# Robin Users’ Manual

PRELIMINARY

## Introduction

Robin represents an easy to use graphical interface for Affymetrix microarray analysis functions from R/BioConductor.

## Preconditions and Glossary

### Commonly used Terms

Robin helps in evaluating microarrays using advanced normalization strategies and statistics from R/BioConductor. Nevertheless, please bear in mind that most statistics and most normalization techniques make some strong assumption and have some general terminology.

When dealing with microarray, almost always one will deal with values which have been transformed by taking the logarithm to the base of 2. The reason is, that by logging the data, the data becomes roughly normally distributed (Gauss shaped), which then allows using tests, like student’s t-test, making assumption about standard deviations and so forth. Unlogged data is almost always **NOT** normally distributed, meaning t-tests are NOT applicable (even tough they might still perform reasonably well). Thus, a difference of 1 unit means a two-fold increase or decrease in expression.

Often data is not represented as treatment value and control value, but instead of **M** and **A**. Here, **M** stands for treatment-control (on the log scale, being a division on the normal scale), and **A** stands for (treatment+ control)/2. So M is a measure of your treatment effect and A of the expression level of that gene.

(Actually another reason for using M and A values is that it is easier to see if values deviate from the zero line as if they were deviating from a line with a slope of one.)

### Affymetrix Files

When dealing with Affymetrix chips, you will be confronted with .CEL and .CDF files, the former describes the scanned intensity for every spot (usually there are 2 times ~11 spots per gene). The CDF file describes where the spots for a probe-set are to be found on the chip, since these are not clustered to compensate for local effects such as bubbles, smears, etc.

### Assumptions

The strongest assumption probably being, that not much changes in your experiment. I.e. the assumption is that let’s say not more than 5, 10 % of your genes are changing and that thus everything is comparable.

If this assumption is violated, you may not get satisfactory results, or worse wrong results. To demonstrate this issue, just consider the probably oldest, easiest (and obsolete) normalization, namely median centring. Here, one just subtracts the median of one experiment from each data point. In this extreme example, Gene1 and Gene2 are completely switched off.

|  |  |  |
| --- | --- | --- |
|  | XP1 | XP2 |
| Gene1 | 10.2 | 0 |
| Gene2 | 3.2 | 0 |
| Gene3 | 4.5 | 4.7 |
| Gene4 | 7.8 | 7.9 |
| Gene5 | 9.9 | 9.8 |
| Gene6 | 10 | 10.2 |
| median | 8.85 | 6.3 |

Table 1: Experiment before normalization

|  |  |  |
| --- | --- | --- |
|  | XP1 | XP2 |
| Gene1 | 1.35 | -6.3 |
| Gene2 | -5.65 | -6.3 |
| Gene3 | -4.35 | -1.6 |
| Gene4 | -1.05 | 1.6 |
| Gene5 | 1.05 | 3.5 |
| Gene6 | 1.15 | 3.9 |

Table 2: Experiment after normalization

As and effect, Genes 5 and 6 seem to be upregulated, even though they were unchanged. These effects would disappear in this case, if also some genes were turned on, which often might be the case, but if you have strong suspicions, that very many genes change, and/or that these change in one direction only, you might have to consult statistic counselling.

## What does Robin do?

You can use Robin for a

1. quality assessment of your Affymetrix data
2. normalization of your microarray data
3. detection of differentially expressed genes
4. preparion of the data for an import into MapMan and/or excel
5. generation of informative plots on your experiment

## Using Robin

Firstly, when using Robin, you just have to localize your CEL files. Robin comes preinstalled with a specialized CDF files for a small selection of organisms (arabidopsis maize, etc), when dealing with other plants, you will need an internet connection, so Robin can use the Bioconductor framework to install you plant CDF files. The INFO button can be used to display some details about the imported CEL files such as microarray type, algorithm parameters and all the technical data included in the header section of the cell file.

fig1_celfiles.tiff

Figure : Importing CEL files into Robin

After having selected your CEL files, you are presented with various options to investigate into the quality of your arrays.

fib2_qc_choice.tiff

Figure : Quality control options available in Robin.