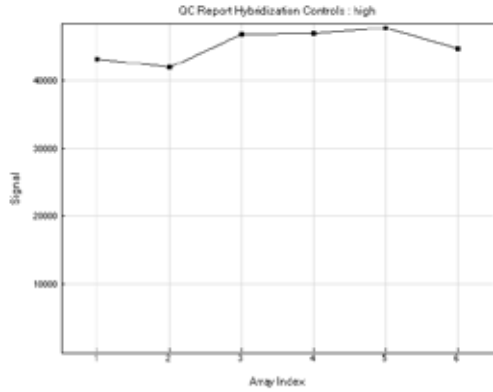


Figure 2: Genomestudio Control Summary Plot



Signal intensity values (y-axis) of hybridization controls are consistently high across all six arrays/samples (x-axis) from a gene expression profiling experiment. This graph does not indicate a data quality problem.

ments from the actual sample of interest. Poor performance of these controls may indicate a problem related to the sample or labeling. Further details of the control plots in GenomeStudio are provided in Table 1, the GenomeStudio Gene Expression Module User Guide, and related Assay Guides.

Normal variations in control plot values can arise due to incidental factors such as system setup, sample origin, and BeadChip type. These factors make it difficult to determine data quality by comparison to a specified expected value for each QC metric. To minimize the influence of these factors, relative—rather than absolute—control values should be used as QC criteria. Relative comparisons of control values can be made by:

1. Identification of outliers by comparison to current and historical data. Outlying samples for any given control metric can be quickly identified in GenomeStudio by using the control summary plot and expanding the plots to view QC values for individual samples (Figure 2). The data can also be exported and examined in other spreadsheet software, enabling further comparisons with historical data.
2. Comparison of samples across two or more control metrics to ensure a consistent ratio between relevant control values. This can be shown by simultaneously displaying and comparing different control plots in GenomeStudio. Alternatively, the control data can be exported and relevant values co-plotted using other software (Figure 3). For example, typical comparisons for the Direct Hyb assay include:
 - Housekeeping & Background. While housekeeping genes are known to fluctuate as a function of tissue type, they should be fairly consistent across arrays when from a similar sample source. Housekeeping genes should produce a higher signal than background.
 - Perfect Match (PM) & Mismatch (MM2). The PM probe signal is expected to be higher than the MM2 probe signal (Figure 3). Deviations in the ratio of PM to MM2 signal for a given sample could indicate a problem with specificity in the experiment.

Table 1: Control Plots in Genomestudio

A: Direct Hyb Control Plots

Control Metric	Expected value
Hybridization Controls*	High > Medium > Low
Low Stringency*	PM > MM2
Biotin and High Stringency*	High
Negative Controls (Background and Noise)	Low
Gene Intensity (Housekeeping and All Genes)	Higher than Background (Housekeeping > All Genes)
Labeling and Background	If used, Labeling > Background; Otherwise, Labeling ≈ Background.

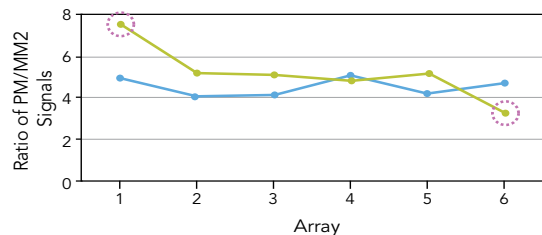
B: WG DASL Control Plots

Control Metric	Expected value
Hybridization Controls*	High > Medium > Low
Contamination	One Code High, Others Low
Stringency† (Low and High)	Low: Red > Green High: Green > Red
Negative Controls (Background and Noise)	Low
Genes (Genes and Variation)	Higher than Background
Gap‡	Higher than Background

*Sample-independent control metrics.

†DASL stringency and gap controls are designed against the glutaminyI-tRNA-synthetase gene (QARS) and are therefore dependent on the expression level of QARS in the sample of interest.

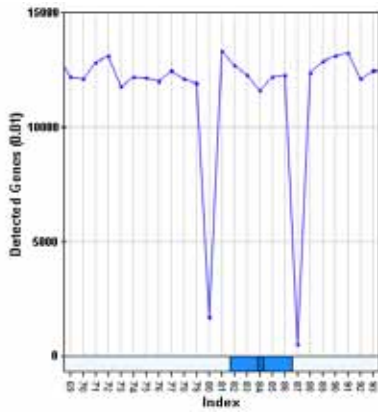
Figure 3: Comparison Across Control Metrics, PM/MM2 Ratio Line Plot



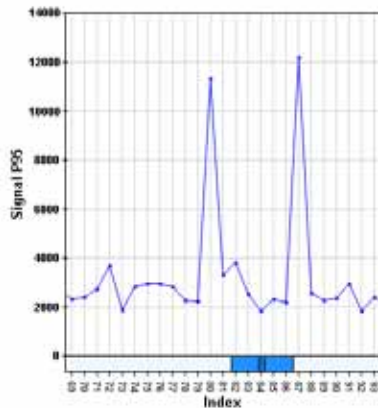
A plot of the ratio of PM/MM2 probe signals across several samples from two different BeadChips (blue and green). In the case of the blue BeadChip, all six samples have similar ratios approximately 4–5 PM/MM2. However, some arrays from the green BeadChip (circled) exhibit deviating ratios, indicating a possible difference in stringency between arrays 1 and 6.

Figure 4: Samples Table Plots

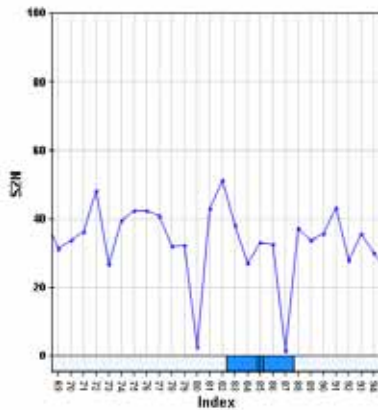
A: Number of Genes Detected for Each Sample



B: P95 Intensity for Each Sample



C: Signal-to-Noise Ratio for Each Sample



Three example line plots of control metrics that can be found in the metrics table. Samples 80 and 87 are detected as outliers in all three plots. This is an indication that these two samples may need to be repeated or excluded from further analysis.

Samples Table Metrics

Several metrics in the Samples Table allow rapid assessment of sample quality. These include the number of genes detected, the p95 intensity, and signal-to-noise ratios.

Number of Genes Detected

GenomeStudio calculates and reports a detection p-value, which represents the confidence that a given transcript is expressed above the background defined by negative control probes. This detection score determines whether a transcript on the array is called detected. A value below the user-defined p-value threshold of either 0.01 or 0.05 indicates a gene is detected. An unusually low number of detected transcripts could result from a number of causes such as high background on the array, low signal, or poor stringency. Thus, the number of detected transcripts is a good overall QC indicator. All samples on a given BeadChip, prepared from the same sample source, should have a similar number of detected transcripts (Figure 4a). However, this may not be the case with samples from highly heterogeneous tissues or from necrotic tissue.

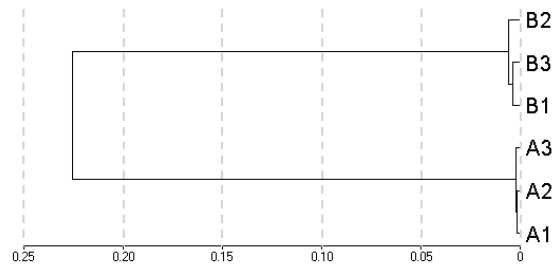
95th Intensity Percentile (p95)

This metric provides a quick way to examine the high-end intensity variation across samples (Figure 4b). Outlier arrays can be easily identified using these plots.

Signal-to-Noise Ratios (p95/p05 and p95/Background)

These metrics can be calculated using the User Defined Function (🔗) in the Samples Table, and provide a quick way to visualize the overall strength of measured signal, compared to the background (Figure 4c). This calculation should be performed on data that have not been normalized.

Figure 5: Dendrogram Showing Clustering Of Similar Samples



Correlations between six samples are plotted as a dendrogram. As expected, biological replicates A1–A3 and B1–B3 cluster tightly together within groups as sub-trees, and there is lower correlation between samples from different groups.

