

Notes for BioConductor: pre-process gene expression

Guiyuan Lei*

14 March, 2006

BioConductor is an open source and open development software project for the analysis and comprehension of genomic data.

1 Install of BioConductor

The website of <http://www.bioconductor.org/docs/install-howto.html> describe how to install BioConductor. I choose to install it by biocLite.R

For me, in order to connect internet in linux terminal. I have to set http proxy by add

```
setenv http_proxy "http://wwwcache.ncl.ac.uk:8080"
```

into file .cshrc under my home directory.

Then follow the following steps:

1. Run R (the version should be higher than 2.1.0). I type `"/usr/bin/R"` because the R in my local linux is verions 2.0.1, the adiminstor did something so that I can use version 2.1.0 by typing this command. But something wrong with this.
2. `source("http://www.bioconductor.org/biocLite.R")`
3. `biocLite(lib="/data/BioConductor",destdir="/dataBioConductor/downloaded")`
argument "lib" is the directory where the BioConductor will be installed, the argument "destdir" is for the directory where the install files/packages be downloaded, you can delete files under this directory

*E-mail: Guiyuan.Lei@ncl.ac.uk

after the installation is finished.

Error during installation:

The version of R should be higher than 2.1.0. I type “/usr/bin/R” because the R in my local linux is verions 2.0.1, the adiminstor did something so that I can use version 2.1.0 by typing this command. But something wrong with this. After the installation, if I type

library(affy), R reported “Error in library(affy) : there is no package called ‘affy’”. If I run R by typing

R

instead of

/usr/bin/R

When I use “library(affy)”, I got a warning:

“warning: incorrect permissions to edit package database

/usr/local/R.2.0.1/lib/library/liblisting.Rda”

The argument “lib” in command biocLite() is the directory where the BioConductor will be installed, the argument “destdir” is for the directory where the install files/packages be downloaded, you can delete files under this directory after the installation is finished.

Error during installation:

1. affyPLM not successfully installed
2. /usr/bin/ld: cannot find -lblas-3

2 Use BioConductor to deal with affymetrix microarray data

1. Put all .CEL files in one directory, like /Huang
2. Run R under the directory /Huang
3. library(affy)
4. Read in data by typing “Huang< – ReadAffy()”
5. Do Normalization by typing “eset< –justRMA()”

6. Save data, `write.exprs(eset,file="mydata.txt")`
7. Write to excel format, `exprs2excel(eset,file="mydata.CSV")`

3 Resources on BioConductor

- Downloading All Packages From A Repository
To download all of the packages from a repository into a directory, use the `download.packages2` function. To do this, select the repository you wish to download the packages from (using repositories for instance) and run `download.packages2(repEntry=REP, destDir=DIR)`, where REP is your `ReposEntry` object and DIR is the directory you'd like them downloaded to (e.g. `"/"`).
`/data/cdfenvs/repos`: A repository for the Bioconductor CDF data packages.
- `affyDemo`: Introduction to Bioconductor `affy` Package
<http://www-edlab.cs.umass.edu/cs691k/conlon/readings/affyDemo.pdf>
- vignettes
<http://www.bioconductor.org/docs/vignettes.html>

4 To do

1. Re-install BioConductor after
 - (a) Upgrade R to version 2.1.0
 - (b) Install blas
2. How to tell `ReadAffy()` which CDF file to be used for loaded `.CEL` files?