Probe intensities, chip2

Probe intensities, chip1

Log-log plot

MVA plot

(log chip1 + log chip2)/2
Normalization

Figure 1: Comparison of linear normalization (A) with signal dependent normalization (B). Axes: Log of ratio ($M = \log(\text{chip1}/\text{chip2})$) versus average log intensity ($A = (\log \text{chip1} + \log \text{chip2})/2$).

Linear Normalization ($M' = M + c$):
- Total signal constant ($\Sigma (M + c) = 0$)
- Housekeeping gene constant ($M_{\text{housek}} + c = 0$)

Non-linear Normalization ($M' = M + f(A)$):
- Curve fitting through all points ($f_{\text{lowess}}(M)$)
- Curve fitting through invariant ranking genes ($f(M_{\text{invariantrank}})$)
- Curve fitting through distribution quantiles ($f_{\text{qspline}}(M_{\text{quantiles}})$)

All methods are available at www.bioconductor.org
Logit Normalization
Not performed on the ratio between two chips or channels but performed on each chip separately. For each probe-target hybrid intensity $y$, the following transformation is performed:

$$\text{logit}(y) = \log\left( \frac{y - \text{background}}{\text{saturation} - y} \right),$$

where $\text{background}$ is calculated as the minimum intensity measured on the chip minus 0.1% of the signal intensity range: $\text{min} - 0.001 \times (\text{max} - \text{min})$, and $\text{saturation}$ is calculated as the maximum intensity measured on the chip plus 0.1% of the signal intensity range: $\text{max} + 0.001 \times (\text{max} - \text{min})$. The result is then $Z$-transformed to mean zero and standard deviation 1.
Expression index calculation
From up to 40 probe intensities to one expression value per gene

• Average Difference (Affymetrix MAS 4.0)

$$\text{AvgDiff} = \frac{\sum_n (PM_n - MM_n)}{N}$$

where \(N\) is the number of probe pairs used

• Signal (Affymetrix MAS 5.0)

$$\text{Signal} = \text{Tukeybiweight} \left[ \log(PM_n - CT_n) \right],$$

where CT < PM.

• Li-Wong Model-based expression index

$$\hat{\theta} = \frac{\sum_n (PM_n - MM_n) \phi_n}{N}$$

where \(\phi_n\) is a scaling factor that is specific to probe pair \(PM_n - MM_n\) and is obtained by fitting a statistical model to a series of experiments. Works even better if MM is omitted.

• Robust Multiarray Average

\[
\text{RMA} = \text{Medianpolish}[\log(\text{PM}_n - \alpha_n)]
\]

where \( \alpha \) is a scaling factor that is specific to probe \( \text{PM}_n \) and is obtained by fitting a statistical model to a series of experiments.


• Position-dependent Nearest Neighbor Model.
A thermodynamic model of probe-target interaction.

• Logit-t.
Do not condense into one number. Instead, perform a \( t \)-test for each \( \text{PM} \) probe in the set. Then use the median \( P \)-value of all \( \text{PM} \) probes to determine whether gene is differentially expressed.
Chip variance

Replicate variance

Experiment variance
Hypothesis testing

For each gene we test the hypothesis that sample and control were sampled from the same population. This we call our NULL hypothesis:

\[
\bar{X}_1, \bar{X}_2, \bar{X}_3, \bar{X}_4
\]

\[
n_1 = 3 \quad S_1^2 \quad n_2 = 3 \quad S_2^2
\]

To calculate the probability that they were sampled from the same population we calculate the t-statistic:

\[
T = \frac{\bar{X}_2 - \bar{X}_1}{\sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}
\]

where \(\bar{X}_2 - \bar{X}_1\) is the difference in means and \(\sigma\) is the pooled variance of \(S_1^2\) and \(S_2^2\). This statistic has the t distribution:

If the total area is 5%, then \(P = 0.05\) and there is a 5% probability that the observed difference occurs when \(X_1\) and \(X_2\) were sampled from the same population (i.e. the NULL hypothesis is true). If \(P < 0.05\) we reject the NULL hypothesis and accept the alternative, that the samples come from two populations with different means.
False positives

Note that if \( P = 0.05 \) and we reject the NULL hypothesis, we have a 5% chance of making a mistake (false positive conclusion).

Multiple testing

If we test 10,000 genes and use a cutoff in \( P \)-value of 0.05, we expect to make 0.05 \times 10,000 = 500 false positive conclusions. If we only accept 10 false positive conclusions, we have to use a cutoff of 10/10,000 = 0.001.

Example: In an experiment we find 100 genes that have a \( P \)-value below the cutoff of 0.001. We expect 10 false positives, so 10/100 or 10% of the genes are expected false positive. This is called the false discovery rate (FDR, Benjamini & Hochberg 1995).

\[
\text{ANOVA}
\]

If we have more than two categories, we cannot use the \( t \)-test directly. Instead, we perform an ANalysis Of VAriance (ANOVA):

\[
F = \frac{\hat{S}_b^2}{\hat{S}_w^2}
\]

where \( \hat{S}_b^2 \) is the variance between groups and \( \hat{S}_w^2 \) is the pooled variance within groups (based on \( S_1^2 \), \( S_2^2 \) and \( S_3^2 \). The \( F \)-statistic has the \( F \)-distribution and we calculate the
\[ P \text{-value} \] and correct for multiple testing as above.

**Significance testing by permutation**

We can also test significance by permutation. If we permute the data by mixing samples and controls before we perform the \( t \)-test or ANOVA we can determine how many genes have a \( P \)-value below the chosen cutoff. Comparing that number of genes to the number of genes in the original (non-permuted) data set that scored below cutoff will give the false discovery rate (FDR; Tusher, Tibshirani & Chu, 2001).