EPA
United States
Environmental Protection
Agency

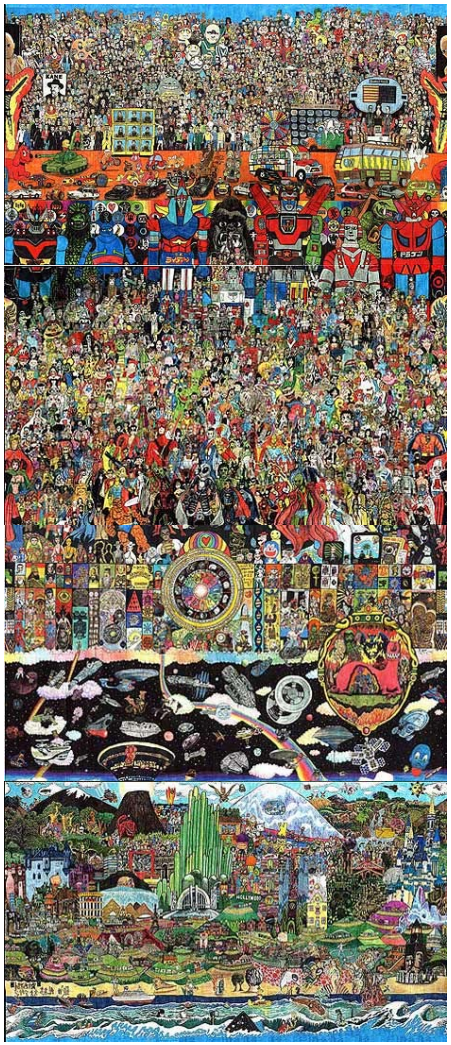
High Throughput Transcriptomics @ USEPA

Matt Martin
Research Biologist

USEPA – National Center for Computational Toxicology
Chemical Safety for Sustainability Research Program

Toxicology Forum
Salt Lake City, Utah
July 28, 2016

Toxicology is About Identifying What Can Go Wrong and Why

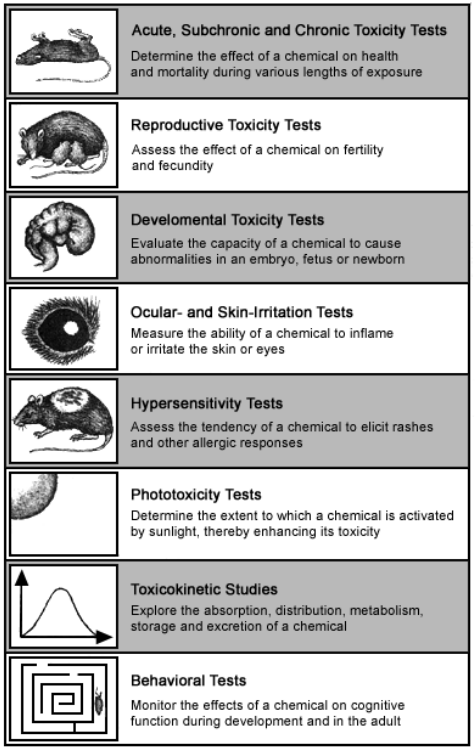


Picture of Everything - Howard Hallis

- The ideal chemical testing approach will provide complete coverage of all relevant toxicological responses
- It should be sensitive and specific
- It should identify the mechanism/mode-of-action (with dose-dependence)
- It should identify responses relevant to the species of interest
- Responses should ideally be translated into tissue-, organ-, and organism-level effects
- It must be economical and scalable

Traditional Animal Studies Are Less Than Ideal

Apical Responses With Limited Mechanistic Insight



Goldberg and Frazier (1989)

Variable Cross-Species Concordance

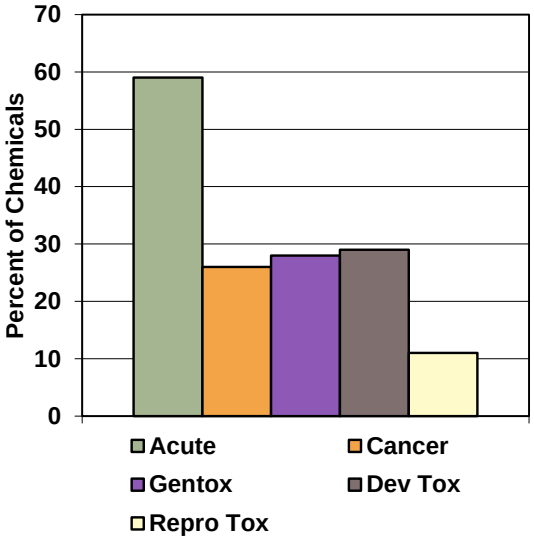
Table 3. Predictability of laboratory animal models for putative human teratogens.[a]

Teratogen/group	Mouse	Rat	Rabbit	Hamster	Primate	Dog	Cat	Pig	Ferret	Guinea pig
Alcohol	+	+	−		+	+		+		+
Androgenic hormones[b]	+	+	+	+	+	+		+		+
Anticancer antimetabolites[c]	±	+	+	−	±	±	±	+		+
Anticancer alkylating agents[d]	+	+	+		+				+	
Anticonvulsants[e]	+	±	+	−	±	−	−			
Coumarin anticoagulants	−	−	−					−		
Antithyroid agents[b]	+	+	+							+
Progestrogenic hormones[b]	+	±	±		−	+				+
DES[b]	+	+	−	±	+				+	
Methylmercury	+	+	−	+	−	−	+	−		
Thalidomide	±	±	+	±	+	±	±	±	+	−
Lithium	+	−	−		−			−		
D-Penicillamine		+		+						
Streptomycin antibiotics[b]	−	+	−							−
Vitamin A analogs	+	+	+	+	+	+		+		+

[a] Legend: (+) teratogenic; (±) variably teratogenic; (−) not teratogenic.
[b] Defects related to functional activity.
[c] Includes azauridine, aminopterin, fluorouracil, methotrexate and cytarabine.
[d] Includes busulfan, chlorambucil, cyclophosphamide, and mechlorethamine.
[e] Includes hydantoin and dione groups, and valproate.

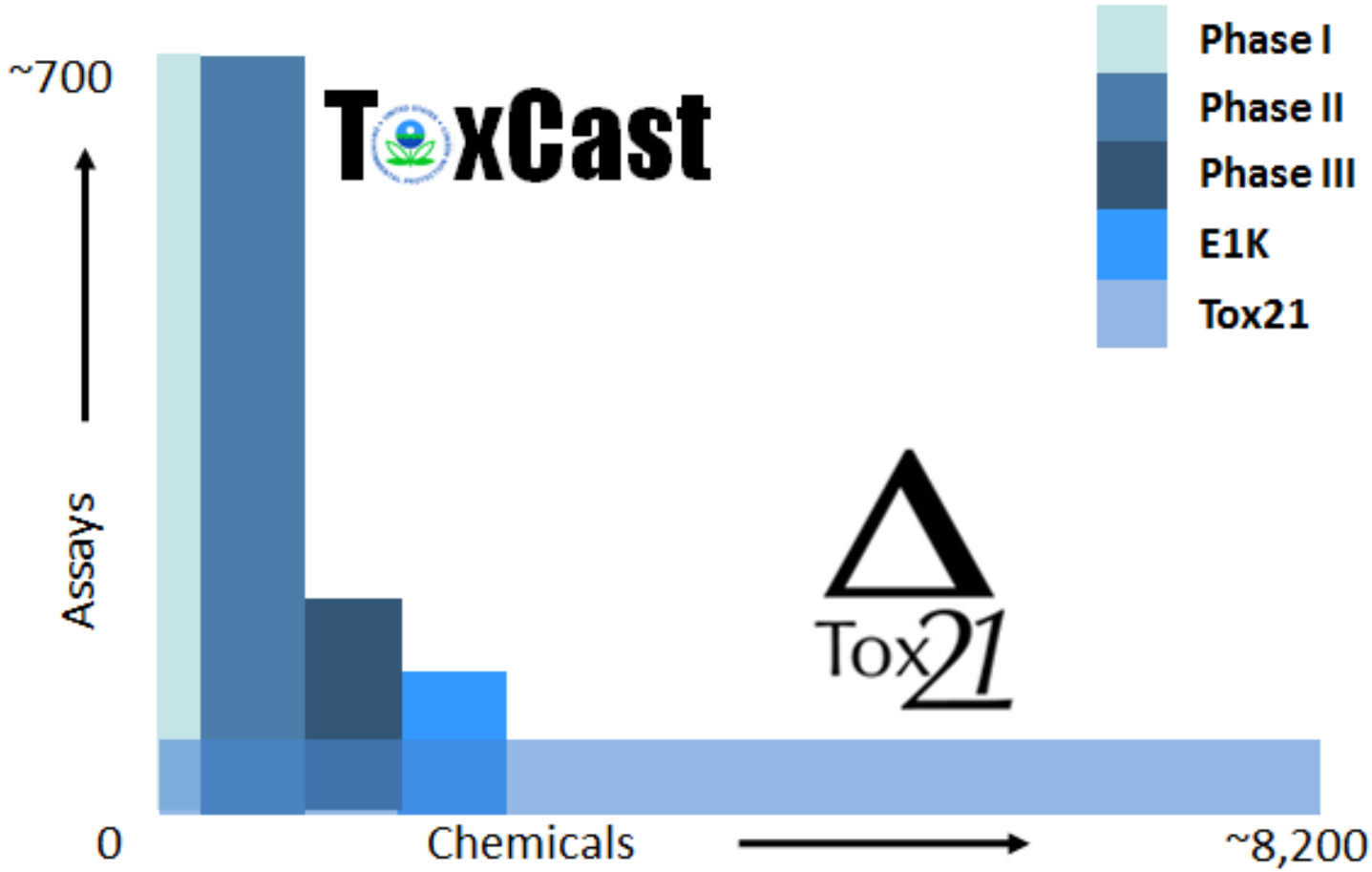
Schardein, et al *EHP* (1985)

High Cost and Limited Scalability



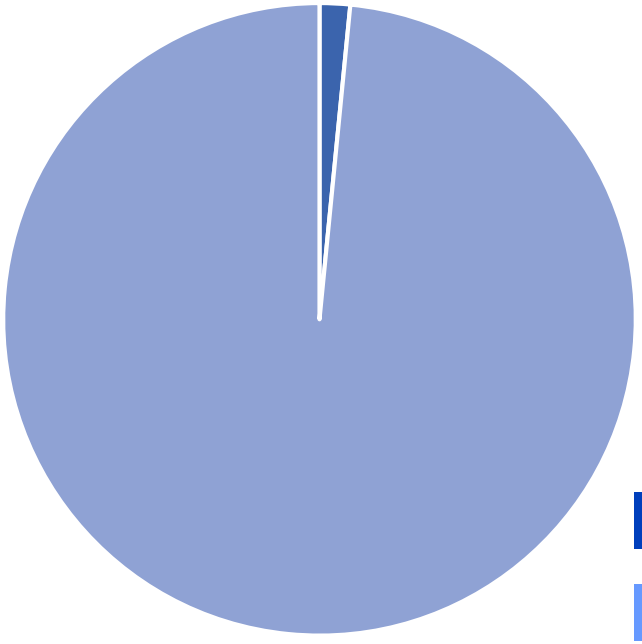
Judson, et al *EHP* (2010)

High-Throughput Screening Efforts Have Attempted to Fill Gaps

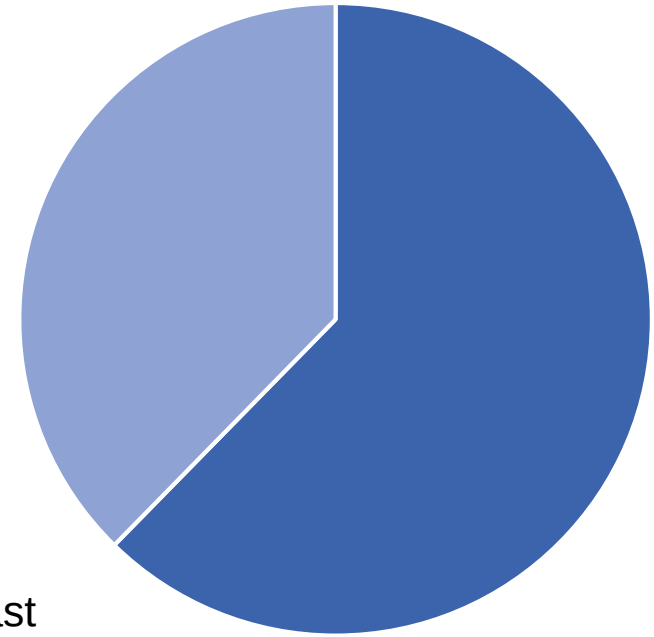


But, Current Coverage of Biological Space is Less Than Optimal

ToxCast Gene Coverage



ToxCast Pathway Coverage*

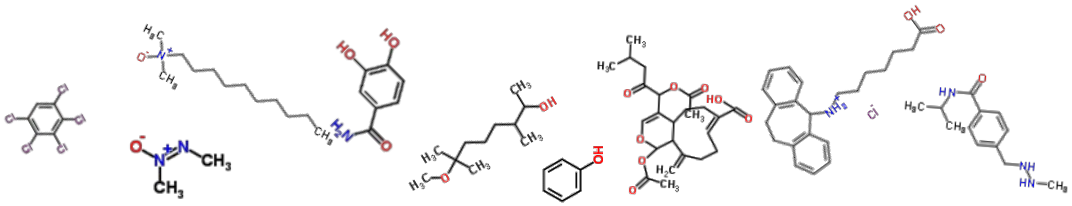


ToxCast

Not in ToxCast

*At least one gene from pathway represented

Incorporating a Comprehensive Biological Screening Platform



Broad Primary Screen for Bioactivity/MOA

HTTr: High Throughput Transcriptomics

Secondary Confirmation Screen

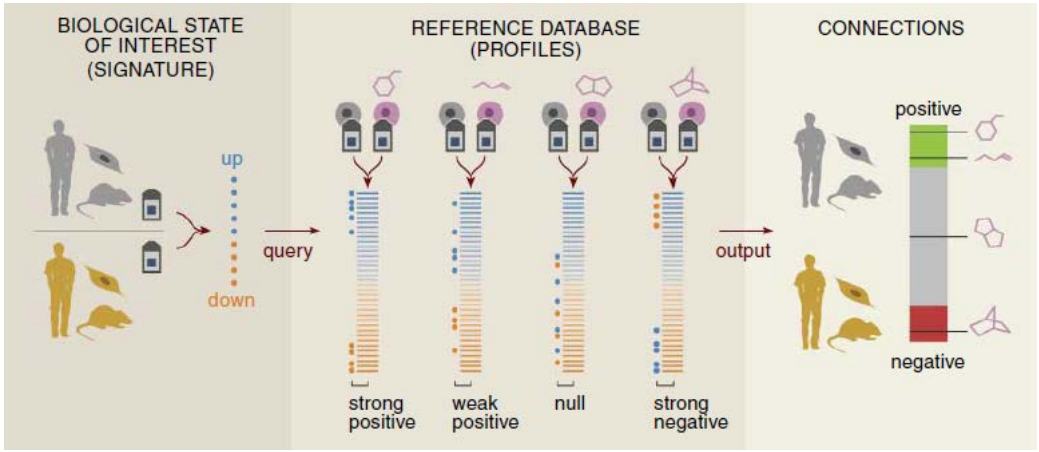
Tertiary Screen to Identify Likely Tissue/Organ Effects

Operationalizing & Deploying

HTTr Assay

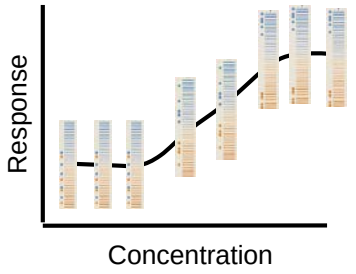
Tier 0

- Identify predominant mechanisms as a function of concentration
- Group chemicals by similar mechanism/bioactivity
- Identify a concentration that results in no transcriptional effects

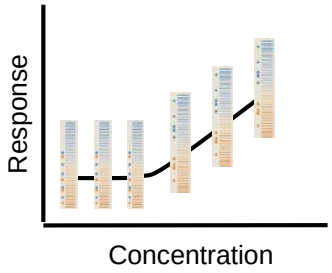


Lamb et al. *Science* (2006)

Broad CMAPdb: 7,000 profiles; 1,309 compounds
NIH LINCs CMAPdb: 9,000 shRNAs, 3,000 over expression ORFs, and 4,000 compounds in 20 cell types/lines (cell lines and primary cells)



Cell Type #1



Cell Type #2

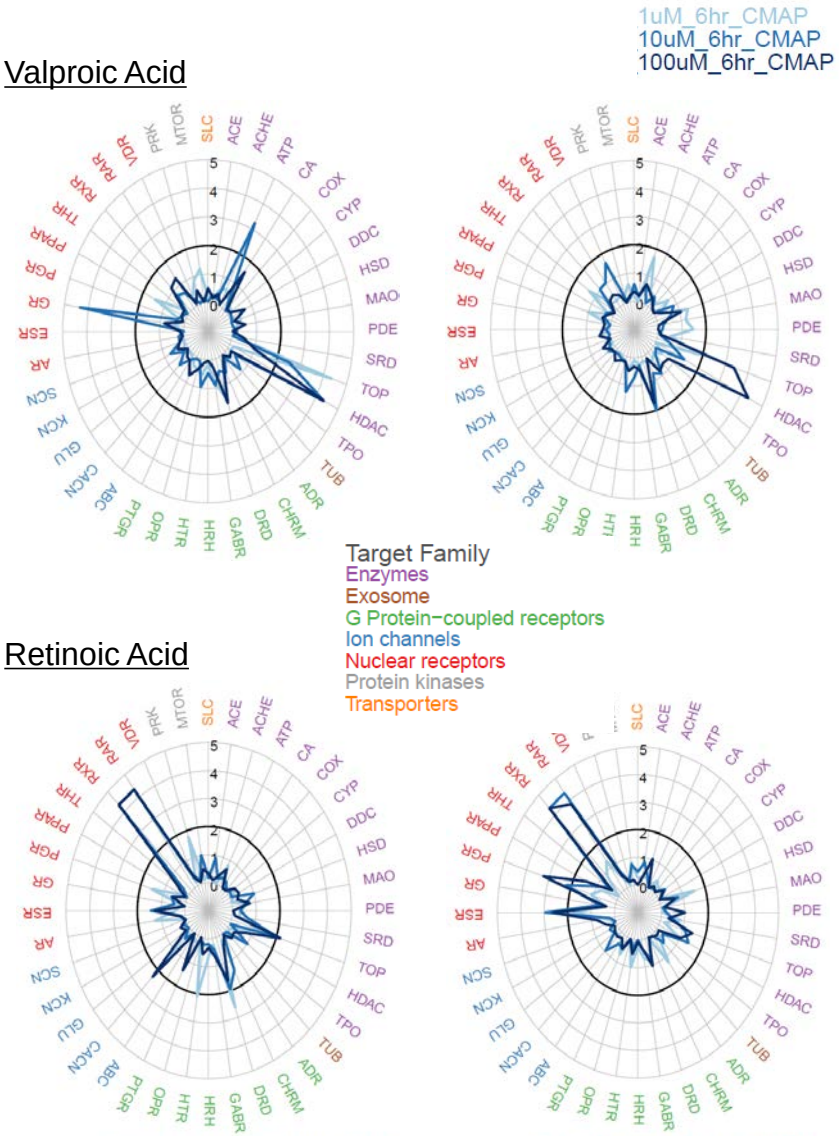
Cell Type #3

…

Identify Mode-of-Action

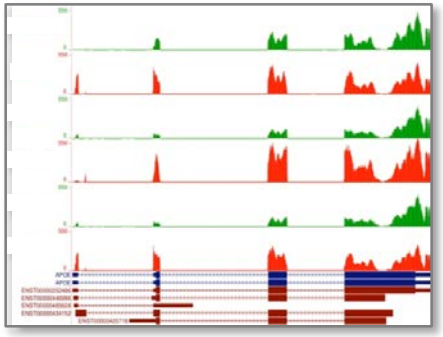
Target Family	Total Profiles	Target Genes	Chemicals	Cell Lines
Cytokine receptors	3	1	1	3
Enzymes	336	40	112	5
Exosome	14	1	4	4
G protein-coupled receptors	585	16	192	4
Ion channels	194	8	65	3
Nuclear receptors	227	10	71	5
Protein kinases	19	8	6	4
Transporters	102	2	35	3

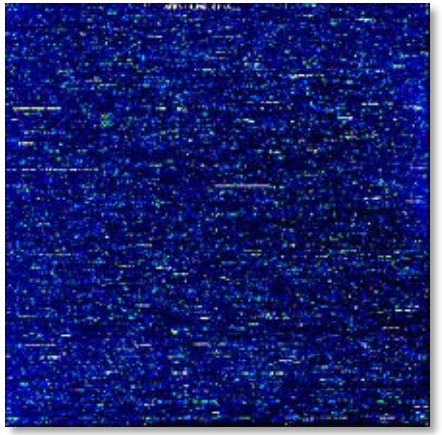
- Developed local database of Broad's CMAP data (~3,000 profiles)
- Annotated targets using KEGG (1,571 profiles)
- Significant genes identified using a z-score cutoff of 2
- Incorporated "JG" scoring method (Jiang and Gentleman 2007)
- Determine significance using a permuted rank approach across target family



Martin and Karmaus, Unpublished

Requirements and Potential Platforms for HT Transcriptomics





Requirements

- Measure or infer transcriptional changes across the whole genome (or very close to it)
- Compatible with 96- and 384-well plate formats (maybe 1536?) and laboratory automation
- Work directly with cell lysates (no separate RNA purification)
- Compatible with multiple cell types and culture conditions
- Low levels of technical variance and robust correlation with orthogonal measures of gene expression changes
- Low cost ($30 - $45 per sample or less)

Potential Platforms

- Low coverage whole transcriptome RNA-seq (3 – 5 million mapped reads)
- Targeted RNA-seq (e.g., TempO-seq, TruSeq, SureSelect)
- Microarrays (e.g., Genechip HT)
- Bead-based (e.g., L1000)

Proposed Plate Map

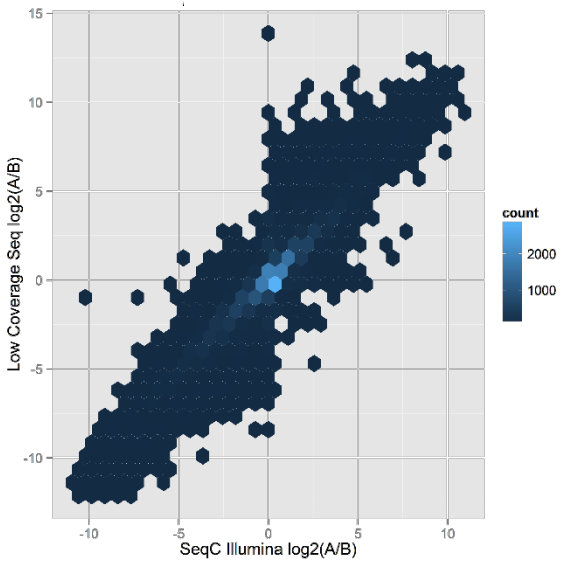
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	cndx
A	MAQC-A (Us)	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	DMSO	8
B	MAQC-A (Us)	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	DMSO	7
C	MAQC-B (Us)	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	DMSO	6
D	MAQC-B (Us)	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	DMSO	5
E	Bulk Lysate (DMSO)	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	Trichostatin	4
F	Bulk Lysate (DMSO)	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	Trichostatin	3
G	Bulk Lysate (DMSO)	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	Trichostatin	2
H	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	Trichostatin	1
I	Bulk Lysate (Trichostatin)	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	Genistein	8
J	Bulk Lysate (Trichostatin)	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	Genistein	7
K	Bulk Lysate (Trichostatin)	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	Genistein	6
L	Bulk Lysate (Trichostatin)	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	Genistein	5
M	Lysis Buffer Only	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	Sirolimus	4
N	Lysis Buffer Only	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	Sirolimus	3
O	MAQC-A (Them)	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	Sirolimus	2
P	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	Sirolimus	1

Replicates (n=3) from separate freeze/thaw

Technical Performance of the Three Sequencing Platforms

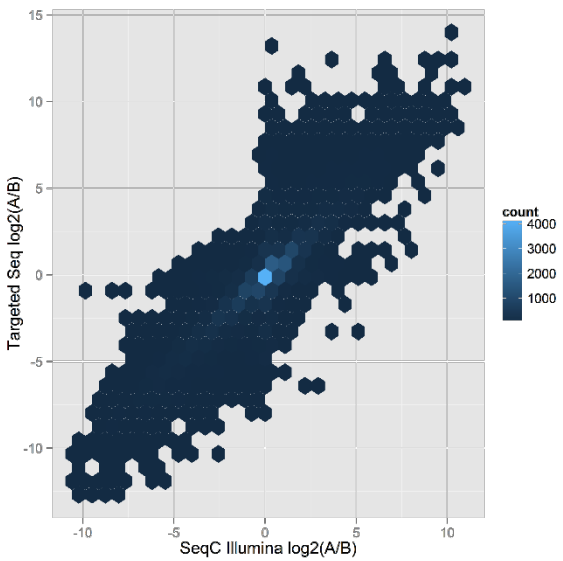
Low Coverage

r^2 0.83



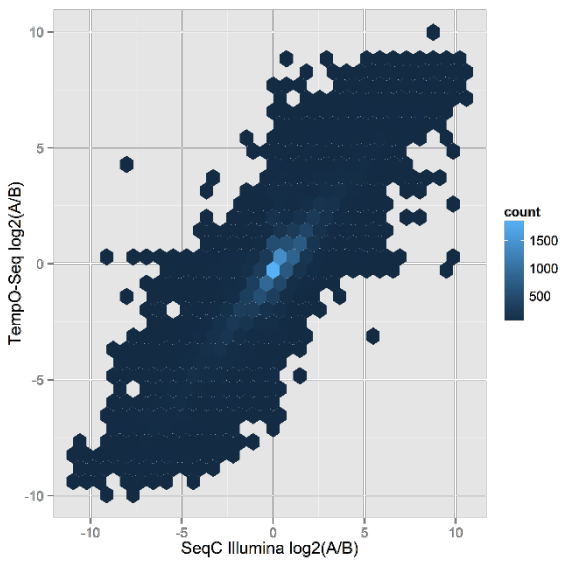
TruSeq

r^2 0.74



TempO-Seq

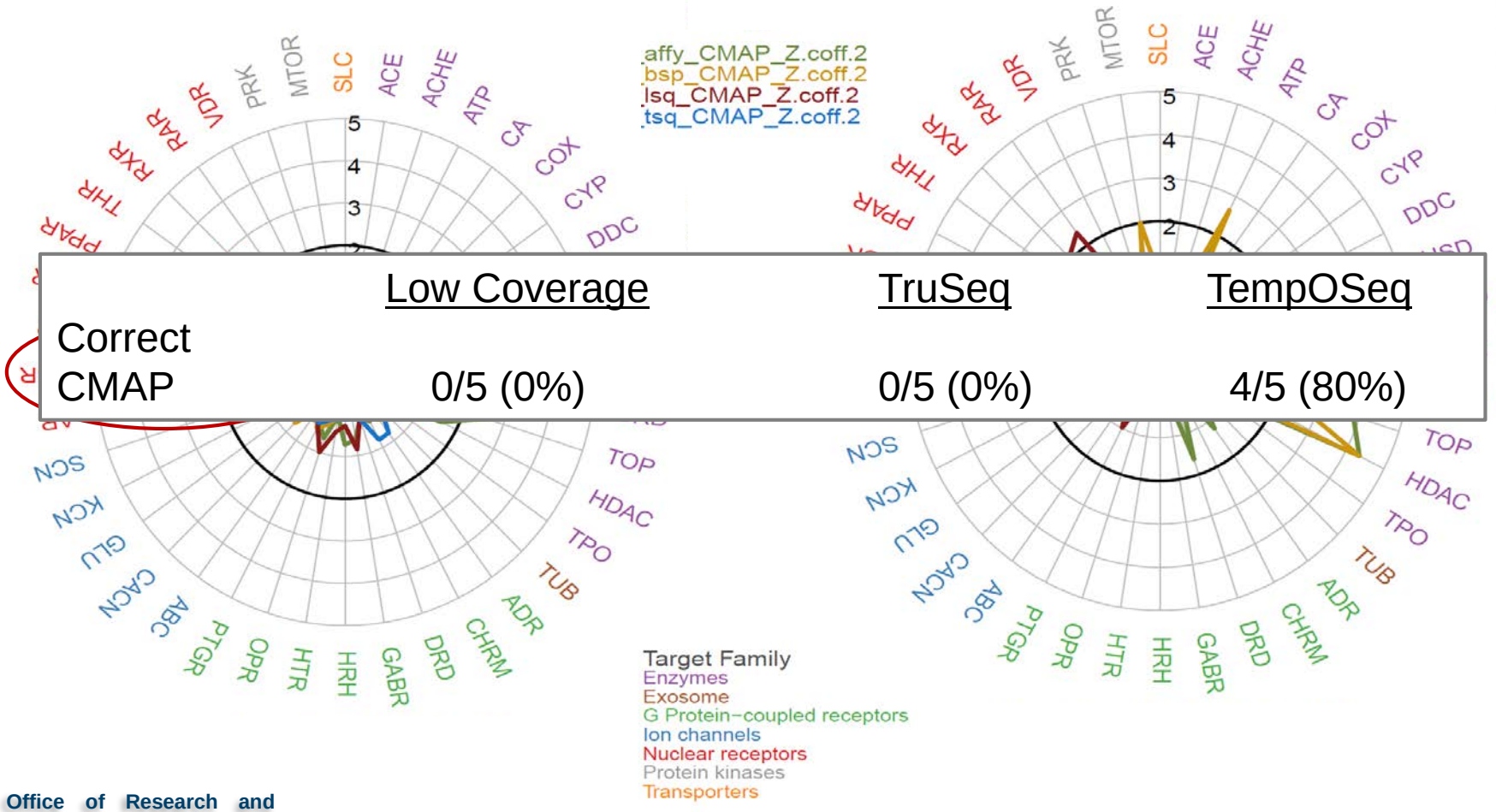
r^2 0.75



Functional Performance of the Three Sequencing Platforms

Genistein

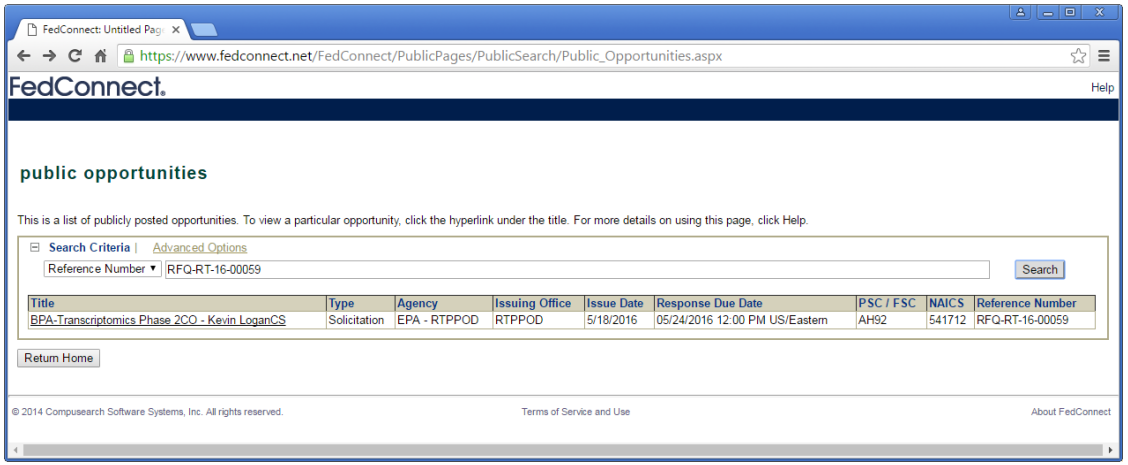
Trichostatin



	Low Coverage	TruSeq	TempOSeq
Correct CMAP	0/5 (0%)	0/5 (0%)	4/5 (80%)

Current Status

- Procurement underway for high-throughput transcriptomic services



- Strategy is to obtain services starting with submission of cell-lysates and delivery of raw and normalized transcriptomic data
    - This will allow multiple collaborative partners to use the same platform (i.e., harmonized) and contribute data from different cell types/models and chemicals

Anticipated Next Steps

- Perform pilot study (Summer) to validate workflow and refine experimental design
- Initiate large scale screen (Fall/Winter)
    - Cell type: MCF7
    - Compounds: 1,000+ (ToxCast Phase I/II + Reference Chemicals)
    - Time Point: Single
    - Concentration Response: 8 (Starting @ 300 uM)
- Perform secondary pilot study looking at cell type selection/ pooling strategies (Fall/Winter)
- Integrate HT transcriptomic platform with metabolic retrofit solution to allow screening +/- metabolism (FY17)
- Explore partnerships to build community database of common chemical set across multiple cell types/lines

EPA United States Environmental Protection Agency

Summary

- High-throughput transcriptomics will fundamentally change the way we evaluate chemicals for safety
  - Greater coverage of biological space
  - Reduced cost
  - Ability to leverage large existing databases of gene expression data
  - Fits logically in a tiered testing approach
  - Allows dose response characterization for both selective and non-selective chemicals
- Platform procurement underway
- Cell type/line selection challenges remain

Acknowledgements

Tox21 Colleagues:
- NTP Crew
- FDA Collaborators
- NCATS Collaborators

Hamner Collaborators:
- Barbara Wetmore
- Michael Black

P&G Collaborators:
- George Daston
- Jorge Naciff

EPA's National Center for Computational Toxicology