



WROCLAW UNIVERSITY
OF ENVIRONMENTAL
AND LIFE SCIENCES



The Faculty of Biology
and Animal Science



THETA
Statistical Genetics Group
Institute of Animal Genetics

RNA-seq sequence analysis with R / Bioconductor

Barbara Kosińska-Selbi*, Joanna Szyda
*barbara.kosinska@upwr.edu.pl



What is BIOCONDUCTOR ?



<https://www.bioconductor.org/>

Bioconductor
OPEN SOURCE SOFTWARE FOR BIOINFORMATICS

Version 3.7 (two releases each year),

Uses R programming language,

1560 software packages,

Active user community,

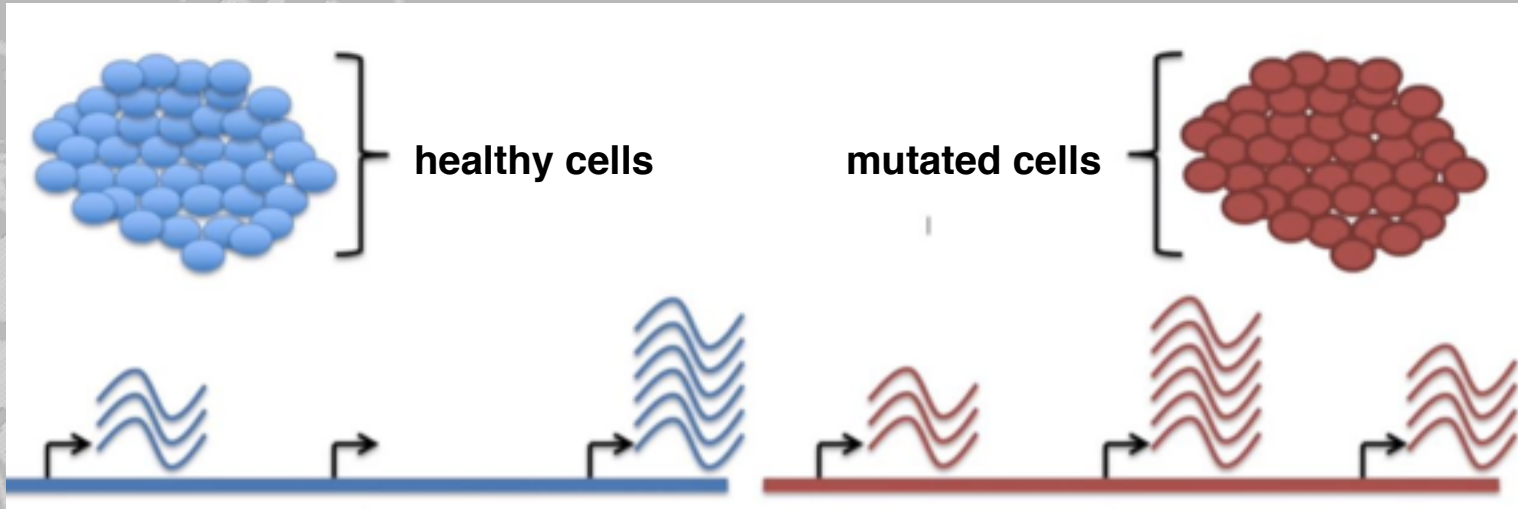
High quality documentation,

Install.packages = *biocLite()*.

AIM:

The availability of Bioconductor software allows for conducting an RNA-seq analysis for persons who are no professional programmers, but are biologists who can then concentrate on the proper interpretation of results.

What is RNA-seq data ?

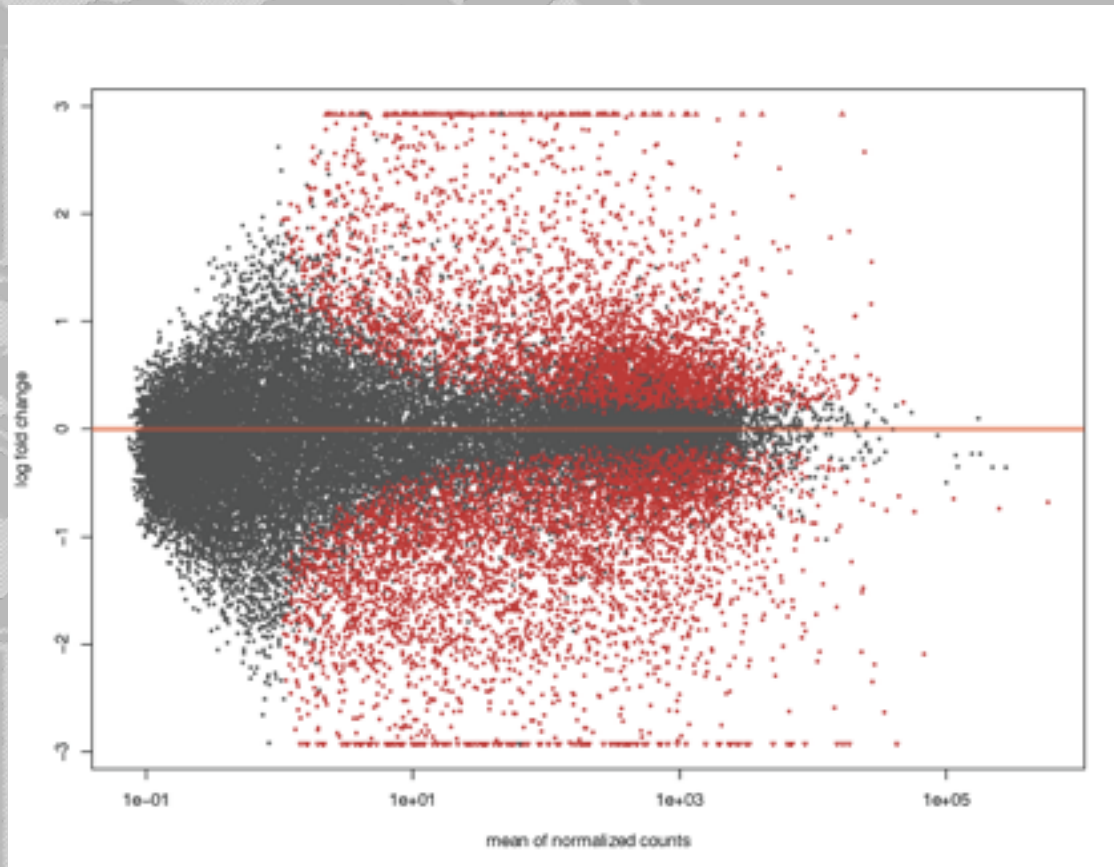


Source : <https://statquest.org/2017/09/01/statquest-a-gentle-introduction-to-rna-seq/>

Identifier	● @SRR566546.970 HWUSI-EAS1673_11067_FC7070M:4:1:2299:1109 length=50
Sequence	● TTGCCTGCCTATCATTTTAGTGCTGTGAGGTGGAGATGTGAGGATCAGT
'+' sign	● +
Quality scores	● hhhhhhhhhghghghhhhhfhhhhfffffe'ee['X]b[d[ed'[Y[~Y
Identifier	● @SRR566546.971 HWUSI-EAS1673_11067_FC7070M:4:1:2374:1108 length=50
Sequence	● GATTTGTATGAAAGTATACAACATAAACTGCAGGTGGATCAGAGTAAGTC
'+' sign	● +
Quality scores	● hhhhgfhhcghghggfcffdhfehhhhcchhdhahchhhdhaehffffde'bVd

Morteza Hosseini M., Pratas D. Armando J. P., 'A Survey on Data Compression Methods for Biological Sequences', 2016, MDPI

RNA-seq Differential Gene Expression in Designed Experiment



source : <http://genviz.org/module%204/0004/02/01/DifferentialExpression/>

Which genes are differentially expressed?

-10,000 of genes X dozens of samples

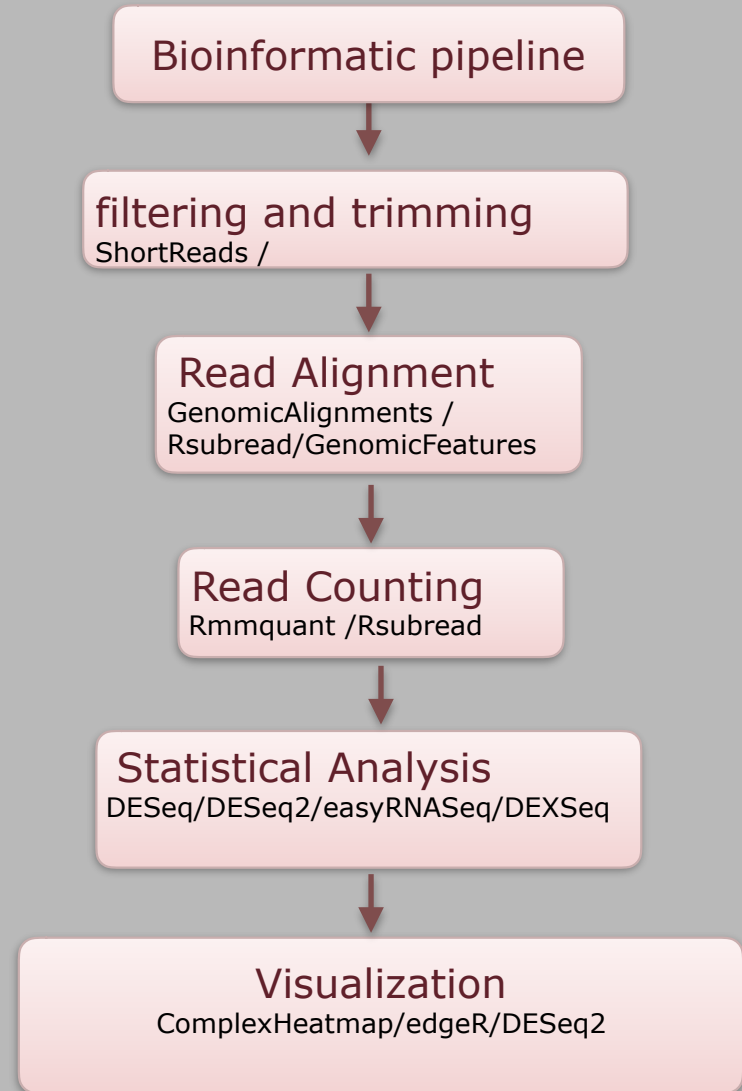
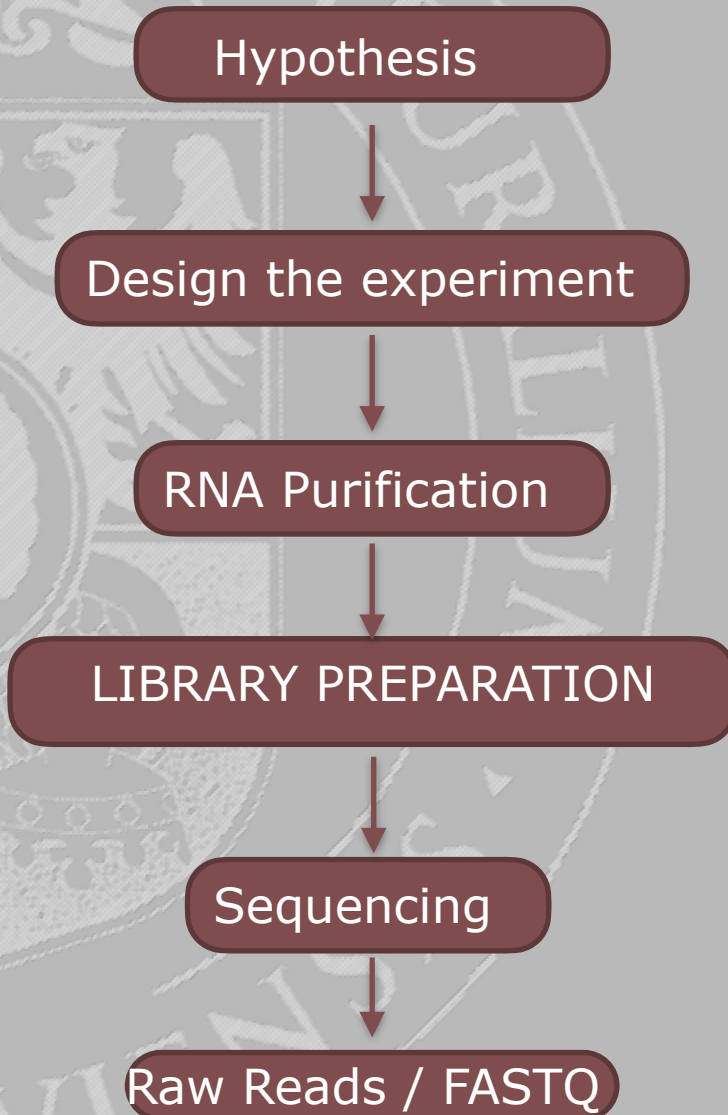
What are the challenges:

-Normalisation of the sample,
-Correct model of statistical error

Packages for measuring gene expression:

-DESeq2,
-edgeR

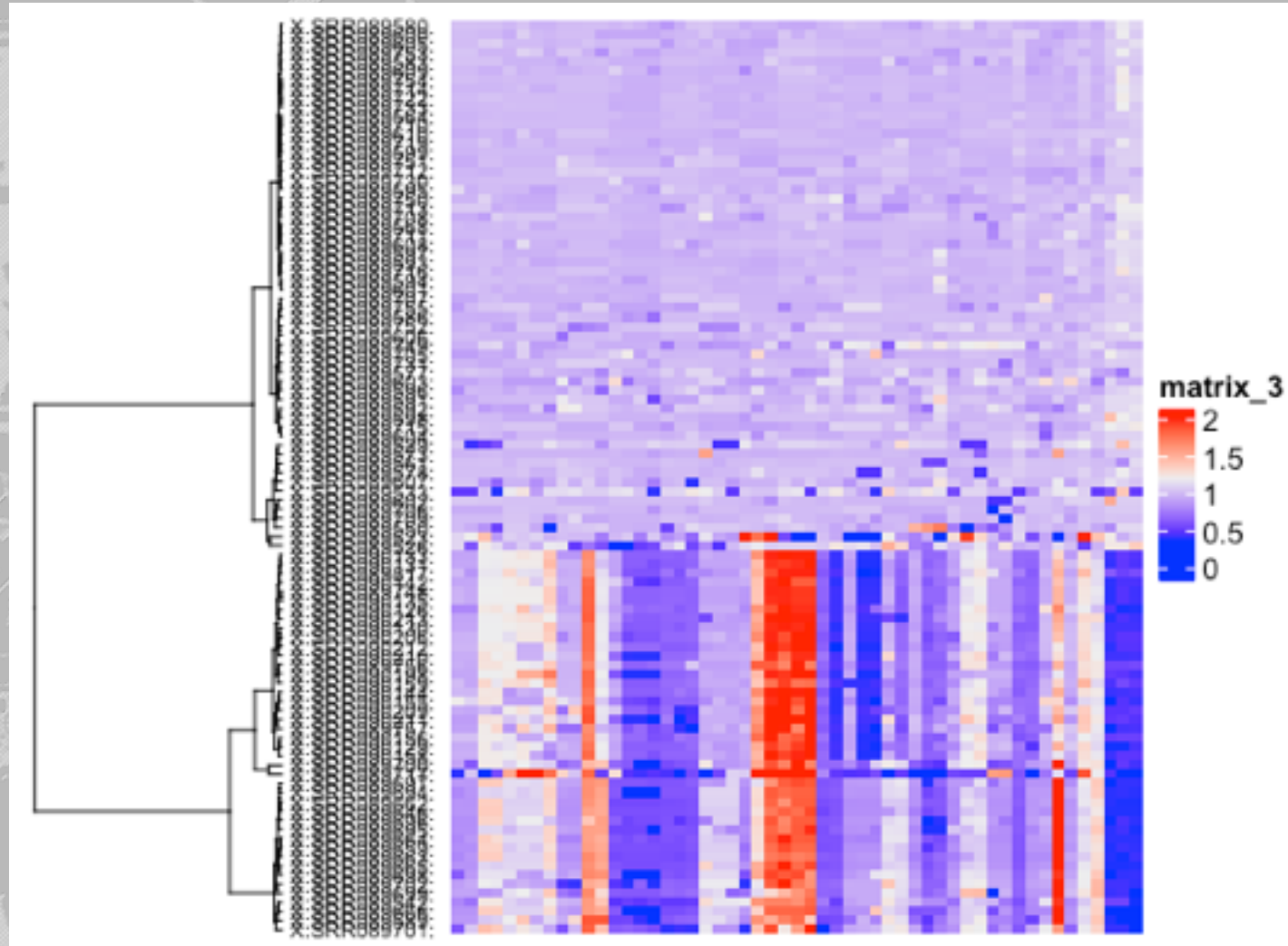
RNA-seq workflow with Bioconductor:



https://www.bioconductor.org/help/course-materials/2016/BiocIntro-May/B3_RNASeq_Workflow.html

GENE EXPRESSION LEVEL

VISUALISATION



Conclusions

There is no standard workflow for RNA-seq analysis. Each step can be done with a different package.

High quality documentation allows biologist to conducting an RNA-seq analysis, who are no professional programmers.

References:

Harmston N. et al, 'GenomicInteractions: An R/Bioconductor package for manipulating and investigating chromatin interaction data', 2015, BMC Genomics, 16:963

Lihua Julie Zhu et al., 'GUIDEseq: a bioconductor package to analyze GUIDE-Seq datasets for CRISPR-Cas nucleases', 2017, BMC Genomics, 18:379

Morteza Hosseini M., Pratas D. Armando J. P., 'A Survey on Data Compression Methods for Biological Sequences', 2016, MDPI

Wolfgang H. et al. , 'Orchestrating high-throughput genomic analysis with Bioconductor', *Nat Methods*. 12(2)

Source : <https://statquest.org/2017/09/01/statquest-a-gentle-introduction-to-rna-seq/>

official webside : <https://www.bioconductor.org/>

Workflow : https://www.bioconductor.org/help/course-materials/2016/BiocIntro-May/B3_RNASeq_Workflow.html

Manuals :

bioconductor.org/packages/release/bioc/manuals/marray/man/marray.pdf

<https://bioconductor.org/packages/release/bioc/manuals/DESeq2/man/DESeq2.pdf>

<https://bioconductor.org/packages/release/bioc/manuals/easyRNASeq/man/easyRNASeq.pdf>

<https://bioconductor.org/packages/release/bioc/manuals/GenomicAlignments/man/GenomicAlignments.pdf>

<https://bioconductor.org/packages/release/bioc/html/biomaRt.html>

<https://bioconductor.org/packages/release/bioc/html/ComplexHeatmap.html>

<https://bioconductor.org/packages/release/bioc/html/GenomicFeatures.html>

<https://bioconductor.org/packages/release/bioc/html/genomeIntervals.html>