United States
Environmental Protection
Agency

High-throughput transcriptomic (HTTr) screening at USEPA: Quality Control, Plate Effects and Concentration-Response Modeling

Joshua A. Harrill, USEPA National Center for Computational Toxicology (NCCT)

EUToxRisk-Tox21 Satellite Meeting
SOT Annual Meeting, Baltimore, MD
March 12th, 2019

Disclaimer

The views expressed in this presentation are those of the author(s) and do not necessarily represent the views or policies of the U.S. Environmental Protection Agency, nor does mention of trade names or products represent endorsement for use.

Disclaimer

- **USEPA HTTr Screening Strategy**
  - Background
  - TempO-Seq Platform
- **Quality Control**
  - Standardized Reference Materials
  - Plate-Based Controls
- **Batch/Plate Effects**
- **Concentration-Response Modeling**

The Next Generation of Computational Toxicology at USEPA

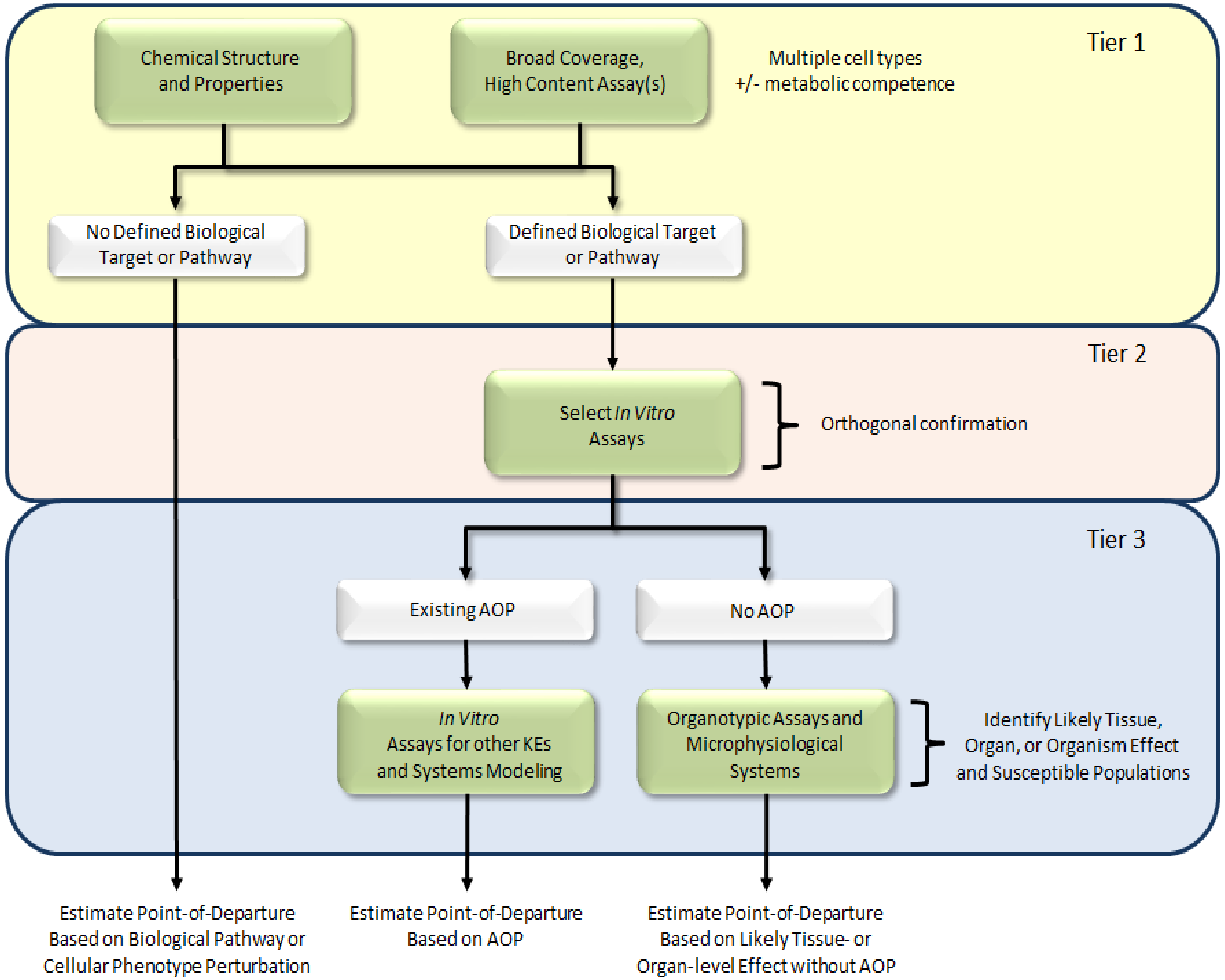
- **Tier 1 assays:**
  - Broad coverage
  - High throughput
  - Conc.-response mode
  - High content outputs
  - Tractable across many cell types / assay formats

- Increasing efficiency and declining cost has made **high-throughput transcriptomics (HTTr)** a practical option for broad coverage *in vitro* chemical screening.

- Bioactivity-based **potency estimates** can be used to identify *in vitro* **bioactivity thresholds**.

- Gene expression **profiles** can potentially be used for **mechanistic prediction** and evaluation of chemical similarity.

Tiered Hazard Evaluation Approach



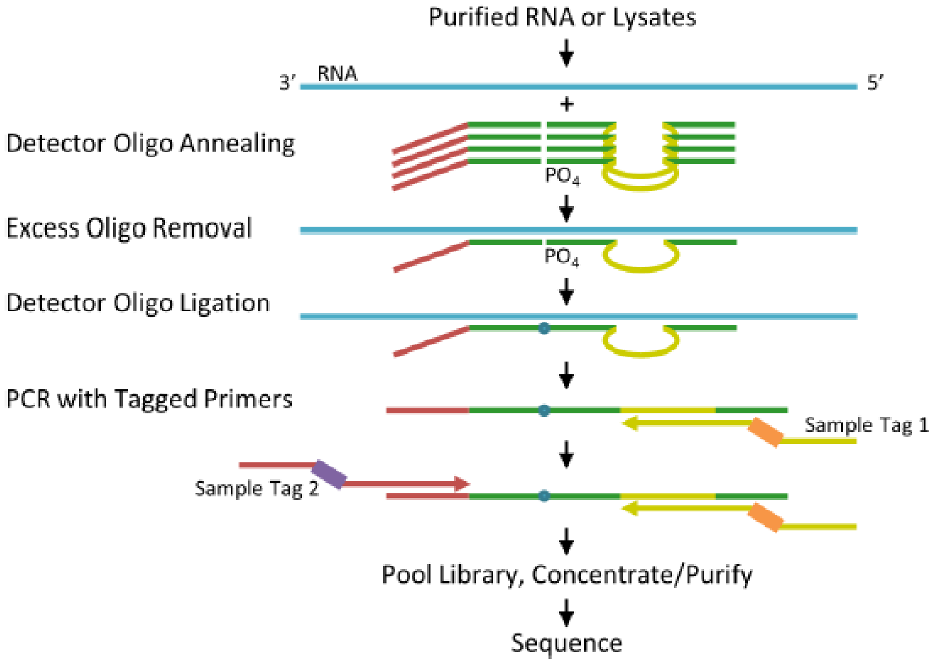
Tier 1
Chemical Structure and Properties
Broad Coverage, High Content Assay(s)
Multiple cell types +/- metabolic competence
No Defined Biological Target or Pathway
Defined Biological Target or Pathway
Tier 2
Select *In Vitro* Assays
Orthogonal confirmation
Tier 3
Existing AOP
No AOP
In Vitro Assays for other KEs and Systems Modeling
Organotypic Assays and Microphysiological Systems
Identify Likely Tissue, Organ, or Organism Effect and Susceptible Populations
Estimate Point-of-Departure Based on Biological Pathway or Cellular Phenotype Perturbation
Estimate Point-of-Departure Based on AOP
Estimate Point-of-Departure Based on Likely Tissue- or Organ-level Effect without AOP

Templated Oligo with Sequencing Readout (TempO-Seq)

Technology

- The **TempO-Seq** human whole transcriptome assay measures the expression of greater than 20,000 transcripts.
- Requires only picogram amounts of total RNA per sample.
- Compatible with purified RNA samples or **cell lysates**.
- Transcripts in cell lysates generated in 384-well format are barcoded according to well position and combined in a single library for sequencing using industry standard instrumentation.
- Scalable, targeted assay:
  - 1) specifically measures transcripts of interest
  - 2) ~50-bp reads for all genes
  - 3) requires less flow cell capacity than RNA-Seq
- Per sample fastq files are generated and aligned to BioSpyder sequence manifest to generate integer count tables.

TempO-Seq Assay Illustration



HTTr Screening in MCF-7 Cells → Experimental Design

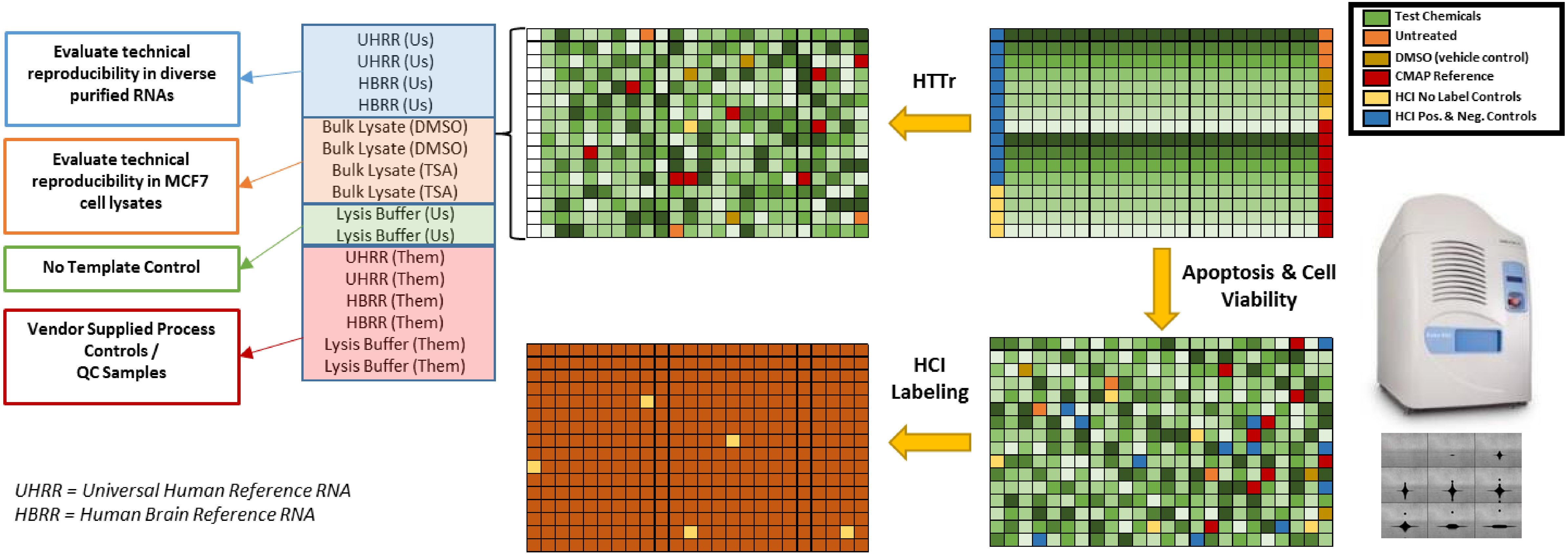
	Study 1: Pilot Screen		Study 2: Large Scale Screen	
Parameter	**Multiplier**	**Notes**	**Multiplier**	**Notes**
Cell Type(s)	1	MCF-7	1	MCF-7
Culture Condition	2	DMEM + 10% HI-FBS PRF-DMEM + 10% CS-HI-FBS	1	DMEM + 10% HI-FBS [a]
Chemicals	44	Mechanistic Diversity w/ Redundancy	2,112 (63)*	ToxCast ph1, ph2, e1k / ph3
Time Points:	3	6, 12, 24 hours	1	6 hours
Assay Formats:	2	TempO-Seq HCI Cell Viability & Apoptosis	2	TempO-Seq HCI Cell Viability & Apoptosis
Concentrations:	8	3.5 log$_{10}$ units; semi log$_{10}$ spacing	8	3.5 log$_{10}$ units; semi log$_{10}$ spacing
Biological Replicates:	3	--	3	--

*63 Chemicals were screened in duplicate.

Treatment Randomization & Quality Control Samples

Treatment Randomization: *Each test plate uniquely randomized with respect to treatment.*

QC Samples: *Quality Control samples included on each plate*



Evaluate technical reproducibility in diverse purified RNAs

- UHRR (Us)
- UHRR (Us)
- HBRR (Us)
- HBRR (Us)

Evaluate technical reproducibility in MCF7 cell lysates

- Bulk Lysate (DMSO)
- Bulk Lysate (DMSO)
- Bulk Lysate (TSA)
- Bulk Lysate (TSA)

No Template Control

- Lysis Buffer (Us)
- Lysis Buffer (Us)

Vendor Supplied Process Controls / QC Samples

- UHRR (Them)
- UHRR (Them)
- HBRR (Them)
- HBRR (Them)
- Lysis Buffer (Them)
- Lysis Buffer (Them)

🟩	Test Chemicals
🟧	Untreated
🟫	DMSO (vehicle control)
🟥	CMAP Reference
🟨	HCI No Label Controls
🟦	HCI Pos. & Neg. Controls

UHRR = Universal Human Reference RNA
HBRR = Human Brain Reference RNA

Block and Plate Group Design

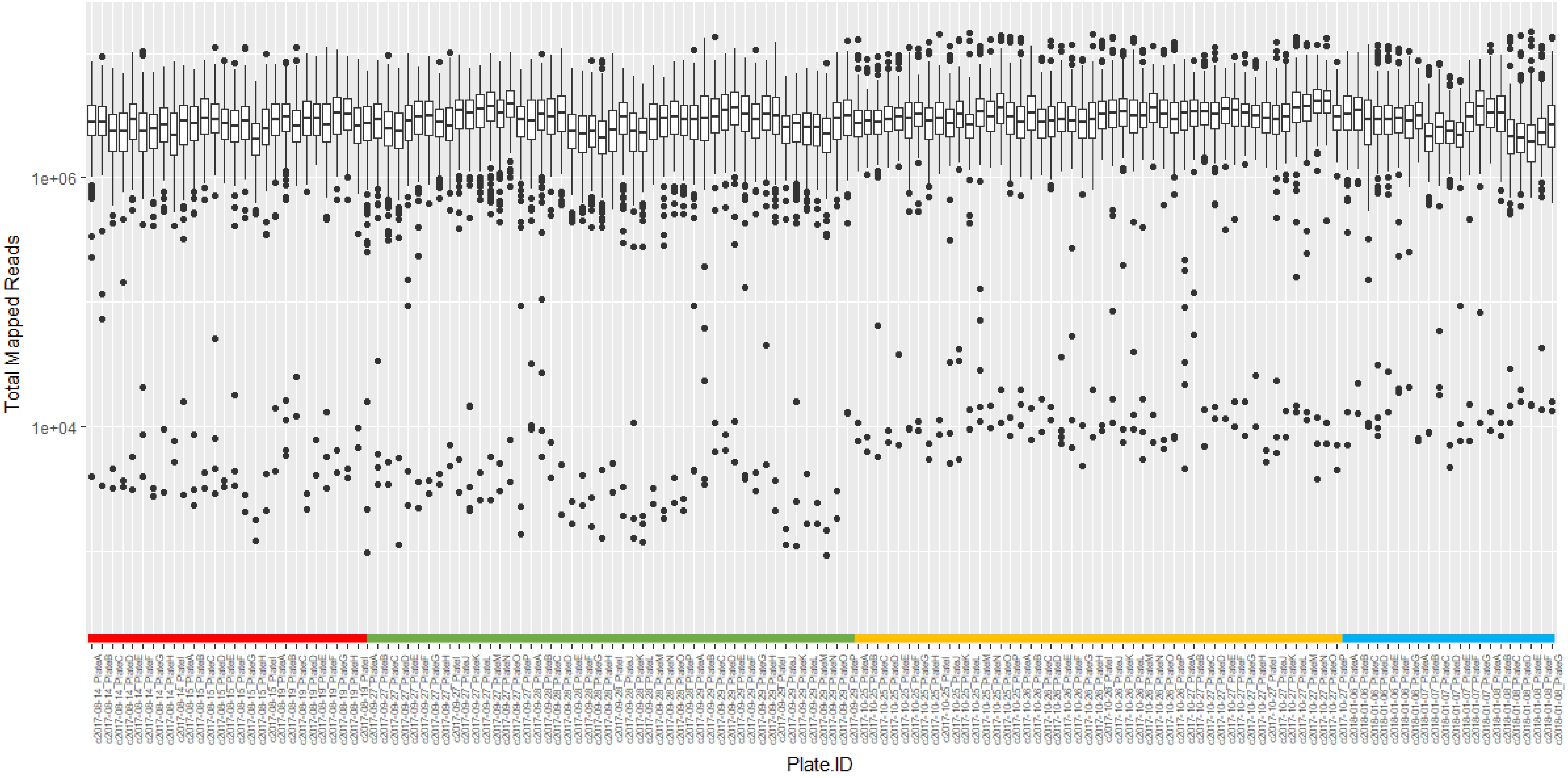
n = 4 study blocks
n = 144 plates
n = 48 plate groups

Plates within a plate group contain the biological replicates for a particular chemical x concentration.

***Note:** Block 4 was aborted due to an instrument malfunction.*

	c2017-08-14	c2017-08-15	c2017-08-19	c2017-09-27	c2017-09-28	c2017-09-29	c2017-10-25	c2017-10-26	c2017-10-26	c2018-01-06	c2018-01-07	c2018-01-08	
pg1	TC00284655	TC00284691	TC00503564										Block 1
pg2	TC00284656	TC00284692	TC00503565										
pg3	TC00284657	TC00284693	TC00503566										
pg4	TC00284658	TC00284694	TC00503567										
pg5	TC00284659	TC00284695	TC00503568										
pg6	TC00284660	TC00284696	TC00503569										
pg7	TC00284661	TC00284697	TC00503570										
pg8	TC00284662	TC00284698	TC00503571										
pg9	TC00284663	TC00284699	TC00503572										
pg10				TC00503636	TC00503868	TC00503900							Block 2
pg11				TC00503637	TC00503869	TC00503901							
pg12				TC00503638	TC00503870	TC00503902							
pg13				TC00503639	TC00503871	TC00503903							
pg14				TC00503640	TC00503872	TC00503904							
pg15				TC00503641	TC00503873	TC00503905							
pg16				TC00503642	TC00503874	TC00503906							
pg17				TC00503643	TC00503875	TC00503907							
pg18				TC00503644	TC00503876	TC00503908							
pg19				TC00503645	TC00503877	TC00503909							
pg20				TC00503646	TC00503878	TC00503910							
pg21				TC00503647	TC00503879	TC00503911							
pg22				TC00503648	TC00503880	TC00503912							
pg23				TC00503649	TC00503881	TC00503913							
pg24				TC00503650	TC00503882	TC00503914							
pg25				TC00503651	TC00503883	TC00503915							
pg26							TC00503932	TC00503964	TC00503996				Block 3
pg27							TC00503933	TC00503965	TC00503997				
pg28							TC00503934	TC00503966	TC00503998				
pg29							TC00503935	TC00503967	TC00503999				
pg30							TC00503936	TC00503968	TC00504000				
pg31							TC00503937	TC00503969	TC00504001				
pg32							TC00503938	TC00503970	TC00504002				
pg33							TC00503939	TC00503971	TC00504003				
pg34							TC00503940	TC00503972	TC00504004				
pg35							TC00503941	TC00503973	TC00504005				
pg36							TC00503942	TC00503974	TC00504006				
pg37							TC00503943	TC00503975	TC00504007				
pg38							TC00503944	TC00503976	TC00504008				
pg39							TC00503945	TC00503977	TC00504009				
pg40							TC00503946	TC00503978	TC00504010				
pg41							TC00503947	TC00503979	TC00504011				
pg42										TC00504082	TC00504100	TC00504118	Block 5
pg43										TC00504083	TC00504101	TC00504119	
pg44										TC00504084	TC00504102	TC00504120	
pg45										TC00504085	TC00504103	TC00504121	
pg46										TC00504086	TC00504104	TC00504122	
pg47										TC00504087	TC00504105	TC00504123	
pg48										TC00504088	TC00504106	TC00504124	

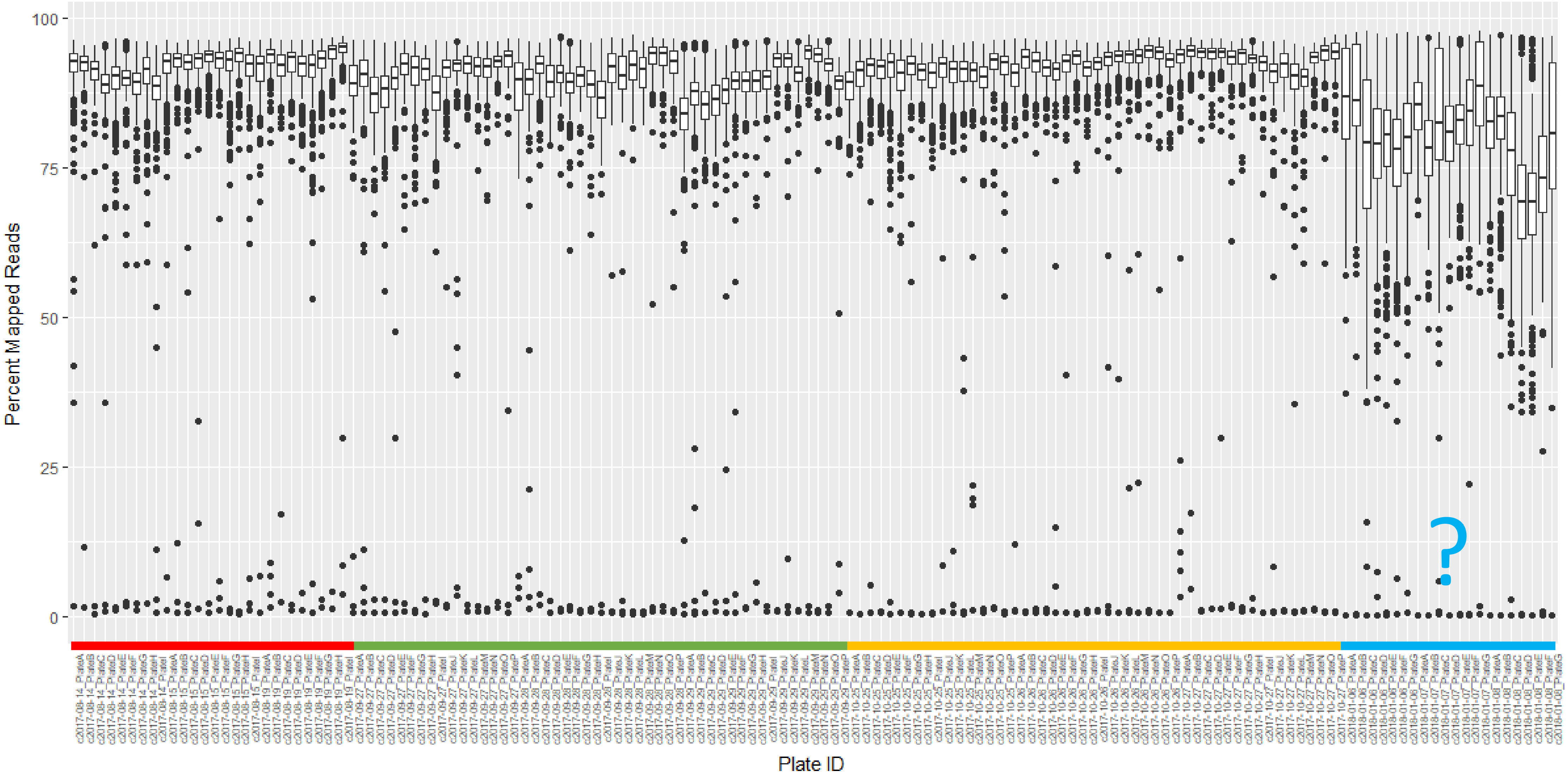
Total Mapped Reads, By Plate



The distribution of total mapped reads is remarkably consistent across plates and blocks.

All Plates (n = 144)

Percent Mapped Reads, By Plate



However, the distribution of mapping rate is very different for Block 5

All Plates (n = 144)

Plate-wise Comparison of FC to Global FC

- Start with data from QC samples of each type:
  - **Standardized Reference Materials:** UHRR, HBRR, BL_DMSO, BL_TSA
  - **Reference Treatments and Vehicle Controls:** DMSO, TSA, SIRO, GEN, Untreated

- For each sample type, create a **matrix of samples x probes** → log2(counts)
  - Eliminate all probes with <95% of samples having non-zero values
  - Replace remaining NULL (i.e. zero) values with 0.5
  - Normalize each sample to 10^6 counts before taking log2

- For each sample type, create a **global median count** profile using data across all plates

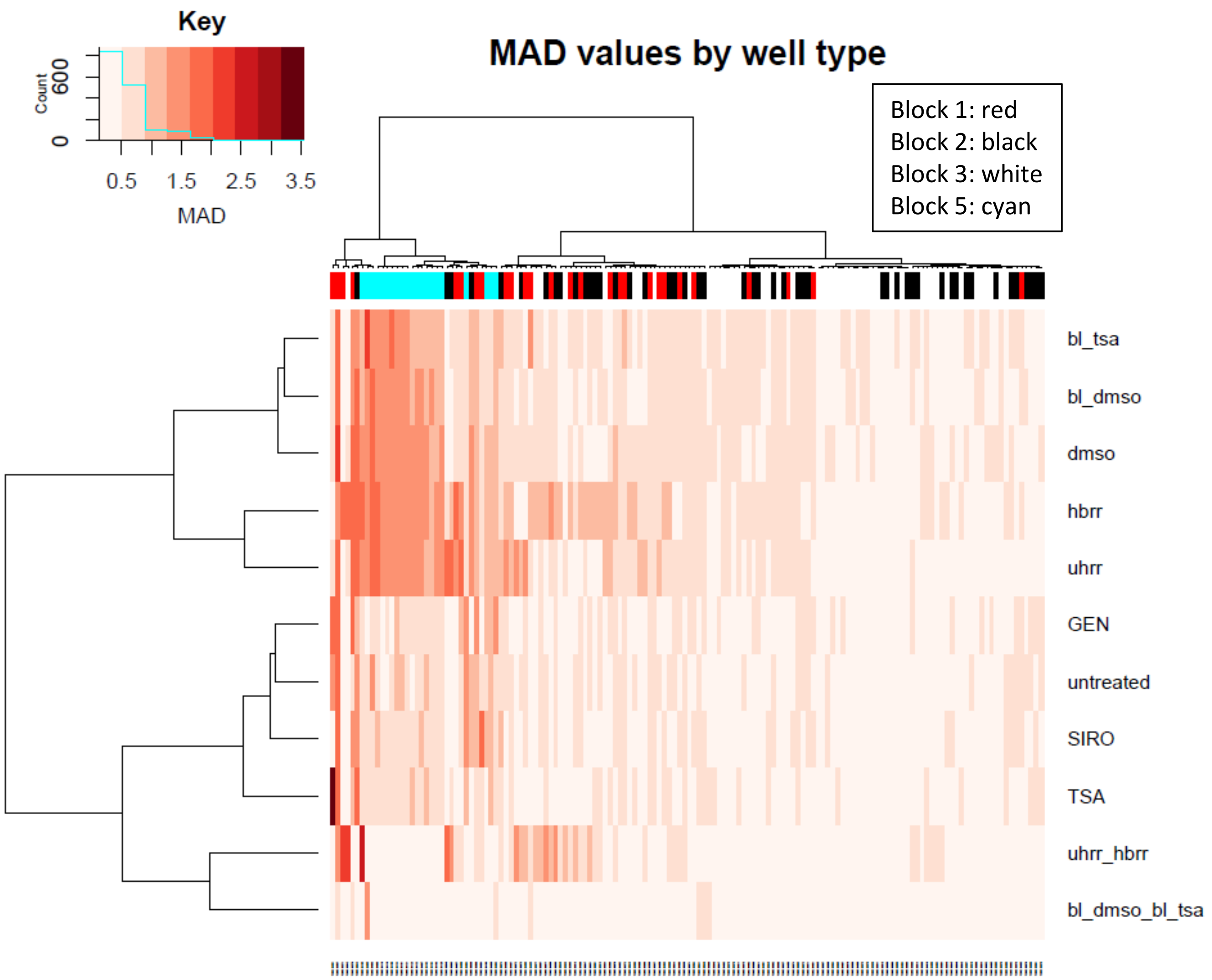
- For each QC sample pairing of interest:
  - Calculate a l2fc matrix on each plate
  - Calculate a global l2fc matrix using the global median count profile
  - Plot platewise l2fc versus global l2fc
  - Calculate median absolute deviation (MAD) of residuals

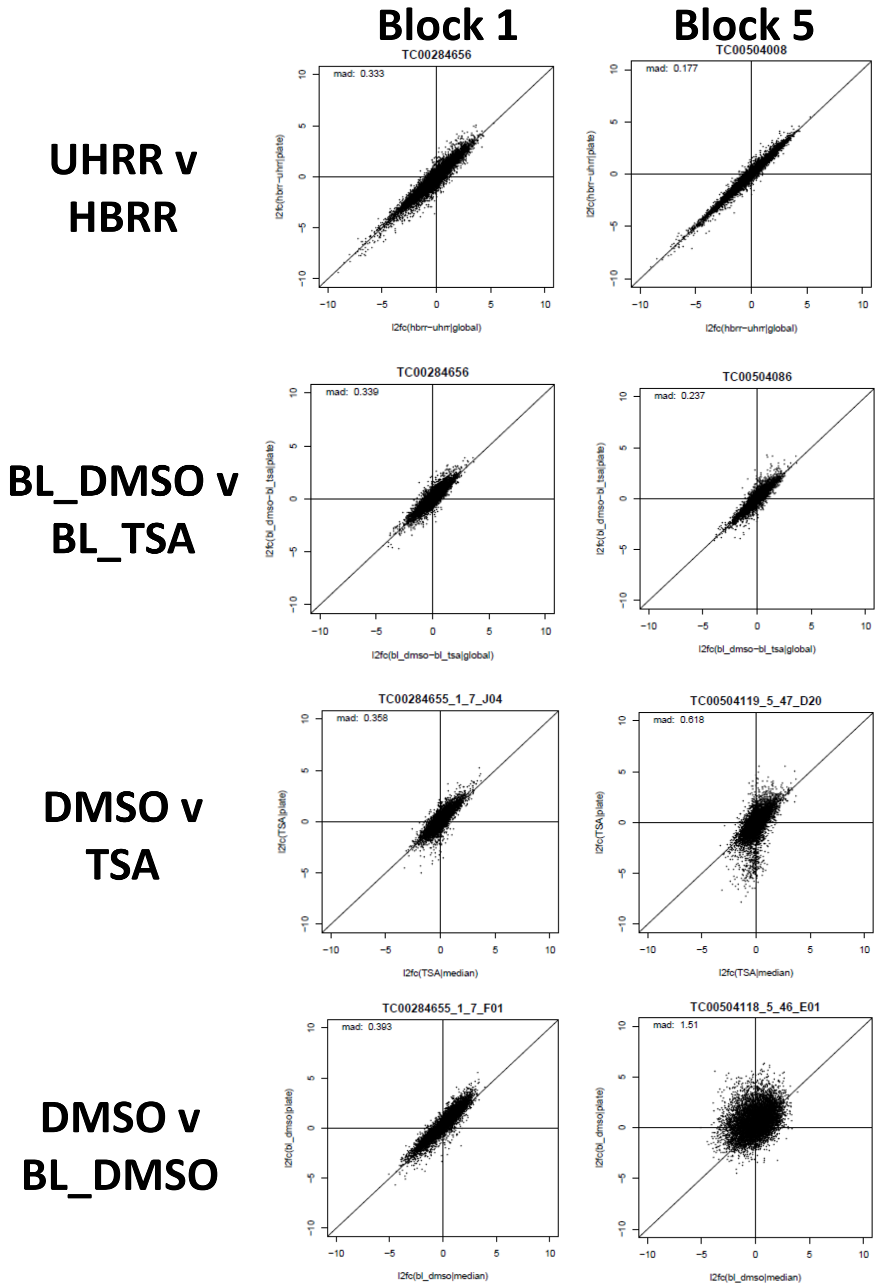
Comparisons of Interest

UHRR	HBRR
BL_DMSO	BL_TSA
DMSO	TSA
DMSO	SIRO
DMSO	GEN
DMSO	Untreated

***Also compared DMSO to each of the Standardized Reference Materials.*

MAD of Residuals as a Plate Level QC Flag



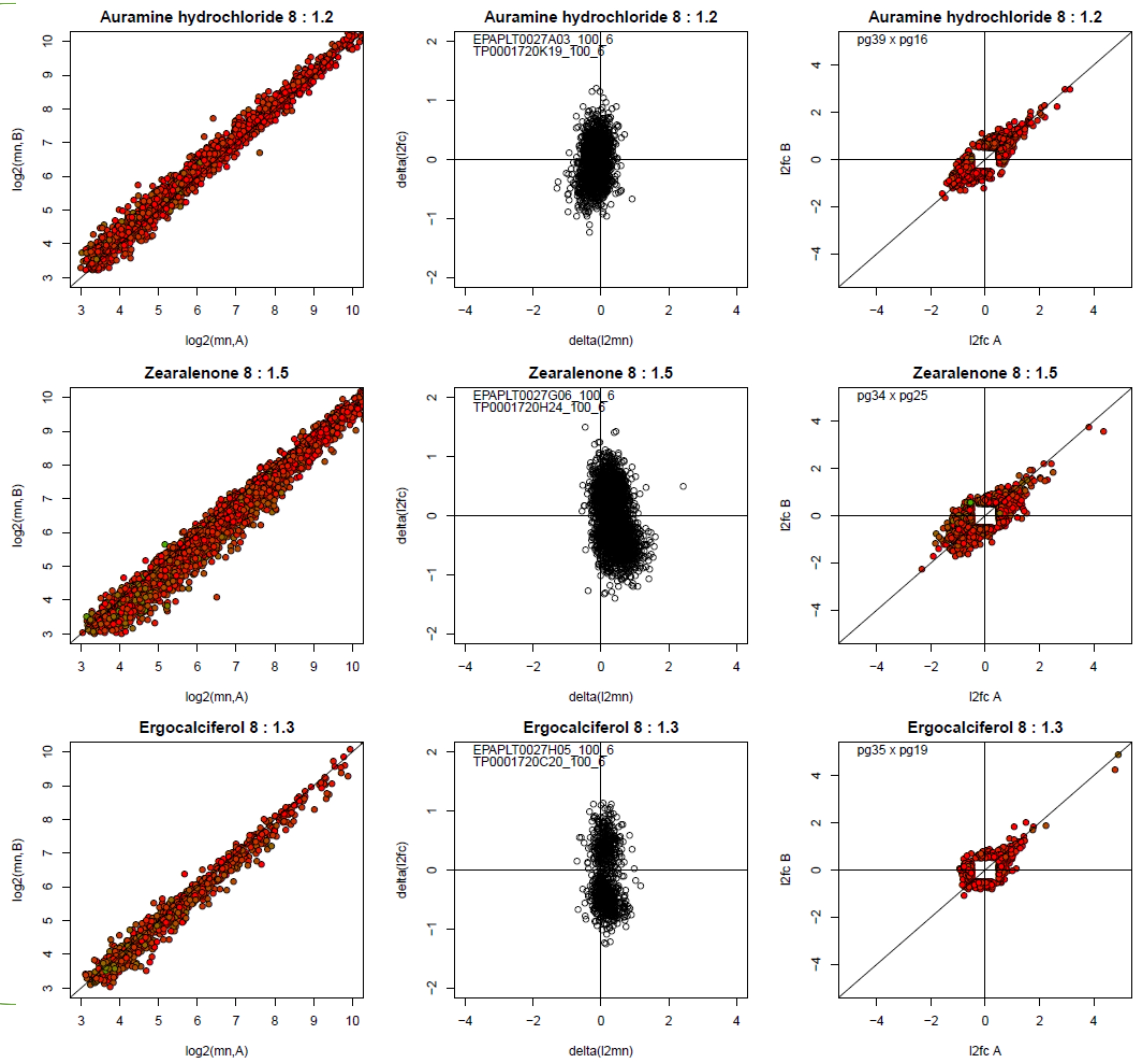


What is the implication of poor performing reference chemical treatments on the reliability of screening results?

Figures courtesy of Richard Judson

Reproducibility of Duplicate Chemicals

	pg02	pg08	pg15	pg16	pg18	pg19	pg20	pg21	pg24	pg25
pg02										
pg08	B									
pg15										
pg16										
pg18										
pg19										
pg20										
pg21										
pg24										
pg25										
pg30			GG	G	B			B	G	
pg31				G		G			G	
pg34				G		G		G		G
pg35			B			G	M			G
pg39				G					G	M
pg40						G	G		GG	G
pg41			G			G		GG		
pg42					B	B	B	B		
pg43			B							
pg44							G		M	B
pg45				B			BBB	B	B	BM
pg46			BB	BB	B		M	B		
pg47					G	BM		B	B	
pg48				M			G	B		

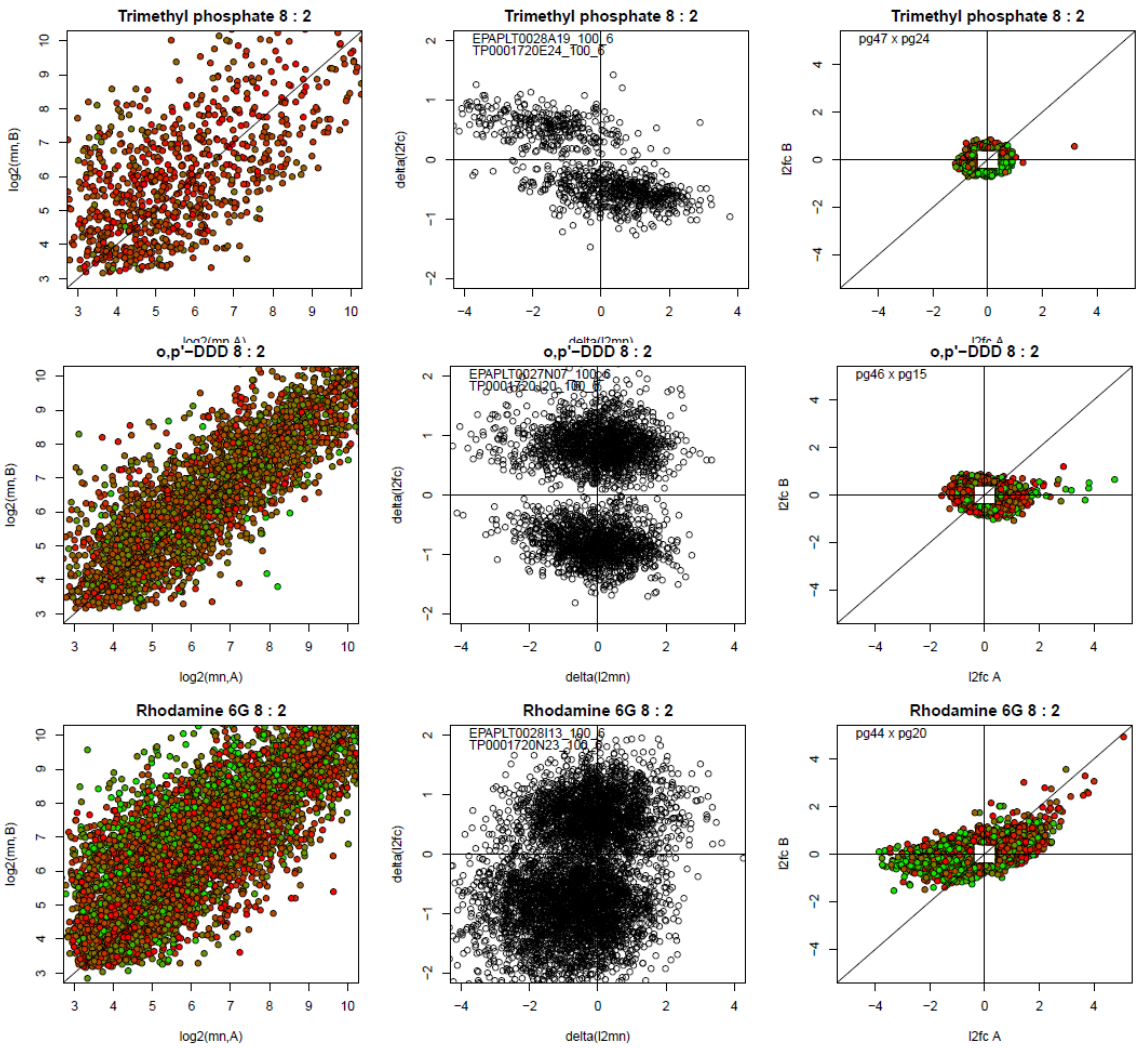


- “Good” performance of plate-based reference chemical treatments in blocks 2 & 3.
- Screening results of duplicates across blocks 2 & 3 were highly correlated

Figures courtesy of Richard Judson

Reproducibility of Duplicate Chemicals

	pg02	pg08	pg15	pg16	pg18	pg19	pg20	pg21	pg24	pg25
pg02										
pg08	B									
pg15										
pg16										
pg18										
pg19										
pg20										
pg21										
pg24										
pg25										
pg30			GG	G	B			B	G	
pg31				G		G			G	
pg34				G		G		G		G
pg35			B			G	M			G
pg39				G					G	M
pg40						G	G		GG	G
pg41			G			G		GG		
pg42					B	B	B	B		
pg43			B							
pg44							G		M	B
pg45				B			BBB	B	B	BM
pg46			BB	BB	B		M	B		
pg47					G	BM		B	B	
pg48				M			G	B		



- “Poor” performance of plate-based reference chemical treatments in Block 5.
- Duplicate comparisons involving block 5 were poorly correlated

Figures courtesy of Richard Judson

MAQC Replacement Project

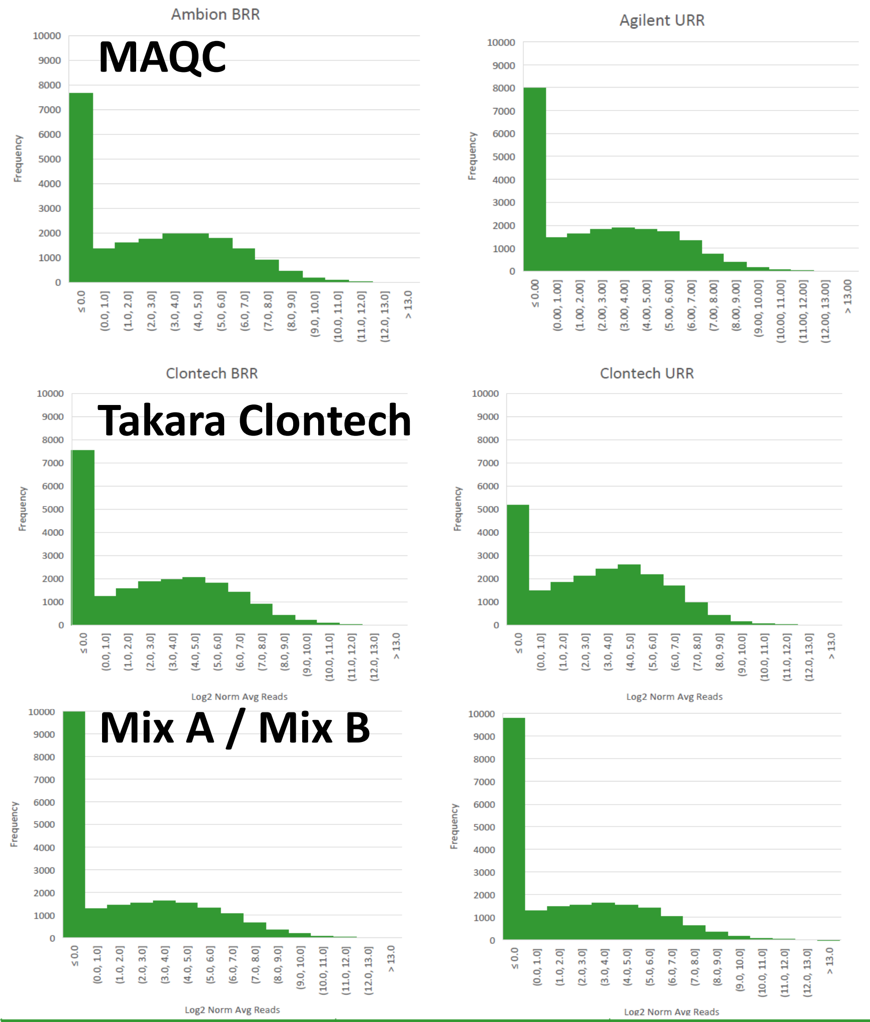
Background: NCCT initially envisioned using MAQC human reference mRNAs as a QC sample pairing
One of the commercially-available MAQC human reference mRNA samples has been discontinued.

Objective: Produce two human-derived purified mRNA products as replacement for MAQC reference mRNAs.
Produce two analogous lysate products as a commercially available lysate standards.

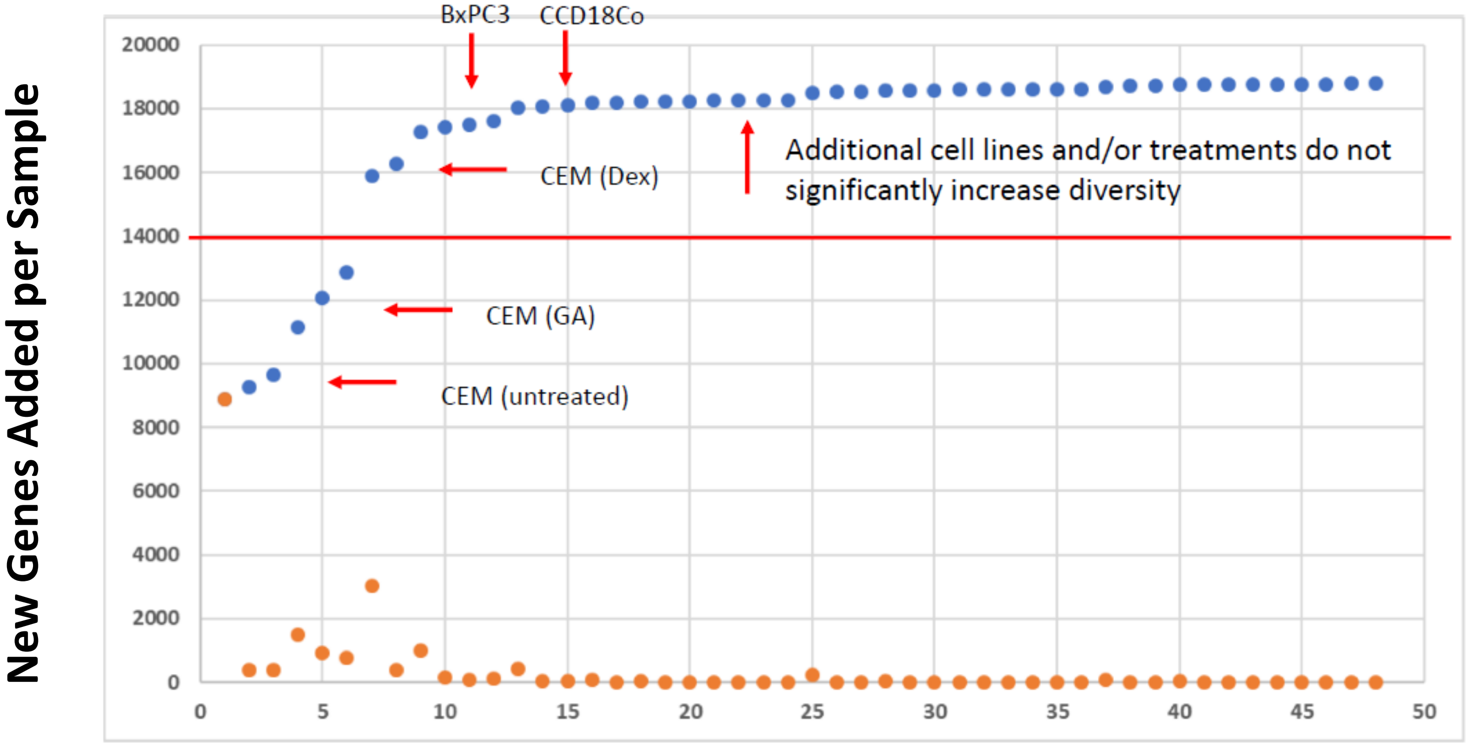
Approach: Use a combination of human-derived cell lines (with pharmacological treatments) to produce two reference RNA / lysate pools with similar number of expressed genes and dynamic range of fold-change as the MAQC samples.

Demonstrate performance using the TempO-Seq whole transcriptome assay.

MAQC Replacement Project



	max log2 fold difference	min log2 fold difference
Thermo Brain/ Agilent URR	13.297	-16.425
Lysate Mix A/ Lysate Mix B	13.760	-13.941
Clontech Brain/ Clontech URR	16.162	-7.035



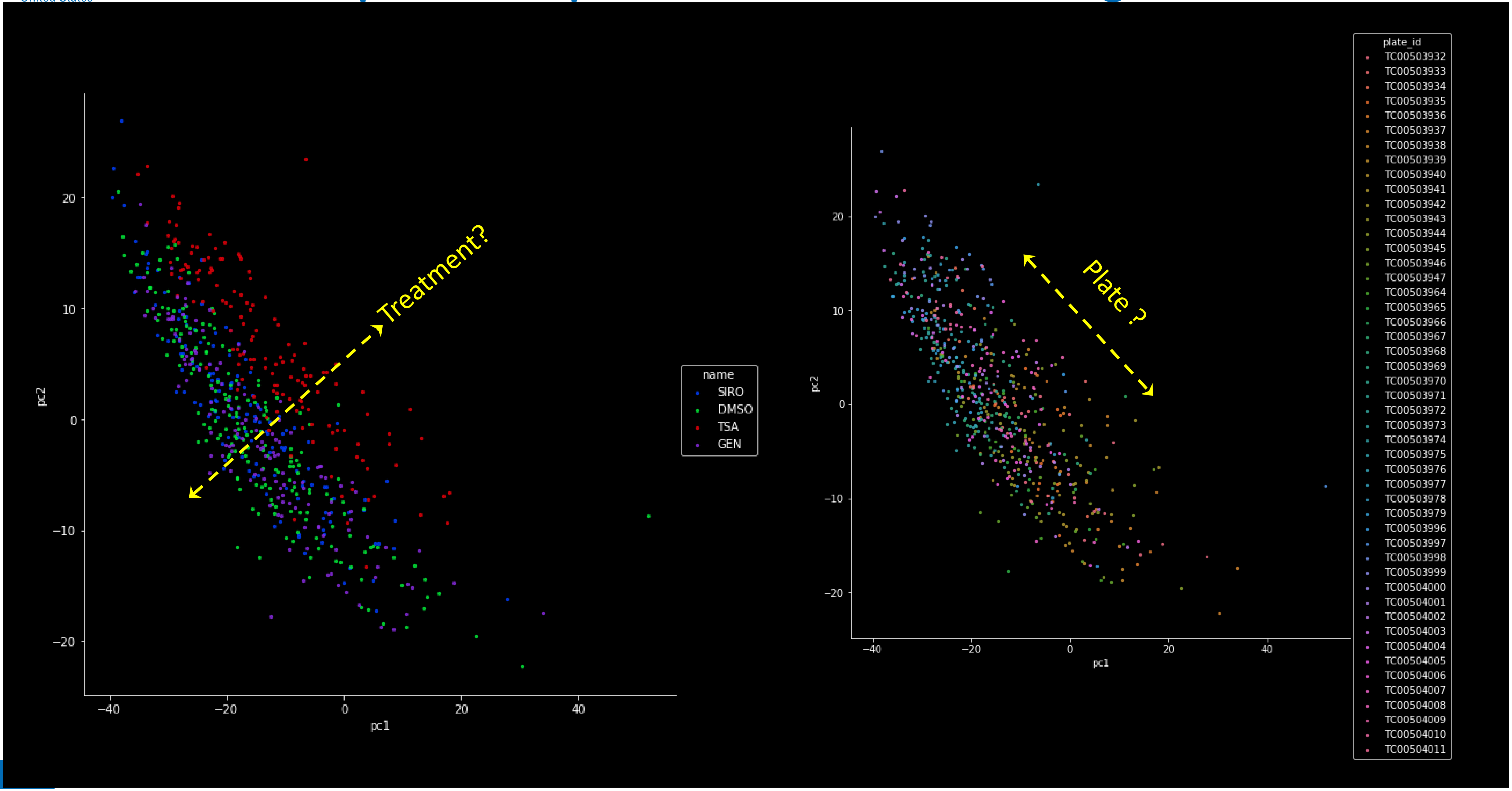
Additional cell lines and/or treatments do not significantly increase diversity

Mix A = Equal volume of Untreated CEM, Untreated CCD18Co, Dex treated CEM, GA treated CEM

Mix B = Equal volume of Untreated CEM, Untreated BxPC3, Dex treated CEM, GA treated CEM

Figures courtesy of BioSpyder

Plate (i.e. Batch) Effects in HTTr Screening Data?



Figures courtesy of Imran Shah

Accounting for Plate Effects in CR Modeling

- To date, it is unknown how potential plate / batch effects in HTTr screening data influences hit.call determinations and potency estimation.

- NCCT has started exploring methods to remove / account for plate (i.e. batch) effects in the HTTr concentration-response modeling pipeline.

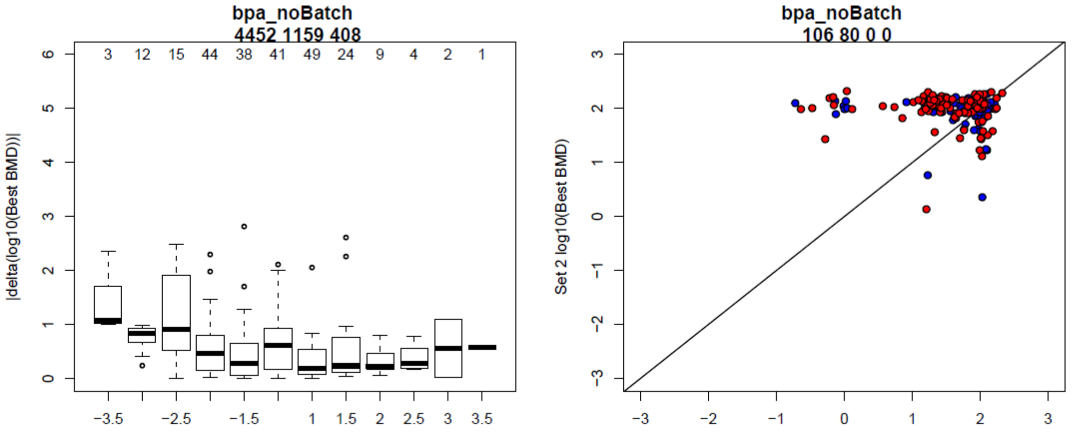
- BMDExpress2.0
  - Currently not configured to address plate effects during the concentration-response modeling process.
  - A pre-processing step is required prior to loading data in BMDExpress to accout for / remove batch effects (ex. limma removeBatchEffect)

- ToSCR (i.e. TempO-Seq Concentration Response)
  - A novel developed by NCCT for concentration-response modeling of count data.
  - Utilizes the shrinkage estimation for dispersions functionality of DESeq2.
  - Includes “plate” as a covariate in the concentration-response modeling procedure (i.e. variable intercept).
  - Shrinkage and plate effect functionalities can be turned ON or OFF.

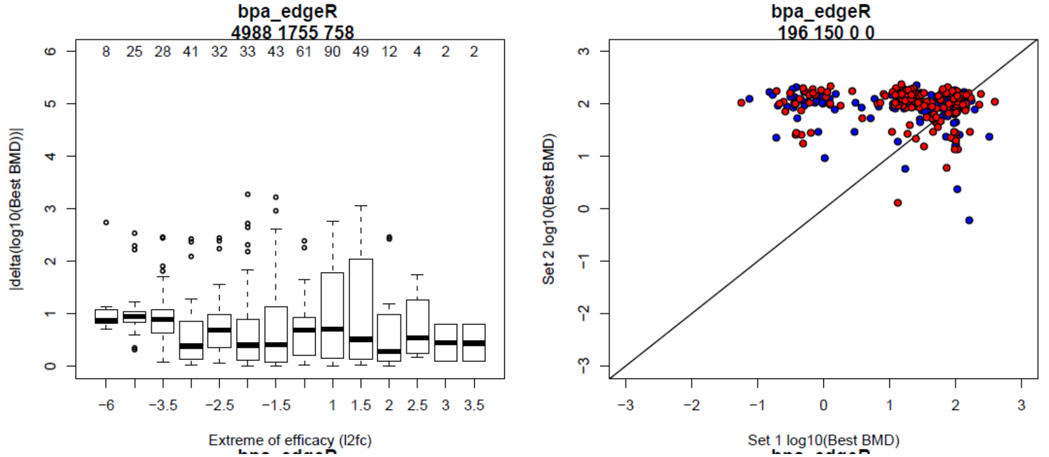
The Effect of Batch Correction on Concordance of Screening Data

- **Bisphenol A was tested in both the MCF-7 HTTr Pilot and MCF-7 HTTr Screen**

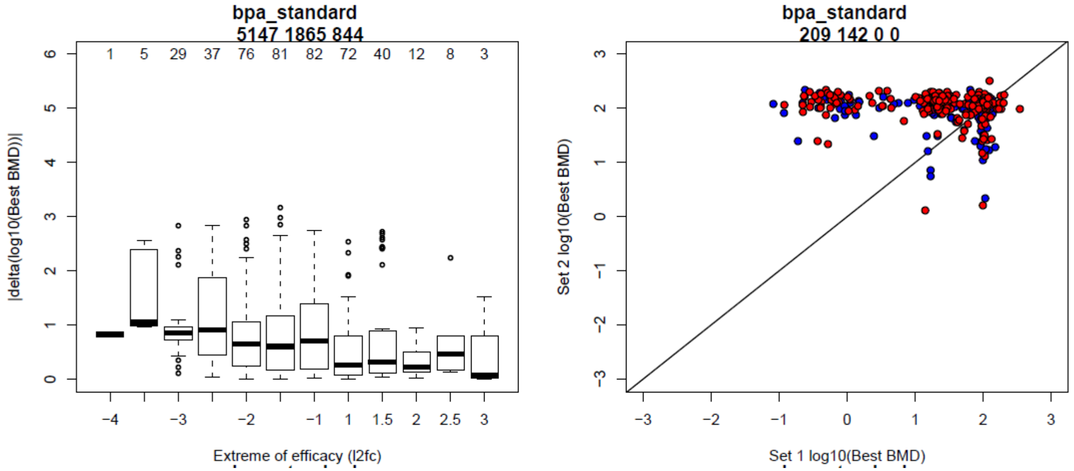
Standard Normalization, No Batch Correction



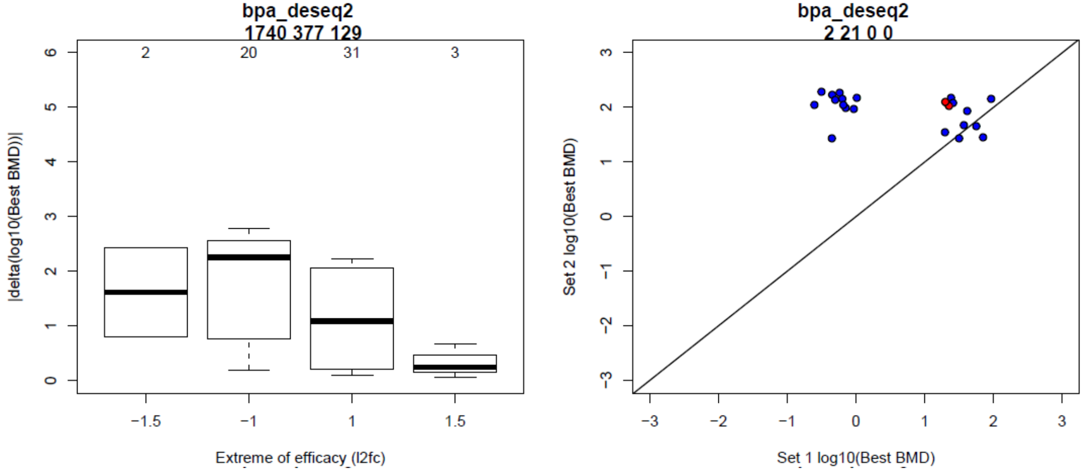
edgeR Normalization, With Batch Correction



Standard Normalization, With Batch Correction



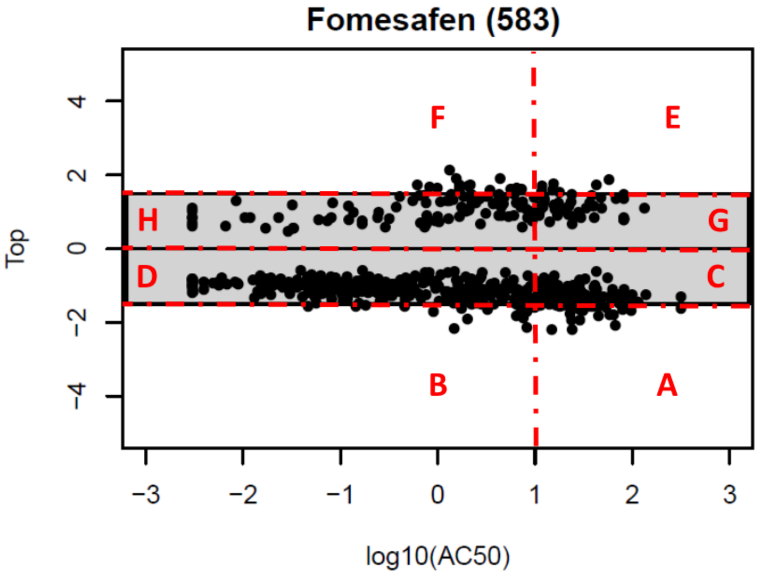
DESeq2 Normalization, With Batch Correction

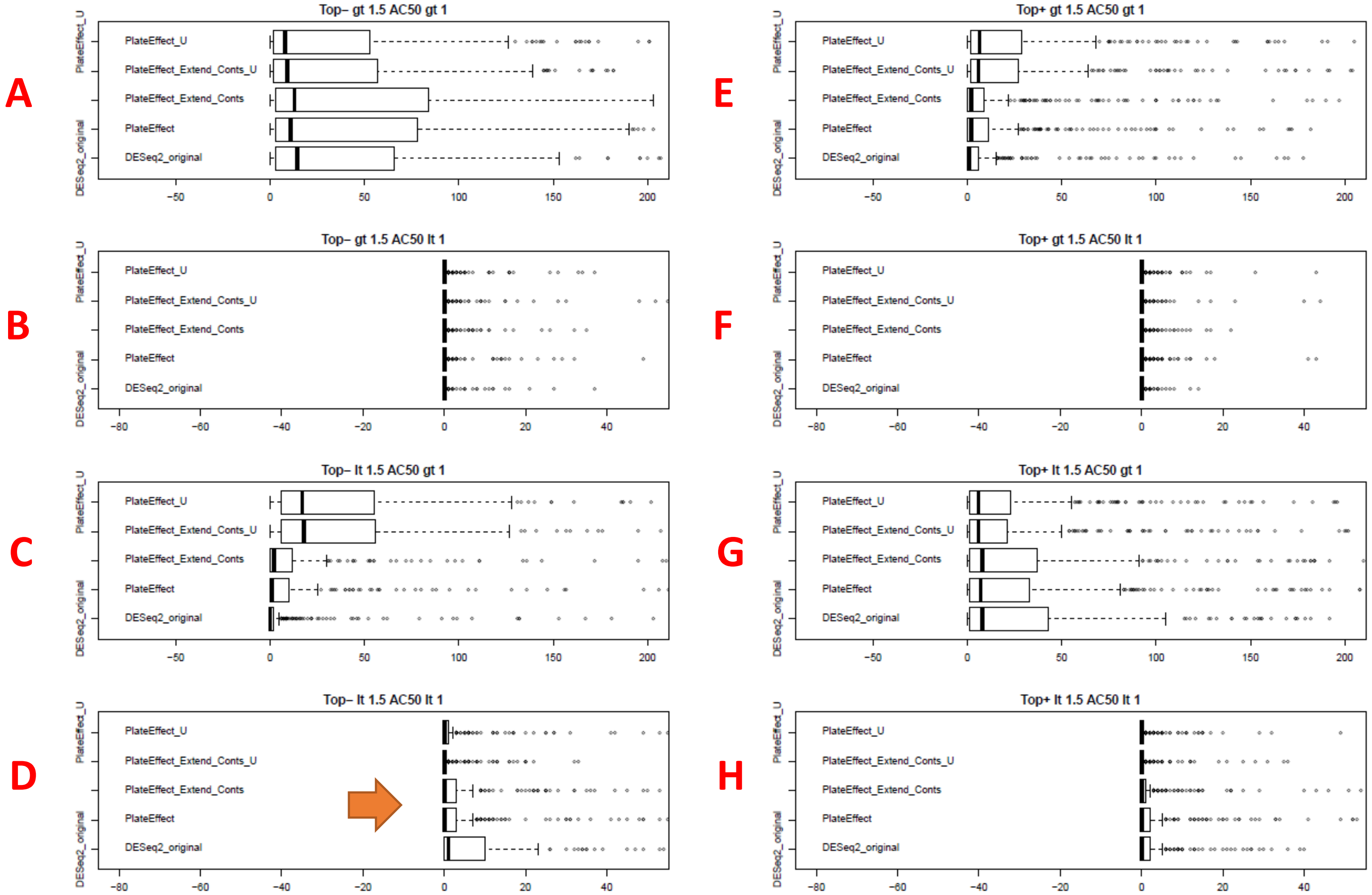


Figures courtesy of Derik Haggard

Accounting for Plate Effects in CR Modeling

CR Modeling with DESeq2 FC Estimates





- Incorporation of **plate effects** in the DESeq2 model reduces the abundance of low potency / low efficacy hitcalls.

Figures courtesy of Richard Judson

Questions / Discussion