

Trabajo Práctico 6: Lectura de Datos de Microarreglos

0.0. Actualización del R

<http://cran.r-project.org/bin/windows/base/R-2.10.1-win32.exe>

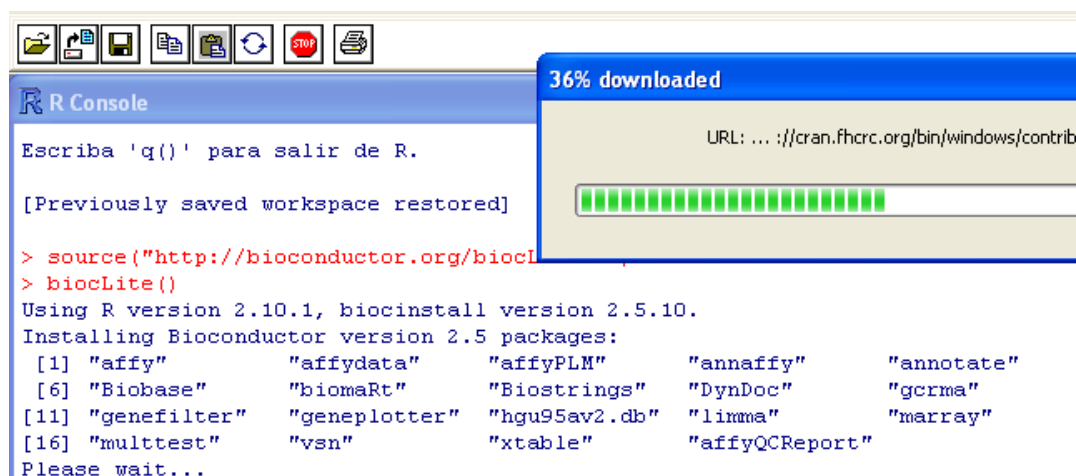
0.1 Obtención de paquetes de Bioconductor

Las diferentes versiones de Bioconductor suelen utilizar diferentes versiones de R.



Como vimos en la práctica 1 y como sugiere la página <http://www.bioconductor.org/docs/install/>, la forma más sencilla de obtener paquetes de Bioconductor es directamente desde la consola del R ejecutar los siguientes comandos

```
> source("http://www.bioconductor.org/biocLite.R")
> biocLite()
```



Siguen apareciendo mensajes:

also installing the dependencies 'DBI', 'RSQLite', 'bitops', 'affyio', 'preprocessCore', 'GO.db', 'KEGG.db', 'AnnotationDbi', 'XML', 'RCurl', 'IRanges', 'RColorBrewer', 'org.Hs.eg.db', 'simpleaffy'

probando la URL 'http://cran.fhcrc.org/bin/windows/contrib/2.10/DBI_0.2-5.zip'

Content type 'application/zip' length 392218 bytes (383 Kb)

URL abierta

downloaded 383 Kb

```
.....  
package 'DBI' successfully unpacked and MD5 sums checked  
package 'RSQLite' successfully unpacked and MD5 sums checked  
package 'bitops' successfully unpacked and MD5 sums checked  
.....
```

```
package 'vsn' successfully unpacked and MD5 sums checked  
package 'xtable' successfully unpacked and MD5 sums checked  
package 'affyQCReport' successfully unpacked and MD5 sums checked
```

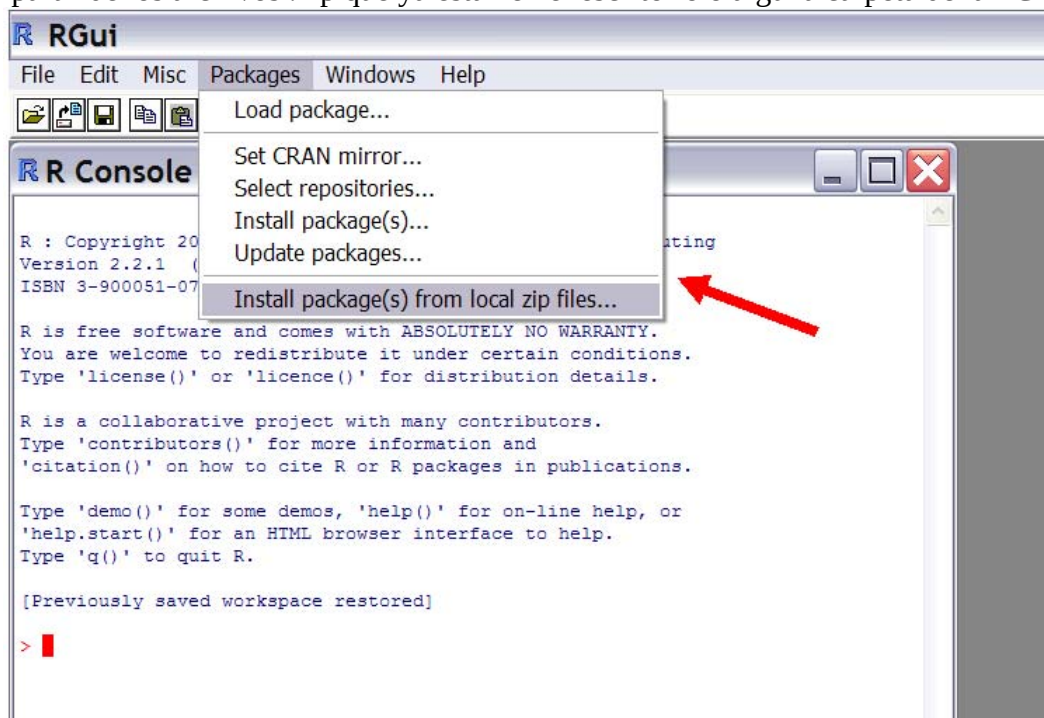
The downloaded packages are in

C:\Documents and Settings\x\Configuración local\Temp\Rtmpp4vAsf\downloaded_packages

Algunas veces la forma anterior da error y no funciona. En ese caso, otra alternativa consiste en ingresar en la página:

- <http://www.bioconductor.org/packages/1.8/bioc/> en el 2006
- <http://www.bioconductor.org/packages/release/BiocViews.html> en el 2007
- **<http://www.bioconductor.org/packages/release/bioc/> en el 2010**

bajar los archivos .zip de las versiones para windows de los paquetes que interesan: affy, affyio, arrayQuality, Biobase, beta7, convert, hexbin, limma, marray, vsn, etc. Luego desde R instalar los paquetes a partir de los archivos .zip que ya están en el escritorio o alguna carpeta de la PC



Ya estamos listos para procesar datos de experimentos de microarreglos.

1. Lectura utilizando el paquete **marray**

Inicie R en el directorio en un directorio de trabajo específico y cargue el paquete **marray**

```
> library(marray)
```

1.1 Lectura de datos “swirl” incluidos en el paquete **marray**

Para ver una descripción del experimento y los datos:

```
> ?swirl    #mire el help
```

¿Qué nombre tiene el objeto que contiene las intensidades?
¿A qué clase pertenece dicho objeto?
¿Sobre qué organismo fue realizado el experimento?
¿Cuál es el objetivo del experimento?
¿Qué significa “dye-swap experiments”?
¿Cuántos microarreglos tiene el lote?
¿Cuántas sondas tiene cada microarray, cuántas son controles?
Describa detalladamente la geometría del arreglo

Compare sus respuestas anteriores con la información que obtiene con:

```
> summary(slot(swirl, "maLayout"))  
> summary(slot(swirl, "maTargets"))
```

¿Qué tipo de objeto es **swirl**?

```
> class(swirl)  
[1] "marrayRaw"
```

```
> slotNames(swirl)  
[1] "maRf"      "maGf"      "maRb"      "maGb"      "maW"      "maLayout"  
[7] "maGnames"  "maTargets" "maNotes"
```

```
> class(slot(swirl, "maLayout"))  
> class(slot(swirl, "maGf"))
```

1.2 Acceso a diferentes partes de **swirl**

a)

a1) Acceda a los slots de **swirl** utilizando la función **slot**
a2) Describa qué muestras han sido hibridizadas a cada arreglo y con qué tinte. ¿Cual es el slot que tiene esta información?

b) Acceda a las primeras 10 filas de la matriz **"maRf"**

c) Repita lo anterior utilizando el operador **@**

d) Arme una tabla que contenga la identificación de los genes que se encuentran en las filas 1330 a 1335 y las correspondientes intensidades para el canal Rojo, para todos los microarrays.

Sugerencia: Vea primero que obtiene con

```
> slotNames (swirl)
> slotNames (swirl@maGnames)
> swirl@maGnames@maLabels[1:5]
```

e) Repita el punto anterior para los 2 primeros microarrays.

f) Obtenga un histograma de las intensidades del canal Rojo para cada uno de los microarrays.

```
par(mfrow=c(2,2))
for(i in 1:4)
  hist( slot(swirl,"maRf")[,i] )
```

g) Obtenga un boxplot de las intensidades del canal Rojo para cada uno de los microarrays, en escala original y transformados por logaritmo en base 2.

```
par(mfrow=c(2,2))
for(i in 1:4)
  boxplot( slot(swirl,"maRf")[,i] ,ylim=c(0,60000 ))

par(mfrow=c(2,2))
for(i in 1:4)
  boxplot( log2(slot(swirl,"maRf")[,i]) , ylim=c(5,17))
```

Repita lo anterior para el canal Verde y para el background

h) Obtenga los histogramas de las intensidades para los genes correspondientes a las muestras "wild type".

i) Obtenga los boxplots de las intensidades para los genes correspondientes a las muestras "wild type".

j) Modifique los gráficos anteriores de manera que tengan las mismas escalas en ambos ejes

k) Obtenga un diagrama de dispersión de las intensidades de los canales del tercer microarray. ¿Qué muestras corresponden a cada canal?

l) Repita h) -k) tomando logaritmo en base 2 de las intensidades.

m) Obtenga un gráfico M-A para el tercer microarray.

m1) con corrección de fondo
 $M = \log_2(Rf) - \log_2(Gf)$
 $A = 1/2 (\log_2(Rf) - \log_2(Gf))$

m1) sin corrección de fondo
 $MS = \log_2(Rf)$
 $AS = 1/2 (\log_2(Rf))$

n) Obtenga un gráfico M-A para cada arreglo dispuestos en una matriz de 2 x 2

o)

o1) Obtenga los boxplots de M para cada uno de los arreglos

o2) Compare con `boxplot(swirl)`. ¿Qué hace este boxplot?

1.3 Obteniendo ayuda en bioconductor

```
> library(Biobase)
```

Al cargar la librería **Biobase** obtenemos acceso al las viñetas (**Vignettes**) que contienen material introductorio de los diferentes paquetes. Podemos acceder a ellas tipeando `openVignette()`, se despliega un menú de opciones. Para obtener información sobre que realiza la función `boxplot` cuando es aplicada a un objeto de clase `marrayRaw` elegimos la opción 12.

```
> openVignette()
```

Please select a vignette:

```
1: Biobase - An introduction to Biobase and ExpressionSets
2: Biobase - Bioconductor Overview
3: Biobase - esApply Introduction
4: Biobase - Notes for eSet developers
5: Biobase - Notes for writing introductory 'how to' documents
6: Biobase - quick views of eSet instances
7: limma - Limma Vignette
8: marray - marray Normalization
9: marray - marray Overview
10: marray - marrayClasses Overview
11: marray - marrayClasses Tutorial (short)
12: marray - marrayInput Introduction
13: marray - marrayPlots Overview
```

Selección: 13

Opening C:/ARCHIV~1/R/R-210~1.1/library/marray/doc/marrayPlots.pdf

```
>
```

Se abre un documento describiendo los gráficos del paquete **marray**.

1.4 Generación de un objeto `marrayInfo`

1.4.1 Lectura de los archivos de salida del procesamiento de imágenes

Se colocan todos los archivos de salida en una carpeta, como ejemplo se han colocado en la carpeta **swirldata**:

C:\Archivos de programa\R\R-2.10.1\library\marray\swirldata

La siguiente instrucción muestra los nombres de los archivos

```
> dir(system.file("swirldata", package = "marray"))  
[1] "fish.gal"      "swirl.1.spot"  "swirl.2.spot"  "swirl.3.spot"  
[5] "swirl.4.spot"  "SwirlSample.txt"
```

Explore los archivos usando el block de notas.

1.4.2 Lectura de información sobre las muestras hibridizadas

La información sobre las muestras se encuentra en el archivo **SwirlSample.txt** y la función **read.marrayInfo** genera un objeto de clase **marrayInfo**

```
> directorio.datos <- system.file("swirldata", package = "marray")  
> muestras.swirl <- read.marrayInfo(file.path(directorio.datos,  
"SwirlSample.txt"))
```

Las siguientes instrucciones permiten la lectura desde una carpeta en particular:

C:\Archivos de programa\R\R-2.10.1\library\marray\swirldata

En este caso es lo mismo, dé un ejemplo en que las instrucciones anteriores no darán el resultado deseado.

```
> dirección.swirl <-  
"C:\\Archivos de programa\\R\\R-2.10.1\\library\\marray\\swirldata"  
  
> muestras.swirl <- read.marrayInfo(file.path(dirección.swirl,  
"SwirlSample.txt"))
```

Describe qué muestras han sido hibridizadas a cada arreglo y con qué tinte a partir de **muestras.swirl**

1.4.3 Lectura de los datos de intensidades de fluorescencia, generación de un objeto de clase `marrayRaw`

Por defecto, se supone que los nombres de los archivos que contienen las intensidades y otros estadísticos se encuentran en la primera columna del archivo de las muestras (target file). En este ejemplo los archivos han sido generados por el programa SPOT.

```
> setwd(dirección.swirl ) # para leer desde esa dirección
```

```
> crudos <- read.Spot(targets = muestras.swirl) # crea el marrayRaw
```

```
Reading ... ./swirl.1.spot
Reading ... ./swirl.2.spot
Reading ... ./swirl.3.spot
Reading ... ./swirl.4.spot
```

1.4.4 Lectura de las anotaciones sobre los microarreglos y los probes

```
> galinfo <- read.Galfile("fish.gal", path=dirección.swirl)
> class(galinfo)
[1] "list"
> crudos@maLayout <- galinfo$layout
> crudos@maGnames <- galinfo$gnames
```

```
> summary(crudos@maLayout)
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 8448spots.

Control spots:

Notes on layout:

```
> summary(crudos@maTargets)
Object of class marrayInfo.
```

	maLabels	Names	slide number	experiment	Cy3
1	swirl.1.spot	swirl.1.spot	81		swirl
2	swirl.2.spot	swirl.2.spot	82	wild type	
3	swirl.3.spot	swirl.3.spot	93		swirl
4	swirl.4.spot	swirl.4.spot	94	wild type	

	experiment	Cy5	date	comments
1	wild type		2001/9/20	NA
2	swirl		2001/9/20	NA
3	wild type		2001/11/8	NA
4	swirl		2001/11/8	NA

```
Number of labels: 4
Dimensions of maInfo matrix: 4 rows by 6 columns
```

Notes:

```
C:\Archivos de programa\R\R-
2.10.1\library\marray\swirl\data/SwirlSample.txt
>
```