ab quality metrics report

Summary

<table>
<thead>
<tr>
<th>Array #</th>
<th>Array Name</th>
<th>MA-plot</th>
<th>Spatial distribution</th>
<th>Boxplots/Density plots</th>
<th>Heatmap</th>
<th>RLE</th>
<th>NUSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Aug04_no_phe_A.cel</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Aug04_no_phe_C.cel</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Aug04_phe_B.cel</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>*</td>
</tr>
<tr>
<td>4</td>
<td>Jun04_no_phe.cel</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>*</td>
</tr>
<tr>
<td>5</td>
<td>Jun04_phe.cel</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>*</td>
</tr>
</tbody>
</table>

*array identified as having a potential problem or as being an outlier.

Index

- **Individual array quality**
  - MA-plot
  - Spatial distribution of features intensites
  - Row-Column effect
- **Homogeneity between arrays**
  - Boxplots
  - Density plots
- **Between array comparison**
- **Variance mean dependency**
- **Affymetrix specific plots**
  - RNA degradation plot
  - RLE
  - NUSE
  - Affymetrix QC stats
  - PM MM plot
Section 1: Individual array quality

Figure 1 represents MA plot for each array. M and A are defined as:

\[
M = \log_2(I_1) - \log_2(I_2)
\]

\[
A = \frac{1}{2} \left( \log_2(I_1) + \log_2(I_2) \right)
\]

where \(I_1\) is the intensity of the array studied and \(I_2\) is the intensity of a "pseudo"-array, which have the median values of all the arrays. Typically, we expect the mass of the distribution in an MA plot to be concentrated along the \(M = 0\) axis, and there should be no trend in the mean of \(M\) as a function of \(A\). Note that a bigger width of the plot of the M-distribution at the lower end of the A scale does not necessarily imply that the variance of the M-distribution is larger at the lower end of the A scale: the visual impression might simply be caused by the fact that there is more data at the lower end of the A scale. To visualize whether there is a trend in the variance of \(M\) as a function of \(A\), consider plotting \(M\) versus rank\((A)\).
Figure 2: False color representations of the arrays' spatial distributions of feature intensities. The color scale is shown in the panel on the right, and it is proportional to the ranks. These plots may help in identifying patterns that may be caused, for example, spatial gradients in the hybridization chamber, air bubbles, spotting or plating problems.
Figure 3 shows the boxplots of the log₂ intensities grouped by row (left panel) and column (right panel) of the array. If there is no spatial effect, the boxes should be homogeneous in width and y position.

Section 2: Homogeneity between arrays
**Figure 4** presents boxplots of the log$_2$(Intensities). Each box corresponds to one array. If the arrays are homogeneous, the boxes should have similar widths and y position.
Figure 5 shows density estimates (histograms) of the data. Arrays whose distributions are very different from the others should be considered for possible problems.

Section 3: Between array comparison
Figure 6 shows a false color heatmap of between arrays distances, computed as the median absolute difference of the M-value for each pair of arrays.

\[ d_{xy} = \text{median}|M_{xi} - M_{yi}| \]

Here, \( M_{xi} \) is the M-value of the \( i \)-th probe on the \( x \)-th array, without preprocessing. This plot can serve to detect outlier arrays.

Consider the following decomposition of \( M_{xi} \): \( M_{xi} = z_i + \beta_{xi} + \epsilon_{xi} \), where \( z_i \) is the probe effect for probe \( i \) (the same across all arrays), \( \epsilon_{xi} \) are i.i.d. random variables with mean zero and \( \beta_{xi} \) is such that for any array \( x \), the majority of values \( \beta_{xi} \) are negligibly small (i.e. close to zero). \( \beta_{xi} \) represents differential expression effects. In this model, all values \( d_{xy} \) are (in expectation) the same, namely 2 times the standard deviation of \( \epsilon_{xi} \). Arrays whose distance matrix entries are way different give cause for suspicion. The dendrogram on this plot also can serve to check if, without any probe filtering, the arrays cluster accordingly to a biological meaning.
Section 4: Variance mean dependency

Figure 7

For each feature, the plot on Figure 7 shows the empirical standard deviation of the intensities of all the arrays on the y-axis versus the rank of the mean of intensities of the arrays on the x-axis. The red dots, connected by lines, show the running median of the standard deviation. After vsn normalization, this should be approximately horizontal, that is, show no substantial trend.

Section 5: Affymetrix specific plots
RNA degradation plot

![RNA degradation plot](image)

**Figure 8**
RLE plot

Figure 9
In this section we present diagnostic plots based on tools provided in the affyPLM package. In **Figure 8** a RNA digestion plot is computed on normalized data (so that standard deviation is equal to 1). In this plot each array is represented by a single line. It is important to identify any array(s) that has a slope which is very different from the others. The indication is that the RNA used for that array has potentially been handled quite differently from the other arrays. **Figure 9** is a Relative Log Expression (RLE) plot and an array that has problems will either have larger spread, or will not be centered at \( M = 0 \), or both. **Figure 10** is a Normalized Unscaled Standard Error (NUSE) plot. Low quality arrays are those that are substantially elevated or more spread out, relative to the other arrays. NUSE values are not comparable across data sets. Both RLE and NUSE are performed on preprocessed data (background correction and quantile normalization). **Figure 11** represents the diagnostic plot recommended by Affymetrix. It is fully describe in the simpleaffy.pdf vignette of the package simpleaffy. Any metrics that is shown in red is out of the manufacturer's specific boundaries and suggests a potential problem, any metrics shown in blue is fine.
Figure 12 shows the density distributions of the log₂ intensities grouped by the matching of the probes. Blue, density estimate of intensities of perfect match probes (PM) and gray the mismatch probes (MM). We expect that, MM probes having poorer hybridization than PM probes, the PM curve should be shifted to the right of the MM curve.