

Open Source Open Science (OSOS) workshop

Next Generation Sequencing - Data processing and gene expression

Luciano V. Cosme
Department of Entomology - Texas A&M University

Software packages you will need

1. Read quality check, trimming, adapter removal and filtering

- FastQC <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>
- NGS QC Toolkit v2.3.3 - <http://59.163.192.90:8080/ngsqctoolkit/>

2. To align reads back to reference genome

- STAR <https://code.google.com/p/rna-star/>
- Optional TopHat <http://ccb.jhu.edu/software/tophat/index.shtml>

3. To count how many reads mapped to each gene

- HTSeq <http://www-huber.embl.de/users/anders/HTSeq/doc/overview.html>
- Optional Cufflinks <http://cufflinks.cbcb.umd.edu/>

Software packages you will need

4. SAM/BAM files manipulation

- SAM tools <http://sourceforge.net/projects/samtools/files/samtools/0.1.18/>
- Bedtools <https://github.com/arq5x/bedtools2>

5. A genome viewer

- IGV <http://www.broadinstitute.org/igv/home>

R packages to install

First install R. Then install Bioconductor

You can find more about Bioconductor here <http://www.bioconductor.org/install/>

How to install Bioconductor:

Open R via command prompt (type "R") and use the command below:

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

R packages to install

Start R via command prompt (by typing "R") and paste the command below:

1. To calculate the size of each gene - GenomeFeatures

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

2. Statistical Analysis - DESeq2

```
source("http://bioconductor.org/biocLite.R")  
biocLite("DESeq2")
```

R packages to install

We will need some packages that are not in Bioconductor. You can install these packages via command prompt in R or RStudio using the command below:

```
install.packages("name of the package")
```

What to install:

"RColorBrewer", "plyr", "lattice", "latticeExtra", "scatterplot3d"

Other packages

Linux, MacOS and Windows

It will be easier to edit and comment your R scripts if you use RStudio

<http://www.rstudio.com/products/RStudio/#Desk>

There are many text/code editors, one of my favorites is Sublime Text

<http://www.sublimetext.com/>

Linux

If you are using Linux (for example Ubuntu) a must have is Kile, especially if you work with Latex files

<http://kile.sourceforge.net/>