Quick start guide for marray

Yee Hwa Yang

August 25, 2004

1. Department of Medicine, University of California, San Francisco,
http://www.biostat.ucsf.edu/jean

1 Overview

This document provides a brief guide to the *marray* package, which is packages for diagnostic plots and normalization of cDNA microarray data. Information on the other packages can be found in the other vignettes.
There are three main components to this package. These are:

- Reading in data.
- Perform simple diagnositc plots to access quality.
- Normalization.

2 Getting started

Installing the package. To install the *marray* package for Windows operating systems, first start R and make sure you are connected to the internet. Next, select "**Packages**" from the menu and click on " **Install package(s) from Bioconductor...**". Lastly, select *marray* from the pop–up windows and click click on "**OK**".

Getting started Below is an outline to get you started:

1. Create a directory and move all the relevant image processing output files (e.g. `.gpr` files) and a file containing gene (or spot) descriptions (e.g. `.gal` file) to that directory.
2. Start R in that directory. Under the windows system, click on the menu **File** and select **Change dir...**. This will result in a pop–up window for you to enter the location of the desired directory containing all the relevant image processing files. Alternatively, click on **browse** to nevigate and select the location of the directory. You can double check that you are in the correct directory by selecting **Display File(s)...** under the **File** menu.
3. To load the package in your R session, type:

```
> library(marray)
```

4. If your working directory contains *GenePix* files (`.gpr`), run the following command.

   ```
   > data <- read.GenePix()
   ```

5. If your working directory contains *Spot* files (`.spot`), run the following command.

   ```
   > data <- read.Spot()
   ```

6. Perform diagnostic plots with package `arrayQuality`.

7. Perform print-tip normalization.

   ```
   > normdata <- maNorm(data)
   ```

8. Perform simple diagnostic plots

   ```
   > image(data)
   > boxplot(data)
   > plot(data)
   ```

With the default arguments, these two functions will read in all the files residing in the directory, create a sub–directory containing the corresponding diagnostic plots (in JPEG format), create a sub–directory with basic quality information and four excel files. The four excel files are normA.xls, normM.xls, normMA.xls and quality.xls which has the normalized log–intensity (A) , log-ratios(M), both M and A values and quality information respectively for all image processing files.

This wrapper function automatically produces nine plots of

- pre– and post–normalization cDNA microarray data;
- MA–plots of pre– and post–normalization log–ratios M;
- color images of pre– and post–normalization log–ratios M;
- color images of average log–intensities A;
- histogram and overlay density of the signal to noise log–ratio for Cy5 and Cy3 channels; where the signal to noise ratios is defined as the foreground intensity (without background adjustment) over the background intensity; and
- dot–plots of M and A values for replicate controls probes.

In addition, this function automatically saves the figures to a file, in jpeg format. More detailed descriptions of all the arguments and options can be found in the help files. For instance, to view the help file for the function `gpTools` in a browser, use `help.start()` followed by `?gpTools`.

Please contact us if you have other image processing output formats and would like a similar wrapper functions.

3 Other vignettes and packages

Greater details can be found in other vignettes. These are:

`marrayClasses`. This vignette describes basic class definitions and associated methods for pre– and post–normalization intensity data for batches of arrays.

`marrayInput`. This vignette describes functionality for reading microarray data into R, such as intensity data from image processing output files (e.g. `.spot` and `.gpr` files for the `Spot` and `GenePix` packages, respectively) and textual information on probes and targets (e.g. from gal files and god lists). `tcltk` widgets are supplied to facilitate and automate data input and the creation of microarray specific R objects for storing these data.

`marrayPlot`. This vignette provides descriptions to functions for diagnostic plots of microarray spot statistics, such as boxplots, scatter–plots, and spatial color images. Examination of diagnostic plots of intensity data is important in order to identify printing, hybridization, and scanning artifacts which can lead to biased inference concerning gene expression.

`marrayNorm`. This vignette describes various location and scale normalization procedures, which correct for different types of dye biases (e.g. intensity, spatial, plate biases) and allow the use of control sequences spotted onto the array and possibly spiked into the mRNA samples. Normalization is needed to ensure that observed differences in intensities are indeed due to differential expression and not experimental artifacts; fluorescence intensities should therefore be normalized before any analysis which involves comparisons among genes within or between arrays.

Note: Sweave. This document was generated using the `Sweave` function from the R *tools* package. The source file is in the `/inst/doc` directory of the package *marray*.