OECD Extended Advisory Group for Molecular Screening and Toxicogenomics (EAGMST)

Discussion Session: Performance Criteria/Parameters for Developing Reference Materials for Transcriptomics and Metabolomics Technologies.

EAGMST Meeting, Boulogne FR
June 21st, 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.*

Background

- **New Approach Methods (NAMs)** is a broadly descriptive term for any non-animal technology, methodology or approach or combination thereof that can be used to provide information on chemical hazard and risk assessment. **Inclusive of 'omics technologies.**

- **Reliability** is "*the extent of reproducibility of results from a test within and among laboratories over time, when performed using the same standardized protocol*" (OECD GD 34) has been identified as a key criteria for the acceptance and use of NAMs data regulatory settings (USEPA, 2018).

- The **need** and **means** for demonstrating **reliability** of 'omics data prior to use in regulatory settings has been recognized in the scientific literature:

  - **Microarray Quality Control (MAQC) Consortium, 2006:** *"Concerns have been raised regarding the **reliability** and consistency, and hence potential application of microarray ['omics] technology in the clinical and regulatory settings…It follows that before this technology can be applied in clinical practice and **regulatory decision making**, microarray standards, quality measures and consensus on data analysis methods need to be developed….microarray studies need **unified metrics and standards**, which can be used to identify suboptimal results and monitor performance in microarray facilities"*

  - **Sauer et al., 2017:** *"The **establishment of calibrated RNA samples** and **reference datasets** were identified as crucial for an objective assessment of the performance of different microarray [i.e 'omics] platforms"*

What are Standardized Reference Materials for ‘Omics?

- Standardized biological samples that are either commercially manufactured or generated in bulk in a research laboratory and contain a mixture of biological molecules in varying amounts that can be measured using an ‘omics platforms

  - Transcriptomics → RNA species
  - Metabolomics → Small molecules
  - Proteomics → Proteins / peptides

- “*Expression values generated on different [‘omics] platforms cannot be directly compared because unique labeling methods and probe sequences will result in variable signals for probes that hybridize to [measure] that same target. Alternatively, the* ***relative expression*** *between a pair of sample types should be maintained across platforms*.” (MAQC, 2006). This concept is generalizable to other types of ‘omics.

- Repeated testing and contrasting of **pairs** of biologically diverse standardized reference materials can be used as an approach for assessing the **reliability** of an ‘omics assay for reproducible detection and quantification of biomolecules of interest.

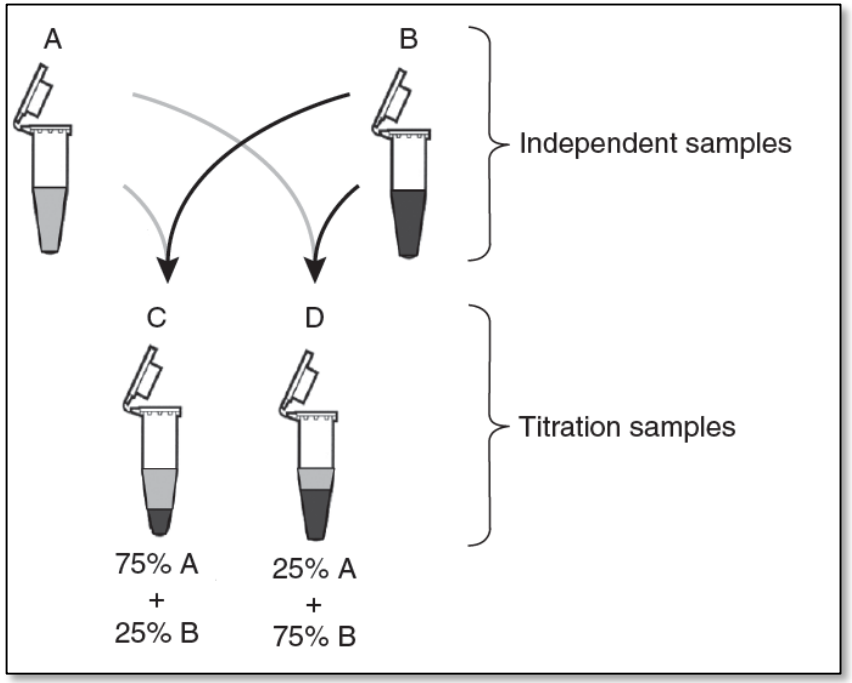
- The combination of biologically different RNA sources and known titration differences provides a method for assessing the relative accuracy of an ‘omics platform based on differential detection. (MAQC, 2006, adapted)

Key Characteristics of Standardized Reference Materials

Key Characteristics:

- Paired samples from diverse sources or with diverse gene expression profiles.
- Substantial overlap in the species of biomolecules contained within each sample.
- Large dynamic range in expression levels / constituent concentrations upon comparison of samples.
- Compatibility with ‘omics platform
- Qualitatively similar to test samples (i.e. sample matrix)
- Widespread availability.
- Suitability for use across ‘omics technologies

MAQC ID	Description
Sample A	Universal Human Reference RNA (UHRR)
Sample B	Human Brain Reference RNA (HBRR)
Sample C	75 % UHRR | 25 % HBRR
Sample D	25 % UHRR | 75 % HBRR

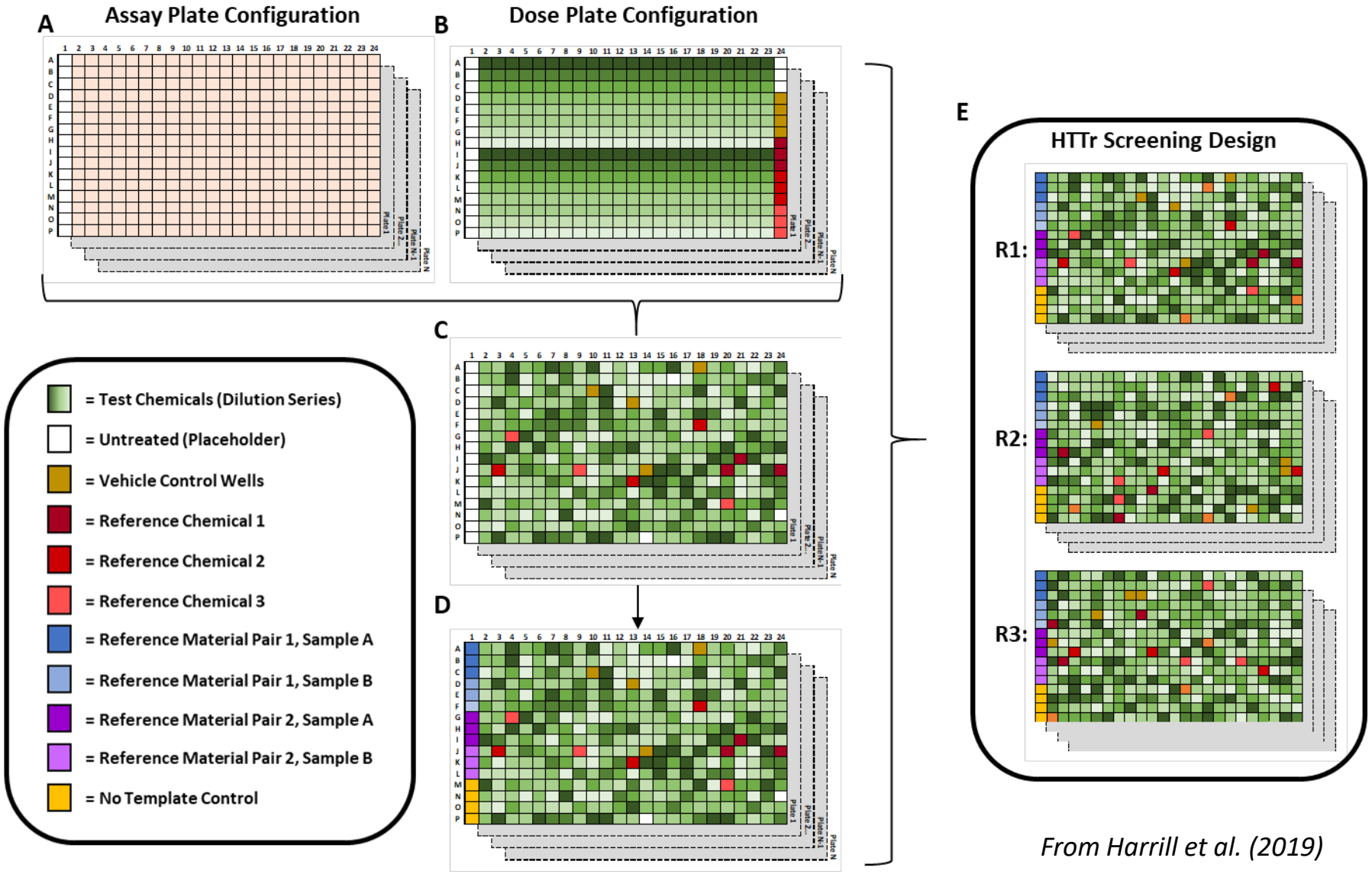


From Shippey et al (2006)

Implementation of Standardized Reference Materials

- Standardized reference materials are designed to evaluate the technical performance of an ‘omics assay…**NOT** the biological response of an *in vivo* or *in vitro* test system.
  - Use reference treatments for this latter purpose.
- Processed in parallel with test samples → they should be subject to the same manipulations and assay conditions as test samples.
- For each batch of test samples processed and assayed, standardized reference materials are processed and assayed in parallel.
- Implemented in a manner that facilitates monitoring of comparability and consistency of results generated across batches and over time.

Example of Implementation for High Throughput 'Omics



From Harrill et al. (2019)

Metrics for Assessing 'Omics Reproducibility

- **Sample-Based Metrics**
  - Concordance in Detection Call
  - Concordance in Detection Level

- **Contrast-Based Metrics**
  - Differential Gene List Overlap**
  - Log FC Ratio Compression
  - Log FC Ratio Rank Correlation (Spearman)
  - Log FC Ratio Correlation
  - "Biology-based" Pathway Enrichment Concordance / Overlap

***Lesson Learned from MAQC*

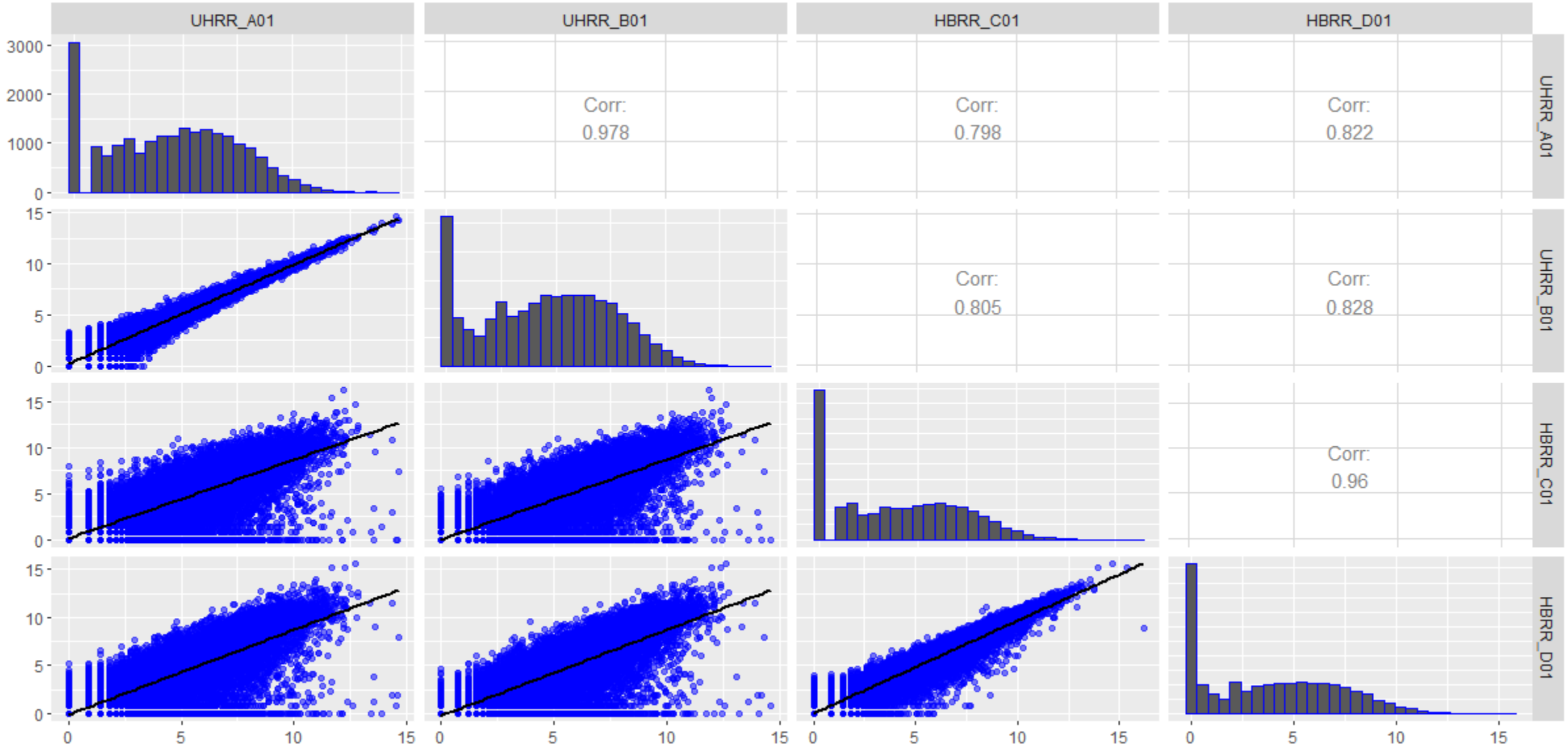
A straightforward approach of FC ranking plus a non-stringent p-cutoff can be successful in identifying reproducible gene lists, whereas ranking and selecting differentially expressed genes solely by the t-test statistic **predestine a poor concordance in results**, in particular for shorter gene lists, due to the relatively unstable nature of the variance (noise) estimate in the t-statistic measure.

Furthermore, the impact of normalization methods on the reproducibility of gene lists becomes minimal when the fold change, instead of the p-value, is used as the ranking criterion for gene selection

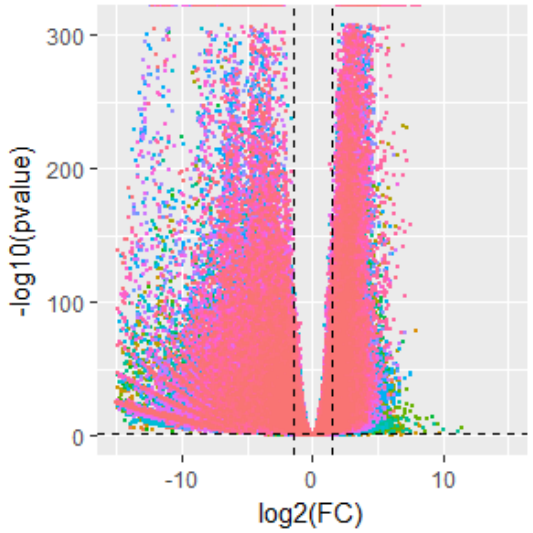
- Each of these metrics are either qualitative, quantitative, comparative or some combination of each.

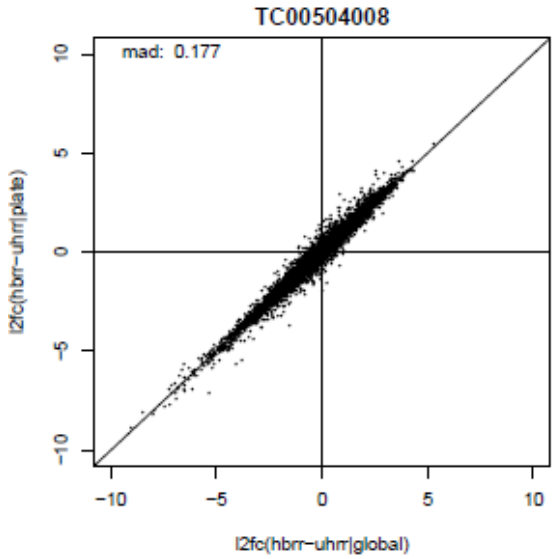
- Interpretation requires defining typical performance ranges.

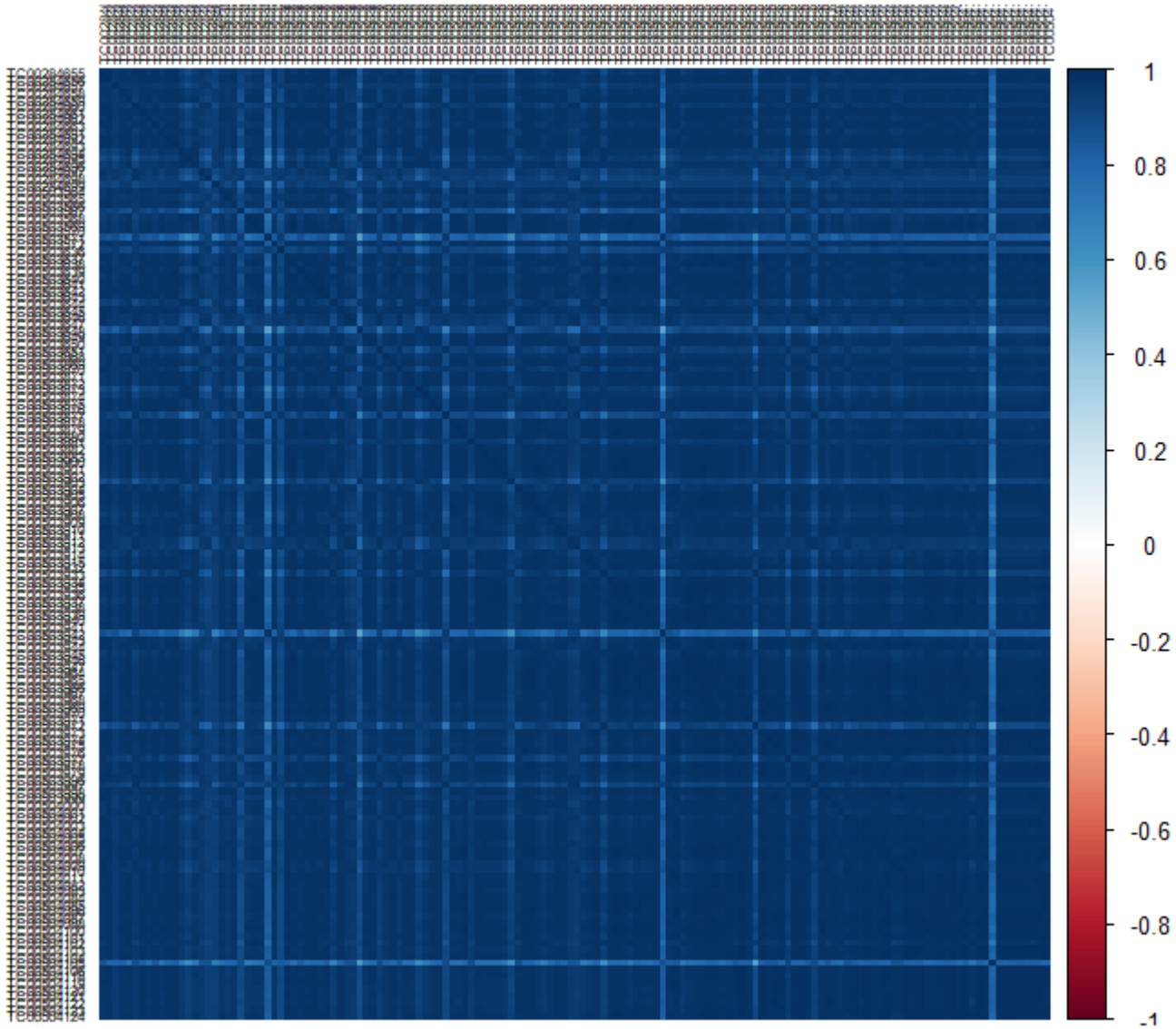
Concordance of Detection Level

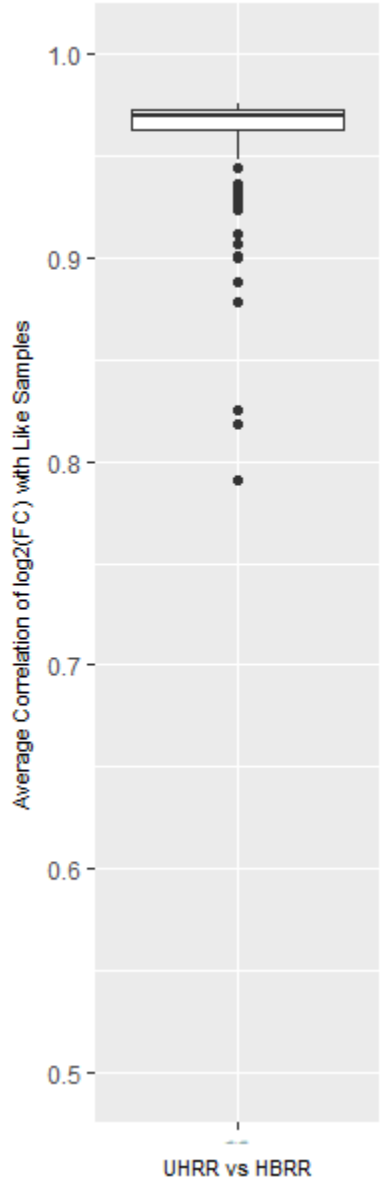


Log FC Ratio Correlation









Pathway Enrichment as a QC Metric

name	pathway	pathset	pathclass	size	pathway_score	null.pvalue	null.madsaway	exp.pvalue	exp.madsaway
Genistein	HALLMARK_ESTROGEN_RESPONSE_EARLY	Hallmark	estrogen	153	0.256717	0	8.162929604	0.028169014	5.79448377
Genistein	HALLMARK_ESTROGEN_RESPONSE_LATE	Hallmark	estrogen	144	0.208794	0	6.553971056	0.036619718	3.345591245
Genistein	Parkinson's disease	Bioplanet	other	82	-0.08271	0.004	-2.508000266	0.492957746	-0.6815578
Genistein	TGF-beta regulation of skeletal system development	Bioplanet	development	49	-0.15166	0.006	-2.620384546	0.005633803	-3.322186167
Genistein	Developmental biology	Bioplanet	development	232	-0.05668	0.009	-2.369383434	0.03943662	-2.081477457
Genistein	Proteasome degradation	Bioplanet	other	46	-0.09579	0.011	-2.285588354	0.304225352	-1.079089245

name	pathway	pathset	pathclass	size	pathway_score	null.pvalue	null.madsaway	exp.pvalue	exp.madsaway
Sirolimus	HALLMARK_MYC_TARGETS_V2	Hallmark	myc	52	-0.31101	0	-7.729808416	0.008450704	-3.508067272
Sirolimus	HALLMARK_MTORC1_SIGNALING	Hallmark	other	161	-0.18133	0	-6.90666279	0.011267606	-3.007041858
Sirolimus	HALLMARK_MYC_TARGETS_V1	Hallmark	myc	167	-0.18344	0	-5.657061193	0.118309859	-1.6447404
Sirolimus	HALLMARK_UNFOLDED_PROTEIN_RESPONSE	Hallmark	other	93	-0.10966	0	-4.877611065	0.005633803	-2.566874463
Sirolimus	Cytosolic tRNA aminoacylation	Bioplanet	RNA	22	-0.25641	0	-4.473527993	0	-4.175819507
Sirolimus	Transfer RNA aminoacylation	Bioplanet	RNA	37	-0.22651	0	-4.077163271	0	-4.526749026

name	pathway	pathset	pathclass	size	pathway_score	null.pvalue	null.madsaway	exp.pvalue	exp.madsaway
Trichostatin A	Gene expression	Bioplanet	other	716	-0.12662	0	-5.781144946	0.064788732	-1.979270755
Trichostatin A	HALLMARK_MYC_TARGETS_V1	Hallmark	myc	167	-0.1749	0	-5.345172308	0.143661972	-1.548821639
Trichostatin A	Messenger RNA splicing: major pathway	Bioplanet	RNA	62	-0.20614	0	-5.178901041	0	-3.975512366
Trichostatin A	HALLMARK_E2F_TARGETS	Hallmark	other	166	-0.16181	0	-5.024173468	0.030985915	-2.590013296
Trichostatin A	HALLMARK_MYC_TARGETS_V2	Hallmark	myc	52	-0.17587	0	-4.916757527	0.036619718	-2.266269214
Trichostatin A	Cleavage of growing transcript in the termination region	Bioplanet	other	39	-0.23322	0	-4.762299303	0.002816901	-2.58003793

- Comparing consistency of top ranked enriched pathways is another potential QC approach.
- Top ranked enriched pathway are consistent with the known bioactivity of the reference chemicals or known DEG profile from standardized reference material pairs.

Data courtesy of Thomas Sheffield

Summary

- Use of standardized reference materials for evaluation of ‘omics assay performance would increase confidence that ‘omics assays provide reliable and repeatable results.

- Results from testing of standardized reference materials assayed in parallel with study samples would be a contributing factor in decisions regarding whether an ‘omics data set is suitable for use in regulatory decision making processes.

- Ideally, a set of standardized reference materials would be compatible for use with multiple ‘omics platforms to facilitate evaluation of cross-platform performance.

- The MAQC studies may serve as a guide for designing standardized reference materials for ‘omics studies and determining what metrics could be used to assess ‘omics assay performance. However…

- Currently, there is no for guidance or consensus standard regarding the biological characteristics, content, derivation or source for standardized reference materials for ‘omics studies.

Discussion Topics for Standardized Reference Materials

- **General Topics**
  - Value for the research community?
  - Value for the regulatory community?
  - Parallels / lessons from other scientific disciplines
    - Analytical chemistry
    - Laboratory proficiency testing
    - High-throughput screening
- **Technical Topics**
  - Sample Characteristics
  - Implementation / Use
  - Performance Metrics
  - Sourcing (commercial vs. other)

Acknowledgements

NCCT HTTr Team

Richard Judson

Imran Shah

Logan Everett

Derik Haggard

Thomas Sheffield

Clinton Willis

Rusty Thomas

Massive Analysis and Quality Control Society

Questions / Discussion