

Normalizing Affy microarray data

All product names are given as examples only and they are not endorsed by the USDA or the University of Illinois.

INTRODUCTION

The following is an interactive demo describing a set of steps that we run to normalize Affymetrix oligonucleotide array data. We use either the RMA (Robust Multi-Array) normalization (retains probe level information; requires large amounts of RAM memory) or GCRMA (uses GC content of probes in normalization with RMA; gives one value for each probe set instead of keeping probe level information) normalization in the R packages *affy* and *gcrma*. The final data file is ready to be used as an input file for SAS. The SAS programs we run are explained on another page.

Click here for use SAS to analyze normalized Affymetrix data after RMA normalization.

Click here for use SAS to analyze normalized Affymetrix data after GCRMA normalization.

Please feel free to contact me with any questions or comments:

Steve Clough (sjclough@uiuc.edu)

DOWNLOAD DEMO SET

Click here to obtain a demo data set of .CEL files that you may use to test and learn how to normalize Affymetrix data.

The CEL file describes the intensities determined for every feature on a chip, without providing information about which probes correspond to which probe sets (such information provided by the CDF file). *Click for Affymetrix description.*

DOWNLOAD AFFY AND GCRMA IN R

To run these analyses you will need to download the FREE *affy* and *gcrma* package in R for the Affymetrix oligonucleotide array probe level data analysis, developed as part of the Bioconductor project.

The Bioconductor project website (<http://www.bioconductor.org/>) has links to various documents related to R/*affy* and R/*gcrma* and the statistical analyses.

R/*affy*. *(Click for explanations on how to download and install)*

R/*gcrma*. *(Click for explanations on how to download and install)*

RUNNING R/affy TO NORMALIZE THE DATA MAINTAINING PROBE INFORMATION

Note: the following descriptions and demo have been developed based on R version R 2.1.1.

Click here for the R/affy functional codes. Once you are familiar with R/affy this set of codes (called “.Rhstory”) is all you’ll need to run the normalization.

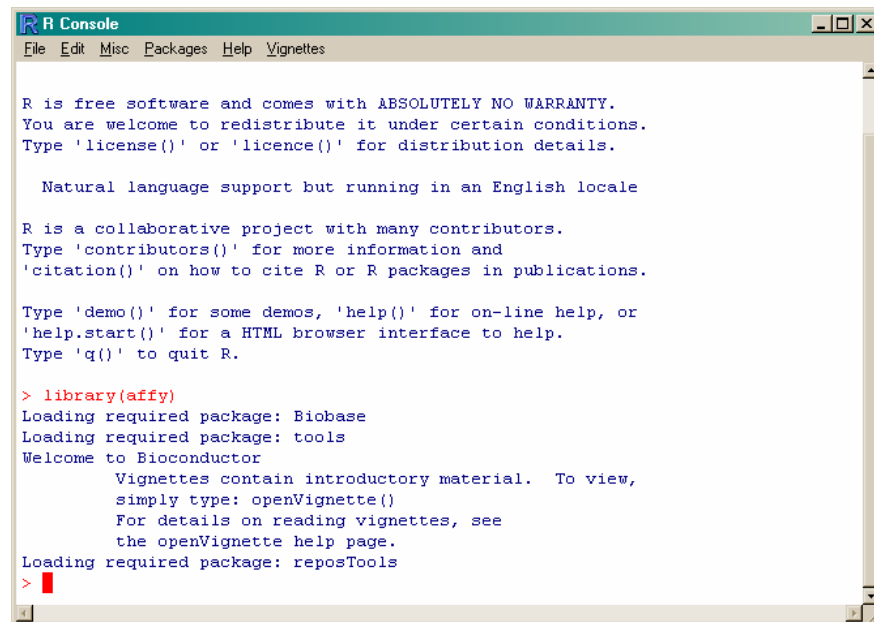
1. PUT FILES INTO SINGLE FOLDER/DIRECTORY.

To run R, you need to have all the .CEL files in the same folder/directory (i.e. C:\temp\Demo\CEL_Folder).

2. RUN AFFY PACKAGE IN R.

- The first step is to load the affy library by opening R and simply typing:

```
>library(affy)
```

A screenshot of the R Console window. The title bar says "R Console". The menu bar includes "File", "Edit", "Misc", "Packages", "Help", and "Vignettes". The console text shows the standard R welcome message, followed by the command "> library(affy)" in red. The output shows "Loading required package: Biobase", "Loading required package: tools", and "Welcome to Bioconductor". It then lists vignettes and provides instructions on how to view them. Finally, it says "Loading required package: reposTools" and ends with a red prompt character ">".

```
R Console
File Edit Misc Packages Help Vignettes

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

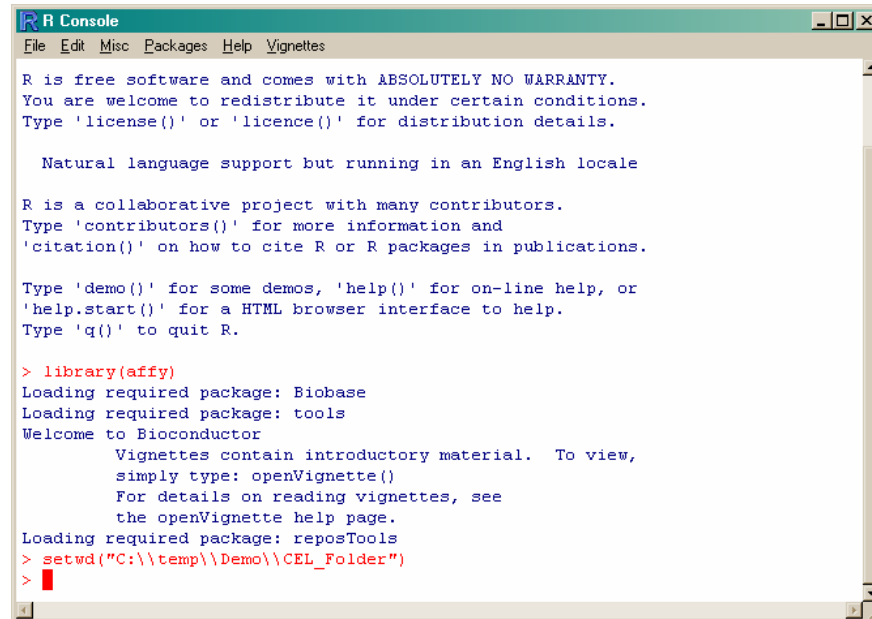
Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for a HTML browser interface to help.
Type 'q()' to quit R.

> library(affy)
Loading required package: Biobase
Loading required package: tools
Welcome to Bioconductor
  Vignettes contain introductory material. To view,
  simply type: openVignette()
  For details on reading vignettes, see
  the openVignette help page.
Loading required package: reposTools
>
```

Loads the entire required packages to run the affy package.

- Set (identify) the working folder/directory where the data are located using double backslashes (i.e. C:\\temp\\Demo\\CEL_Folder)

```
>setwd("C:\\temp\\Demo\\CEL_Folder")
```



```
R Console
File Edit Misc Packages Help Vignettes

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

Natural language support but running in an English locale

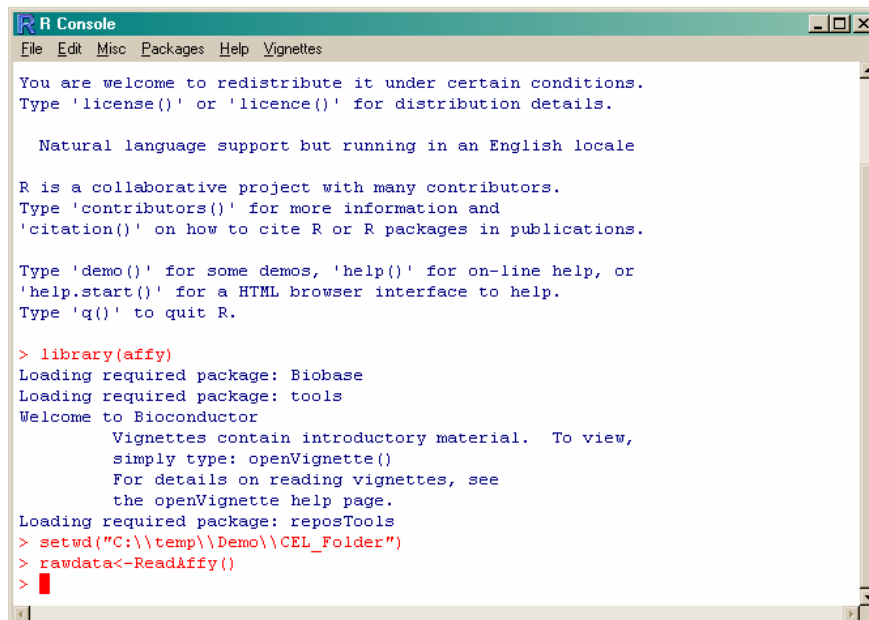
R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for a HTML browser interface to help.
Type 'q()' to quit R.

> library(affy)
Loading required package: Biobase
Loading required package: tools
Welcome to Bioconductor
  Vignettes contain introductory material. To view,
  simply type: openVignette()
  For details on reading vignettes, see
  the openVignette help page.
Loading required package: reposTools
> setwd("C:\\temp\\Demo\\CEL_Folder")
>
```

- Read the raw data into the file “rawdata”.

```
>rawdata<-ReadAffy()
```



```
R Console
File Edit Misc Packages Help Vignettes

You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

Natural language support but running in an English locale

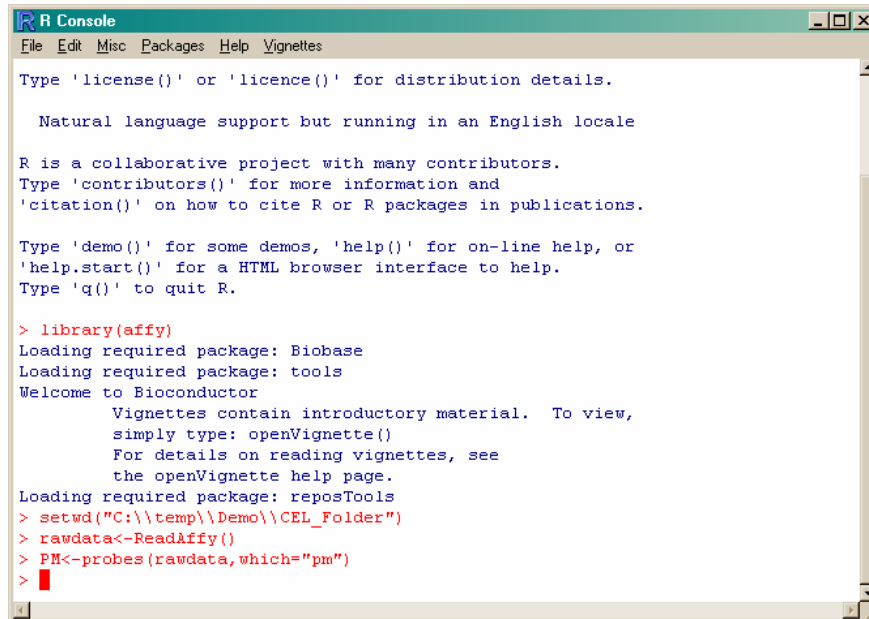
R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for a HTML browser interface to help.
Type 'q()' to quit R.

> library(affy)
Loading required package: Biobase
Loading required package: tools
Welcome to Bioconductor
  Vignettes contain introductory material. To view,
  simply type: openVignette()
  For details on reading vignettes, see
  the openVignette help page.
Loading required package: reposTools
> setwd("C:\\temp\\Demo\\CEL_Folder")
> rawdata<-ReadAffy()
>
```

- Extract perfect match probe intensities from the “rawdata” file into a new file called “PM”.

```
>PM<-probes(rawdata,which="pm")
```



```
R Console
File Edit Misc Packages Help Vignettes

Type 'license()' or 'licence()' for distribution details.

Natural language support but running in an English locale

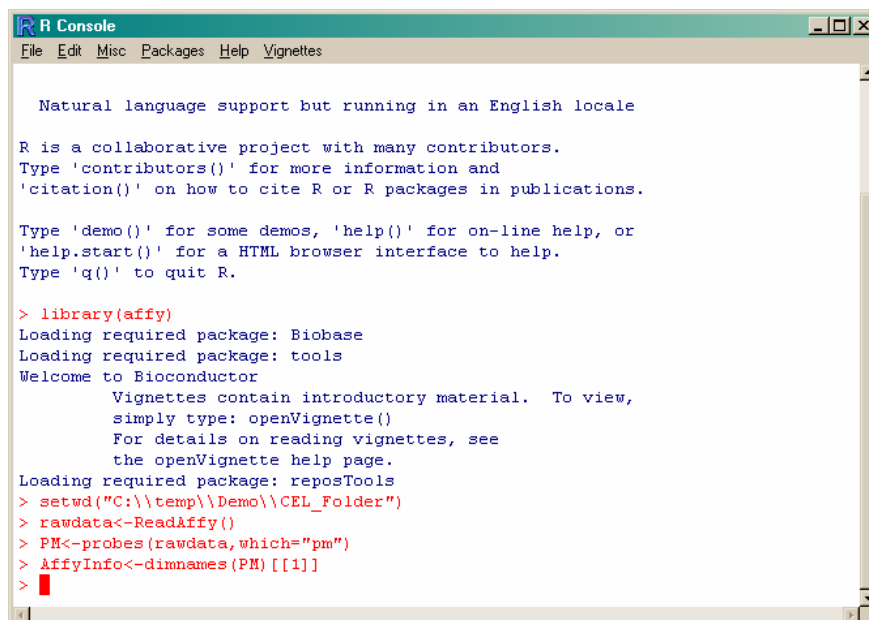
R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for a HTML browser interface to help.
Type 'q()' to quit R.

> library(affy)
Loading required package: Biobase
Loading required package: tools
Welcome to Bioconductor
  Vignettes contain introductory material. To view,
  simply type: openVignette()
  For details on reading vignettes, see
  the openVignette help page.
Loading required package: reposTools
> setwd("C:\\temp\\Demo\\CEL_Folder")
> rawdata<-ReadAffy()
> PM<-probes(rawdata,which="pm")
>
```

- Retrieve the gene IDs from the first column of the “PM” file and save in “AffyInfo” file.

```
>AffyInfo<-dimnames(PM)[[1]]
```



```
R Console
File Edit Misc Packages Help Vignettes

Natural language support but running in an English locale

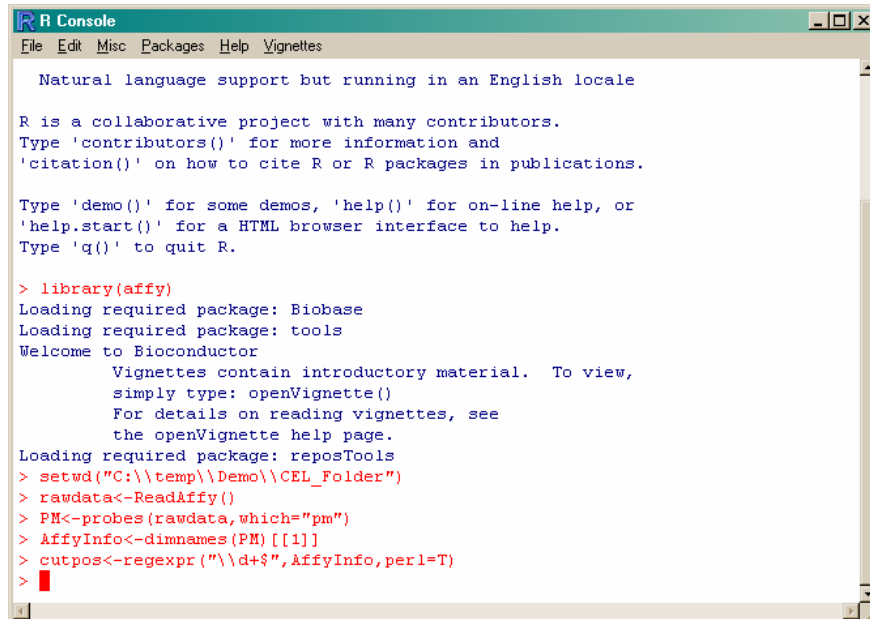
R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for a HTML browser interface to help.
Type 'q()' to quit R.

> library(affy)
Loading required package: Biobase
Loading required package: tools
Welcome to Bioconductor
  Vignettes contain introductory material. To view,
  simply type: openVignette()
  For details on reading vignettes, see
  the openVignette help page.
Loading required package: reposTools
> setwd("C:\\temp\\Demo\\CEL_Folder")
> rawdata<-ReadAffy()
> PM<-probes(rawdata,which="pm")
> AffyInfo<-dimnames(PM)[[1]]
>
```

- Look for number of digits following each probe name within “AffyInfo” data set and save into “cutpos” file. Gives a -1 if no digits follow the name.

```
>cutpos<-regexpr("\\d+$",AffyInfo,perl=T)
```



```
R Console
File Edit Misc Packages Help Vignettes

Natural language support but running in an English locale

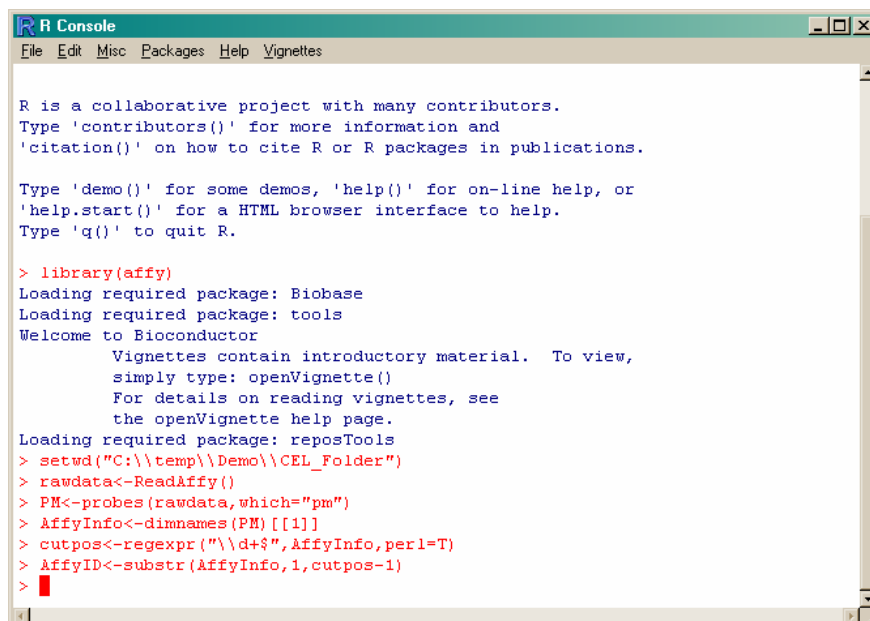
R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for a HTML browser interface to help.
Type 'q()' to quit R.

> library(affy)
Loading required package: Biobase
Loading required package: tools
Welcome to Bioconductor
  Vignettes contain introductory material. To view,
  simply type: openVignette()
  For details on reading vignettes, see
  the openVignette help page.
Loading required package: reposTools
> setwd("C:\\temp\\Demo\\CEL_Folder")
> rawdata<-ReadAffy()
> PM<-probes(rawdata,which="pm")
> AffyInfo<-dimnames(PM)[[1]]
> cutpos<-regexpr("\\d+$",AffyInfo,perl=T)
>
```

- Extract the digits following the probe names in “AffyInfo” data set and save these IDs as “AffyID” file.

```
>AffyID<-substr(AffyInfo,1,cutpos-1)
```



```
R Console
File Edit Misc Packages Help Vignettes

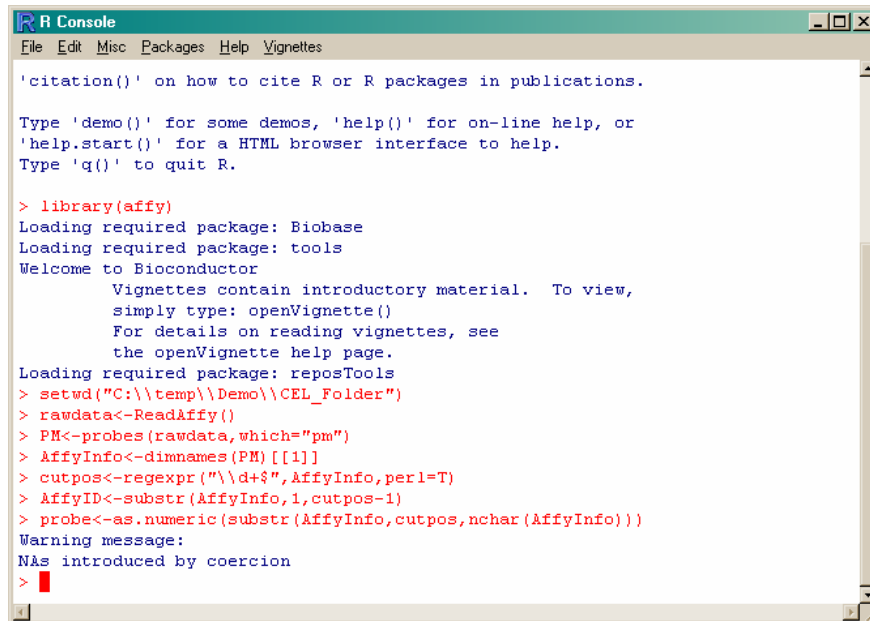
R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for a HTML browser interface to help.
Type 'q()' to quit R.

> library(affy)
Loading required package: Biobase
Loading required package: tools
Welcome to Bioconductor
  Vignettes contain introductory material. To view,
  simply type: openVignette()
  For details on reading vignettes, see
  the openVignette help page.
Loading required package: reposTools
> setwd("C:\\temp\\Demo\\CEL_Folder")
> rawdata<-ReadAffy()
> PM<-probes(rawdata,which="pm")
> AffyInfo<-dimnames(PM)[[1]]
> cutpos<-regexpr("\\d+$",AffyInfo,perl=T)
> AffyID<-substr(AffyInfo,1,cutpos-1)
>
```

- Take numeric objects from the “AffyInfo” data set, which are the probe IDs (most are 1-11) and save as “probe” file.

```
>probe<-as.numeric(substr(AffyInfo,cutpos,nchar(AffyInfo)))
```



```
R Console
File Edit Misc Packages Help Vignettes

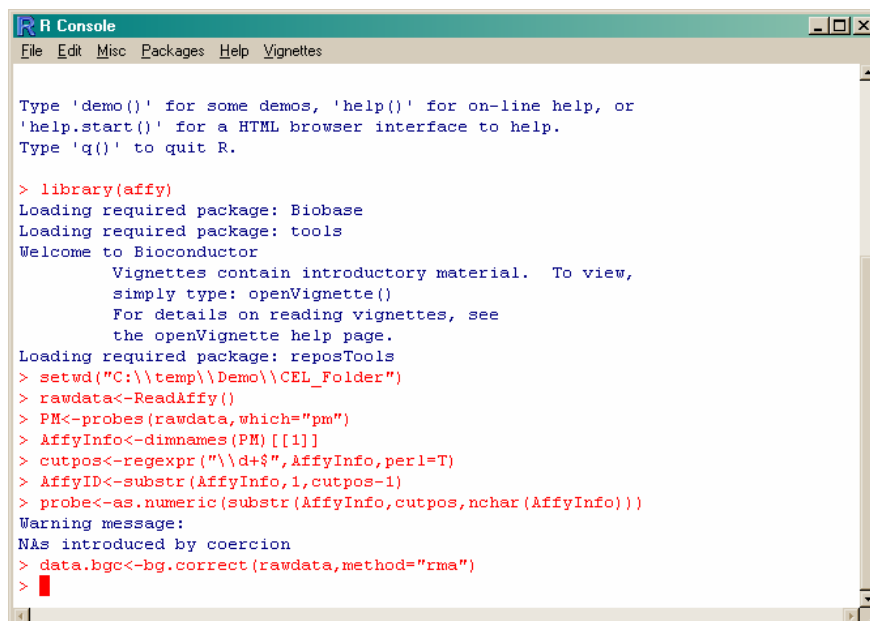
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for a HTML browser interface to help.
Type 'q()' to quit R.

> library(affy)
Loading required package: Biobase
Loading required package: tools
Welcome to Bioconductor
  Vignettes contain introductory material. To view,
  simply type: openVignette()
  For details on reading vignettes, see
  the openVignette help page.
Loading required package: reposTools
> setwd("C:\\temp\\Demo\\CEL_Folder")
> rawdata<-ReadAffy()
> PM<-probes(rawdata,which="pm")
> AffyInfo<-dimnames(PM)[[1]]
> cutpos<-regexpr("\\d+$",AffyInfo,perl=T)
> AffyID<-substr(AffyInfo,1,cutpos-1)
> probe<-as.numeric(substr(AffyInfo,cutpos,nchar(AffyInfo)))
Warning message:
NAs introduced by coercion
> █
```

- Raw data background corrects probe intensity values with the RMA method.

```
>data.bg<-bg.correct(rawdata,method="rma")
```



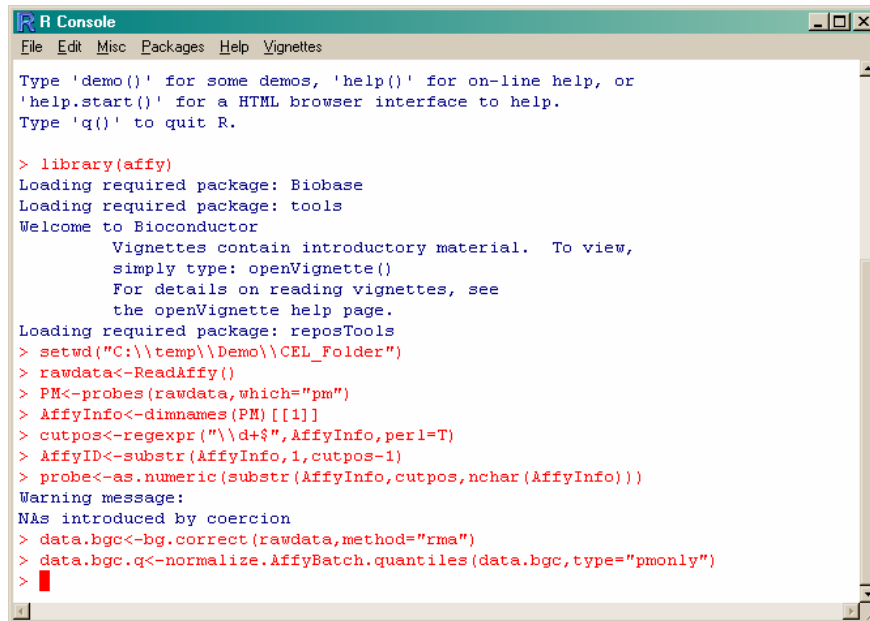
```
R Console
File Edit Misc Packages Help Vignettes

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for a HTML browser interface to help.
Type 'q()' to quit R.

> library(affy)
Loading required package: Biobase
Loading required package: tools
Welcome to Bioconductor
  Vignettes contain introductory material. To view,
  simply type: openVignette()
  For details on reading vignettes, see
  the openVignette help page.
Loading required package: reposTools
> setwd("C:\\temp\\Demo\\CEL_Folder")
> rawdata<-ReadAffy()
> PM<-probes(rawdata,which="pm")
> AffyInfo<-dimnames(PM)[[1]]
> cutpos<-regexpr("\\d+$",AffyInfo,perl=T)
> AffyID<-substr(AffyInfo,1,cutpos-1)
> probe<-as.numeric(substr(AffyInfo,cutpos,nchar(AffyInfo)))
Warning message:
NAs introduced by coercion
> data.bg<-bg.correct(rawdata,method="rma")
> █
```

- Normalize the perfect match probe level intensities based upon quantiles method.

```
>data.bgc.q<-normalize.AffyBatch.quantiles(data.bgc,type="pmonly")
```



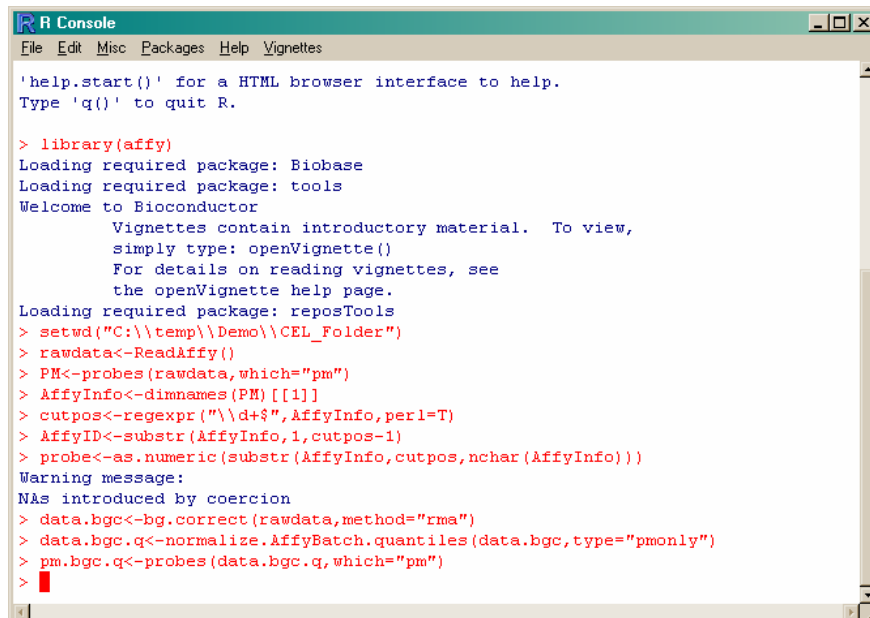
```
R Console
File Edit Misc Packages Help Vignettes

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for a HTML browser interface to help.
Type 'q()' to quit R.

> library(affy)
Loading required package: Biobase
Loading required package: tools
Welcome to Bioconductor
  Vignettes contain introductory material. To view,
  simply type: openVignette()
  For details on reading vignettes, see
  the openVignette help page.
Loading required package: reposTools
> setwd("C:\\temp\\Demo\\CEL_Folder")
> rawdata<-ReadAffy()
> PM<-probes(rawdata,which="pm")
> AffyInfo<-dimnames(PM)[[1]]
> cutpos<-regexpr("\\d+$",AffyInfo,perl=T)
> AffyID<-substr(AffyInfo,1,cutpos-1)
> probe<-as.numeric(substr(AffyInfo,cutpos,nchar(AffyInfo)))
Warning message:
NAs introduced by coercion
> data.bgc<-bg.correct(rawdata,method="rma")
> data.bgc.q<-normalize.AffyBatch.quantiles(data.bgc,type="pmonly")
> █
```

- Extract perfect match probe intensities from the “data.bgc” file into a new file called “pm.bgc.q”.

```
>pm.bgc.q<-probes(data.bgc.q,which="pm")
```



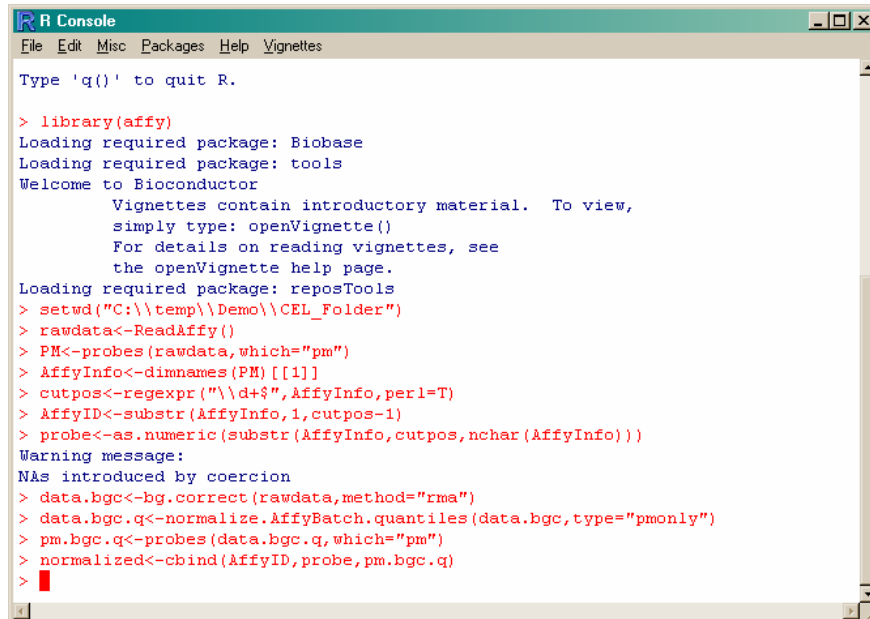
```
R Console
File Edit Misc Packages Help Vignettes

'help.start()' for a HTML browser interface to help.
Type 'q()' to quit R.

> library(affy)
Loading required package: Biobase
Loading required package: tools
Welcome to Bioconductor
  Vignettes contain introductory material. To view,
  simply type: openVignette()
  For details on reading vignettes, see
  the openVignette help page.
Loading required package: reposTools
> setwd("C:\\temp\\Demo\\CEL_Folder")
> rawdata<-ReadAffy()
> PM<-probes(rawdata,which="pm")
> AffyInfo<-dimnames(PM)[[1]]
> cutpos<-regexpr("\\d+$",AffyInfo,perl=T)
> AffyID<-substr(AffyInfo,1,cutpos-1)
> probe<-as.numeric(substr(AffyInfo,cutpos,nchar(AffyInfo)))
Warning message:
NAs introduced by coercion
> data.bgc<-bg.correct(rawdata,method="rma")
> data.bgc.q<-normalize.AffyBatch.quantiles(data.bgc,type="pmonly")
> pm.bgc.q<-probes(data.bgc.q,which="pm")
> █
```

- Combine “AffyID”, “probe”, and “pm.bgc.q” file into a new file called “normalized”, which contains the normalized PM data.

```
>normalized<-cbind(AffyID,probe,pm.bgc.q)
```



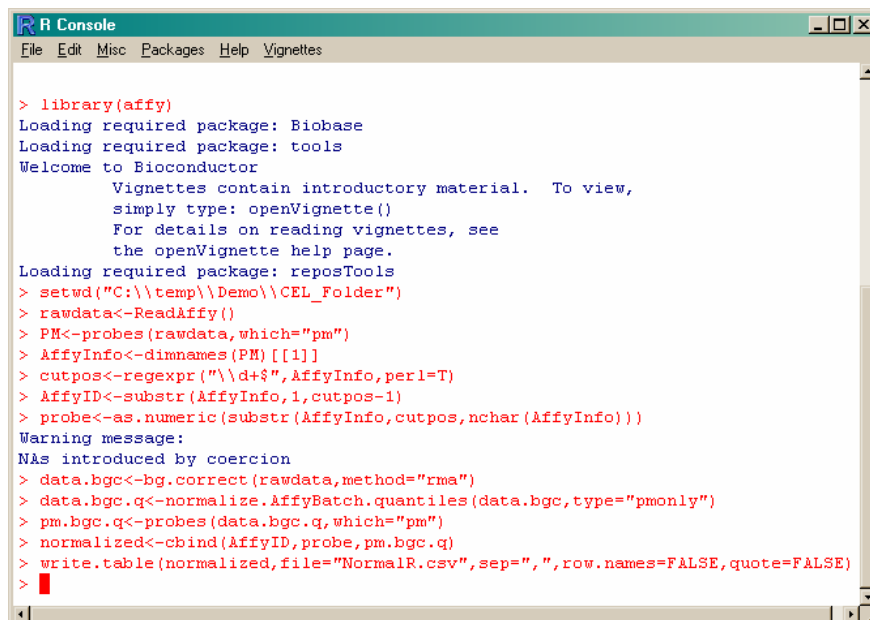
```
R Console
File Edit Misc Packages Help Vignettes

Type 'q()' to quit R.

> library(affy)
Loading required package: Biobase
Loading required package: tools
Welcome to Bioconductor
  Vignettes contain introductory material. To view,
  simply type: openVignette()
  For details on reading vignettes, see
  the openVignette help page.
Loading required package: reposTools
> setwd("C:\\temp\\Demo\\CEL_Folder")
> rawdata<-ReadAffy()
> PM<-probes(rawdata,which="pm")
> AffyInfo<-dimnames(PM)[[1]]
> cutpos<-regexpr("\\d+$",AffyInfo,perl=T)
> AffyID<-substr(AffyInfo,1,cutpos-1)
> probe<-as.numeric(substr(AffyInfo,cutpos,nchar(AffyInfo)))
Warning message:
NAs introduced by coercion
> data.bgc<-bg.correct(rawdata,method="rma")
> data.bgc.q<-normalize.AffyBatch.quantiles(data.bgc,type="pmonly")
> pm.bgc.q<-probes(data.bgc.q,which="pm")
> normalized<-cbind(AffyID,probe,pm.bgc.q)
>
```

- To have the expression measure in an Excel readable format, save the “normalized” file as a .csv file (i.e. NormalR.csv)

```
>write.table(normalized,file="NormalR.csv",sep=" ",row.names=FALSE,
quote=FALSE)
```



```
R Console
File Edit Misc Packages Help Vignettes

> library(affy)
Loading required package: Biobase
Loading required package: tools
Welcome to Bioconductor
  Vignettes contain introductory material. To view,
  simply type: openVignette()
  For details on reading vignettes, see
  the openVignette help page.
Loading required package: reposTools
> setwd("C:\\temp\\Demo\\CEL_Folder")
> rawdata<-ReadAffy()
> PM<-probes(rawdata,which="pm")
> AffyInfo<-dimnames(PM)[[1]]
> cutpos<-regexpr("\\d+$",AffyInfo,perl=T)
> AffyID<-substr(AffyInfo,1,cutpos-1)
> probe<-as.numeric(substr(AffyInfo,cutpos,nchar(AffyInfo)))
Warning message:
NAs introduced by coercion
> data.bgc<-bg.correct(rawdata,method="rma")
> data.bgc.q<-normalize.AffyBatch.quantiles(data.bgc,type="pmonly")
> pm.bgc.q<-probes(data.bgc.q,which="pm")
> normalized<-cbind(AffyID,probe,pm.bgc.q)
> write.table(normalized,file="NormalR.csv",sep=" ",row.names=FALSE,quote=FALSE)
>
```