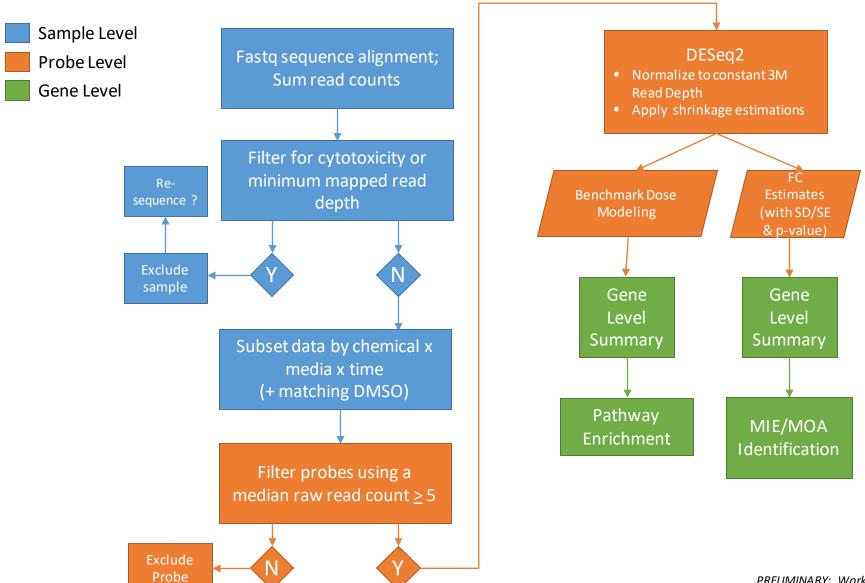
**Test Plate** 



#### **Echo Dispensing**



#### **DRAFT Data Analysis Pipeline**



# **Assay Performance Metrics**

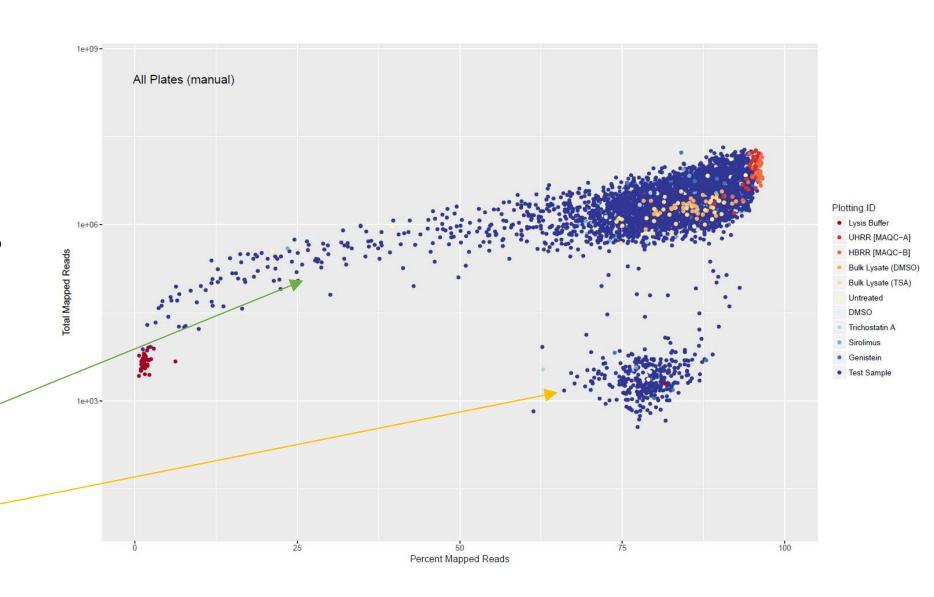
• Total Mapped Reads vs. Percent Mapped Reads

Correlation and Variation in Technical Replicates [within plate]

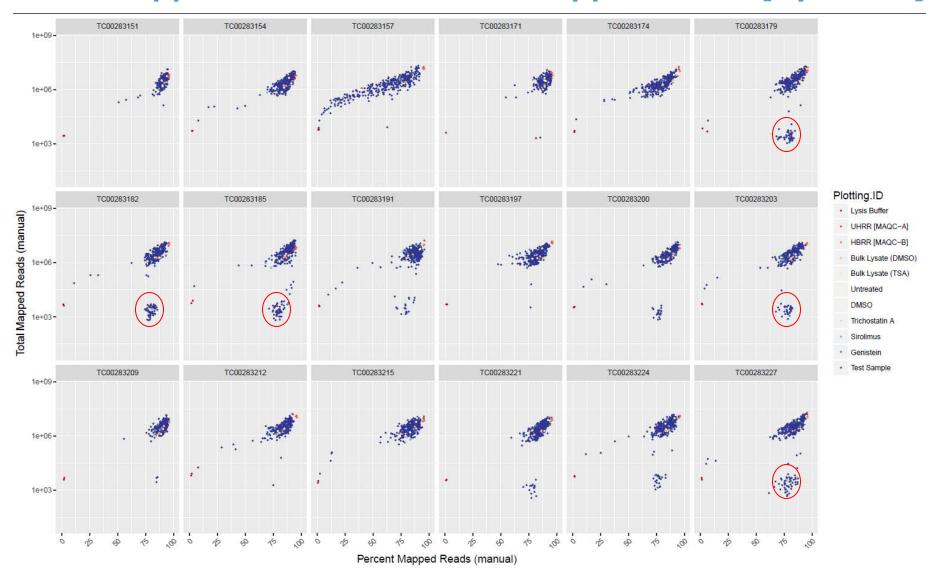
- Correlation and Variation in Biological Replicates [across plates]
- Detection of Biological Signal
  - Transcriptional Biomarkers
  - Connectivity Mapping

#### **Total Mapped Reads vs. Percent Mapped Reads [All Plates]**

- Average total mapped reads of test samples ~ 3.0x10<sup>6</sup>
- Average mapped read count per gene ~150
- Percent mapped reads > 75%
- Lysis Buffer blanks have low total reads, but not zero.
- Purified RNAs clustered at upper left.
- Comet tail?
- Off-set cluster?

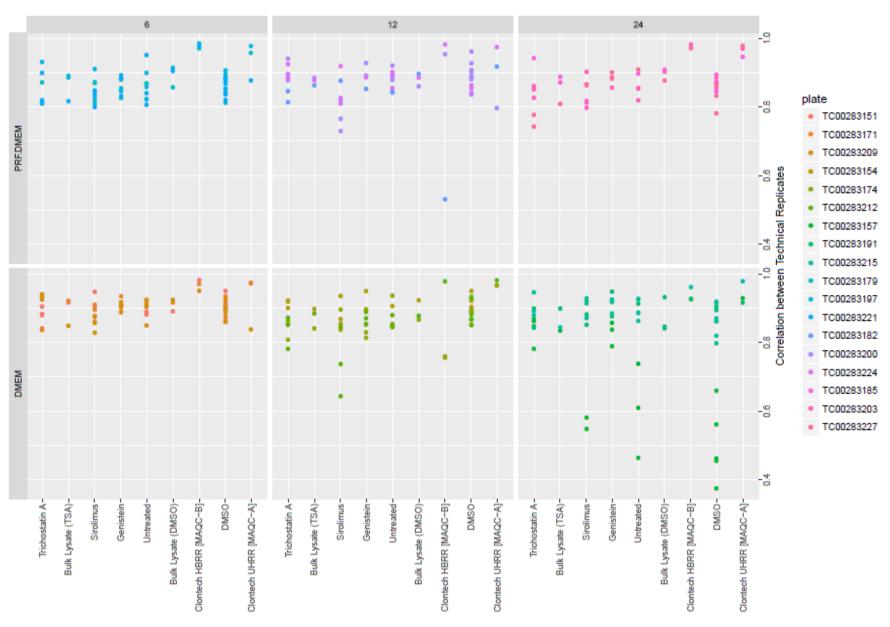


#### **Total Mapped Reads vs. Percent Mapped Reads [By Plates]**



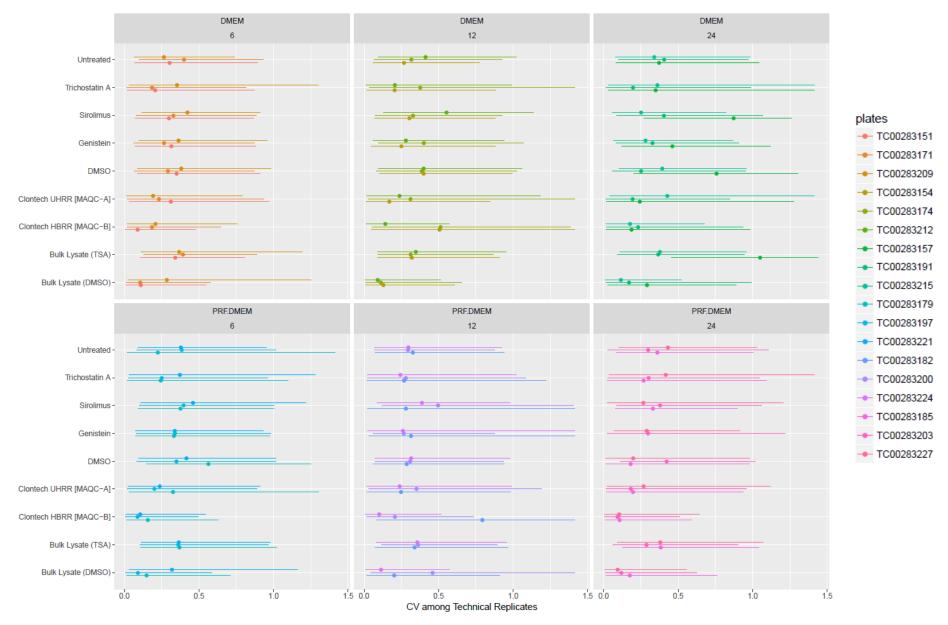
- Comet tail → Due to one "poor performing" plate
- Offset cluster → Low read count samples across many plates (red circles) → Candidates for resequencing.

# **Correlation Among Technical Replicates**



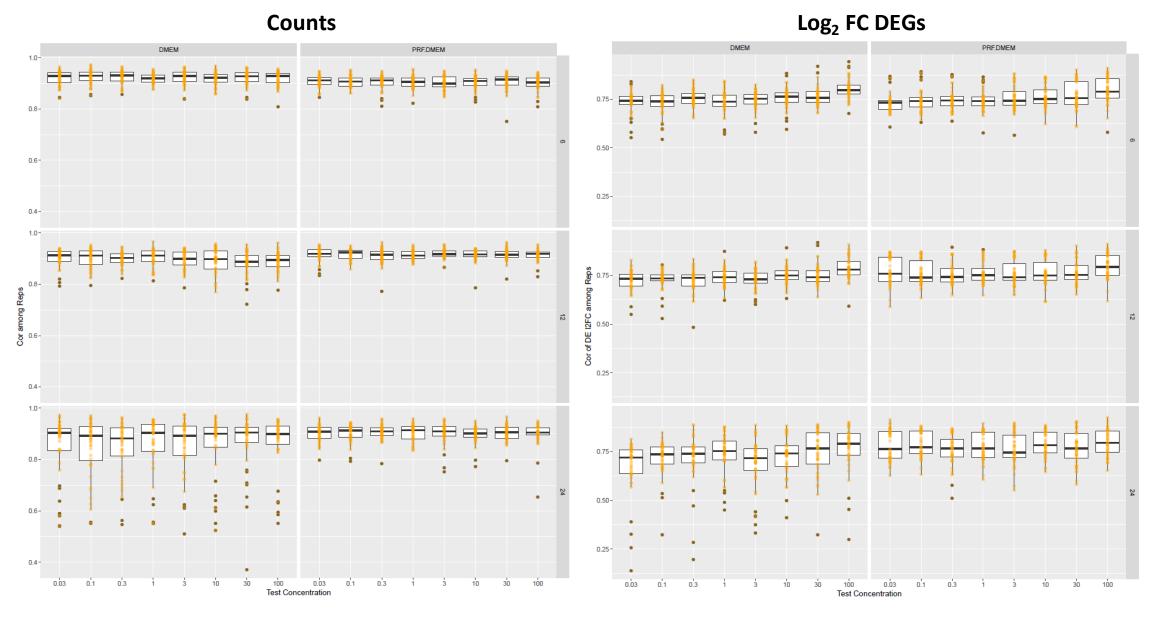
• Correlation among technical replicates is high (> 0.85 %).

# Coefficient of Variation (CV) Among Technical Replicates



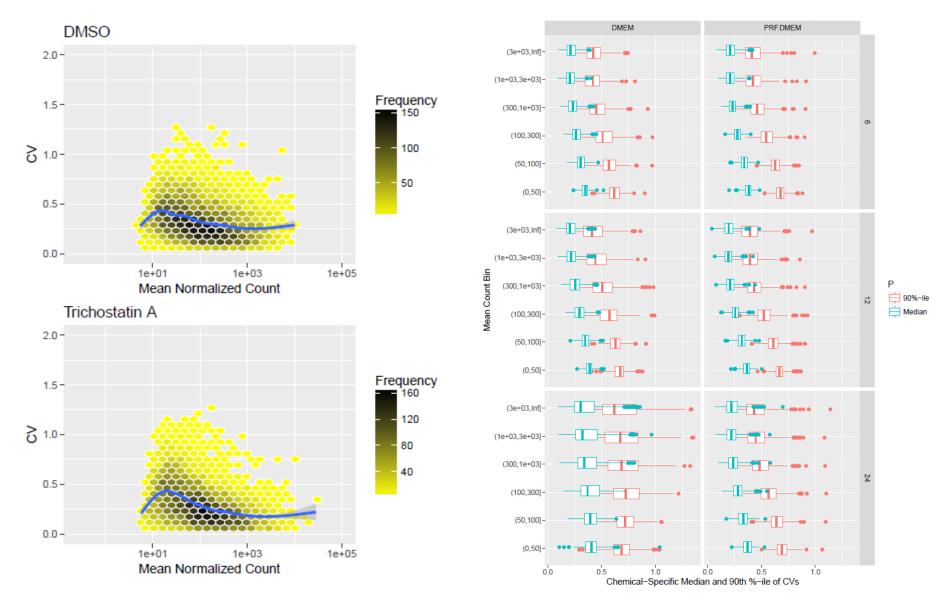
Coefficient of variation in gene expression values is low (median ~30 %).

## Correlations in Biological Replicates, Stratified by Expression Level



• Correlations of raw counts and  $log_2$  FC of DEGs is high ( $\geq 0.85$ ) for most conditions.

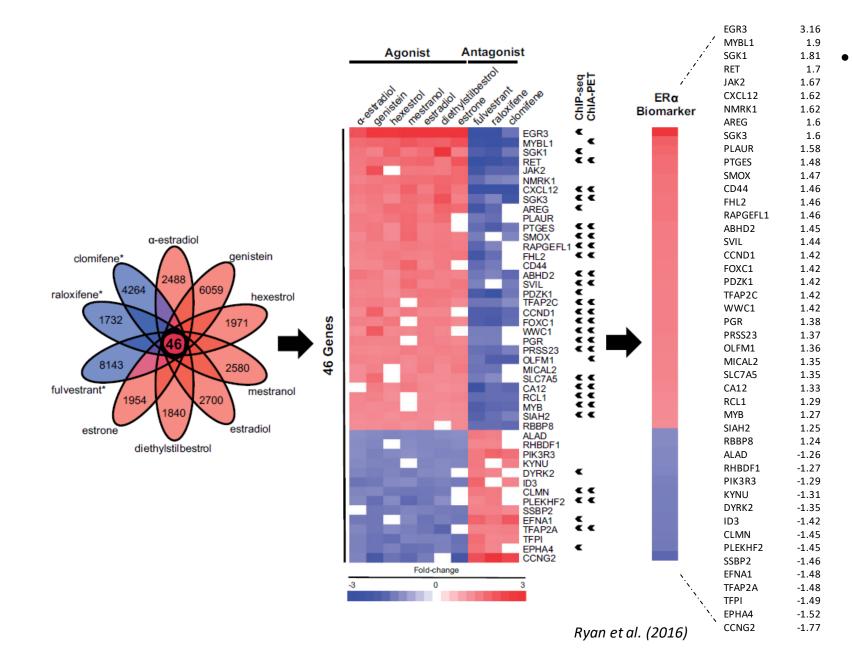
## Coefficient of Variation (CV) Stratified by Expression Level



CVs decrease as a function of mean expression level.

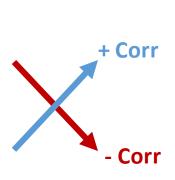
#### **ER**α Biomarker Signature

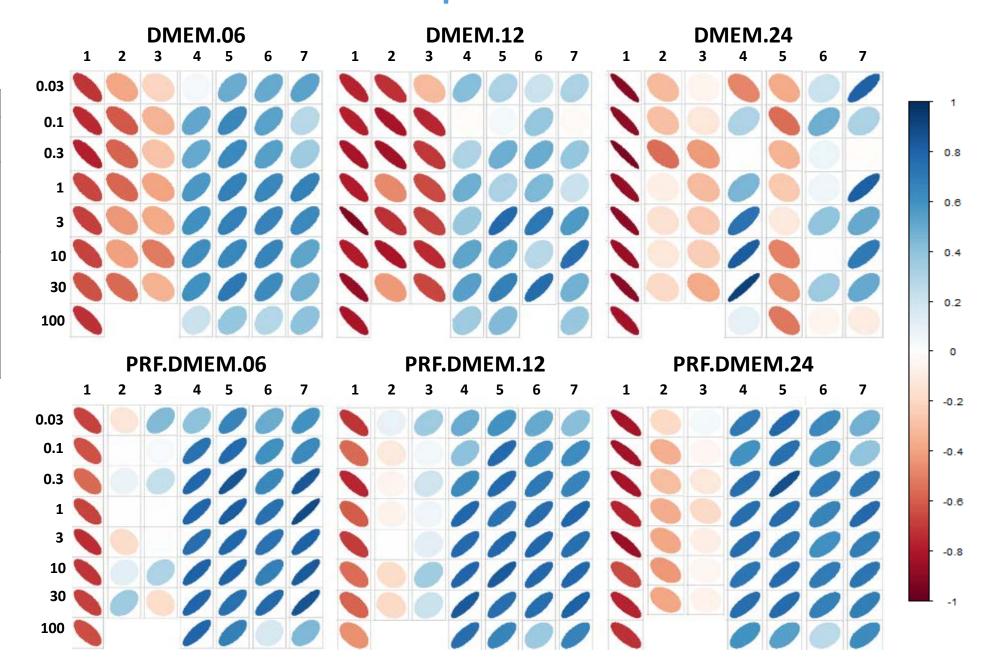
- Biomarker signature determined by treating MCF7 cells with various  $ER\alpha$  agonists and antagonists.
- Can we use this to detect biologically meaningful signal in the BioSpyder data?



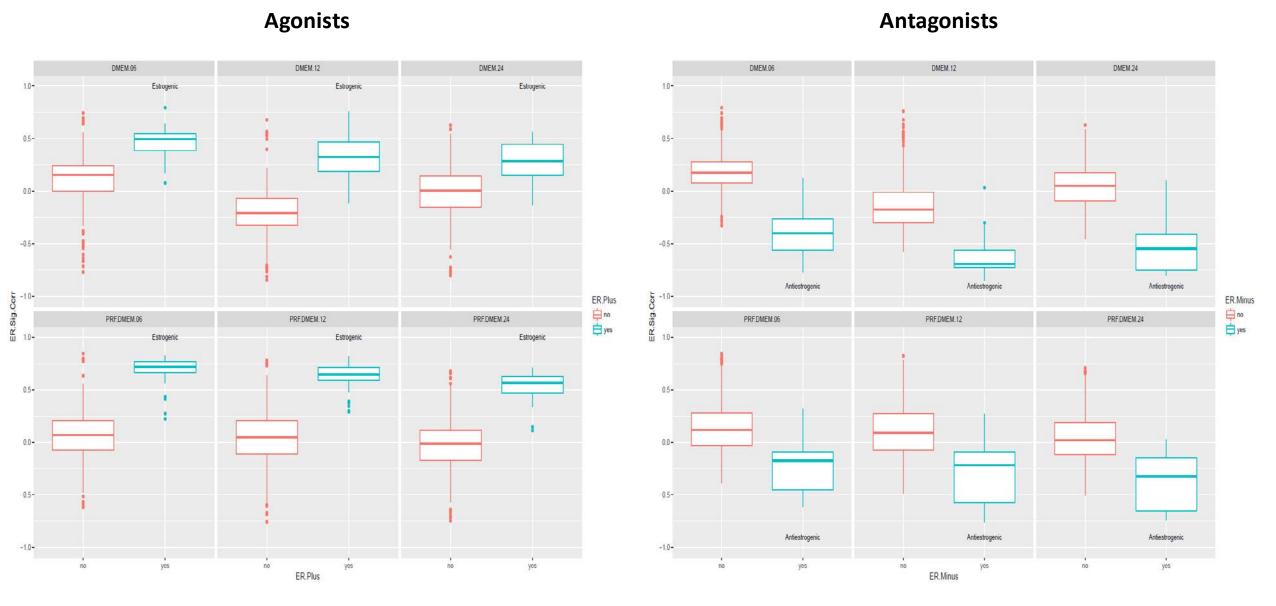
#### **Correlation with ER** a Transcriptional Biomarker

	Chemical	MOA
1	Fulvestrant	Antiestrogen (SERD)
2	4- Hydroxytamoxifen	Antiestrogen
3	Clomiphene Citrate	(SERM)
4	Bisphenol A	
5	Bisphenol B	
6	4-Nonylphenol, branched	Estrogenic
7	4-Cumylphenol	



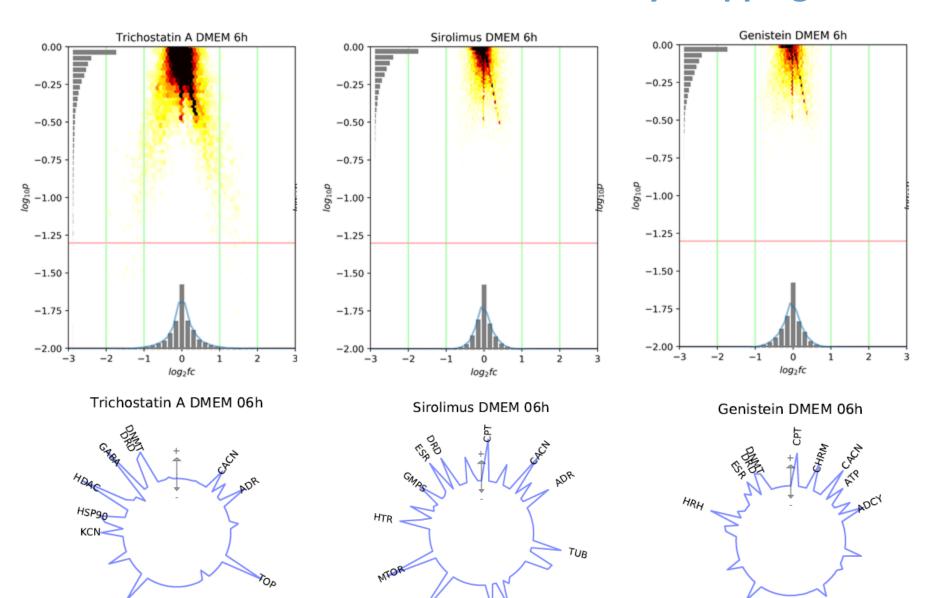


## Correlation with ERa Transcriptional Biomarker - Antagonists



• The ability to detect ERa antagonists (particularly SERMs) was decreased by use of charcoal stripped serum.

#### **Connectivity Mapping**



- Differential gene expression observed with reference chemicals.
- Putative targets identified using Connectivity
   Mapping
- Large degree of promiscuity of predicted targets observed.
- Currently evaluating additional methods for MIE prediction

#### **Benchmark Dose Modeling**



Parameter	Criteria <sup>a</sup>
Pre-filter:	ANOVA $(p_{raw} < 0.05 \&  FC  \ge 2)$
Models	Hill, Exponential 2, poly2, power, linear
BMR Factor:	1.349 (10 %)
Best Model Selection:	Lowest AIC
Hill Model Flagging <sup>b</sup> :	'k' < 1/3 Lowest Positive Dose Retain Flagged Models
Pathway Analysis:	Genes with BMD <= Highest Dose ≥ 3 ≥ 5% Gene Set Coverage Fisher's Exact Two Tailed ≤ 0.05
Gene Set Collections <sup>c</sup> :	MSigDB_C2 MSigDB_H Reactome

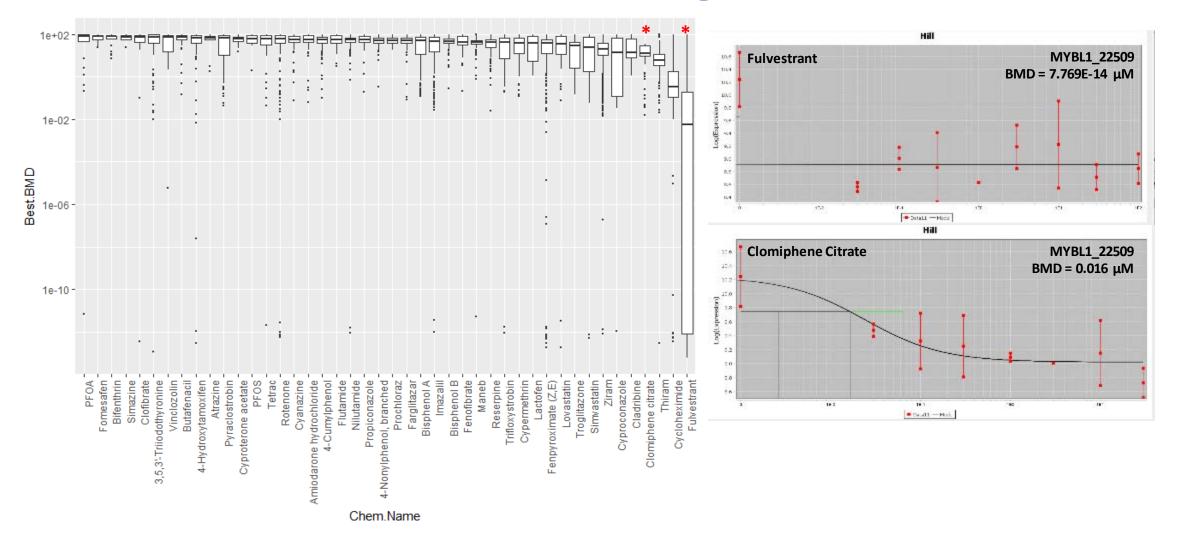
<sup>&</sup>lt;sup>a</sup> Exploratory analysis – modeling criteria not finalized

<sup>b</sup> Flagged Hill Models were retained to illustrate a specific point regarding concentration range selection

#### <sup>c</sup> Gene Set Collections:

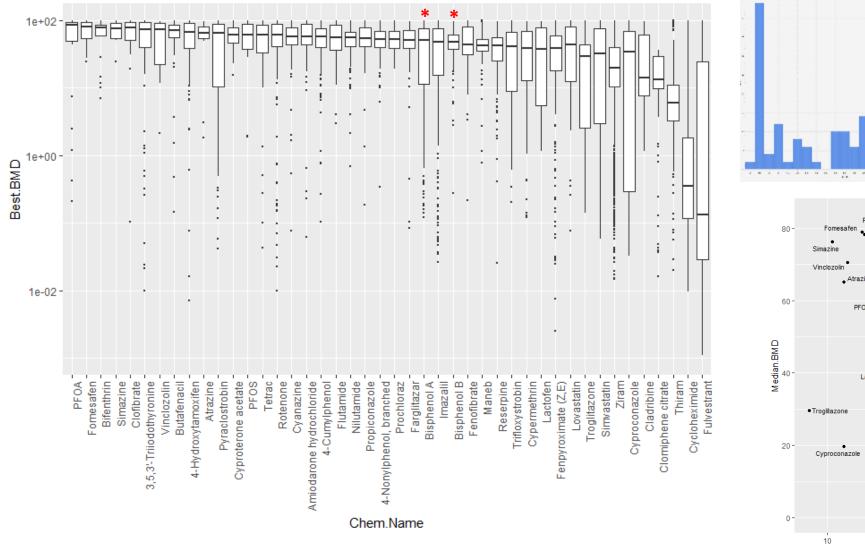
- MSigDB\_C2: Curated gene sets from online pathway databases, publications and knowledge of domain experts (n = 4738).
- MSigDB\_H: Coherently expressed signatures derived by aggregating many MSigDB gene sets to represent well-defined biological states or processes (n = 50).
- **Reactome:** Open-source, curated and peer reviewed pathway database with hierarchical pathway relationships in specific domains of biology. (n = 1764). Some pathways included in MSigDB\_C2.

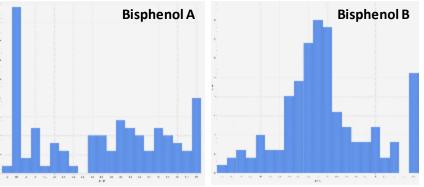
#### **Benchmark Dose Modeling Results**

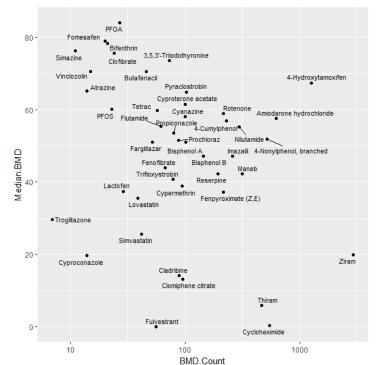


- A high occurrence of flagged Hill fits with unreasonably low BMDs may indicate the concentration range was not low enough.
- Flagged BMDs were observed with low frequency in this dataset.
- The identify of genes with flagged hill models was inconsistent across chemicals. Not driven by DMSO controls.

#### **Benchmark Dose Modeling Results**







- Wide range of chemical potencies at the probe level.
- The distribution of probe level BMDs vary from chemical to chemical.
- No apparent relationship between potency and number of probes affected (?).

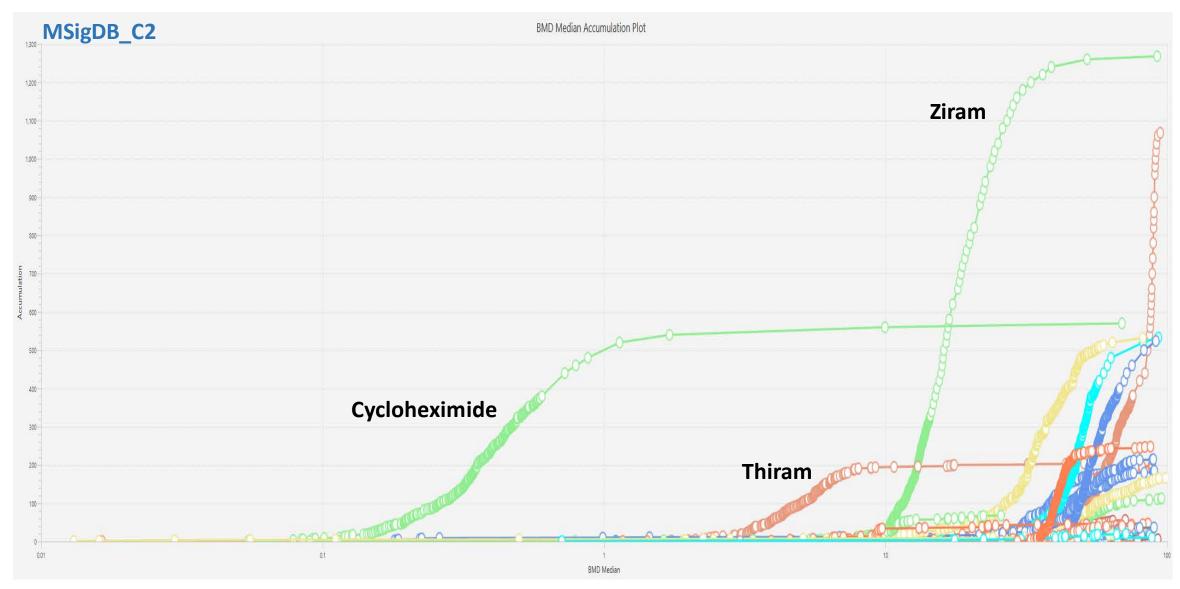
## **Pathway Enrichment**

#### **Numbers of Pathways Enriched**

Chemical Name	MSigDB_C2	MSigDB_H	Reactome	Chemical Name	MSigDB_C2	MSigDB_H	Reactome
Ziram	1268	26	314	Propiconazole	20	1	2
4-Hydroxytamoxifen	1068	14	331	3,5,3'-Triiodothyronine	18	0	1
Cycloheximide	570	24	126	Fenofibrate	17	0	1
4-Nonylphenol, branched	533	7	127	Cyanazine	16	0	1
Amiodarone hydrochloride	524	12		Flutamide	10	0	1
Reserpine	523	11	80	Fulvestrant	9	1	0
Maneb	248	3	75	Cypermethrin	7	0	1
Rotenone	215	5	22	Lovastatin	6	0	0
Thiram	204	5	64	Simvastatin	5	0	0
4-Cumylphenol	198	4	27	Butafenacil	3	0	0
Bisphenol B	185	2	31	Vinclozolin	2	0	0
Fenpyroximate (Z,E)	183	5	14	Tetrac	2	0	1
Cyproterone acetate	166	5	4	Lactofen	2	0	0
Prochloraz	113	2	10	Cyproconazole	0	0	0
Clomiphene Citrate	68	3	0	Clofibrate	0	0	0
Nilutamide	56	0	29	PFOS	0	0	0
Trifloxystrobin	47	1	2	Simazine	0	0	0
Cladribine	47	0	71	Fomesafen	0	0	0
Bisphenol A	45	1	5	Troglitazone	0	0	0
Imazalil	41	0	4	PFOA	0	0	0
Pyraclostrobin	37	0	1	Atrazine	0	0	0
Farglitazar	22	1	0	Bifenthrin	0	0	0

- Heterogeneity in the amount and type of pathways enriched.
- Changing filtering stringency and BMD modeling strategy affects these results.

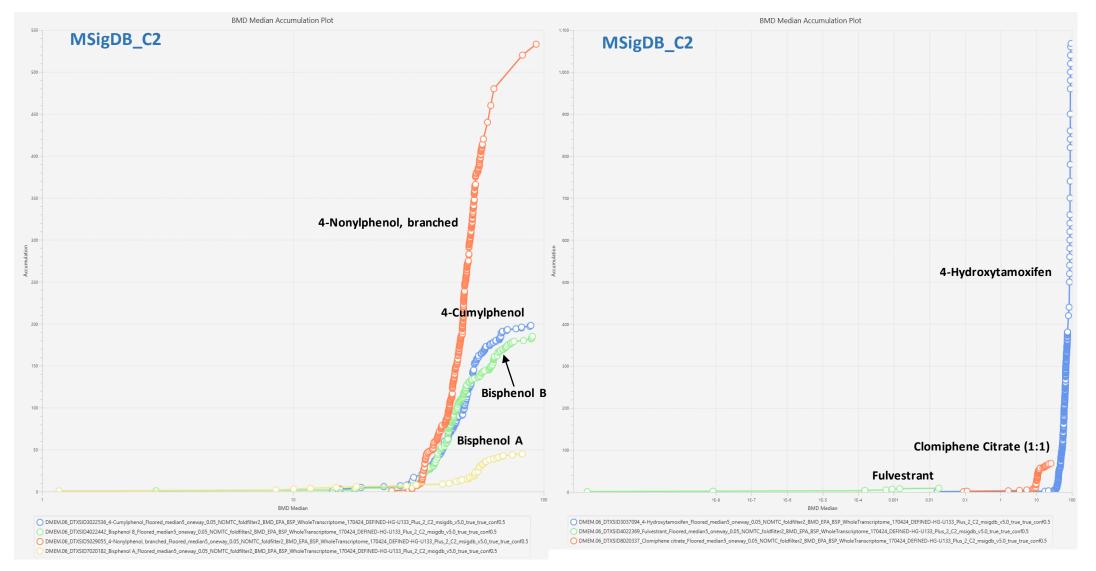
# **Pathway Potencies**



Broad range of pathway level potency estimates and number of pathways affected across chemicals.

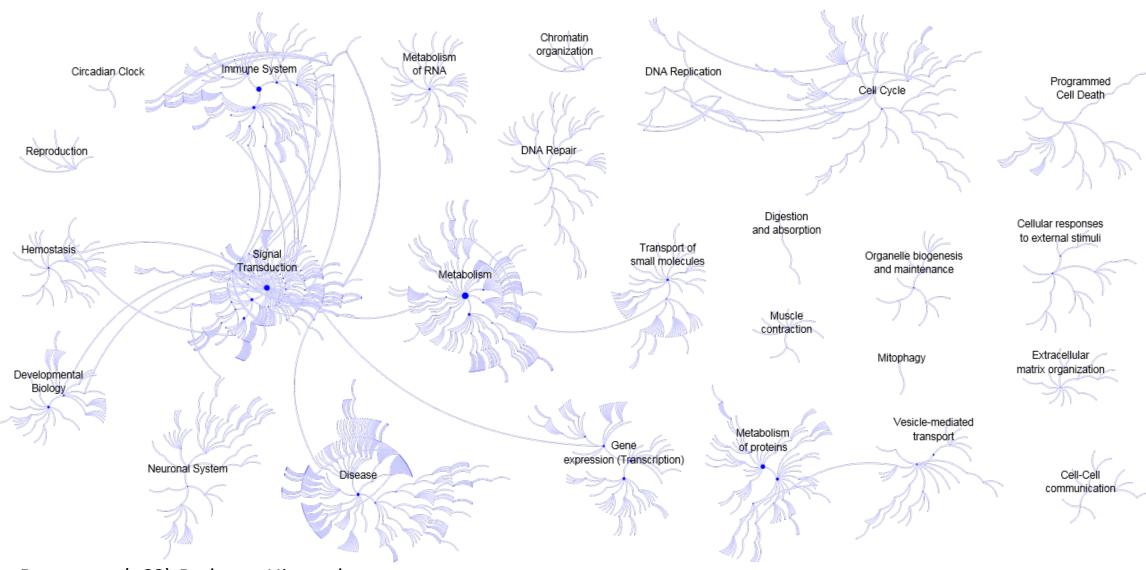
# **Pathway Potencies**

ER Agonist ER Antagonist



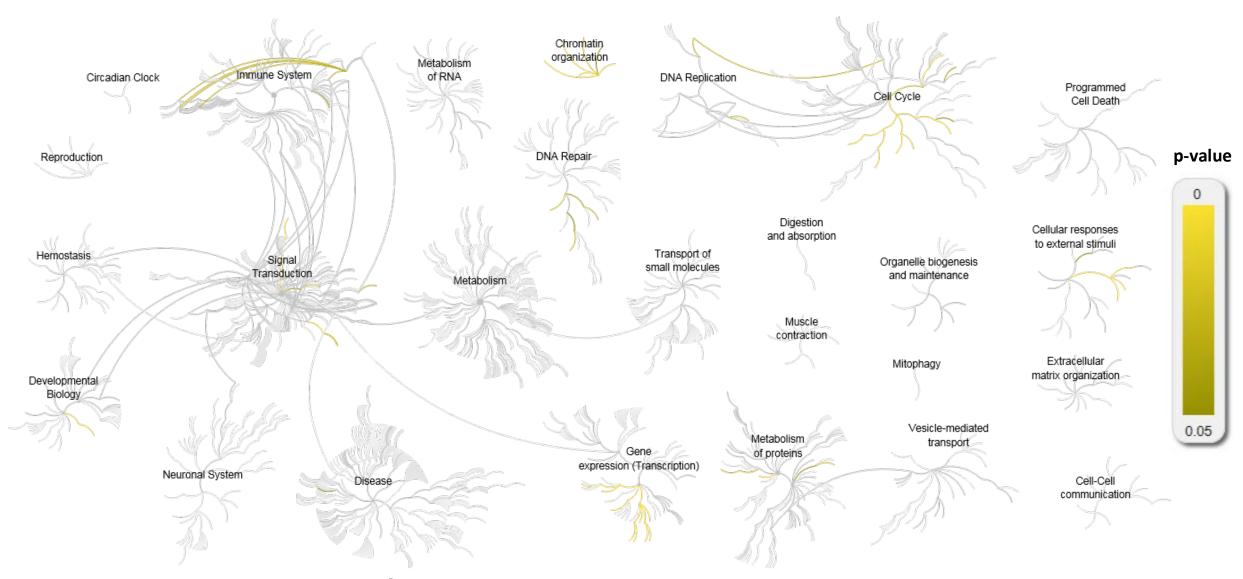
• Heterogeneity in pathway levels potency estimates and number of pathways affected within chemical class.

# **Network Mapping**



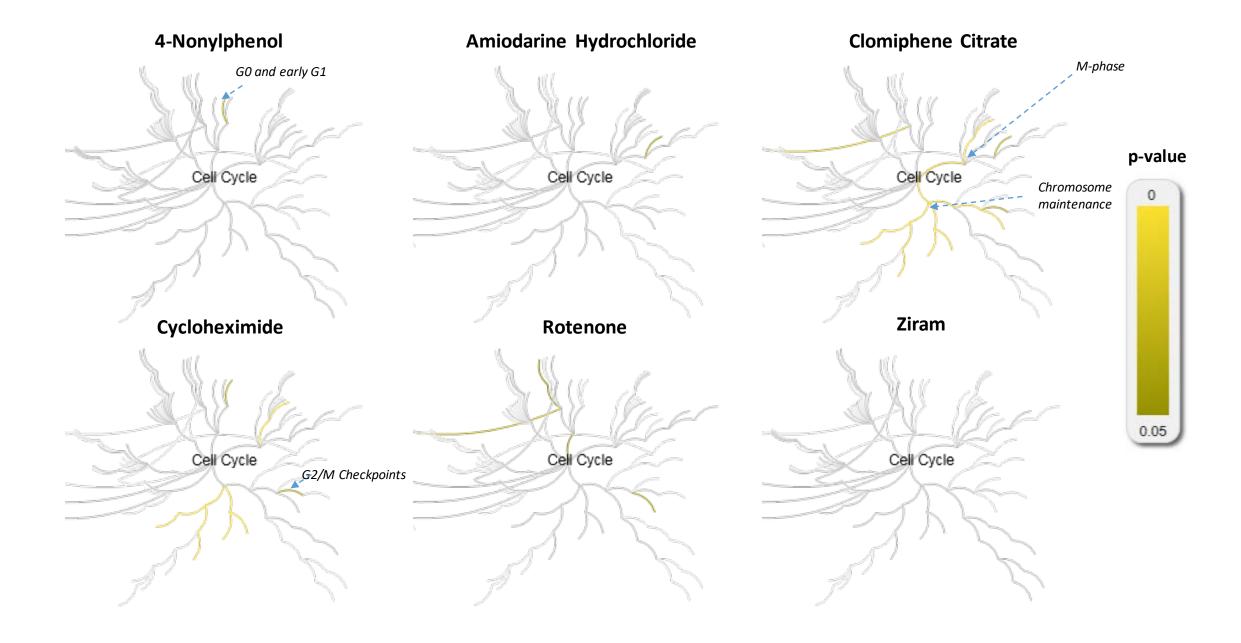
• Reactome (v60) Pathway Hierarchy

## **Network Mapping [Clomiphene Citrate]**

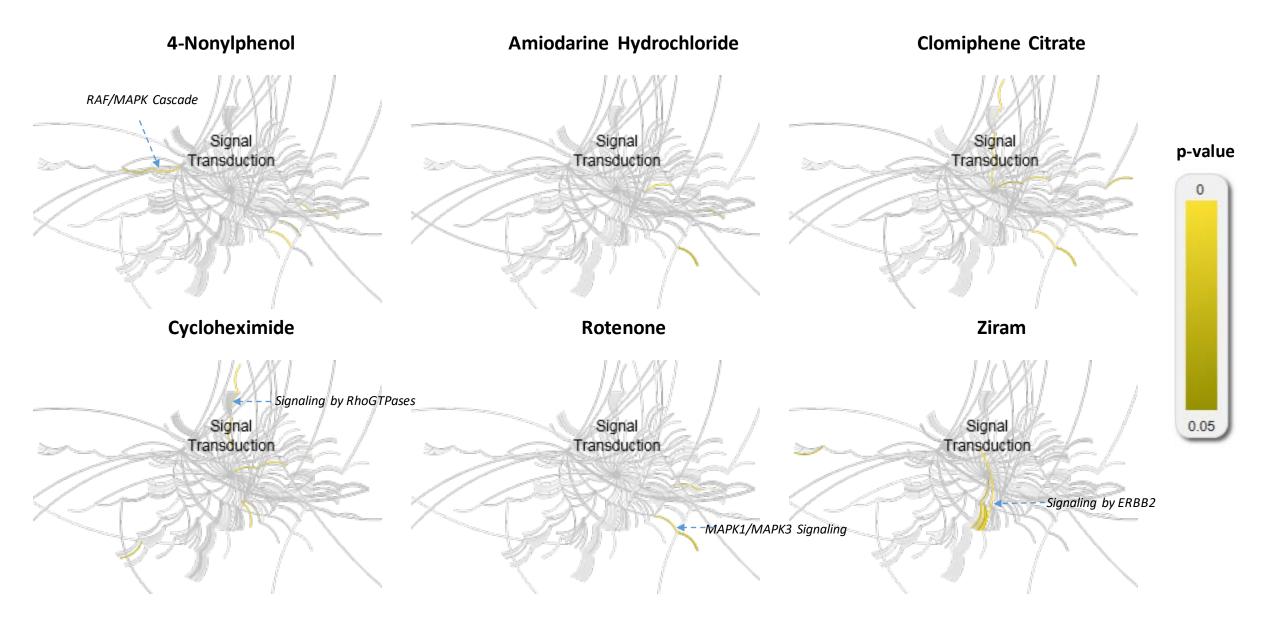


- Reactome (v60) Pathway Hierarchy -> Overlaid with enrichment scores based on probes with acceptable BMD model fit
- Highlights different areas of biology affected by a chemical

# **Diversity in Response of Cell Cycle Networks**



# **Diversity in Response of Signal Transduction Networks**



# **Current Activities & Future Directions**

#### • Fall 2017:

- Refining data analysis pipeline and BMD modeling approach.
- Exploring methods for MIE prediction & characterization of biological responses.
- Prepping initial publication.
- Conducting concentration-response screening of 2,200 chemicals in MCF7 cell model (8 conc., 6 HR exposure).

#### Beyond 2017:

- Tox21 reference chemical partner project
- Screening in additional cell lines.
- Coupling with image-based phenotypic screening assay.

# Acknowledgements

#### NCCT

- Clinton Willis
- Danielle Suarez
- Chad Deisenroth
- Imran Shah
- Woody Setzer
- Derek Haggard
- Matt Martin
- Richard Judson
- Rusty Thomas

#### BioSpyder

- Joel McComb
- Jo Yeakley
- Jason Downing
- Milos Babic
- Kyle LeBlanc
- Harper Van Steenhouse

# Questions?