

Quick Reference Card



QC Metrics for Exon and Gene Design Expression Arrays

A summary based on the Affymetrix Quality Assessment of Exon and Gene Arrays White Paper

Purpose

Affymetrix upholds its dedication to providing the industry's most comprehensive and informative QC metrics. For GeneChip® Exon and Gene Arrays, Affymetrix has created a wide range of controls to ensure your experiment is successful. The concepts presented here are used to aid researchers in determining if their experimental and control samples prepared with the Affymetrix whole-transcript (WT) assay and hybridized on Affymetrix expression microarrays pass a minimum level of quality control.

Required Software Tool

The metrics discussed in this document are generated by the Affymetrix® Expression Console™ software.

Hybridization Quality Control

The process of performing quality control for hybridizations within a microarray experiment is often simplified to determining if the distribution of various quality metrics associated with individual hybridizations in a group display any distinct outliers. Unfortunately, hard cutoffs are difficult to apply because of sample-to-sample variability and are more likely to result in incorrectly removing acceptable samples. In more than two years of direct customer support of these assays, Affymetrix has found that the process of monitoring the distributions of several metrics functions well to ensure that a microarray experiment has passed a minimum level of quality control.

Identifying Outliers with Affymetrix® Expression Console™ Software

Identifying outliers with Expression Console enables you to remove problematic samples prior to downstream statistical analysis, saving both time and effort. This document provides guidance on how you can identify outliers. At a high level, monitor the distributions of the metrics outlined below within experiments in a study. Look for hybridizations which consistently have metric values at the tails of the distribution. A good rule of thumb is to flag outliers which have values two standard deviations away from the mean of the metric value for that experiment.

Accounting for Differences in Samples When Monitoring these QC Metrics

When comparing the distribution of quality metrics, it's logical to assume that values may group according to sample type. If samples are significantly different, it may be necessary to independently monitor distributions according to sample type.

Determining the Number of Replicates Needed to Get an Accurate View of the Distribution of QC Metrics

Three replicates for each sample type should be sufficient to monitor the distribution of the quality metrics. A minimum of five arrays is recommended to perform probe summarization using RMA.

Understanding What Different Metrics Measure

Metrics have been divided into three sections: sample metrics, hybridization metrics and labeling metrics.

Understanding Expression Console Analysis Options

- Affymetrix recommends using RMA (due to its speed) for the purposes of monitoring quality control.
- Analyze your arrays at the gene level, not the exon level, when monitoring quality. You can always go back and reanalyze the data at the exon level for a complete analysis.
- Analyze your whole study at one time to ensure that the same probe summarization model is uniformly applied across your experiment.
- Limit your analysis only to core-level probe sets for exon arrays. Using extended or full content will increase variability in your metrics since there is a higher percentage of unexpressed content in these categories.

Monitoring Sample Quality

A sample/hybridization should be flagged and possibly removed when several metrics are outliers within the distribution of samples. If none of these metrics consistently denotes an outlier sample, then a researcher should confidently proceed with downstream analysis.

Table 1: Sample Quality Metrics

Metric	What Does It Measure?	How can it be used for QC?
Pos_vs_neg_auc	This is the “area under the curve,” or AUC value for a ROC curve which plots the detection of positive controls against the false detection of negative controls. For a definition of what constitutes negative and positive controls, please see FAQ section.	This metric is a robust measurement for overall data quality. Typical values range between 0.8 and 0.9, with a value of 1.0 being perfect and a value of 0.5 illustrating no discernable difference between the positive and negative controls. Flag arrays as outliers when the value significantly drops below 0.8. This metric is not sensitive to subtle array quality differences, but a value significantly below 0.8 is a strong indicator of a poor-quality data. Values may be slightly lower with FFPE data.
All Probe Set Mean	The mean of the signal of all probe sets included in the analysis.	This metric can be used to detect bright or dim arrays. The value is computed after probe summarization and normalization for the probe sets included in the analysis (core content, in this case). The values should be more consistent for biological replicates than for different sample types. Flag outliers in overall study or within same sample type.
All Probe Set RLE Mean	Mean absolute relative log expression (RLE). The signal of each probe set is compared to the median signal value of this probe set in the study. The metric is the mean of these differences from all the probe sets.	Unusually high values indicate that the signals on the array are different from others in the study, i.e., big values are bad. This metric is most useful for studies with similar sample types to detect outlier arrays. For a set of different tissues, for example, this metric is less useful. On the gene array, values ranged from 0.27 to 0.61 in a diverse tissue panel set and between 0.1 and 0.23 for non-outlier technical replicates. Flag outlier arrays and monitor other metrics.

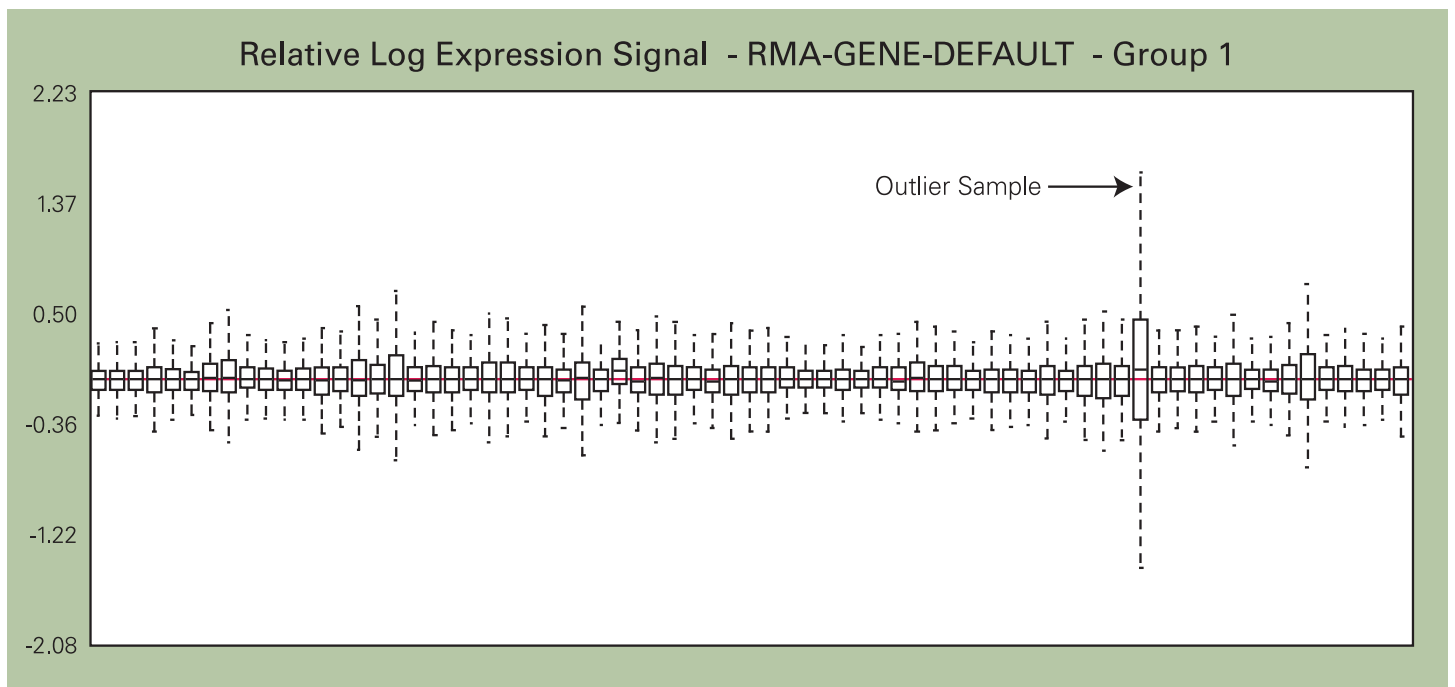


Figure 1: RLE Box Plot The distribution of RLE values is easily examined by using the RLE box plots available within Expression Console. Box plots of the relative log expression for all the probesets analyzed indicate that one of the samples is a clear outlier.

Table 2: Additional Sample Quality Metric (Exon-level Analysis Only)

Metric	What Does It Measure?	How can it be used for QC?
%P	Percent of exon-level probe sets detected based on the DABG (detection above background) algorithm.	The value should be similar for same tissue types. Samples are potentially problematic when %P is more than 10 percentage points lower than that of similar sample types. Absolute number varies greatly between tissue types.

Monitoring Hybridization Quality

The bacterial spikes (*BioB*, *BioC*, *BioD* & *Cre*) are best for monitoring the quality of the hybridization, similar to 3'-based expression. These spikes are added to the hybridization cocktail but are labeled independently of the rest of the sample. The four spikes are input with increasing concentration. Signal values of the spikes can be plotted in EC using the report metrics line graph. A plot of the signal values associated with these spikes should show a trend where $BioB < BioC < BioD < Cre$.

Monitoring Labeling Quality (this is an active area of research and a lack of rank order should not be over interpreted)

The polyA-control RNAs (*Lys*, *Phe*, *Thr* and *Dap*) are best for monitoring the quality of the labeling reaction. These polyA RNAs are spiked into the sample prior to amplification and labeling. Consequently, these molecules are amplified and labeled with the rest of the biological sample. The rank order of the signal values for these probe sets should show $Lys < Phe < Thr < Dap$. Note that impurities within the sample RNA can impact the efficiency of the labeling of these controls.

Frequently Asked Questions (FAQ)

What exactly are positive and negative controls? I see some signal in my negative controls; are they really negative?

Positive controls are probe sets designed against putative exons of about 100 housekeeping genes shown to be expressed at detectable levels across a variety of tissues. Since the extent of alternative splicing and transcript expression is not known for all tissues, not all exons are expected to be expressed in all tissues.

Negative controls are putative intronic regions of the same 100 housekeeping genes used for positive controls. These probe sets may be expressed in certain tissues through intron retention. They are not true negative controls. Overall, the positive and negative control probe sets provide a medium-size dataset with expected high and low signal values, respectively. This data set is useful to estimate overall data quality through the Pos_vs_neg_auc value.

I think I have a dim (or bright) hybridization; what other metrics can I examine?

All_mean is a good metric to assay for hybridization intensity, but PM_mean can also be used. PM_mean is a probe-level metric while all_mean is a probe set metric. PM_mean is the mean of perfect match raw intensities prior to any transformations such as normalization or probe summarization. PM_mean and all_mean can be contrasted to understand the effect that data processing steps have on the average intensity of an array as all_mean has been subject to any data transformations that have been performed during signal estimation and normalization. Apparent outliers only based on PM_mean can be ignored when corrected through data normalization in all_mean.

How is the bgrd_mean metric calculated and why is the value sometimes higher than that of PM_mean?

The mean of the background probe signal values is based on background probes as defined in the background probe file, which are by default the antigenomic probes. Antigenomic probes consist of about 1,000 probes for each level of GC content (0 to 25) without any homology to most studied organisms. This set has a higher GC content than the average probe on the array, and can therefore have relatively higher signal values than the mean of all probes (PM_Mean).

Can I measure the quality of a single hybridization without the rest of the experiment?

Affymetrix does not recommend quality monitoring using single-array performance metrics without consideration for the rest of the experiment. In large-scale expression experiments using similar sample types, researchers are likely to develop their own single-array guidelines on what metric values are predictive of high- or poor-quality samples. However, said guidelines are likely to be dependent on sample type and Affymetrix is unable to recommend such guidelines for all possible situations. It is also important to note that the trend toward favoring model-based signal estimation algorithms (for all microarray experiments even beyond the Affymetrix platform) makes single-array quality determination very difficult due to the necessity of analyzing multiple arrays at once to calculate signal estimates.

I have a sample with an "all_rle_mean" value that is an outlier relative to the rest of my samples. Is there another metric that I can examine to confirm this?

"All_rle_mean" is an averaged value across all probe sets. A user can view the non-averaged values by examining the relative log expression box plots. Furthermore, the all_rle_mean often correlates with all_mad_residual_mean.

Both all_rle_mean and all_mad_residual_mean show outliers in the samples at the last time point in my time course experiment. Should I toss this time point?

It's important to remember that both of these metrics measure how different a sample is relative to the consensus of many samples. It is common for some samples in an experiment to be "highly affected," such as the last time point in a time course or the highest dose in a dose-response experiment. The user should be more tolerant of samples with outliers in all_rle_mean or all_mad_residual_mean if the biological treatment would predict that the expression profile would be highly altered.

Is there one quality metric that is most predictive of the quality of a hybridization?

Pos_vs_Neg_AUC is a great first-pass metric. If it is below 0.8, then that's a pretty good indicator that sample problems exist. However, having it above 0.8 isn't a guarantee that the sample is good.

Why isn't DABG available in my gene-level analyses?

The DABG algorithm evaluates detection by combining probe-level p-values that are assumed to be monitoring the same region of a transcript. This assumption is met for an exon-level probe. However, when combining all probes across a transcript in a gene-level analysis, this assumption is not guaranteed. Some of the probes may hit parts of the gene which are expressed while others may not, and yet the gene is still expressed. Since this type of scenario could generate misleading detection calls, DABG is not considered a robust gene-level metric.

One of my hybridizations results in skewed values for several of the metrics, but the values aren't grossly out of line. I feel like I could make a case to keep it in or toss it. What should I do?

If a sample appears "on the line," it is best to leave the sample in the experiment for analysis and simply flag it as "questionable." During the initial analysis, treat all samples uniformly. Once candidate genes have been identified, review how the questionable sample displays data relative to the other replicates within the sample. It's always possible to remove a sample later on the analysis workflows. Microarray data is frequently more robust than one might anticipate.

Contact Information

Affymetrix, Inc.

3420 Central Expressway
Santa Clara, CA 95051 USA
E-mail: support@affymetrix.com
Tel: 1-888-362-2447 (1-888-DNA-CHIP)
Fax: 1-408-731-5441

Affymetrix UK Ltd

Voyager, Mercury Park,
Wycombe Lane, Wooburn Green,
High Wycombe HP10 0HH
United Kingdom
E-mail: supporteurope@affymetrix.com
UK and Others Tel: +44 (0) 1628 552550
France Tel: 0800919505
Germany Tel: 01803001334
Fax: +44 (0) 1628 552585

Affymetrix Japan, K. K.

Mita NN Bldg, 16 Floor
4-1-23 Shiba, Minato-ku
Tokyo 108-0014, Japan
E-mail: supportjapan@affymetrix.com
Tel: +81-(0)3-5730-8200
Fax: +81-(0)3-5730-8201