Microarray analysis with R: a review

Lluís Revilla Sancho, Facultat de Biociències, Universitat Autònoma de Barcelona

UAB
Universitat Autònoma
de Barcelona

Introduction

Microarrays are used in laboratories to analyse the presence of RNA of known elements using fluorescence. The scanned image is processed with a wide range of software and algorithms. Several studies propose standards for microarray experiments , how to share the results and how to design them.
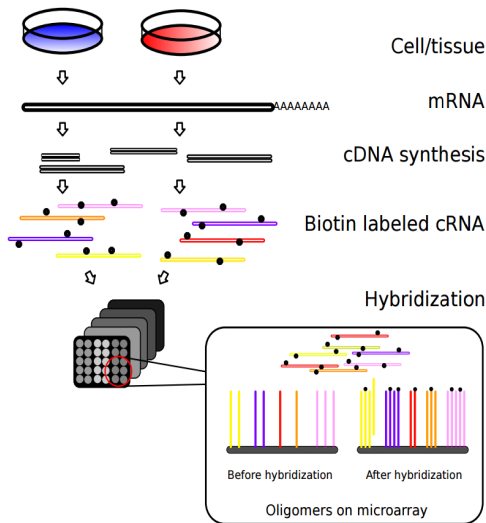

Figure 1: Laboratory steps of microarra experiments. Adapted from Hänzelmann, S. (2012) *Pathway-centric approaches to the analysis of high-throughput genomics data*, Universitat Pompeu Fabra ,Ph.D. Spain

Package	Background Correction	Microarray normalization	Microarray summarization	Differentially Expressed Genes	Gene Set Enrichment	Observations
limma	✓✓✓	✓✓	✓✓	✓✓		
genefilter					✓✓✓	Subseting
affy	✓✓✓	✓✓	✓✓			
affyio				✓✓✓		Reader for affy package
multtest						Useful statistical tests
GEOquery					✓✓	Querying
GO.db					✓✓	Queries Gene Ontology Consortium
impute						Impute missing data
gcrma	✓					For Affymetrix
Gviz						Plotts genes in chromosomes
vsn	✓	✓✓	✓			
affyPLM	✓					For Affymetrix
GOstats					✓✓✓	
siggenes				✓		
simpleaffy	✓✓	✓✓	✓	✓✓		Depends on affy
DNAcopy						Microarray design
marray	✓✓	✓✓		✓		
oligo	✓✓	✓✓	✓	✓		
minfi	✓✓		✓			Illumina methylation
annaffy						Iconnection between affymetrix and the web
topGO					✓✓✓	
methylumi	✓✓	✓	✓			Illumina methylation
sva	✓✓					Batch correction
beadarray	✓					BeadScan

Table 1: Packages sorted by unique IP downloads, classified by step where they supply functions

Objective:

Study the best packages of R project Bioconductor, which provides tools for the analysis and comprehension of high-throughput genomic data, to analyse microarrays. Compare the algorithms behind them and propose a workflow to improve the quality of the information.

Methods:

The top downloaded Bioconductor packages related to microarrays were investigated. The vignettes of these packages, where the functions and algorithms are described, were analysed and the number of citations they received was studied to analyse their usage.

Bioconductor
OPEN SOURCE SOFTWARE FOR BIOINFORMATICS
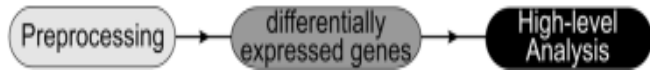
Results:

Packages and tools were listed by capabilities, citations and downloads and the main functions of the relevant packages described.

Discussion

Packages share similar schema but differ on the algorithms. They also differ on the scope of their functions; many of them are addressed to a single step, while few provide functions for several analysis steps.

Conclusion:

With just two packages a complete analysis of microarrays can be done. But the diversity of the experiments is reflected on the diversity of the packages. A unique workflow using these packages can't handle all the considerations needed for deep scientific analysis.
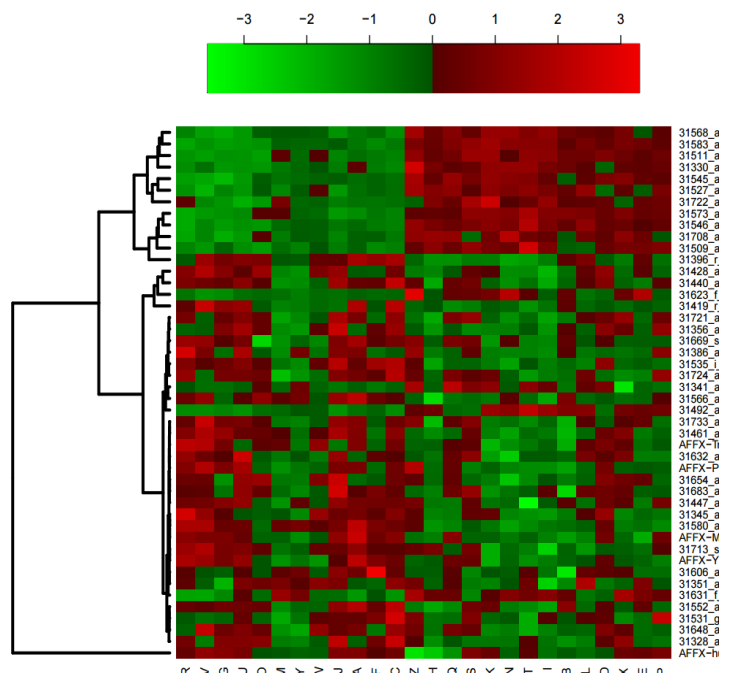
Figure 2 :Example of a MIAME compliant heatmap. Adapted from Heatplus vignette, consulted at http://www.bioconductor.org/packages/release/bioc/vignettes/Heatplus/inst/doc/annHeatmap.pdf the 8/05/2015.

References

•Brazma, A. (2009). "*Minimum information about a microarray experiment (MIAME)–successes, failures, challenges.*" The Scientific World Journal, 9, 420-423.

•Gentleman, Robert C., et al. (2004) *"Bioconductor: open software development for computational biology and bioinformatics."* Genome biology 5.10 : R80.

•Hänzelmann, S. (2012) "*Pathway-centric approaches to the analysis of high-throughput genomics data",* Universitat Pompeu Fabra ,Ph.D. Spain