Quick start guide for marray

Yee Hwa Yang

October 13, 2014

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

Contents

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.

After the two main pre–processing tasks, image analysis and normalization, the next steps in the statistical analysis depend on the biological question for which the microarray experiment was designed. Thus, different Bioconductor packages may be applicable. For example, for identifying differentially expressed genes, functions in the packages `limma`, `EBarrays`, `siggenes`, and `multtest` may be used.

2 Getting started

To load the `marray` package in your R session, type `library(marray)`. We demonstrate the functionality of this R packages using gene expression data from the Swirl zebrafish experiment which is included as part of the package. To load the swirl dataset, use `data(swirl)`, and to view a description of the experiments and data, type `?` `swirl`.

1. To begin, users will create a directory and move all the relevant image processing output files (e.g. `.spot` files) and a file containing target (or samples) descriptions (e.g. `SwirlSample.txt` file) to that directory. For this illustration, the data has been gathered in the data directory `swirldata`.

2. Start R in the desired working directory and load the *marray* packages:

```
> library(marray)
> dir(system.file("swirldata", package="marray"))

[1] "SwirlSample.txt" "fish.gal"        "swirl.1.spot"    "swirl.2.spot"
[5] "swirl.3.spot"    "swirl.4.spot"
```

3. **Data input:** Read in the target file containing information about the hybridization.

```
> datadir <- system.file("swirldata", package="marray")
> swirlTargets <- read.marrayInfo(file.path(datadir, "SwirlSample.txt"))
```

4. Read in the raw fluorescent intensities data, by default we assume that the file names are provided in the **first** column of the target file.

```
> mraw <- read.Spot(targets = swirlTargets, path=datadir)

Reading ...   /tmp/Rtmp0EJoLe/Rinst70842af77bdb/marray/swirldata/swirl.1.spot
Reading ...   /tmp/Rtmp0EJoLe/Rinst70842af77bdb/marray/swirldata/swirl.2.spot
Reading ...   /tmp/Rtmp0EJoLe/Rinst70842af77bdb/marray/swirldata/swirl.3.spot
Reading ...   /tmp/Rtmp0EJoLe/Rinst70842af77bdb/marray/swirldata/swirl.4.spot
```

If your working directory contains *GenePix* files (`.gpr`), run the following command. By default, the function `read.GenePix` will also set up printer layout and probe annotation information.

```
> data <- read.GenePix(targets=swirlTargets)
```

5. Read in the probe annotation information.

```
> galinfo <- read.Galfile("fish.gal", path=datadir)
> mraw@maLayout <- galinfo$layout
> mraw@maGnames <- galinfo$gnames
```

6. **Array quality assessment:**, the following command generates diagnostic plots for a qualitative assessment of slide quality. The results are saved as png files in the working directory. We uses the wrapper functions provided in the package `arrayQuality`.

```
> library(arrayQuality)
> maQualityPlots(mraw)
```

In addition, you can perform simple diagnostic plots with

```
> image(mraw)
> boxplot(mraw)
> plot(mraw)
```

7. **Normalization:** Perform print-tip normalization for each arrays and take a look at the data summary.

```
> normdata <- maNorm(mraw)
> summary(normdata)
```

8. Output the normalized log–ratios M data.

```
> write.marray(normdata)
```

9. **Identify DE genes:** Using the linear model package *limma* to identify differential expressed (DE) genes between wildtype and mutant. Perform fold-chance estimation as well as apply Bayesian smoothing to the standard errors.

```
> library(limma)
> LMres <- lmFit(normdata, design = c(1, -1, -1, 1), weights=NULL)
> LMres <- eBayes(LMres)
```

10. Show the top 50 genes and write it out into a clickable html file.

```
> restable <- toptable(LMres, number=50, genelist=maGeneTable(normdata), resort.by="M")
> table2html(restable, disp="file")
```

11. To utilize other bioconductor packages for downstream analysis, it is also possible to convert objects of class `marrayNorm` into objects of class `ExpressionSet` (see definition in the `Biobase` package), see package `convert` package for more details.

```
> library(convert)
> as(normdata, "ExpressionSet")
```

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*.