

---

# arrayQualityMetrics report for GSE28703\_rawData

---

- [Section 1: Between array comparison](#)
  - Distances between arrays
  - Principal Component Analysis
- [Section 2: Array intensity distributions](#)
  - Boxplots
  - Density plots
- [Section 3: Variance mean dependence](#)
  - Standard deviation versus rank of the mean
- [Section 4: Individual array quality](#)
  - MA plots

---

+ [\*\*Array metadata and outlier detection overview\*\*](#)

---

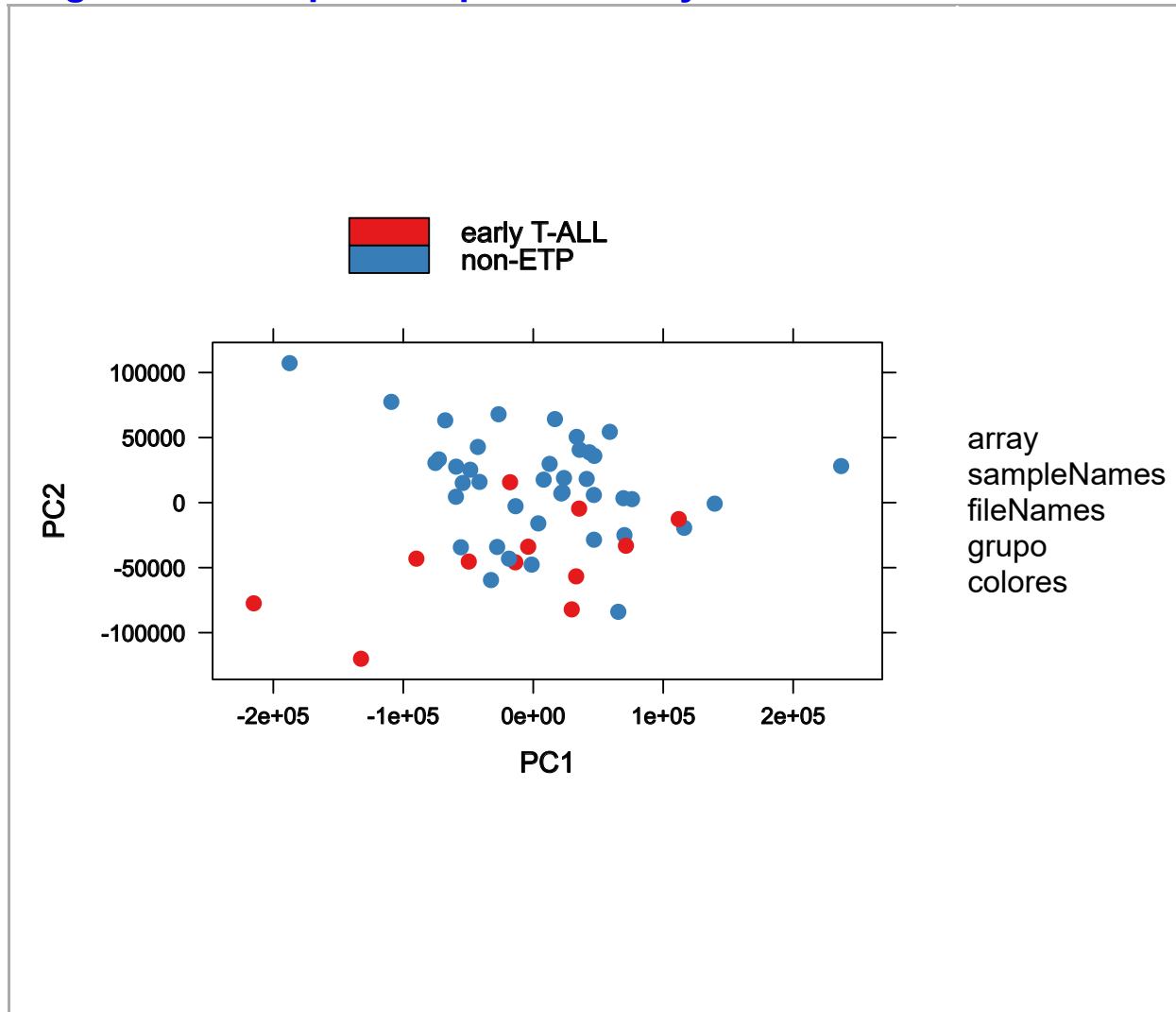
## Section 1: Between array comparison

- [\*\*Figure 1: Distances between arrays.\*\*](#)



+ Figure 2: Outlier detection for Distances between arrays.

- Figure 3: Principal Component Analysis.

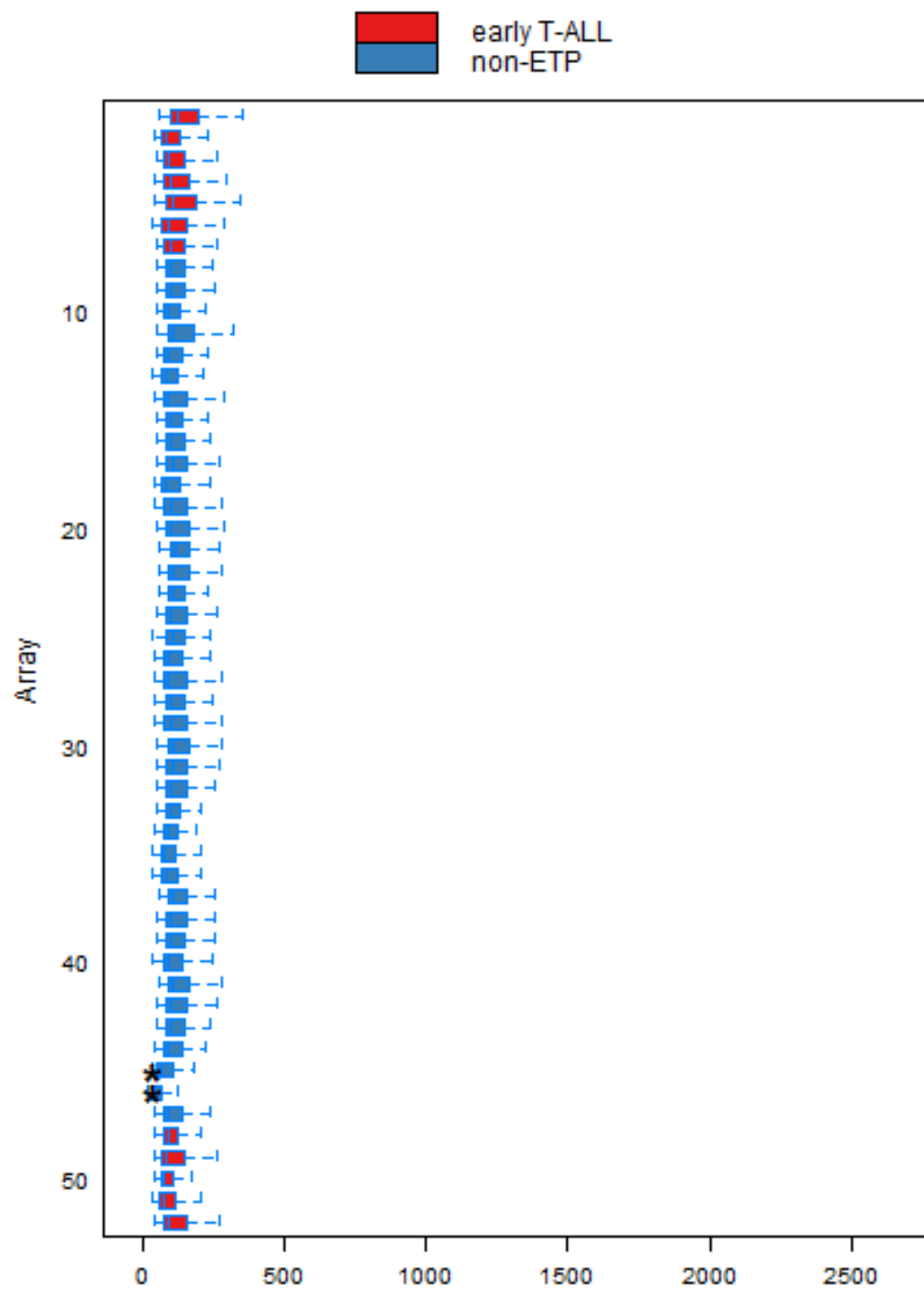


**Figure 3** ([PDF file](#)) shows a scatterplot of the arrays along the first two principal components. You can use this plot to explore if the arrays cluster, and whether this is according to an intended experimental factor, or according to unintended causes such as batch effects. Move the mouse over the points to see the sample names.

Principal component analysis is a dimension reduction and visualisation technique that is here used to project the multivariate data vector of each array into a two-dimensional plot, such that the spatial arrangement of the points in the plot reflects the overall data (dis)similarity between the arrays.

## Section 2: Array intensity distributions

- Figure 4: Boxplots.

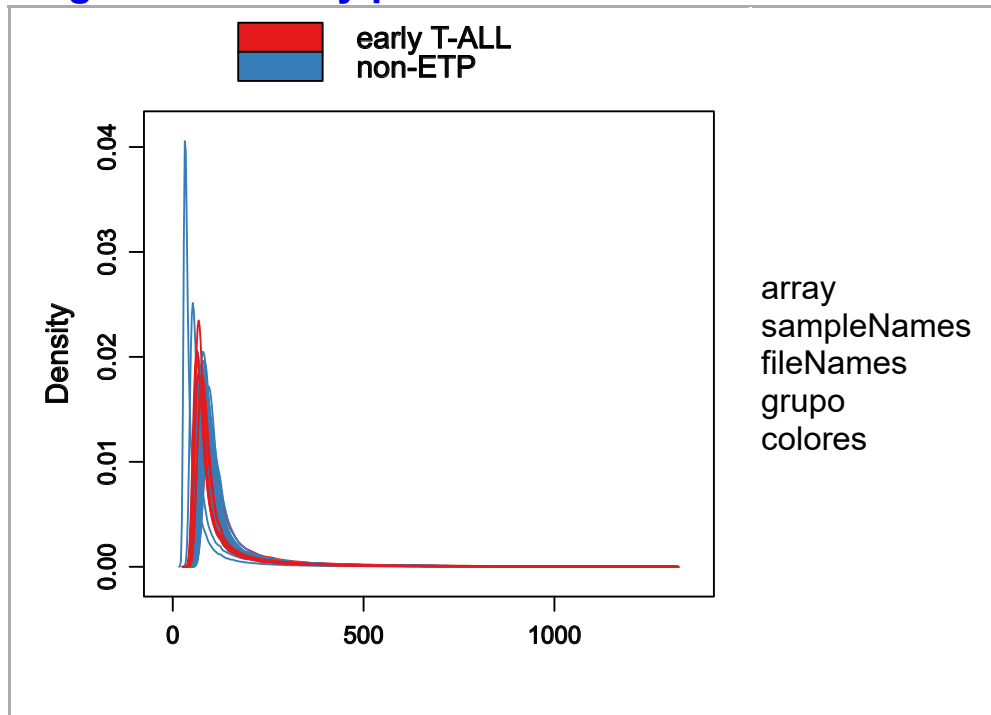


**Figure 4** ([PDF file](#)) shows boxplots representing summaries of the signal intensity distributions of the arrays. Each box corresponds to one array. Typically, one expects the boxes to have similar positions and widths. If the distribution of an array is very different from the others, this

may indicate an experimental problem. Outlier detection was performed by computing the Kolmogorov-Smirnov statistic  $K_a$  between each array's distribution and the distribution of the pooled data.

+ **Figure 5: Outlier detection for Boxplots.**

- **Figure 6: Density plots.**

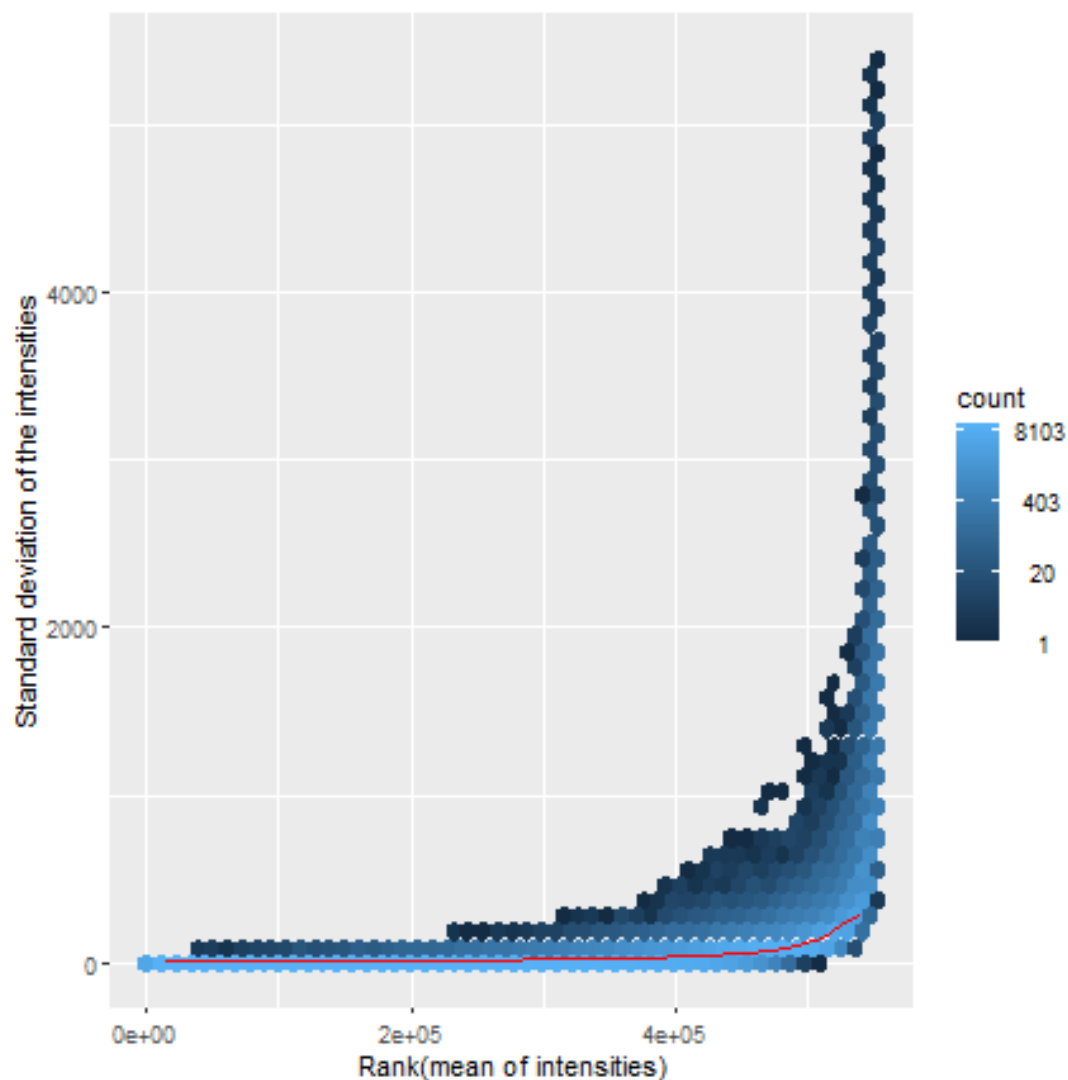


**Figure 6** [\(PDF file\)](#) shows density estimates (smoothed histograms) of the data. Typically, the distributions of the arrays should have similar shapes and ranges. Arrays whose distributions are very different from the others should be considered for possible problems. Various features of the distributions can be indicative of quality related phenomena. For instance, high levels of background will shift an array's distribution to the right. Lack of signal diminishes its right right tail. A bulge at the upper end of the intensity range often indicates signal saturation.

---

## Section 3: Variance mean dependence

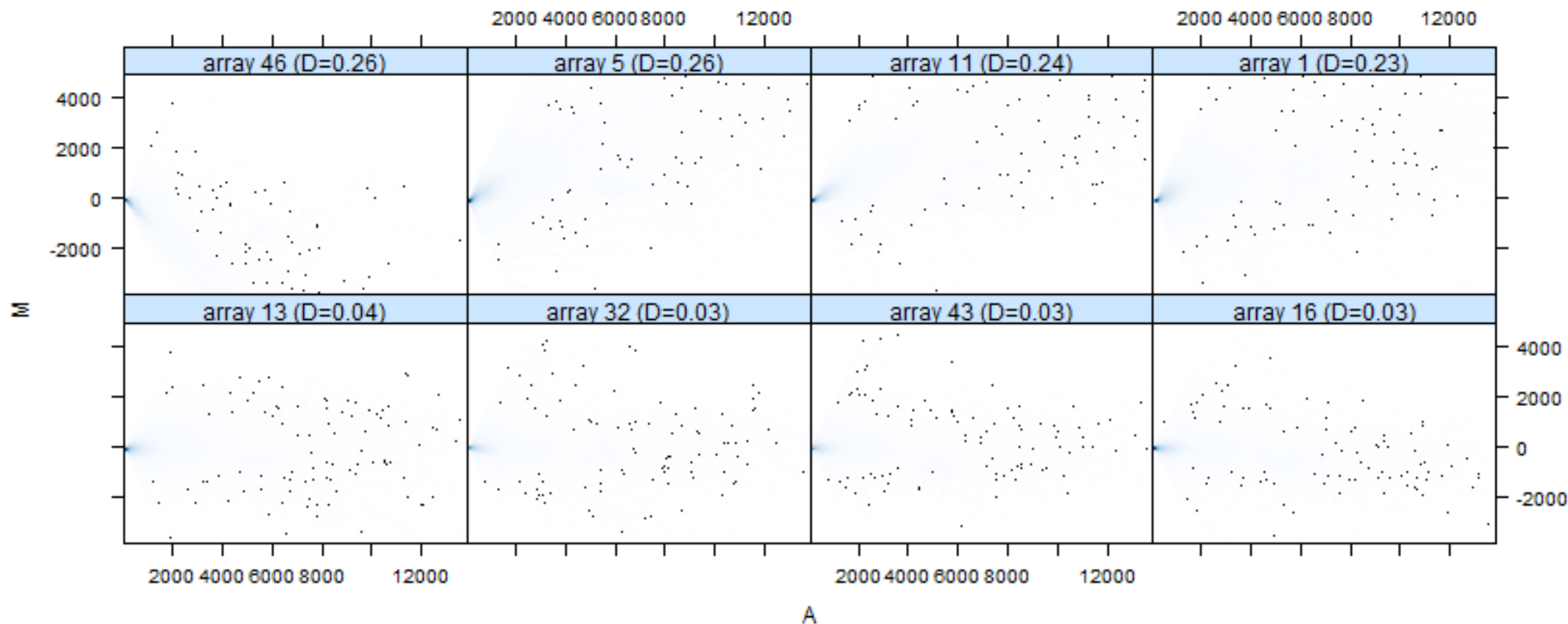
- **Figure 7: Standard deviation versus rank of the mean.**



**Figure 7** [\(PDF file\)](#) shows a density plot of the standard deviation of the intensities across arrays on the y-axis versus the rank of their mean on the x-axis. The red dots, connected by lines, show the running median of the standard deviation. After normalisation and transformation to a logarithm(-like) scale, one typically expects the red line to be approximately horizontal, that is, show no substantial trend. In some cases, a hump on the right hand of the x-axis can be observed and is symptomatic of a saturation of the intensities.

## Section 4: Individual array quality

- **Figure 8: MA plots.**



**Figure 8** [\(PDF file\)](#) shows MA plots.  $M$  and  $A$  are defined as:

$$M = \log_2(I_1) - \log_2(I_2)$$

$$A = 1/2 (\log_2(I_1) + \log_2(I_2)),$$

where  $I_1$  is the intensity of the array studied, and  $I_2$  is the intensity of a "pseudo"-array that consists of the median across 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  $M$  as a function of  $A$ . If there is a trend in the lower range of  $A$ , this often indicates that the arrays have different background intensities; this may be addressed by background correction. A trend in the upper range of  $A$  can indicate saturation of the measurements; in mild cases, this may be addressed by non-linear normalisation (e.g. quantile normalisation).

Outlier detection was performed by computing Hoeffding's statistic  $D_a$  on the joint distribution of  $A$  and  $M$  for each array. Shown are first the 4 arrays with the highest values of  $D_a$ , then the 4 arrays with the lowest values. The value of  $D_a$  is shown in the panel headings. 6 arrays had  $D_a > 0.15$  and were marked as outliers. For more information on Hoeffding's  $D$ -statistic, please see the manual page of the function `hoeffd` in the `Hmisc` package.

**+ Figure 9: Outlier detection for MA plots.**



