

KEY ASPECTS OF A MOCK IND SUBMISSION CONTAINING MICROARRAY DATA

Laura Reid¹, Wendell Jones¹, Rich Cohn¹, Thomas Goralski¹ and Federico Goodsaid²

¹Expression Analysis, Inc., Durham, NC; ²Schering-Plough, Lafayette, NJ



INTRODUCTION

We conducted a mock submission using microarray based data as a collaborative exercise between FDA and industry. The data files were derived from a toxicogenomics study with eighteen rat samples that had been hybridized to Affymetrix U34A GeneChips.

The objectives were to:

- (1) Contribute to building and refining a process in which microarray data may be incorporated into submissions to FDA;
- (2) Provide a suitable framework in which to develop recommendations; and
- (3) Contribute to the development of consensus around the specific elements of applicable recommendations.

SUBMISSION FORMAT

We designed the mock submission to address developing needs in the microarray community, rather than to meet a specific submission format.

The document was divided into seven sections:

- **Laboratory Infrastructure.** hybridization platform, protocol, chip reader
- **Informatics Infrastructure.** data management, IT, statistical system infrastructure and associated quality assurance/quality control measures
- **Study-Specific Sample Processing & Array Performance Characteristics.** the use and interpretation of RNA standards and controls, quality control metrics, data quality, and array performance characteristics
- **Bias Mitigation: Biological & Processing Factors.** detection and mitigation of biases related to processing
- **Data Analysis Methods.** statistical analysis methods and related considerations
- **Toxicology Results.** toxicology results independent of toxicogenomic data
- **Interpretation of Toxicology & Toxicogenomics.** toxicogenomics discussed in the context of the toxicology data to provide an overall interpretation of experimental results

DATA SET

Female rats were treated with a vehicle control or a compound designed to reduce cholesterol levels by inhibiting HMGCoA reductase. The animals were sacrificed after 1, 7 or 30 days of dosing. RNA was extracted from liver samples for gene expression analysis. Additional assays were performed to identify toxicological effects of the compound.

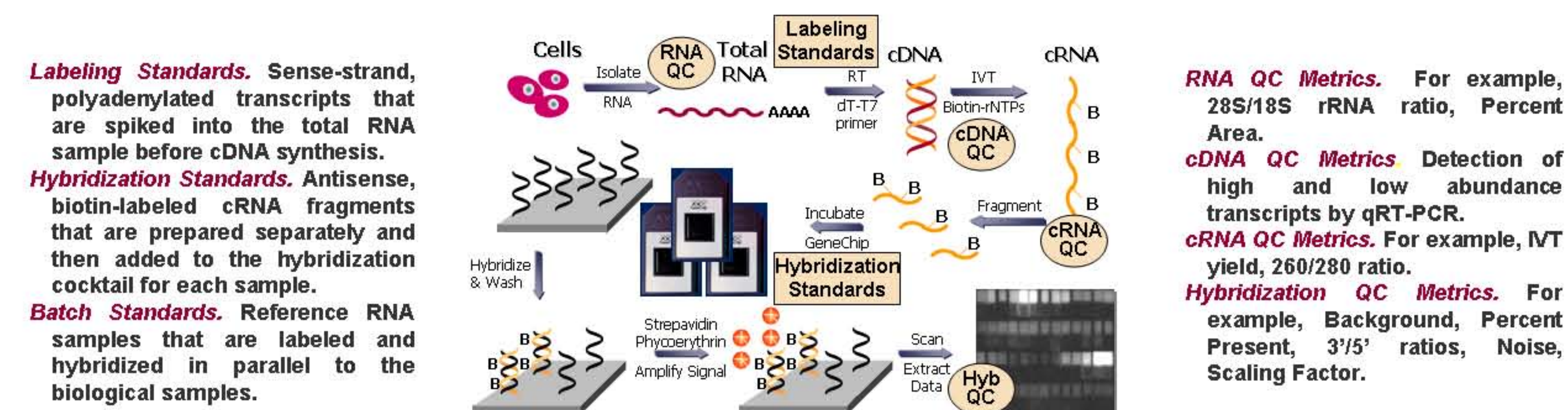
Biotinylated cRNA targets were prepared from three control and three treated animals at each of the time points. A total of eighteen targets were hybridized to Affymetrix Rat Genome U34A GeneChips, which contain short oligonucleotide probes representing more than 7,000 well characterized rat genes.

RNA STANDARDS & QC METRICS

Several microarray controls are available to evaluate performance at multiple steps during target preparation and data collection. For completeness, we choose to describe many quality parameters in the mock submission.

We divided the controls into two complementary categories:

Standards are external reference materials that confirm the performance of the protocols and reagents.
QC Metrics are measures collected during the processing that reflect the quality of individual RNA samples.



Labeling Standards. Sense-strand, polyadenylated transcripts that are spiked into the total RNA sample before cDNA synthesis.
Hybridization Standards. Antisense, biotin-labeled cRNA fragments that are prepared separately and then added to the hybridization cocktail for each sample.
Batch Standards. Reference RNA samples that are labeled and hybridized in parallel to the biological samples.

RNA QC Metrics. For example, 2S/18S rRNA ratio, Percent Area.
cDNA QC Metrics. Detection of high and low abundance transcripts by qRT-PCR.
cRNA QC Metrics. For example, IVT yield, 260/280 ratio.
Hybridization QC Metrics. For example, Background, Percent Present, 3'/5' ratios, Noise, Scaling Factor.

ARRAY PERFORMANCE & VALIDATION

In addition to the RNA standards and QC metrics, we proposed that an FDA submission should include evidence for the validity of the submitted data set. This information was presented in the form of reproducibility data between appropriate sample replicates that suggests a level of confidence in the resulting lists of differentially-expressed genes.

The submission included three biological replicates for each condition. To examine the reproducibility of the data set, we compared the array results generated from different samples representing the same experimental condition.

Three reproducibility statistics were calculated:

- **Correlation.** The concordance of the intensity of each probe cell in one hybridization with its corresponding intensity in a replicate hybridization.
- **Detection Call Agreement.** The number of transcripts called "Present" in both samples.
- **Signal Value Agreement.** The number of transcripts with a less than 2-fold difference in signal values between replicates. Transcripts called "Absent" in both samples are excluded and signal values below 64 are censored (TGT=500).

The table at right demonstrates how the Agreement statistics were calculated using mock data not included in the submission.

Mock Data Illustrating the Calculation of Agreement Statistics

	Detection Call	Signal Level
Replicate 1	P	200
Replicate 2	P	900
Replicate 3	A	800

Discordant values are highlighted.

Reproducibility Measures in All Treatment Groups

Reproducibility Measures within Control, Day 0 Group

Reproducibility Measure	Data Source	Comparison			Average Measure
		S1 to S2	S1 to S3	S2 to S3	
Correlation	Probe	0.971	0.975	0.970	0.972
Detection Call Agreement	Transcript	85.7 %	86.3 %	86.9 %	86.3 %
Signal Value Agreement*	Transcript	95.69 %	96.22 %	96.38 %	96.09 %

S1, S2 and S3 represent samples from three rats in the same treatment group.
*For transcripts detected in at least one sample using a lower limit Signal value of 2⁶

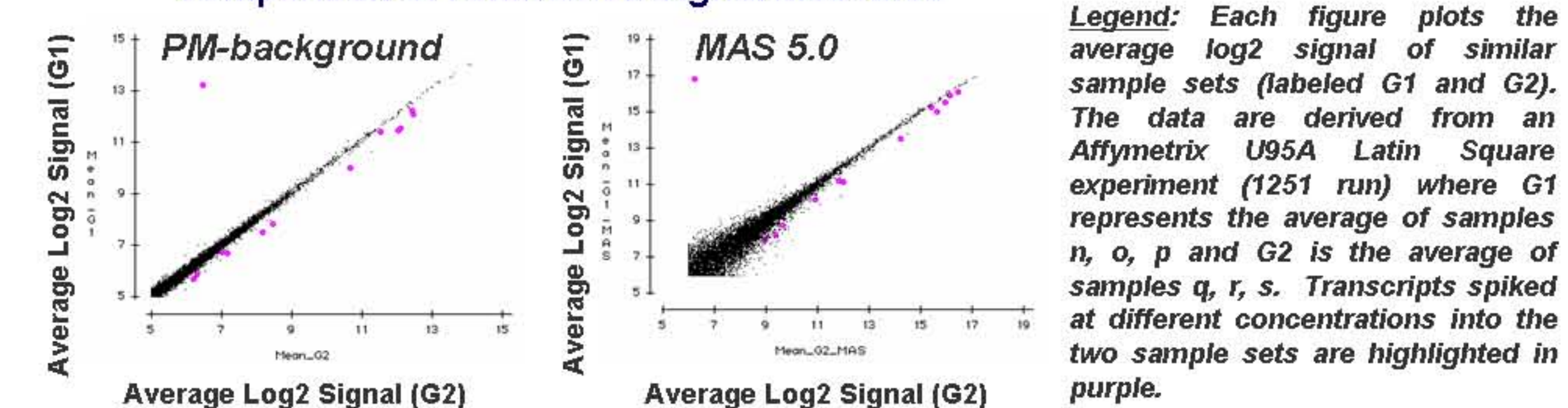
qRT-PCR is often considered the "gold standard" for RNA expression research and has been extensively used to validate microarray platforms. While qRT-PCR confirmation may be useful for some transcripts, there is considerable debate as to whether it is necessary for every study that includes microarray data. For most purposes, and given a validated platform, we propose the reproducibility of the data combined with a set of appropriate QC metrics should be sufficient for judging array performance. Investigators may choose to reserve qRT-PCR confirmation studies for transcripts with borderline significance values or biologically-critical genes.

DATA ANALYSIS

It is likely that future FDA submissions will utilize a variety of data extraction and analysis protocols. As recommended by the Best Practices Working Group (1), we provided two measures of gene expression for the data set in our mock FDA submission: MAS 5.0 and a perfect match only measure, referred to as PM - Background. There are several other Affymetrix signal algorithms that may be suitable (eg, PDNN, RMA, MBEI) relative to the number of samples analyzed. New and better methods for expression summarization will continue to appear.

Both signal algorithms calculate a quantitative summary measure of expression for each transcript. **MAS 5.0** examines hybridization intensities to the PM probes after subtracting background and the estimated non-specific hybridization to the MM probes. It summarizes probe pairs using a robust weighted mean. **PM - Background** uses a similar probe summarization method, except that only local background is subtracted from the PM intensities. This measure also uses different normalization technique and filtering criteria.

Comparison of Alternative Signal Measures

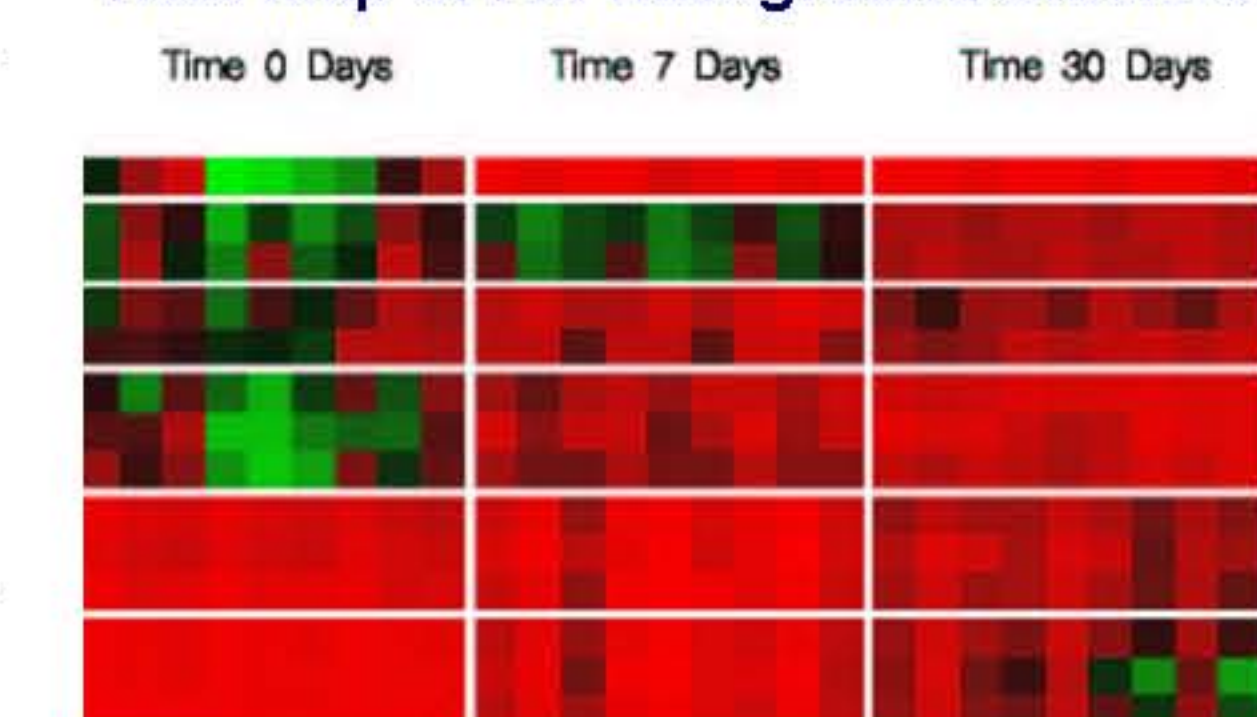


Legend: Each figure plots the average log₂ signal of similar sample sets (labeled G1 and G2). The data are derived from an Affymetrix U95A Latin Square experiment (1251 run) where G1 represents the average of samples n, o, p and G2 is the average of samples q, r, s. Transcripts spiked at different concentrations into the two sample sets are highlighted in purple.

We find that PM-centric measures tend to greatly reduce the variance seen in microarray data, especially at lower abundance levels. In this case, MM probes are used primarily for chip-based normalization and background assessment.

Specificity is a very difficult problem with microarrays. To account for multiple testing issues and the accompanying high false positive rates, up and down regulated genes for both signal measures were identified using the SAM method (2). To facilitate biological interpretation, we used simple clustering methods to identify potentially co-regulated genes.

Heat Map of PM-Background Clusters



TOXICOGENOMIC INTERPRETATION

Although decreases in body weight were observed in the treated animals, this study showed no significant toxicity from the compound in the liver or other tissues at the dose tested. We demonstrated the role of microarray data in a toxicological report by linking some of the expression results identified in the microarray study to the pharmacological mechanism of the drug compound.

We were delighted to note that several features of the microarray data supported the biological conclusions.

- First, the two different data extraction and normalization methods (MAS 5.0 and PM-Background) both generated similar clusters of differentially expressed genes.
- Second, both analyses reported increased expression of the HMGCoA reductase gene in the treated animals, as expected after exposure to an inhibitor compound.
- Third, expression changes in other genes in the cholesterol biosynthesis pathway and related metabolic pathways were detected.

LESSONS LEARNED

- Microarrays represent hugely parallel experiments that are not independent (i.e. testing thousands of gene simultaneously), and so present unique statistical challenges. Novel analysis methods should have appropriate validation and accompanying documentation.
- Continued efforts are required to understand the particulars of microarray platforms (e.g. effects of probe sequence and filtering method).
- FDA reviewers sought straightforward accept/reject criteria based upon appropriate microarray controls. Multiple RNA standards and QC metrics were included for informational purposes. We are not suggesting that each metric must be included in future FDA submissions.
- Measures of reproducibility between replicate samples and of variability (i.e. coefficient of variation) in the RNA standards and QC metrics were useful when judging the quality of a data set.
- Continued studies are necessary to compare data generated in different microarray facilities and to quantify the level of internal processing variability.
- While standard operating procedures, equipment calibration and validation records are required to confirm process capabilities, these details may not be necessary in a submission, but should be maintained by the sponsor, probably in a Drug Master File.
- Microarray tables are too large for paper copies. We originally submitted XML versions of the data set, but later discovered that SAS and Excel formats were easier for some participants to manage.

ACKNOWLEDGEMENT

We extend special thanks to past and present members of the CDER Nonclinical Pharmacogenomics Subcommittee for their careful review and helpful discussions during the submission process.

Siham Biade	Maria Rivera
Paul Brown	Timothy Robison
David Essayan	Lilliam Rosario
Patricia Harlow	Barry Rosenzweig
Wafa Harrouk	Frank Sistare
John Leighton	Karol Thompson
Scott Pine	Josie Yang
Lynnda Reid	Hao Zhang

REFERENCES

1. Best Practices for Data Generation and Interpretation in Clinical Trials, Nature Reviews Genetics, 2004
2. Tusher et al., PNAS 98:5116-5121, 2001