NTP
National Toxicology Program

Tox21 Phase 3: High-Throughput Transcriptomics and the S1500+ Initiative

NTP Board of Scientific Counselors Meeting
December 08, 2017

Richard S. Paules, Ph.D.
Acting Chief, Biomolecular Screening Branch
Division of the National Toxicology Program
National Institute of Environmental Health Sciences

NTP
National Toxicology Program

The Toxicology in the 21st Century ("Tox21") Federal Partnership

NTP Board of Scientific Counselors Meeting
December 14, 2016

Richard S. Paules, Ph.D.
Acting Chief, Biomolecular Screening Branch
Division of the National Toxicology Program
National Institute of Environmental Health Sciences

NTP Tox21 Phase III: Improving on Biological Coverage and Human Relevance

The Assumption:

- Global "Omic" (Whole System) approaches can **link** genomic perturbations with alterations in biological processes that result in toxicity and/or disease

The Goal:

- A rapid and affordable quantitative High-Throughput Transcriptomic (HTT) measurement of expression levels of genes for large numbers of samples (chemicals x doses x times x cells/tissues)

1. Low depth coverage, "**whole transcriptome**" targeted gene expression analysis of ~22k genes (**EPA**-led)
2. Targeted gene expression analysis using a **set of representative or "Sentinel" genes, S1500+,** to determine pathway and network perturbations (**NTP**-led)

Outline

- Development of the S1500+ Gene Set
- Evaluation of Performance of the Human S1500+ Gene Set
- Evaluation of the S1500+ on the TempO-Seq HT-Transcriptomics (HTT) Platform
- HTT Proof of Concept Study with an In Vitro Human Liver Organotypic Model
- Future Directions

Tox21 S1500+ Gene Selection Workgroup Members

- **Scott Auerbach**, Biomolecular Screening Branch, **DNTP**, NIEHS
- **Pierre Bushel**, Biostatistics Branch, **DIR**, NIEHS
- **Jennifer Collins**, Exposure, Response & Technology Branch, **DERT**, NIEHS
- **Agnes Forgacs**, National Center for Computational Toxicology, **US EPA**
- **David Gerhold**, Genomic Toxicology Group, National Center for Advancing Translational Sciences (**NCATS**)
- **Richard Judson**, National Center for Computational Toxicology, **US EPA**
- **Elizabeth Maull**, Biomolecular Screening Branch, **DNTP**, NIEHS
- **Deepak Mav**, **SciOme**, Inc.
- **Alex Merrick**, Biomolecular Screening Branch, **DNTP,** NIEHS
- **Rick Paules**, Biomolecular Screening Branch, **DNTP**, NIEHS
- **Ruchir Shah**, **SciOme**, Inc.
- **Dan Svoboda**, **SciOme**, Inc.
- **Donna Mendrick**, National Center for Toxicological Research, **US FDA**
- **Weida Tong**, National Center for Toxicological Research, **US FDA**
- **Rusty Thomas**, National Center for Computational Toxicology, **US EPA**

Attributes of a Tox21 S1500+ Gene Set

1. **Diversity:** Capture the maximal expression variability and dynamics
2. **Co-Expression:** Capture the **Sentinel** genes with maximal co-expression information to represent members of nodes or networks
3. **Maximal Pathway Coverage:** Genes are included to ensure maximal biological pathway coverage

   1500 "Sentinel" Genes Informatic Derived – S1500

4. **Inclusion of toxicity and disease related genes:** Specific genes selected for their reported roles in toxicity-related and disease-related processes (Includes L1000 gene set being used in the NIH LINCS program)

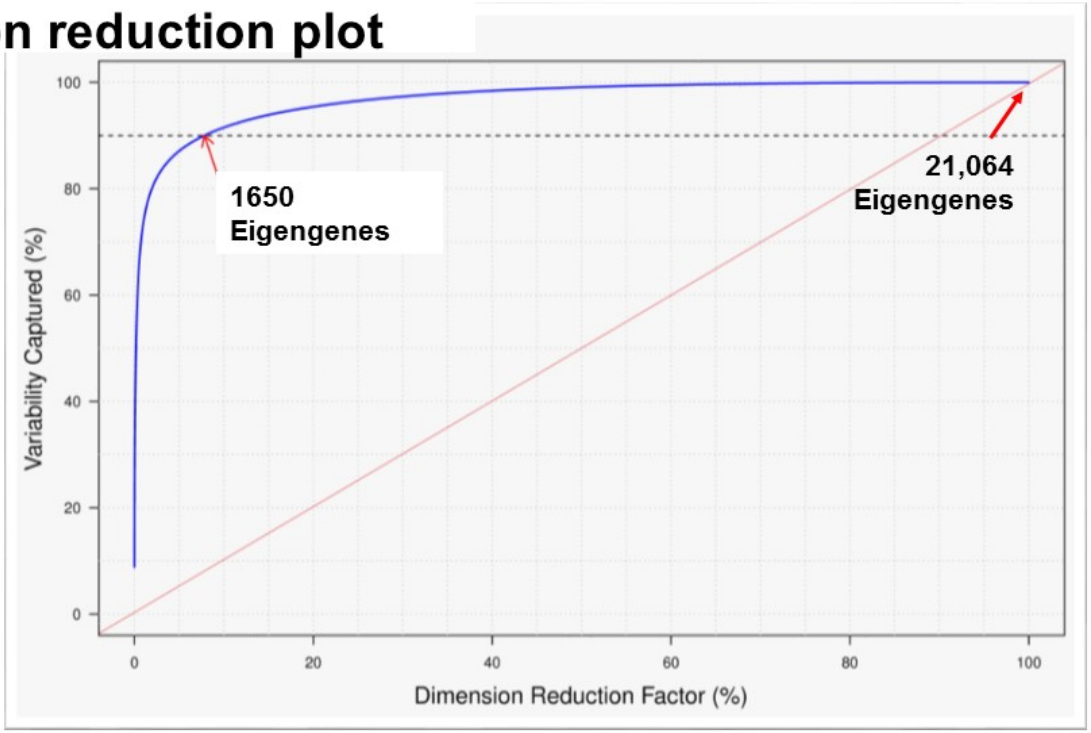
   Hybrid Informatic Derived + Nominated Genes – S1500+

5. **"Extrapolatability":** This property refers to the ability to extrapolate or **infer** or **impute** with some accuracy the expression changes of unmeasured genes from those measured in this reduced set of sentinel genes

Mav *et al*. 2017. A hybrid gene selection approach to create targeted gene sets for Tox21 high-throughput transcriptomics. *PLoS One*. In Revision.

High Degree of Correlated Gene Expression in the Human Transcriptome

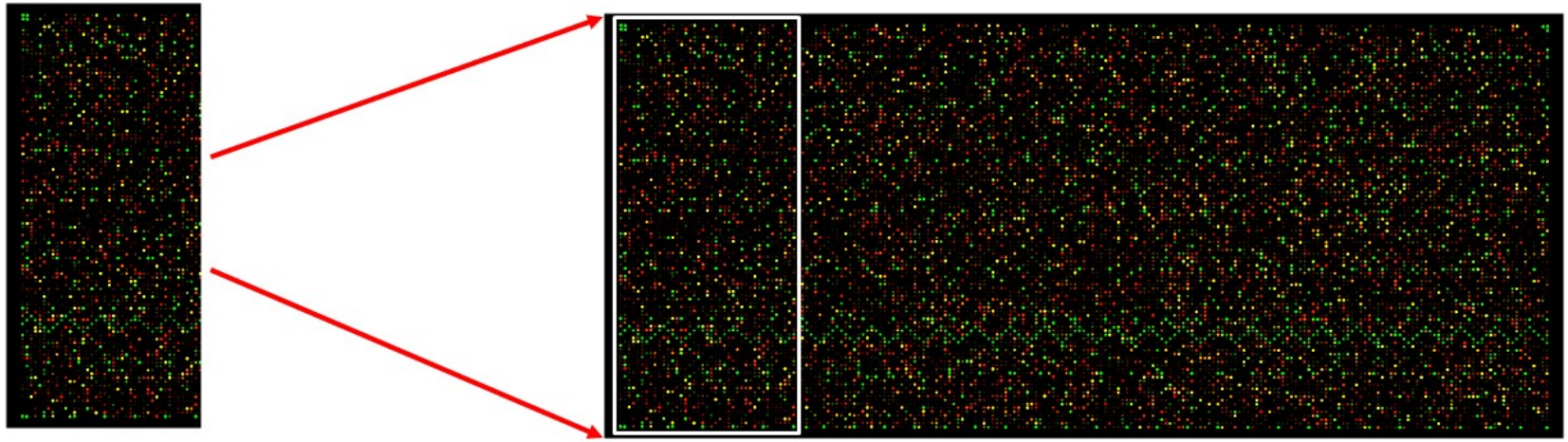
Dimension reduction plot



X-axis – the percentage of the total principal components (eigengenes); **Y-axis** – percentage of variability captured. The red line represents the expected relationship given statistically independent gene expression, whereas the blue curve shows the observed relationship.

Tox21

Attributes of a Tox21 S1500+ Gene Set



5. **"Extrapolatability":** This property refers to the ability to extrapolate or **infer** or **impute** with some accuracy the expression changes of unmeasured genes from those measured in this reduced set of sentinel genes

Mav *et al*. 2017. A hybrid gene selection approach to create targeted gene sets for Tox21 high-throughput transcriptomics. *PLoS One*. In Revision.

Human S1500+ Gene Set Pathway Coverage

Molecular Signature DB Canonical Pathways Coverage

	Pathways Covered (Total 1320)	Mean Coverage (Proportion of Pathway Genes)	Median Coverage (Proportion of Pathway Genes)	Mean Multiplicity
1500 data-driven gene list (after Steps 1-3)	659	0.12	0.08	7.73
S1500 gene set (after Step 1-4)	1320	0.26	0.25	10.38
S1500+ (2739 genes)	1320	0.43	0.43	11.44
Random 1500	541 (443, 695)	0.07 (0.05, 0.09)	0.07 (0.05,0.09)	5.88 (5.60, 6.19)
Random 2739	852 (759, 946)	0.13 (0.11, 0.15)	0.12 (0.11,0.14)	5.87 (5.66, 6.18)
L1000	906	0.17	0.16	12.97

Pathways covered are calculated relative to the **1320 canonical pathways** (MSigDB genesets) in MSigDB version 4.0. The pathway level coverage is defined as fraction of genes from pathway that overlap selected gene set.

Mean and Median coverage values are derived from pathway level coverage of 1320 canonical pathways from MSigDB (v4.0).

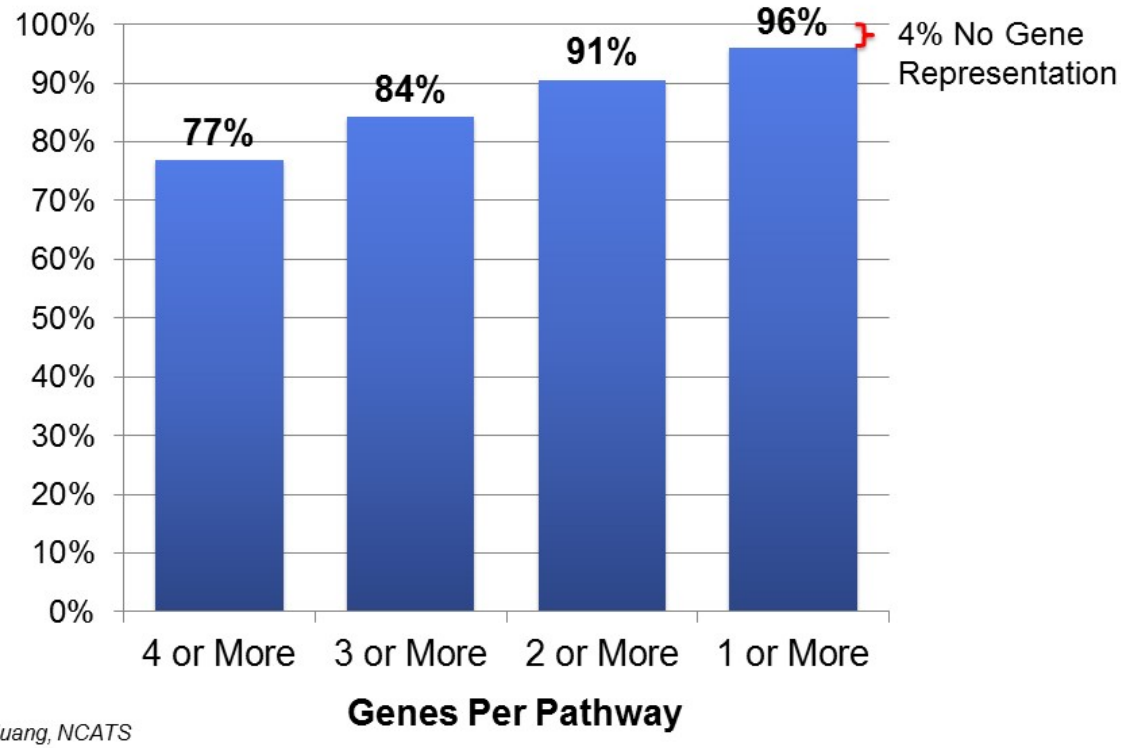
The gene level multiplicity metric represents number of pathways a gene is part of. Mean multiplicity is computed using gene level multiplicity metrics across all selected genes.

For "Random 1500" and "Random 2739" gene sets parenthesized values represent mean and range (min, max) across 20 alternative randomizations.

Tox21

Human S1500+ Pathway Coverage – NCATS BioPlanet

Hosts the universe of public, curated human pathways (~ 2000 Pathways)



From Ruili Huang, NCATS

Outline

- Development of the S1500+ Gene Set
- Evaluation of Performance of the Human S1500+ Gene Set
- Evaluation of the S1500+ on the TempO-Seq HT-Transcriptomics (HTT) Platform
- HTT Proof of Concept Study with an In Vitro Human Liver Organotypic Model
- Future Directions

Tox21

Human S1500+ Gene Set Extrapolation Performance

Performance of Extrapolated Gene Sets Using an Independent GEO Test Set

	Pearson Correlation[a]	Concordance Rate[b]	Significance Overlap[c]	Mean Squared Error[d]
Gene Level Performance (Fold Change Values)				
S1500+ (2739 genes)	0.75	0.94	0.37	0.20
Random 2739	0.76 (0.75, 0.76)	0.93 (0.93, 0.93)	0.38 (0.37, 0.38)	0.22 (0.22, 0.22)
Pathway Level Performance (GSEA Scores)				
S1500+ (2739 genes)	0.87	0.90	0.60	0.05
Random 2739	0.78 (0.77, 0.79)	0.86 (0.86, 0.86)	0.44 (0.42, 0.46)	0.08 (0.08, 0.08)

Random values represent mean and range (min, max) across 20-fold cross validation.

[a] Pearson correlations reflect agreement between extrapolated and measured values

[b] Concordance rates reflect the agreement between the extrapolated and the measured data calculated as (TP + TN)/(TP+TN+FP+FN)

[c] Significance overlap relays the proportion of genes/pathways having values (i.e. fold change or GSEA scores) in the top 1% in both the measured and extrapolated datasets

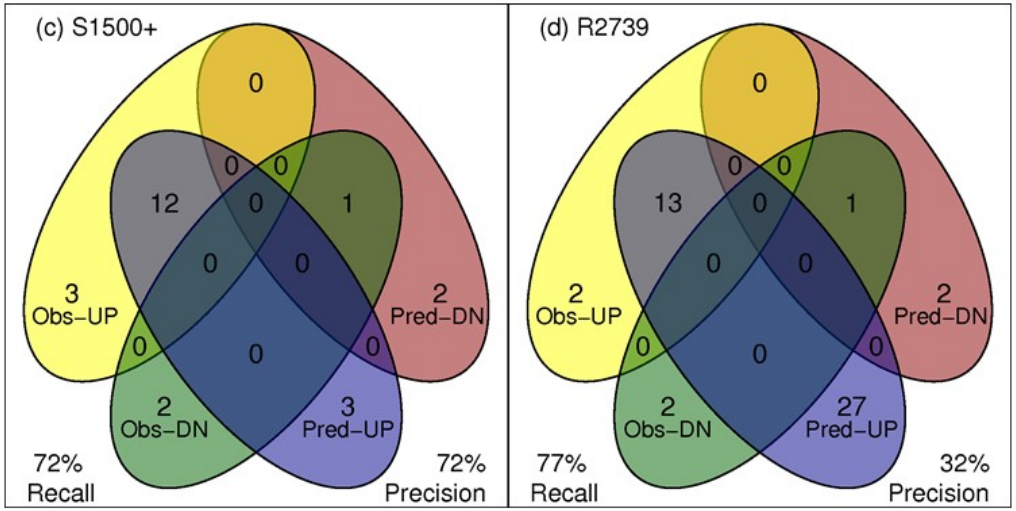
[d] Mean squared error measures the average squared difference between the extrapolated and measured values

Tox21

Human S1500+ Gene Set Extrapolated Performance

Pathway Level Performance on a Case Study

Follicular Lymphoma vs. Tonsillectomy – Concordance Venn Diagrams



All significantly enriched pathways were identified using enrichment score >0.5 and Kolmogorov Smirnov p-value < 0.001 for this analysis.

Recall is the percentage of the observed up-/down-regulated Pathways (Obs-Up and Obs-Down) that were also correctly predicted as up-/down-regulated (Pred-Up/Down).

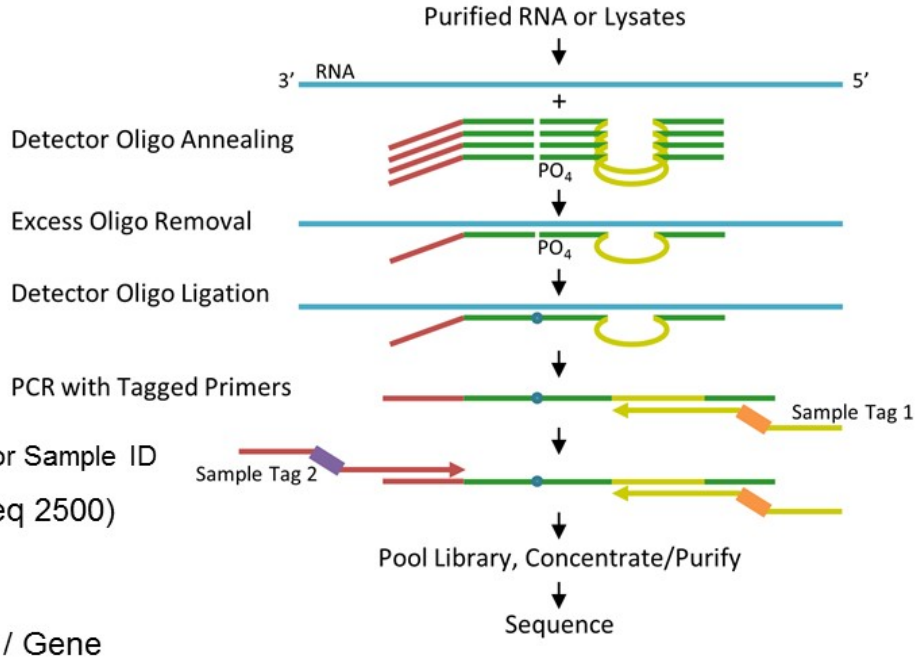
Precision is the percentage of the predicted up- and down-regulated Pathways that were observed as up- and down-regulated.

Outline

- Development of the S1500+ Gene Set
- Evaluation of Performance of the Human S1500+ Gene Set
- Evaluation of the S1500+ on the TempO-Seq HT-Transcriptomics (HTT) Platform
- HTT Proof of Concept Study with an In Vitro Human Liver Organotypic Model
- Future Directions

BioSpyder TempO-Seq™ Technology

RASL-Seq Based Assay



NextGen Sequencing: 300 x 10^6 reads / lane (HiSeq 2500)

- Detector Oligo Target Sequence – Gene ID
- Sample Specific Index Sequence – Sample ID
- NTP Target Average Read Count ~ 500 Reads / Gene
- S1500+ Gene Set – ~ 200 samples / lane
- Whole Transcriptome (22K) – ~ 30 samples / lane
- RNA-Seq – ~ 10 samples / lane

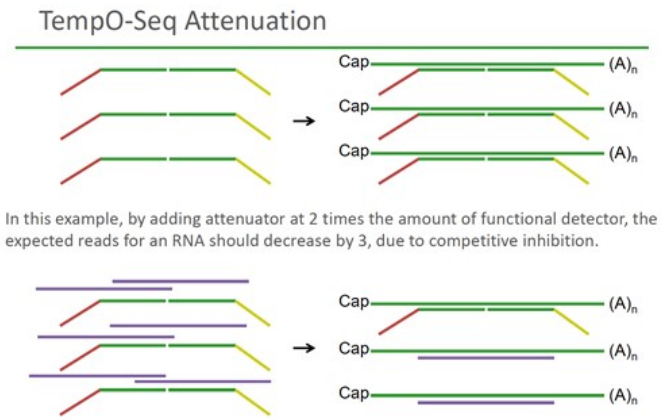
Yeakley JM, *et al*. 2017. *PLoS One*. **12(5)**: e0178302.

TempO-Seq Attenuation

Fold Attenuator Added	Expected Signal
0X	100%
1X	50%
2X	33%
4X	25%
9X	10%

20 Genes Attenuated in Human HepaRG Experiments

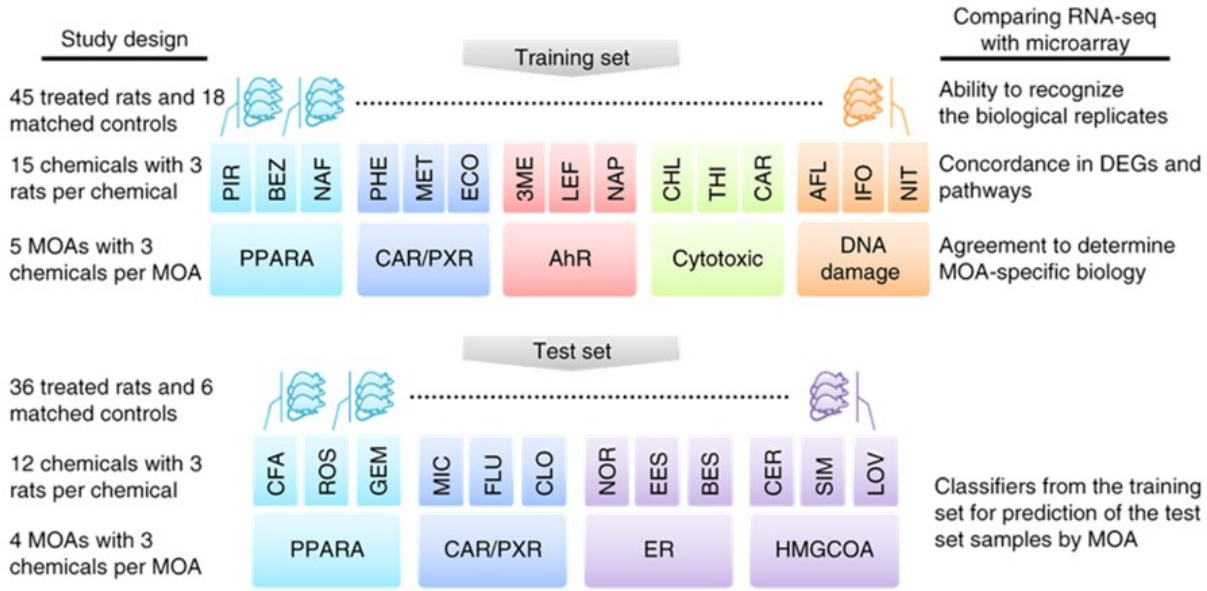
Probe ID	Max fold
IL8_14324	10
NAMPT_10869	4
TTR_7397	4
ALB_217	3
CALR_943	3
CXCL1_28308	3
FN1_27231	3
ND6_4508	3
CYP3A5_27090	2
FN1_2459	2
GDF15_18329	2
HP_3085	2
IFITM3_17816	2
MT1G_26225	2
MT2A_4334	2
NAMPT_28431	2
PGK1_5094	2
PPIA_27705	2
SAT1_6103	2
SERPINE1_6253	2



TempO-Seq Platform Evaluation with Rat SEQC Samples

DrugMatrix Rat Liver RNA Samples

- **Total of 105 RNA samples from livers of rats treated with 27 chemicals with 7 Modes of Action (MOAs) for technology evaluation**



Gene Level Performance

Measured Genes Only: Rat "S1500+ *beta*" - ~ 2300 genes

Extrapolated Genes Only: Rat Transcriptome - ~ 15,000 genes

Stat	Threshold	Pearson Correlation	MSE	Significance Overlap	Sensitivity	Specificity	Concordance Rate
Measured							
Foldchange	2	0.8314	0.3641	0.6875	0.6785	0.9370	0.8857
Nominal.Pvalue	0.05	0.7477	0.0192	0.2270	0.5270	0.8044	0.7448
FDR.Broad (mean)	0.1	0.7242	0.1063	0.5231	0.6000	0.9814	0.9789
Global.Pvalue	0.05	0.8133	0.7095	0.6875	0.8114	0.7792	0.7873
Connection	0.5	0.9799	0.0033	0.7500	0.8795	0.9776	0.9544
Extrapolated							
Foldchange	2	0.6156	0.3562	0.4167	0.2220	0.9876	0.9094
Nominal.Pvalue	0.05	0.6055	0.9746	0.1743	0.4971	0.7694	0.6618
FDR.Broad (mean)	0.1	0.5813	0.1011	0.3726	0.1643	0.9968	0.9953
Global.Pvalue	0.05	0.6588	0.5538	0.4167	0.3950	0.9358	0.8622
Connection	0.5	0.9625	0.0055	0.8333	0.9773	0.9121	0.9202

Gene vs. Pathway Level Performance

All Genes = Measured + Extrapolated Rat genes

Stat	Threshold	Pearson Correlation	MSE	Significance Overlap	Sensitivity	Specificity	Concordance Rate
			Gene Level				
Foldchange	2	0.6922	0.3578	0.4992	0.3675	0.9787	0.9048
Nominal.Pvalue	0.05	0.6315	0.9056	0.1841	0.5040	0.7899	0.6942
FDR.Broad (mean)	0.1	0.6170	0.1021	0.4258	0.3662	0.9938	0.9921
Global.Pvalue	0.05	0.7052	0.5840	0.4992	0.5232	0.9088	0.8477
Connection	0.5	0.9709	0.0041	0.8889	0.9796	0.9238	0.9316
			Pathway Level				
ES	0.5	0.6868	0.1415	0.5556	0.6625	0.8063	0.7619
NES	2	0.7108	0.8612	0.5815	0.5840	0.9887	0.9814
Nominal.Pvalue	0.05	0.6094	0.4232	0.5940	0.6957	0.8909	0.8508
FDR.Broad (mean)	0.1	0.8232	0.1442	0.5795	0.7044	0.9471	0.9303
Global.Pvalue	0.05	0.8198	0.5277	0.5815	0.6722	0.8897	0.8539
Connection	0.5	0.9355	0.0121	0.8056	1.0000	0.9257	0.9373

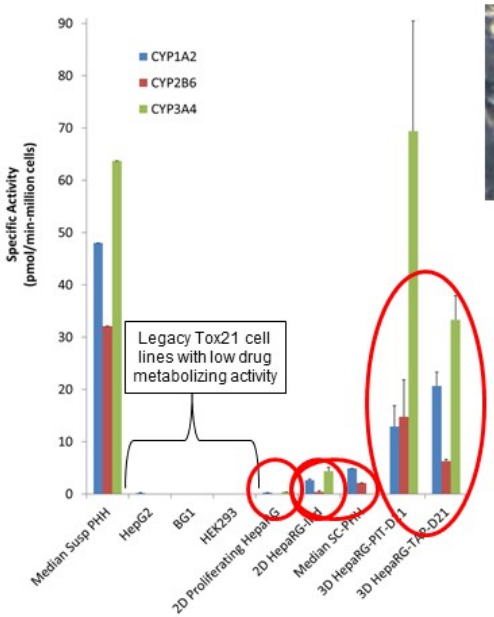
ES = Enrichment Score
NES = Normalized Enrichment Score

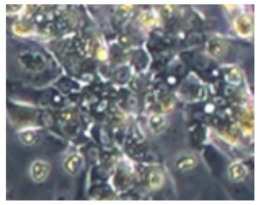
Outline

- Development of the S1500+ Gene Set
- Evaluation of Performance of the Human S1500+ Gene Set
- Evaluation of the S1500+ on the TempO-Seq HT-Transcriptomics (HTT) Platform
- HTT Proof of Concept Study with an *In Vitro* Human Liver Organotypic Model
- Future Directions

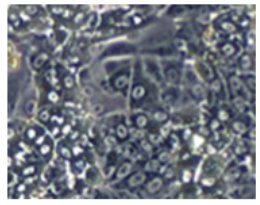
HT Transcriptomics – Initial Human *In Vitro* Model System

Steve Ferguson & Sreeni Ramaiahgari, NTP

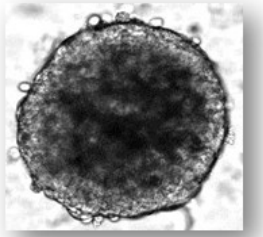




2D HepaRG Proliferating



2D HepaRG Differentiated



3D HepaRG

Ramaiahgari S *et al*. 2017. From the Cover: Three-dimensional (3D) HepaRG Spheroid Model with Physiologically-Relevant Xenobiotic Metabolism Competence and Hepatocyte Functionality for Liver Toxicity Screening. *Toxicol. Sci.*, 159(1): 124-136.

HT Transcriptomics – Initial Study

Steve Ferguson & Sreeni Ramaiahgari, NTP

- **HepaRG *in vitro* Model Treatments – 384 well format**
  - 2D proliferating HepaRG (moderate metabolism)
  - 2D HepaRG differentiated, confluent (induced metabolism)
  - 3D HepaRG (robust metabolism) **– * data yet to come**
  - 24 Compounds
  - 10 Concentration levels, each with triplicate wells
  - 3 Independent runs (different days)
  - 96 hr treatment period with dosing on Day 1 and Day 3 (media changes)
  - Endpoints: morphology, LDH, ROS, metabolomics, targeted P450 metabolism
  - Lysis for gene expression analysis using the Tox21 human S1500+ gene set
  - ~9000 samples analyzed by HTT, generating **~ 25 Million data points**

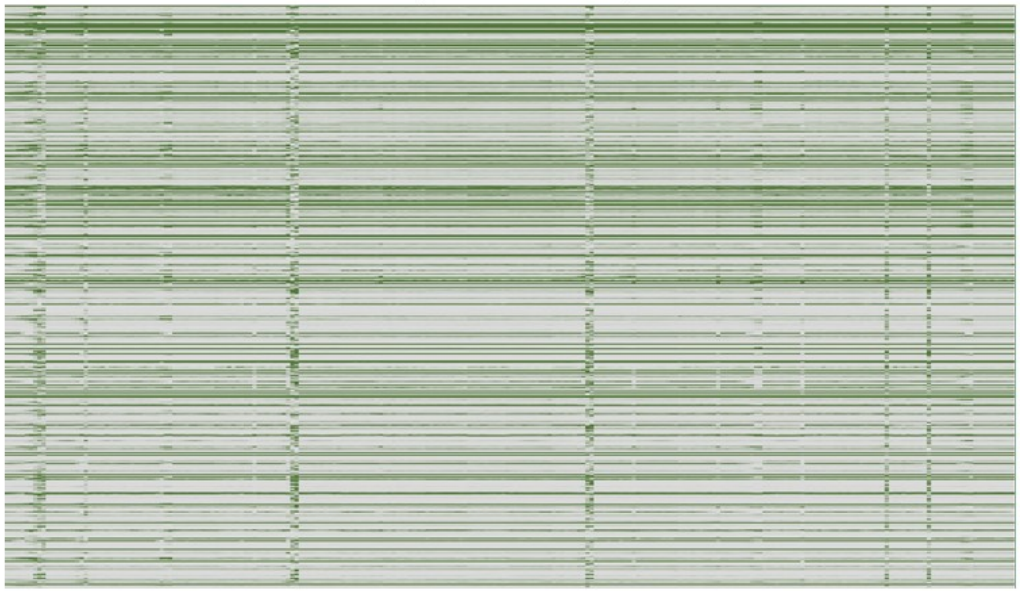
S1500+ HT Transcriptomics – Gene Read Counts

Trey Saddler, NTP

Transcripts Per Million

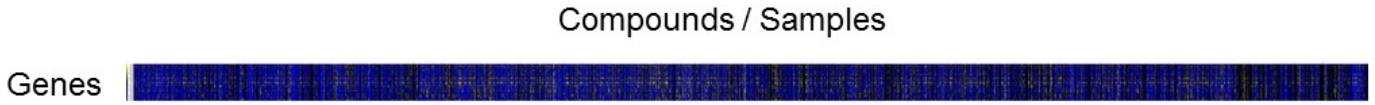
Compounds / Samples

Genes



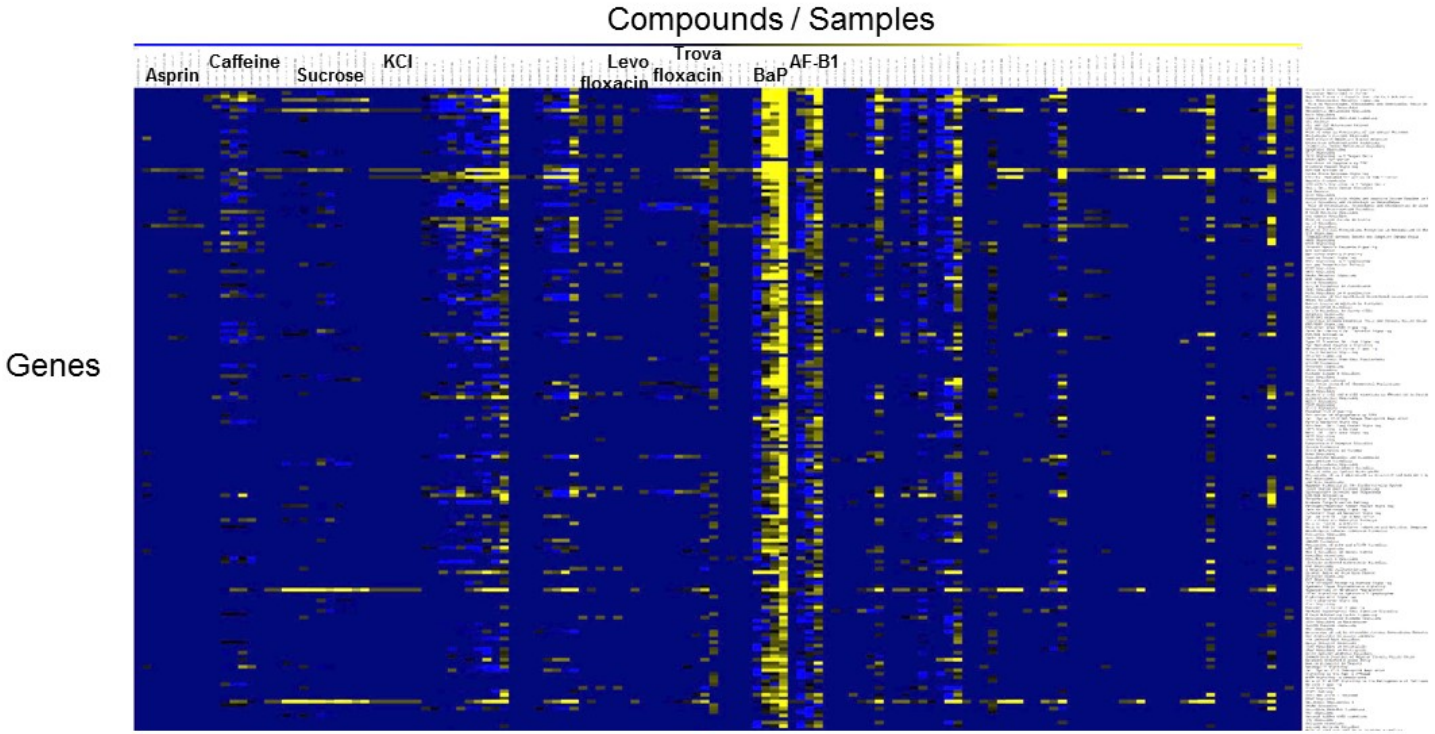
S1500+ HT Transcriptomics – Gene Fold Change

Sreeni Ramaiahgari, NTP



S1500+ HT Transcriptomics – Pathways

Sreeni Ramaiahgari, NTP



High-Throughput Transcriptomics (HTT) at NTP

Moving Tox21 towards Quantitative Systems Toxicology

Hypothesis

- Transcriptome profiling of *in vitro* treated human cells can provide an approximation of human *in vivo* responses to chemical exposures

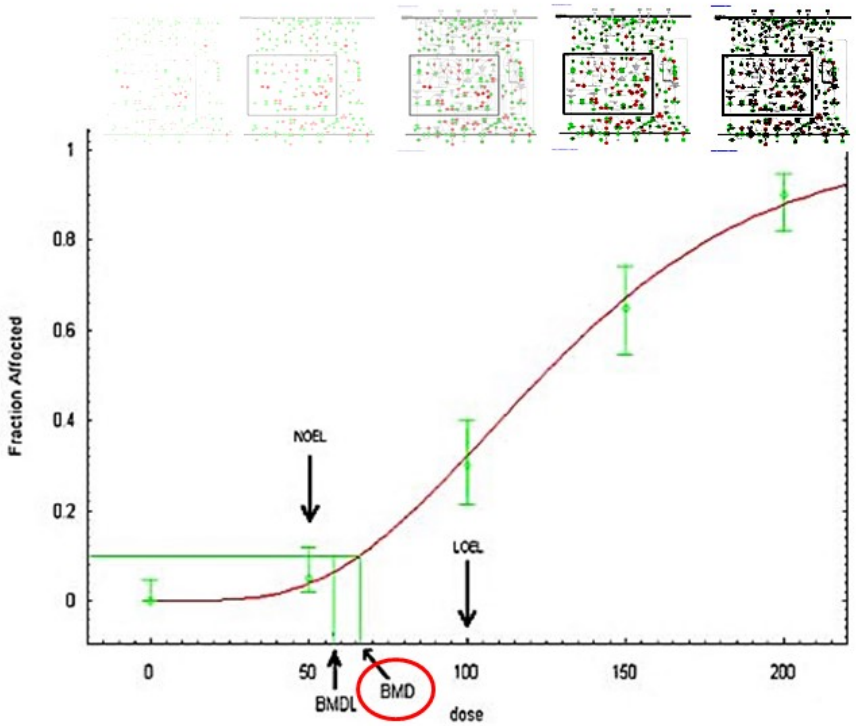
Goal

- To evaluate High-Throughput Transcriptomic analysis of *in vitro* cell models for providing **Bench Mark Dose (BMD)** information relevant to human BMD values and begin to address best practices.

Signature-level Genomic Benchmark Dose

BMDExpress – Rusty Thomas, Hamner, *et al.*

BMDExpress 2.0 – Scott Auerbach, NTP, *et al.*

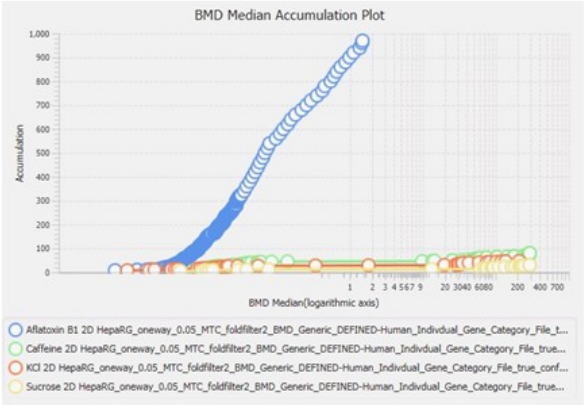


Tox21

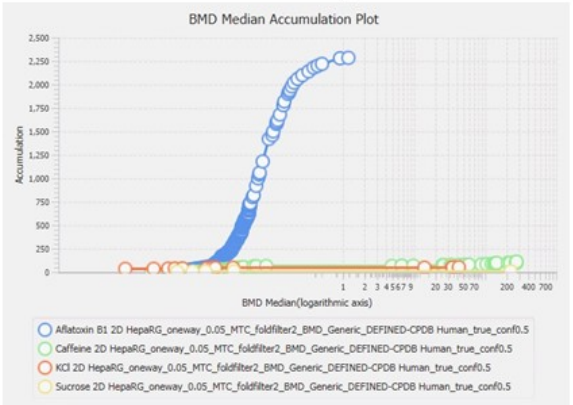
Aflatoxin B1 vs. Controls (2D Confluent) BMDs

Accumulation of significantly affected Genes with their BMD values
Accumulation of significantly affected Pathways with their BMD values

Gene Level BMDs

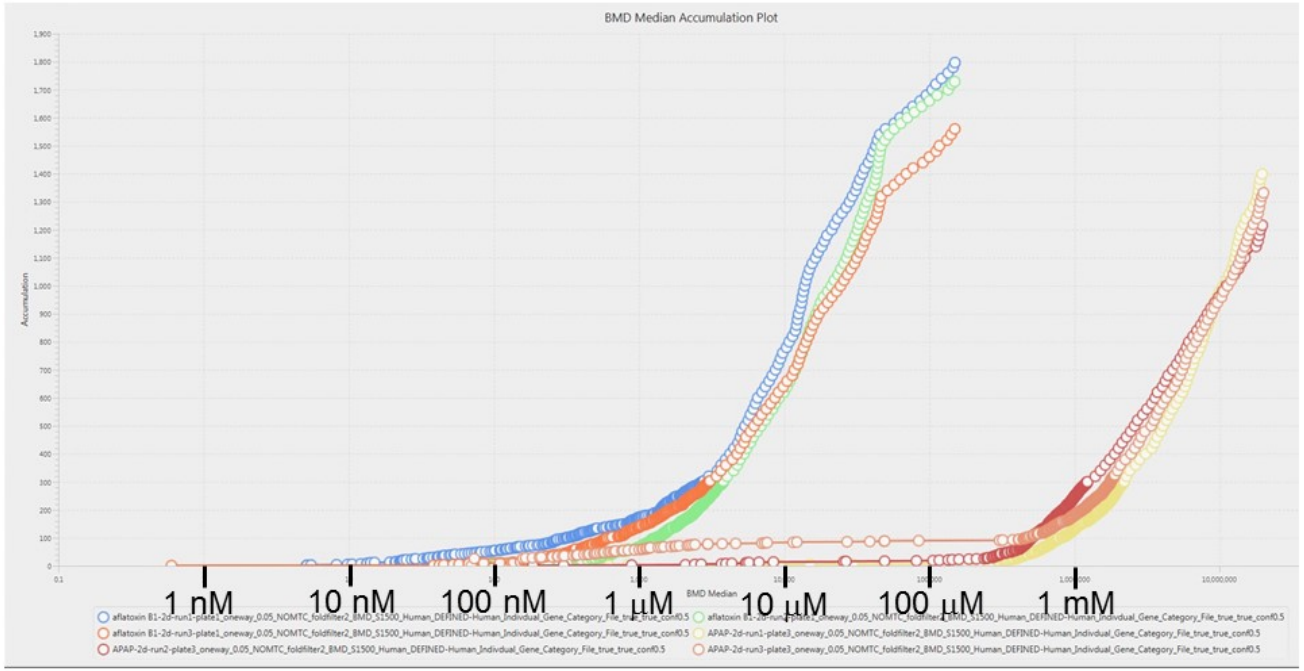


CP-DB Pathway BMDs



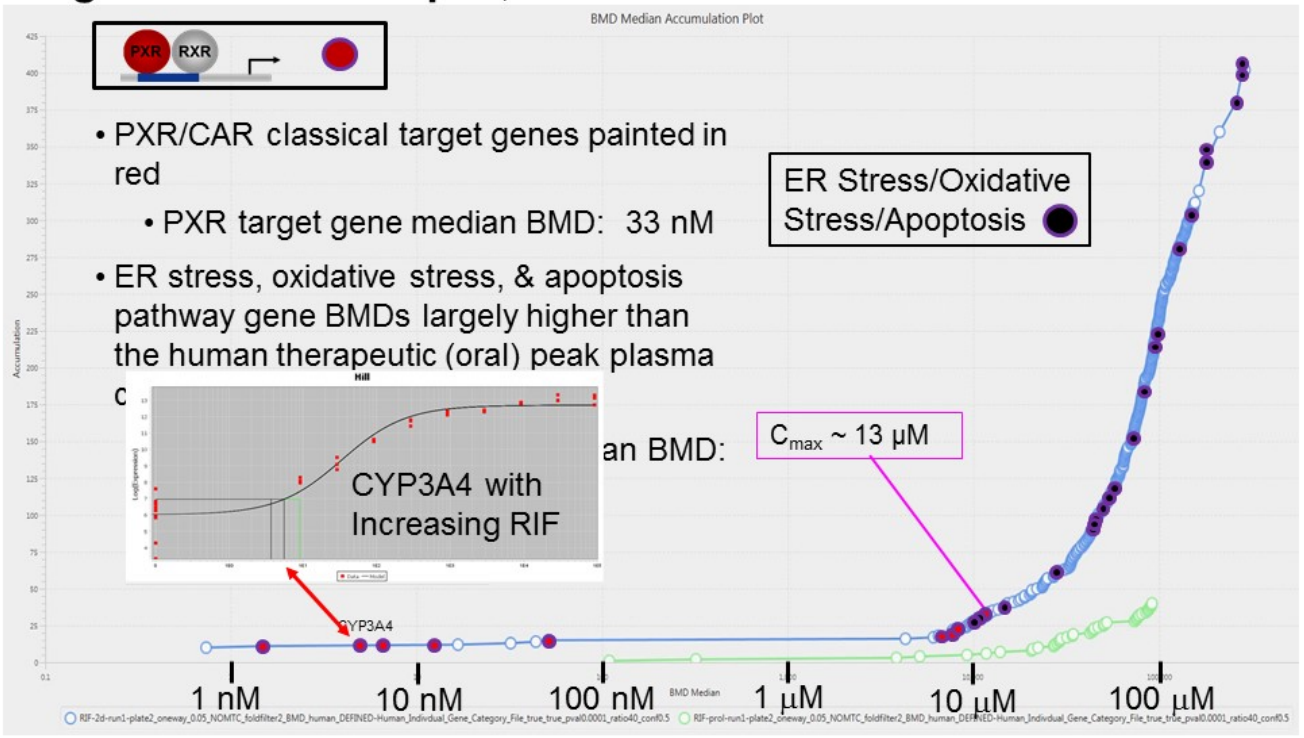
Each dot is a gene or pathway BMD value. The BMDs from a given treatment are linked by a line. The BMDs are ordered from most “potent” to least “potent”. The plot shows the accumulation from rank 1 to the total number of significantly affected genes or pathways. The most “potent” perturbed gene or pathway BMDs are on the left and the least “potent” are on the right

Steve Ferguson, NTP – Gene Level Median BMD Values



Rifampicin Activation of PXR Signaling via CYP3A4 Induction

Steve Ferguson & Nisha Sipes, NTP – Gene Level Median BMD Values

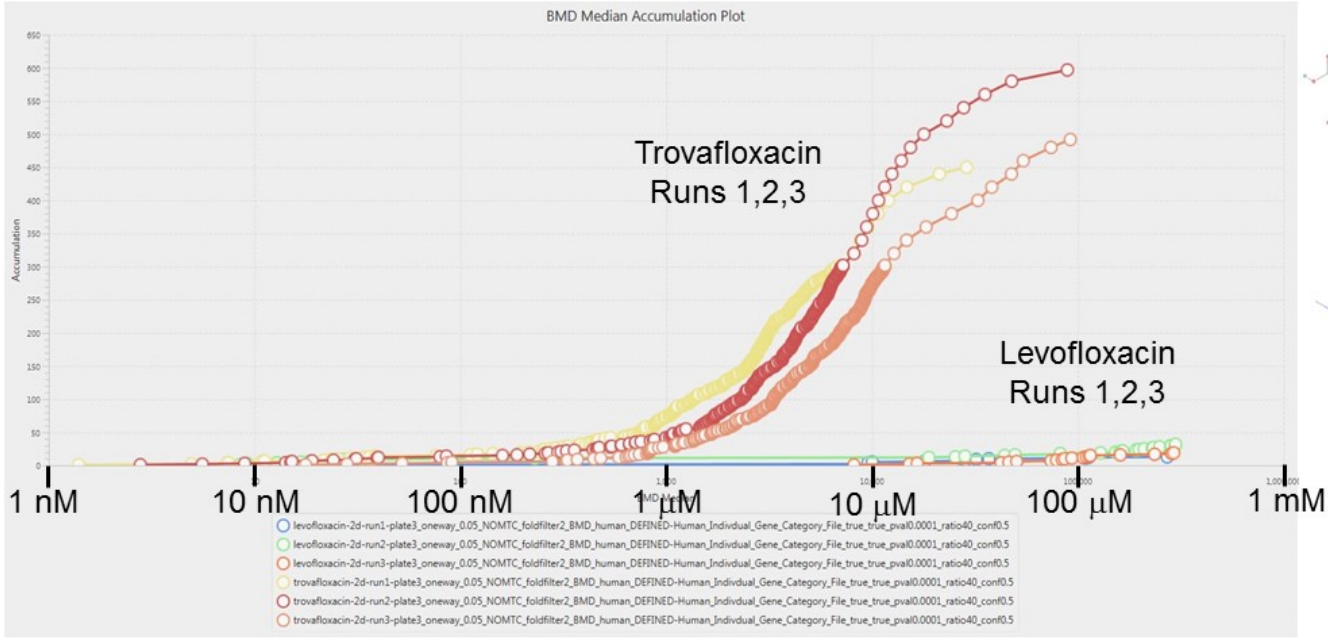


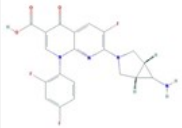
Tox21

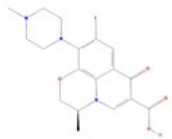
Case Study Comparison of Drug Analogues

Steve Ferguson, NTP – Gene Level Median BMD Values

Trovafloxacin vs. Levofloxacin



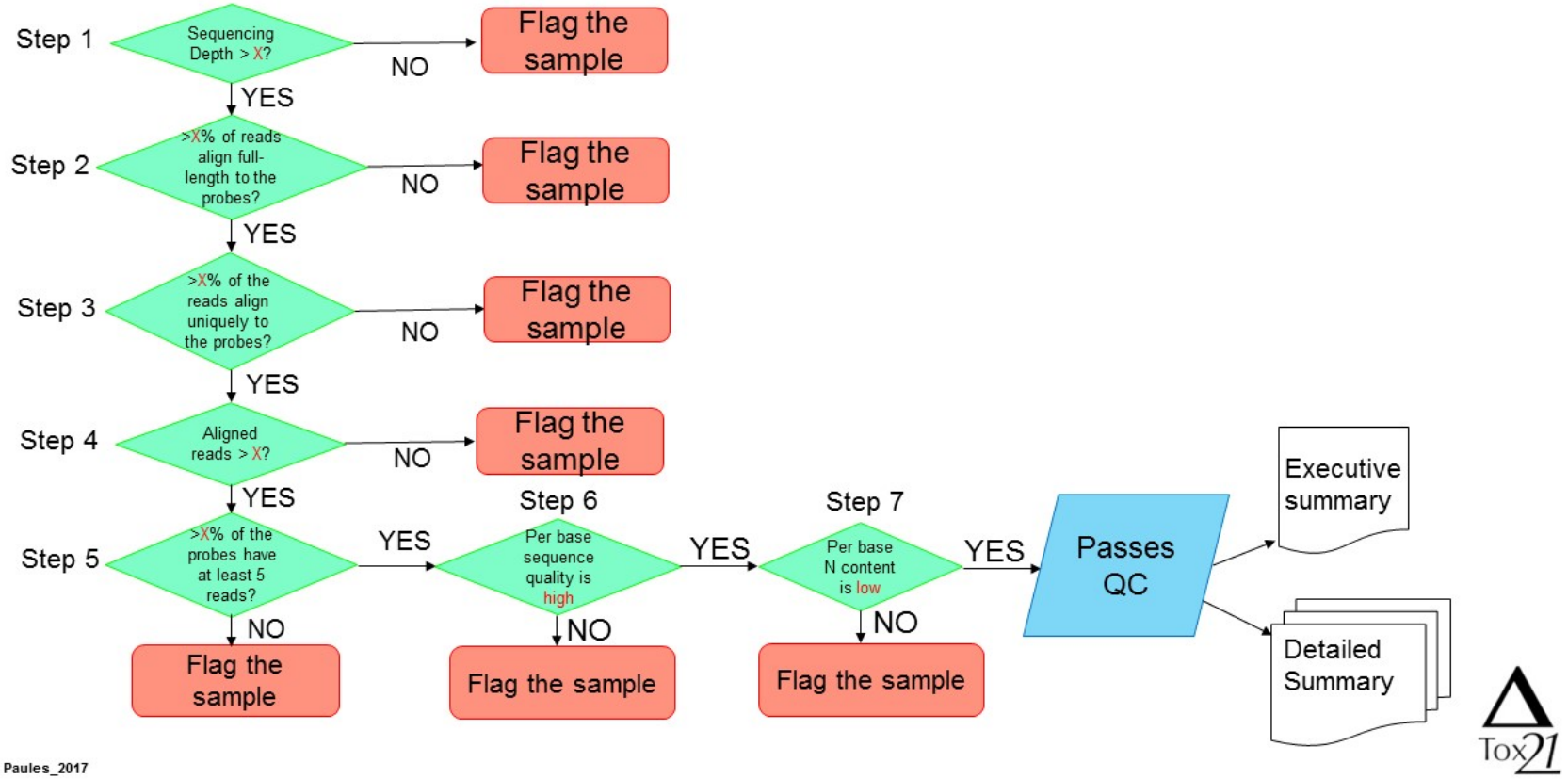




Outline

- Development of the S1500+ Gene Set
- Evaluation of Performance of the Human S1500+ Gene Set
- Evaluation of the S1500+ on the TempO-Seq HT-Transcriptomics (HTT) Platform
- HTT Proof of Concept Study with an In Vitro Human Liver Organotypic Model
- Future Directions

NTP HT Transcriptomics QC Workflow (under construction)



Moving Forward

- S1500+ development for **human** (complete)**, rat & mouse** (near completion)**, zebrafish** (initiated)
- HTT **screening** for **prioritization** – BioSimilarity, Biological Read-Across
  - e.g. Perfluorinated Compounds (PFOA, PFOS), Botanicals, Polycyclic Aromatic Compounds, Flame Retardants, etc.
- Integrate High Content Imaging with qHTS & HT-T – *Phenotypic Anchoring*
- Integrate Metabolomics data from NTP Labs with HT-T data – PBPK, etc.
- Application to biomaterial from NTP rat and mouse studies
  - Acute (5 Day) Rat Transcriptomics Studies
  - Diversity Outbred mouse models for screening for Genetic Variation in responses to chemical exposures
  - Rat and Mouse Archived FFPE material

Tox21: A Collaboration of Many…

NTP BSB

Scott Auerbach
Steve Ferguson
Sreenivasa Ramaiahgari
Nisha Sipes
Alex Merrick
Kristine Witt
Stephanie Smith-Roe
Alison Harrill
Julie Foley
Jui-Hua Hsieh
Fred Parham
Trey Saddler
Ray Tice (Retired)

BioSpyder Team

NTP / NIEHS

Mike DeVito
Warren Casey
Elizabeth Maull
Nicole Kleinstreuer
Suramya Waidyanatha
Windy Boyd
Mamta Behl
Brad Collins
Jennifer Fostel
Keith Shockley
Grace Kissling

John Bucher
Linda Birnbaum

SciOme

Deepak Mav
Ruchir Shah
Dan Svoboda

SciOme

Tox21 Colleagues

NCATS

Anton Simeonov
Chris Austin
Menghang Xia
Ruili Huang
Dave Gerhold
Anna Rossoshek

US EPA

Rusty Thomas
Kevin Crofton
Richard Judson
Ann Richard
Bob Kavlock

US FDA

Suzy Fitzpatrick
Dan Acosta
Donna Mendrick
Weida Tong

Tox21

NTP
National Toxicology Program

Thank you!

Questions?