

# Bioconductor for Gene Expression Analysis



**PRESENTED BY LUISA  
MERCADO**

# Presentation Roadmap



- What is Gene Expression Analysis?
- What is Bioconductor?
- The ALL dataset
- Example 1: Non-specific Filtering
- Example 2: Gene Selection
- Example 3: Multiple Testing Correction
- Summary

# What is Gene Expression Analysis?

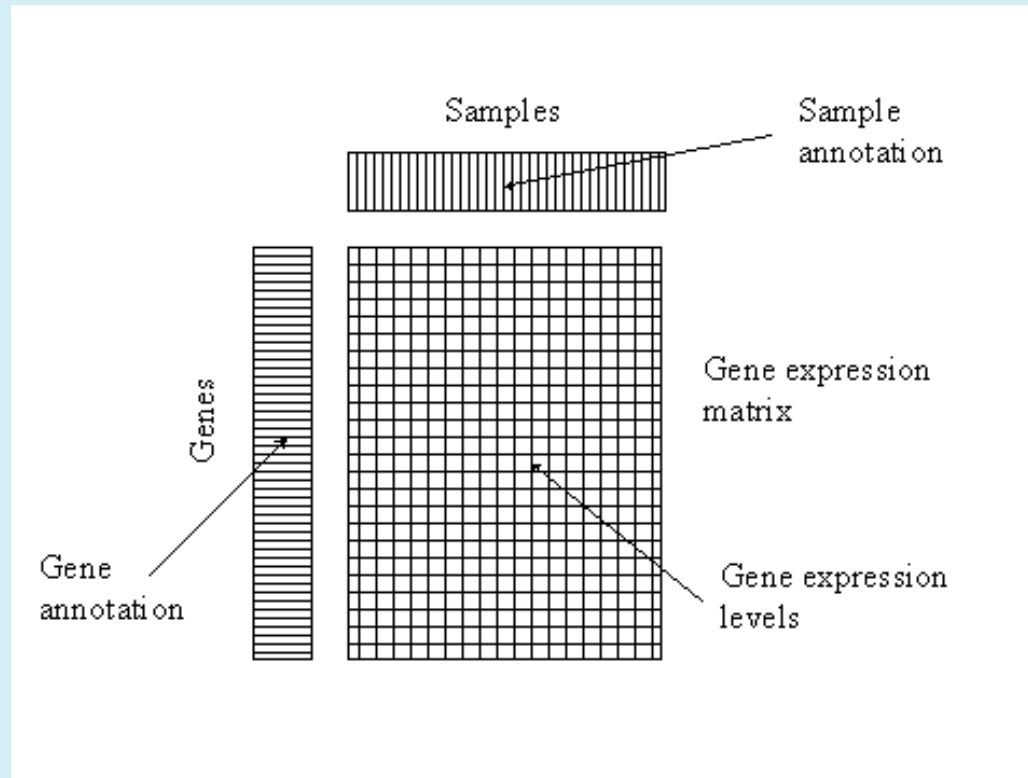


- Gene expression analysis consists on monitoring the expression levels of multiple genes simultaneously under a particular condition.
- Comparisons of the level of expression of the genes could be used to identify prognostic biomarkers, classify diseases or monitor the response to therapy.

# What is Gene Expression Analysis?



- Gene expression data can be represented as a matrix of expression levels



Source: [www.google.com](http://www.google.com)

# What is Gene Expression Analysis?



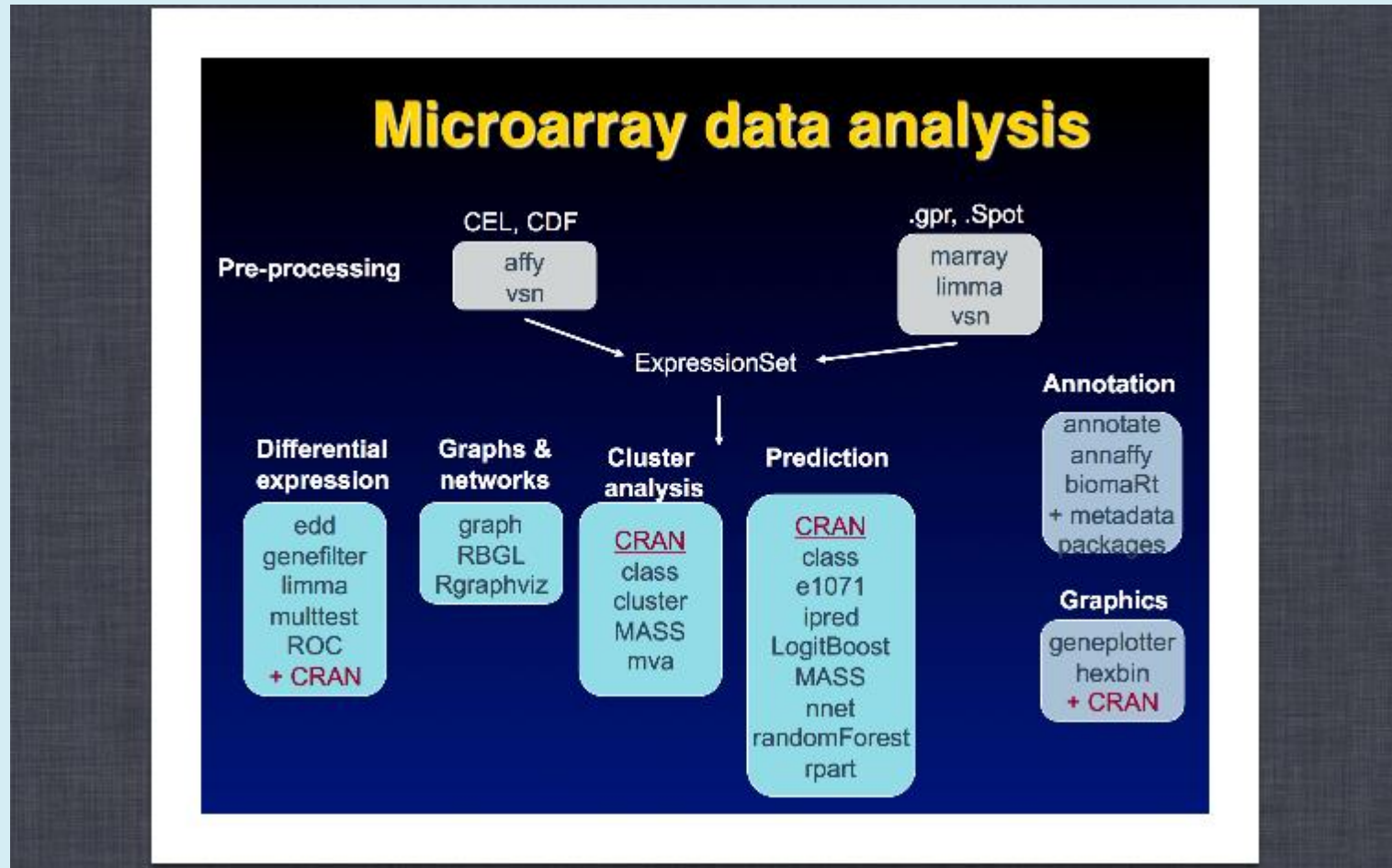
- Gene expression analysis can be summarized into four stages:
- Data Processing/ Quality Control
- Differential Expression
- Clustering and Data Visualization
- Classification and Prediction

# What is Bioconductor?



Bioconductor is an open source and open development software for the analysis of genomic data. It uses the R programming language to design and distribute integrated and interoperable software modules, called packages to provide comprehensive software solutions to relevant problems.

# What is Bioconductor?



Source: Bioconductor.org

# The ALL Dataset



- This dataset come from a study of acute lymphoblastic leukemia (ALL). It consists of microarrays from 128 different individuals with this type of disease. There are 95 samples with B-cell ALL and 33 with T-cell ALL, which refers to two different types of tumors among these samples. The B-cell ALL sample, contains information about individuals carrying the BCR/ABL mutation and individuals that do not display a cytogenetic abnormality. The total number of genes found in the B-cell ALL sample is 12,625.

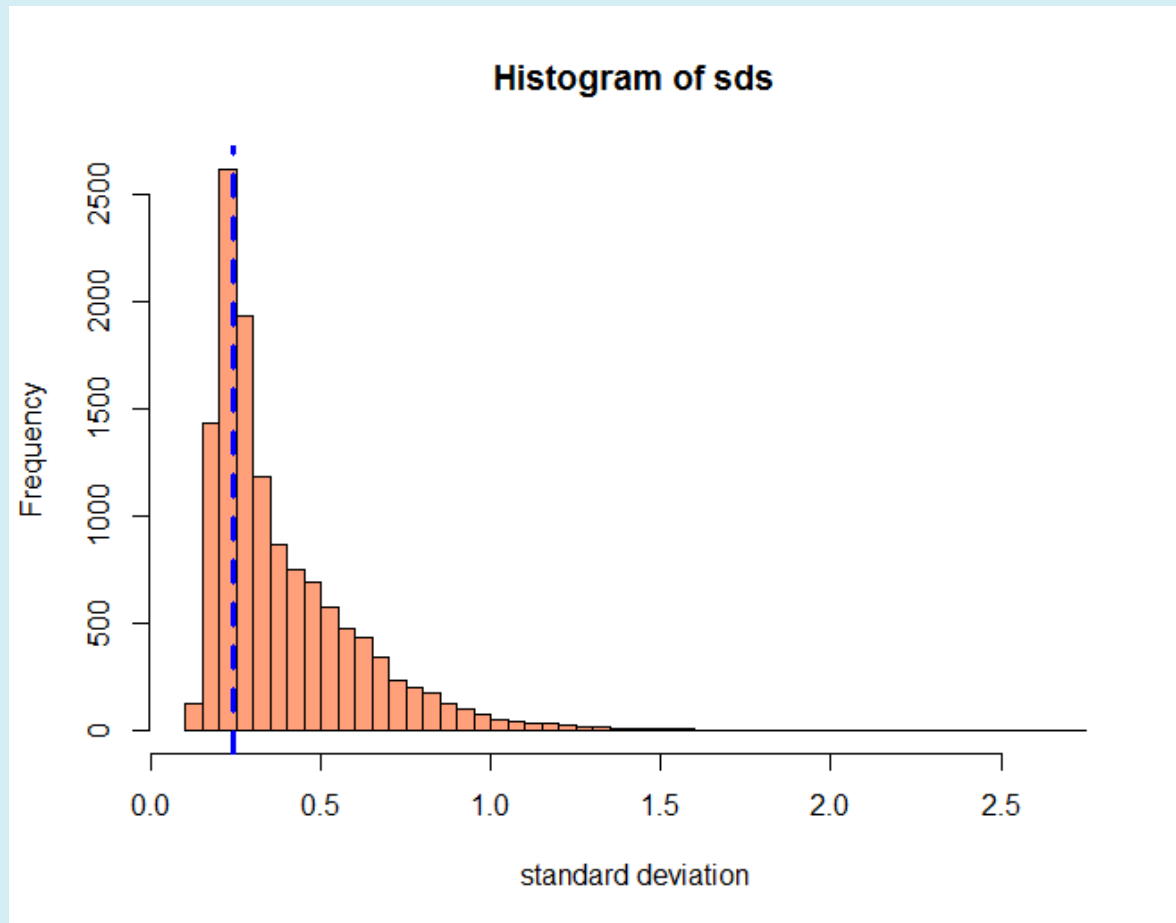


# Example 1: Non-specific Filtering



- Non-specific filtering is used to remove those genes that seen to be low or never expressed under any condition.
- The overall variability is calculated for each probe set regardless to which sample they belong to. Those genes with low variability are removed from the analysis assuming that gene expression is reflected as high variability. The **rowSds** and the **shorth** functions from the **genefilter** package can be used to perform this task.

# Example 1: Non-specific Filtering

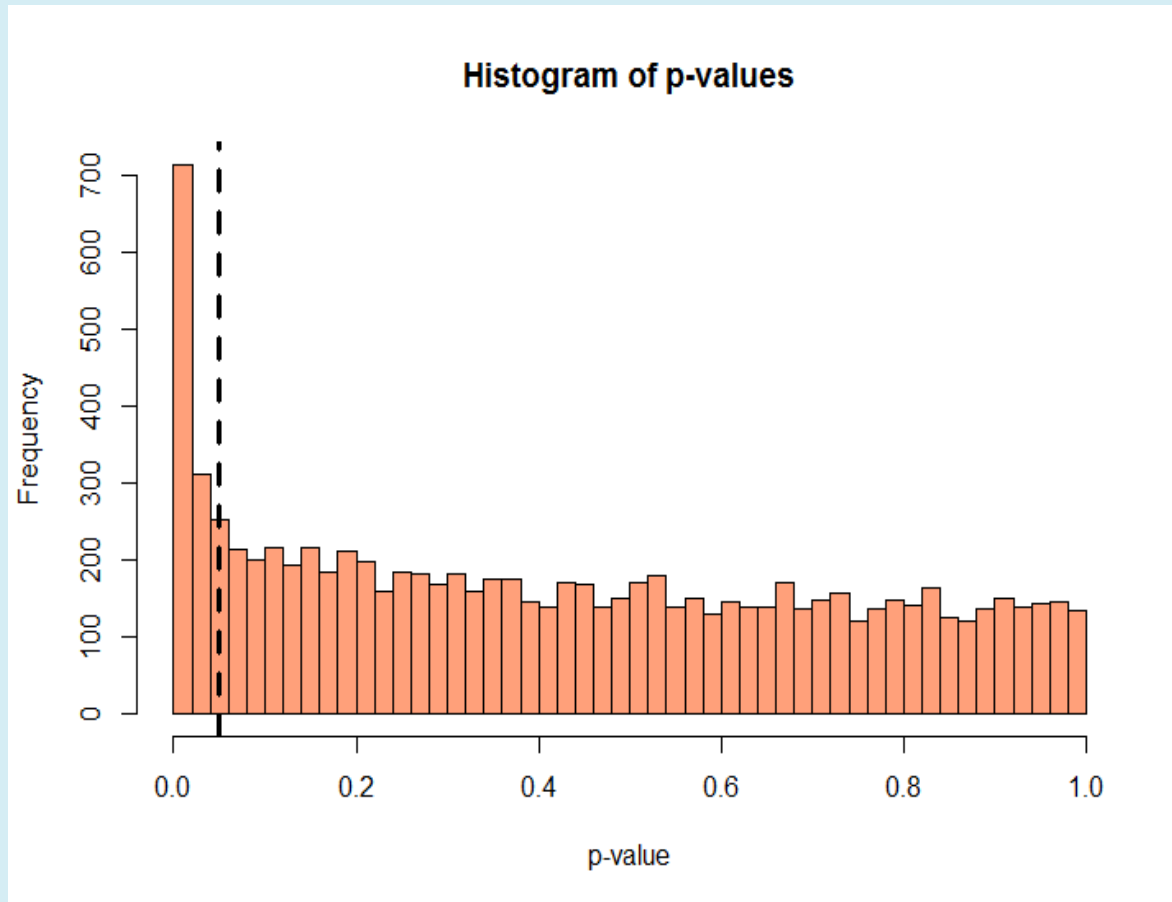


## Example 2: Gene Selection



- Gene selection consists of selecting those genes that are differentially expressed between samples and therefore can be used to discriminate between them.
- A statistical test can be performed for each probe. The null hypothesis is that they are not differentially expressed.
- The function **rowttests** from the `genefilter` package can be used to perform this task.

# Example 2: Gene Selection

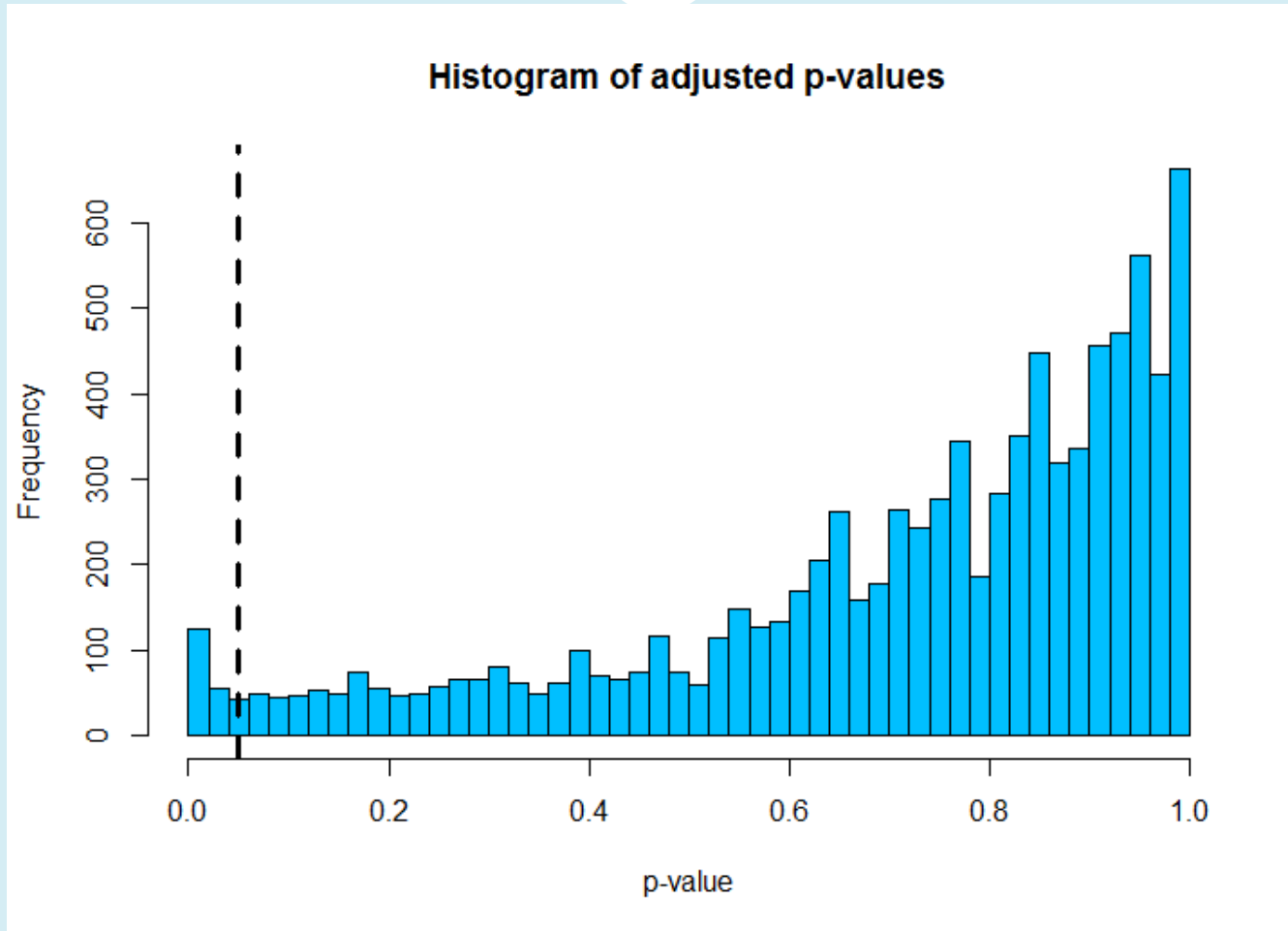


# Example 3: Multiple Testing Correction



- Multiple Testing Correction aims to reduce the rate of type I errors resulted from multiple statistical tests.
- The function **mt.raw2adjp** from the `multtest` package uses the Benjamini & Hochberg Procedure to control the False Discovery Rate (FDR).

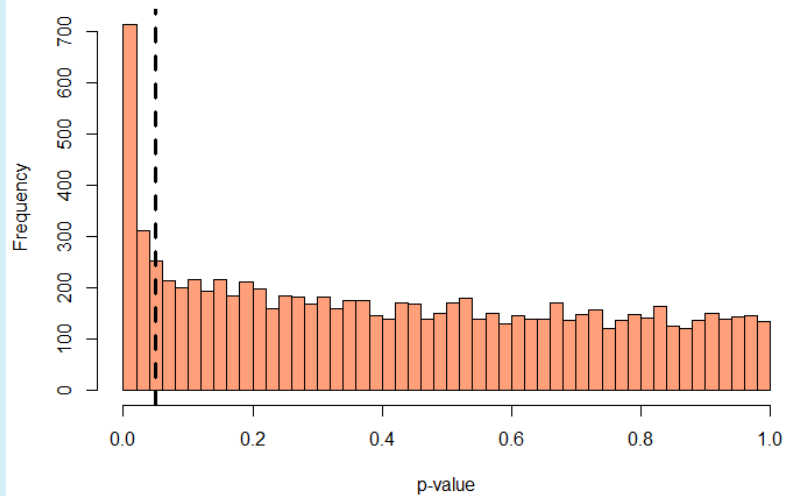
# Example 3: Multiple Testing Correction



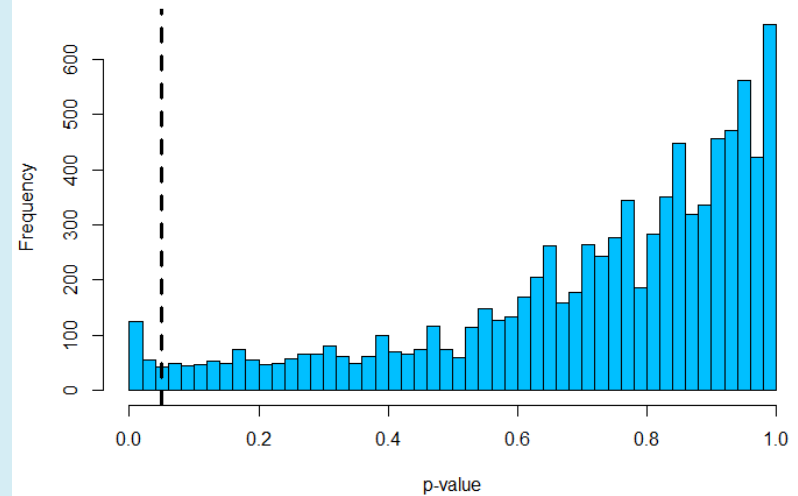
# Example 3: Multiple Testing Correction



Histogram of p-values



Histogram of adjusted p-values



# Presentation Roadmap



- What is Gene Expression Analysis?
- What is Bioconductor?
- The ALL dataset
- Example 1: Non-specific Filtering
- Example 2: Gene Selection
- Example 3: Multiple Testing Correction
- Summary



# Sources:



- Hahne, F., Wolfgang, H., Gentleman, R., & Falcon, S. (2008). Bioconductor Case Studies. Springer.
- Heydebrek, A. v., Wolfgang, H., & Gentleman, R. (2004). Differential Gene Expression with the Bioconductor Project. Bioconductor Working Papers.
- Hofmann, W.-K. (2006). Gene Expression Profiling by Microarrays. Cambridge University Press.
- McLachlan, G. J., Do, K. A., & Ambrose, C. (2004). Analyzing Microarray Gene Expression. John Wiley & Sons, Inc. .
- Tarca, A. L., Romero, R., & Draghici, S. (2006). Analysis of microarray experiments of gene expression profiling. National Institute of Health-Public Access, 373-388.