RNASeqGUI

http://bioinfo.na.iac.cnr.it/RNASeqGUI
Francesco Russo and Claudia Angelini
Istituto delle Applicazioni del Calcolo "Mauro Picone", IAC-CNR, Naples, Italy

MOTIVATIONS

RNASeqGUI is a graphical user interface for the **identification of differentially expressed genes** from RNA-Seq experiments.

1. An open source graphical user interface for the **DE** analysis of RNA-Seq data.
2. It includes several well known RNA-Seq tools, available at **www.bioconductor.org**.
3. ***High usability***, ***portability***, ***modulability***, allowing the possibility to introduce new functionalities in an immediate way.
4. The interface is not just a collection of some known methods, but it is designed to guide the user during the entire analysis process.
5. RNASeqGUI is also helpful for those who are expert R-users since it speeds up the usage of several methods drastically.

RNASeqGUI is freely available at http://bioinfo.na.iac.cnr.it/RNASeqGUI.

METHODS

RNASeqGUI is written in **R**, an open source object oriented language for statistical computing and graphics. It requires the **RGTK2** graphical library [1] to run. The GUI is divided into several sections and interfaces. One can start to analyse the data from any section or interface. There are two main starting points. One can use the GUI beginning from either *Bam Exploration Section* or from *Pre-Analysis Section*.

Data Analysis Section is the core of RNASeqGUI and contains several methods to identify differentially expressed genes.

FUTURE WORKS

In future, we will include new methods, new normalization procedures, the possibility to define more complex experimental designs, the pathway analysis and the Gene Ontology.

STRUCTURE OF THE GUI AND USAGE

RNASeqGUI is divided into five main sections. Each section is dedicated to a particular step of the data analysis process. The first section covers the exploration of the bam files. The second concerns the counting process of the mapped reads against a genes annotation file. The third focuses on the exploration of count-data and on preprocessing of the data, including the normalization procedures. The fourth is about the identification of the differentially expressed genes that can be performed by several methods, **edgeR, DESeq, DESeq2, noiSeq, baySeq**. Finally, the fifth section regards the inspection of the results produced by these methods and the quantitative comparison among them.

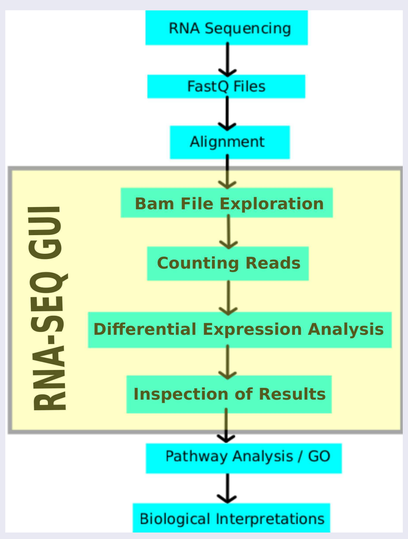

Figure: RNASeq worflow

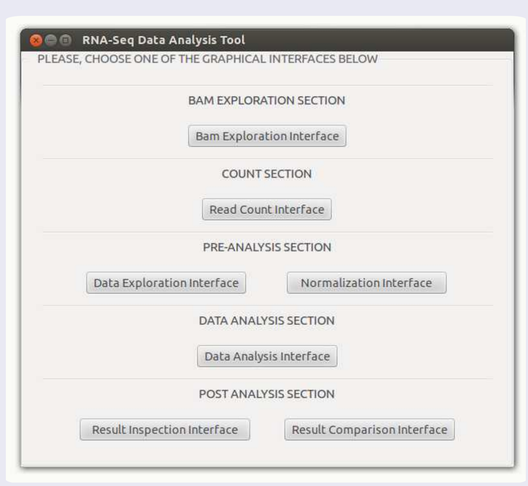

Figure: Sections of the GUI

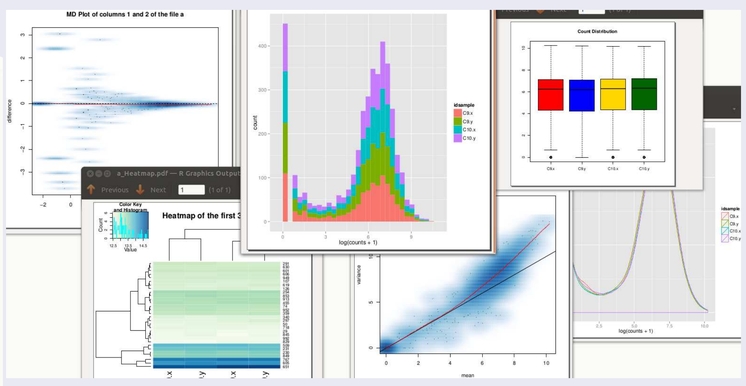

Figure: Some plots from Data Exploration Interface

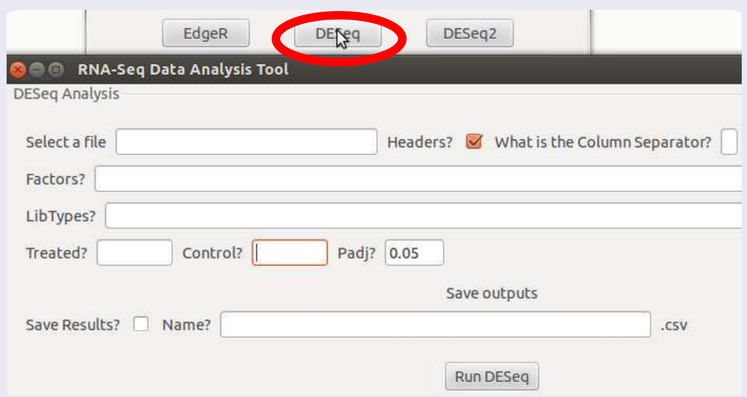

Figure: DESeq interface

DATA AND ANALYSIS

We used the dataset published by [3]. This dataset consists of seven samples. Three samples represent the response to a treatment and four samples are controls. Each sample is a cell culture of *Drosophila melanogaster*. We aligned the *fastq* files by running `tophat2`. Once the bam files were obtained, we performed the analysis with RNASeqGUI. We compared the results of edgeR, DESeq and NOISeq among them. We analysed the dataset published by [3] as a real data working example. We selected the chromosome 2L only to reduce the execution time. We analysed the expression of 2986 genes belonging to 2L chromosome. The methods found 128, 148, 102 DE gene, respectively. Among these, 86 genes were found DE in all the three used methods.

BamFileName	NameOfTheReducedBam	LibraryType	LibraryLayout
CG8144_RNA-1	2L_1	treated	single
CG8144_RNA-3	2L_3	treated	paired
CG8144_RNA-4	2L_4	treated	paired
Untreated-1	2L_U1	untreated	single
Untreated-3	2L_U3	untreated	paired
Untreated-4	2L_U4	untreated	paired
Untreated-6	2L_U6	untreated	single

Figure: Experimental design

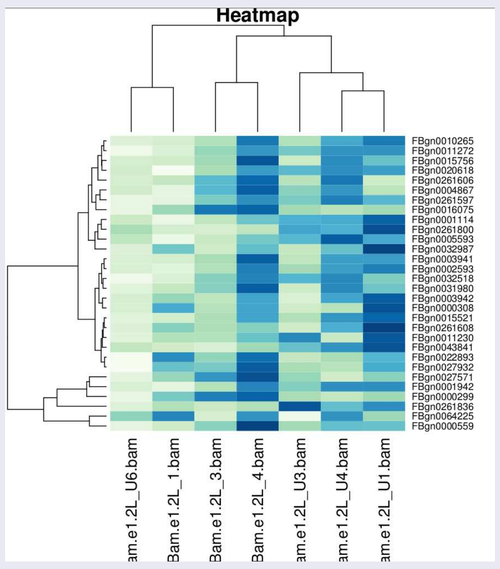

Figure: Heatmap

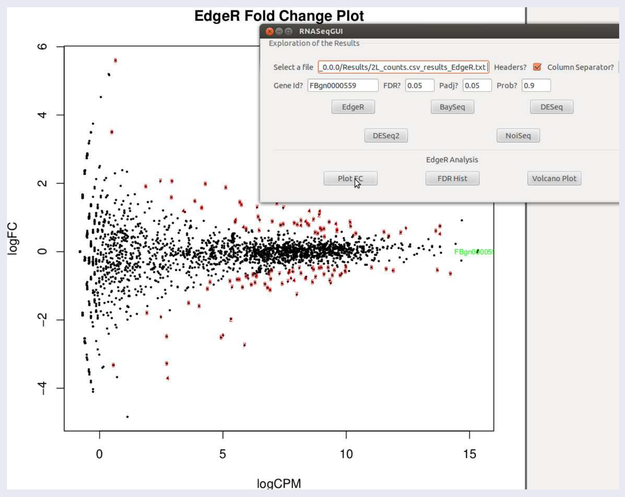

Figure: EdgeR Fold Change Plot

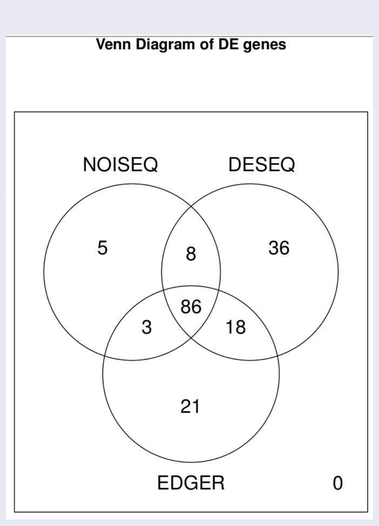

Figure: Venn diagram

REFERENCES

- [1] Michael L and Duncan TL (2010) RGtk2: A Graphical User Interface Toolkit for R. Journal of Statistical Software 37(8).
- [2] Villa-Vialaneix N and Leroux D (2013) sexy-rgtk: a package for programming RGtk2 GUI in a user-friendly manner. In Proceedings of: 2emes rencontres R.
- [3] Brooks AN, Yang L, Duff MO, Hansen KD, Park JW, Dudoit S, Brenner SE and Graveley BR (2011). Conservation of an RNA regulatory map between Drosophila and mammals. Genome Research 21:193-202.

ACKNOWLEDGEMENTS

This work was supported by the Italian Flagship **InterOmics** Project (PB.P05) and by BMBS **COST Action** BM1006.