

Bioconductor

Jean-Yves Sgro

May 9, 2017

Contents

1	Introduction	1
1.1	R and S	1
1.2	Bioconductor	1
2	Installing Bioconductor	2
2.1	Core installation	2
2.2	Specific packages	3
	References	3

1 Introduction

1.1 R and S

R is a general software and language system for data analysis and graphics (Ihaka and Gentleman 1996) and considered a dialect of the S language created by Bell Laboratories (*S-PLUS Programmer's Manual* 1992), now commercialized as TIBCO *Spotfire S+* by TIBCO.

A history of S was published online, last available in 2015 (*A Brief History of S* 2015) (File format `ps`¹ or `pdf`² with more in a list of other S development.³)

R can be downloaded and further extended by installing packages from the “Comprehensive R Archive Network” (CRAN) <https://www.r-project.org/>.

1.2 Bioconductor

Bioconductor is a collection of specialized software packages that run with the R engine.

Packages can run independently within R or call on one or more additional packages.

The cohort of the many hundred of packages dedicated to the analysis of microarrays collectively form the “Bioconductor” software.

The first release of Bioconductor was in May 2001 (R. C. Gentleman et al. 2004)

Wikipedia provides the following definition of what is Bioconductor :

Bioconductor is a free, open source and open development software project for the analysis and comprehension of genomic data generated by wet lab experiments in molecular biology.

The project was started in the Fall of 2001 and is overseen by the Bioconductor core team, based primarily at the Fred Hutchinson Cancer Research Center, with other members coming from various US and international institutions.

¹<http://ect.bell-labs.com/sl/doc/94.11.ps>

²http://www.lcg.unam.mx/~lcollado/R/resources/history_of_S.pdf

³<http://ect.bell-labs.com/sl/doc/comp.html>

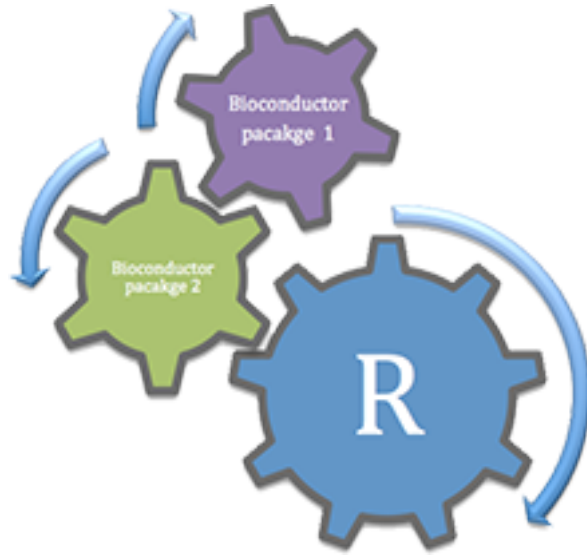


Figure 1:

1.2.1 Packages

Most Bioconductor components are distributed as **R packages**, which are add-on modules for **R**. Package installation is independent of the **CRAN** network website and therefore the **RStudio** “Install” button cannot be used.

Bioconductor packages can be installed within **R** from line-command which obtains the package installer from the Bioconductor web site <http://www.bioconductor.org/> and also installs the package dependencies.

Bioconductor is based primarily on the statistical **R** programming language, but does contain contributions in other programming languages.

While most packages are available from the Bioconductor official web site, some authors prefer to distribute their contributions from their own web site or by request.

2 Installing Bioconductor

For today’s class we need to install Bioconductor, unless you have installed it on this iMac previously.

You can use the **R** application or the **RStudio R console** for this.

The detailed installation instructions are listed on the official web site at: <http://www.bioconductor.org/install/>

You can follow the following summary.

2.1 Core installation

Installation is performed via an **R** script that is retrieved from the web site and executed:

```
source("https://bioconductor.org/biocLite.R")
biocLite()
```

2.2 Specific packages

For today's class we'll need specific packages that we can install now. We don't need to request the `biocLite.R` again since we just downloaded it in the previous command set.

```
# source("https://bioconductor.org/biocLite.R")
biocLite("SpikeInSubset")
biocLite("genefilter")
biocLite("hgu95av2.db")
biocLite("annaffy")
```

This packages will be installed, as well as dependencies for these packages.

Note: the `c()` function can be used to specify packages within one `biocLite()` command. For example:

```
biocLite(c("SpikeInSubset", "genefilter", "hgu95av2.db", "annaffy" )
```

References

A Brief History of S. 2015. Online. Murray Hill, New Jersey: AT&T Bell Laboratories.

Gentleman, R. C., V. J. Carey, D. M. Bates, B. Bolstad, M. Dettling, S. Dudoit, B. Ellis, et al. 2004. "Bioconductor: open software development for computational biology and bioinformatics." *Genome Biol.* 5 (10): R80.

Ihaka, Ross, and Robert Gentleman. 1996. "R: A Language for Data Analysis and Graphics." *Journal of Computational and Graphical Statistics* 5 (3): 299–314. doi:10.1080/10618600.1996.10474713.

S-PLUS Programmer's Manual. 1992. Version 3.1. Seattle, WA, USA: StatSci, a Division of MathSoft, Inc.