

Tutorial for Bioconductor's marrayClasses package

Sandrine Dudoit¹ and Yee Hwa Yang²

October 27, 2003

1. Division of Biostatistics, University of California, Berkeley,

<http://www.stat.berkeley.edu/~sandrine>

2. Department of Statistics, University of California, Berkeley, yeehwa@stat.berkeley.edu

Contents

1	Overview	1
2	Microarray classes	2
3	Creating and accessing slots of microarray objects	2
4	Basic microarray methods	3
4.1	Printing methods for microarray objects	3
4.2	Subsetting methods for microarray objects	5
4.3	Methods for accessing slots of microarray objects	6
4.4	Methods for assigning slots of microarray objects	7
4.5	Methods for coercing microarray objects	8
4.6	Functions for computing layout parameters	8

1 Overview

This document provides a short tutorial for the `marrayClasses` package, which is part of a suite of four packages for diagnostic plots and normalization of cDNA microarray data, with `marrayInput`, `marrayNorm`, and `marrayPlots`. The `marrayClasses` packages contains basic class definitions and associated methods for pre- and post-normalization intensity data for batches of arrays. Greater details on the package are given in the `marrayClasses` document of the `inst/doc` directory.

Help files. As with any R package, detailed information on functions, classes and methods can be obtained in the help files. For instance, to view the help file for the class `marrayRaw` in a browser, use `help.start()` followed by `? marrayRaw` or alternately the dyadic class `? marrayRaw`.

Case study. We demonstrate the functionality of this collection of R packages using gene expression data from the Swirl zebrafish experiment. To load the Swirl dataset, use `data(swirl)`, and to view a description of the experiments and data, type `? swirl`.

```
> library("marrayClasses")
> data(swirl)
```

2 Microarray classes

The three main classes for cDNA microarray data are

marrayLayout: This class is used to keep track of important layout parameters for two-color cDNA microarrays. It contains slots for: the total number of spotted probe sequences on the array, the dimensions of the spot and grid matrices, the plate origin of the probes, information on spotted control sequences (e.g. probe sequences which should have equal abundance in the two target samples, such as housekeeping genes).

marrayRaw: This class represents pre-normalization intensity data for a batch of cDNA microarrays. A *batch* of arrays consists of a collection of arrays with the same layout ("marrayLayout"). The class contains slots for the green (Cy3) and red (Cy5) foreground and background intensities, the layout of the arrays, and descriptions of the target samples hybridized to the arrays and probe sequences spotted onto the arrays.

marrayNorm: This class represents post-normalization intensity data for a batch of cDNA microarrays. The class contains slots for the average log-intensities $A = \log_2 \sqrt{RG}$, the normalized log-ratios $M = \log_2 R/G$, the location and scale normalization values, the layout of the arrays, and descriptions of the target samples hybridized to the arrays and probe sequences spotted onto the arrays.

Other classes are **marrayInfo**, **marraySpots**, and **marrayTwo**. The function **slotNames** can be used to get information on the slots of a formally defined class or an instance of the class. For example, to get information of the slots for the **marrayLayout** class or on the slots for the object **swirl** use

```
> slotNames("marrayLayout")

[1] "maNgr"      "maNgc"      "maNsr"      "maNsc"      "maNspots"
[6] "maSub"      "maPlate"    "maControls" "maNotes"

> slotNames(swirl)

[1] "maRf"      "maGf"      "maRb"      "maGb"      "maW"      "maLayout"
[7] "maGnames"  "maTargets" "maNotes"
```

3 Creating and accessing slots of microarray objects

Creating new objects. The function **new** from the **methods** package may be used to create new objects from a given class. For example, to create an object of class **marrayInfo** describing the target samples in the Swirl experiment, one could use the following code

```
> zebra.RG <- as.data.frame(cbind(c("swirl", "WT", "swirl", "WT"),
+   c("WT", "swirl", "WT", "swirl")))
> dimnames(zebra.RG)[[2]] <- c("Cy3", "Cy5")
> zebra.samples <- new("marrayInfo", maLabels = paste("Swirl array ",
+   1:4, sep = ""), maInfo = zebra.RG, maNotes = "Description of targets for Swirl experiment")
> zebra.samples
```

Object of class `marrayInfo`.

```
      maLabels  Cy3   Cy5
1 Swirl array 1 swirl   WT
2 Swirl array 2    WT swirl
3 Swirl array 3 swirl   WT
4 Swirl array 4    WT swirl
```

Number of labels: 4

Dimensions of `maInfo` matrix: 4 rows by 2 columns

Notes:

Description of targets for Swirl experiment

Slots which are not specified in `new` are initialized to the prototype for the corresponding class. These are usually "empty", e.g., `matrix(0,0,0)`. In most cases, microarray objects can be created automatically using the input functions and their corresponding widgets in the `marrayInput` package. These were used to create the object `swirl` of class `marrayRaw`.

Accessing slots. Different components or slots of the microarray objects may be accessed using the operator `@`, or alternately, the function `slot`, which evaluates the slot name. For example, to access the `maLayout` slot in the object `swirl` and the `maNgr` slot in the layout object `L`

```
> L <- slot(swirl, "maLayout")
> L@maNgr
```

```
[1] 4
```

4 Basic microarray methods

The following basic methods were defined to facilitate manipulation of microarray data objects. To see all methods available for a particular class, e.g., `marrayLayout`, or just the print methods

```
> showMethods(classes = "marrayLayout")
> showMethods("print", classes = "marrayLayout")
```

4.1 Printing methods for microarray objects

Since there is usually no need to print out fluorescence intensities for thousands of genes, the `print` method was overloaded for microarray classes by simple report generators. For an overview of the available microarray printing methods, type `methods ? print`, or to see all print methods for the session

```
> showMethods("print")
```

Function "print":

```
x = "ANY"
```

```
x = "marrayLayout"
```

```

x = "marrayInfo"
x = "marrayRaw"
x = "marrayNorm"
x = "marrayTwo"
x = "marraySpots"
x = "data.frame"
    (inherited from x = "ANY")
x = "array"
    (inherited from x = "ANY")
x = "matrix"
    (inherited from x = "ANY")

```

For example, summary statistics for an object of class `marrayRaw`, such as `swirl`, can be obtained by `print(swirl)` or simply `swirl`.

```
> swirl
```

```
Pre-normalization intensity data:      Object of class marrayRaw.
```

```
Number of arrays:      4 arrays.
```

```
A) Layout of spots on the array:
```

```
Array layout:      Object of class marrayLayout.
```

```
Total number of spots:      8448
```

```
Dimensions of grid matrix:      4 rows by 4 cols
```

```
Dimensions of spot matrices:      22 rows by 24 cols
```

```
Currently working with a subset of 8448 spots.
```

```
Control spots:
```

```
There are      2 types of controls :
```

```
Control      N
      768      7680
```

```
Notes on layout:
```

```
No Input File
```

```
B) Samples hybridized to the array:
```

```
Object of class marrayInfo.
```

	maLabels	# of slide	Names	experiment	Cy3	experiment	Cy5	date
1	81	81	swirl.1.spot	swirl	wild type	swirl	2001/9/20	
2	82	82	swirl.2.spot	wild type	swirl	wild type	2001/9/20	
3	93	93	swirl.3.spot	swirl	wild type	swirl	2001/11/8	
4	94	94	swirl.4.spot	wild type	swirl	wild type	2001/11/8	

	comments
1	NA
2	NA
3	NA
4	NA

Number of labels: 4

Dimensions of maInfo matrix: 4 rows by 6 columns

Notes:

C:/GNU/R/rw1041/library/marrayInput/data/SwirlSample.txt

C) Summary statistics for log-ratio distribution:

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
swirl.1.spot	-2.73	-0.79	-0.58	-0.48	-0.29	4.42
swirl.2.spot	-2.72	-0.15	0.03	0.03	0.21	2.35
swirl.3.spot	-2.29	-0.75	-0.46	-0.42	-0.12	2.65
swirl.4.spot	-3.21	-0.46	-0.26	-0.27	-0.06	2.90

D) Notes on intensity data:

4.2 Subsetting methods for microarray objects

In many instances, one is interested in accessing only a subset of arrays in a batch and/or spots in an array. Subsetting methods "[" were defined for this purpose. For an overview of the available microarray subsetting methods, type `methods ? "["` or to see all subsetting methods for the session `showMethods("[")`. When using the "[" operator, the first index refers to spots and the second to arrays in a batch. Thus, to access the first 100 probe sequences in the second and third arrays in the batch `swirl` use

```
> swirl[1:100, 2:3]
```

Pre-normalization intensity data: Object of class `marrayRaw`.

Number of arrays: 2 arrays.

A) Layout of spots on the array:

Array layout: Object of class `marrayLayout`.

Total number of spots: 8448

Dimensions of grid matrix: 4 rows by 4 cols

Dimensions of spot matrices: 22 rows by 24 cols

Currently working with a subset of 100 spots.

Control spots:

There are 2 types of controls :

```
Control      N
      48      52
```

Notes on layout:
No Input File

B) Samples hybridized to the array:
Object of class `marrayInfo`.

	maLabels	# of slide	Names	experiment	Cy3 experiment	Cy5	date
2	82	82	swirl.2.spot	wild type		swirl	2001/9/20
3	93	93	swirl.3.spot		swirl	wild type	2001/11/8
	comments						
2	NA						
3	NA						

Number of labels: 2
Dimensions of `maInfo` matrix: 2 rows by 6 columns

Notes:
C:/GNU/R/rw1041/library/marrayInput/data/SwirlSample.txt

C) Summary statistics for log-ratio distribution:

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
swirl.2.spot	-1.08	-0.26	-0.12	-0.12	0.0	1.72
swirl.3.spot	-1.78	-0.84	-0.44	-0.40	0.1	1.28

D) Notes on intensity data:

4.3 Methods for accessing slots of microarray objects

A number of simple methods were defined to access slots of the microarray classes. Using such methods is more general than using the `slot` function or `@` operator. In particular, if the class definitions are changed, any function which uses the `@` operator will need to be modified. When using a method to access the data in the slot, only that particular method needs to be modified. Thus, to access the layout information for the array batch `swirl` one may also use `maLayout(swirl)`.

In addition, various methods were defined to compute basic statistics from microarray object slots. For instance, for memory management reasons, objects of class `marrayLayout` do not store the spot coordinates of each probe. Rather, these can be obtained from the dimensions of the grid and spot matrices by applying methods: `maGridRow`, `maGridCol`, `maSpotRow`, and `maSpotCol` to objects of class `marrayLayout`. Print-tip-group coordinates are given by `maPrintTip`. Similar methods were also defined to operate directly on objects of class `marrayRaw`, `marrayNorm`, `marraySpots`, and `marrayTwo`. The commands below may be used to display the number of spots on the array, the dimensions of the grid matrix, and the print-tip-group coordinates.

```
> swirl.layout <- maLayout(swirl)
```

```

> maNspots(swirl)

[1] 8448

> maNspots(swirl.layout)

[1] 8448

> maNgr(swirl)

[1] 4

> maNgc(swirl.layout)

[1] 4

> maPrintTip(swirl[1:10, 3])

[1] 1 1 1 1 1 1 1 1 1 1

```

4.4 Methods for assigning slots of microarray objects

A number of methods were defined to replace slots of microarray objects, without explicitly using the `@` operator or `slot` function. These make use of the `setReplaceMethod` function from the R `methods` package. As with the accessor methods just described, the assignment methods are named after the slots. For example, to replace the `maNotes` slot of `swirl.layout`

```

> maNotes(swirl.layout)

[1] "No Input File"

> maNotes(swirl.layout) <- "New value"
> maNotes(swirl.layout)

[1] "New value"

```

To initialize slots of an empty `marrayLayout` object

```

> L <- new("marrayLayout")
> L

```

Array layout: Object of class `marrayLayout`.

Total number of spots:

Dimensions of grid matrix: rows by cols

Dimensions of spot matrices: rows by cols

Currently working with a subset of spots.

Control spots:

Notes on layout:

```
> maNgr(L) <- 4
```

Similar methods were defined to operate on objects of class `marrayInfo`, `marrayRaw`, `marrayNorm`, `marraySpots`, and `marrayTwo`.

4.5 Methods for coercing microarray objects

To facilitate navigation between different classes of microarray objects, we have defined methods for coercing microarray objects from one class into another. A list of such methods can be obtained by `methods ? coerce`. For example, to coerce an object of class `marrayRaw` into an object of class `marrayNorm`

```
> swirl.norm <- as(swirl, "marrayNorm")
```

It is also possible to convert objects of class `marrayRaw` or `marrayNorm` into objects of class `exprSet` (see definition in the `Biobase` package)

```
> as(swirl, "exprSet")
```

```
Expression Set (exprSet) with
  8448 genes
  4 samples
      phenoData object with 6 variables and 4 cases
  varLabels
      : # of slide
      : Names
      : experiment Cy3
      : experiment Cy5
      : date
      : comments
```

4.6 Functions for computing layout parameters

In some cases, plate information is not stored in `marrayLayout` objects when the data are first read into R. We have defined a function `maCompPlate` which computes plate indices from the dimensions of the grid matrix and number of wells in a plate. For example, the `Swirl` arrays were printed from 384-well plates, but the plate IDs were not stored in the `fish.gal` file. To generate plate IDs (arbitrarily labeled by integers starting with 1) and store these in the `maPlate` slot of the `marrayLayout` object use

```
> maPlate(swirl) <- maCompPlate(swirl, n = 384)
```

Similar functions were defined to generate and manipulate spot coordinates: `maCompCoord`, `maCompInd`, `maCoord2Ind`, `maInd2Coord`. The function `maGeneTable` produces a table of spot coordinates and gene names for objects of class `marrayRaw`, `marrayNorm`, and `marraySpots`.