



Reproducible Research Analysis with RNASeqGUI

<http://bioinfo.na.iac.cnr.it/RNASeqGUI>

Francesco Russo and Claudia Angelini

Istituto delle Applicazioni del Calcolo "Mauro Picone", IAC-CNR, Naples, Italy



AIM

The analysis of RNA-seq data usually requires to carry out several steps, to use different methods and to compare their outputs to obtain more reliable and less biased results. RNASeqGUI is a novel R-package (open-source) that implements a graphical user interface for the DE analysis of RNA-Seq data. It includes several well known RNA-Seq tools, available at www.bioconductor.org. It generates several reports (*html, docx, pdf, odt, txt*) of all steps performed during the analysis of a specific project. Reports are produced in R markdown format via **knitr** library and they include the documentation of the methods used and the R code that has been executed during the RNASeqGUI usage. Results can be viewed and explored on a web browser thanks to **ReportingTools** library that allows the user to navigate through them.

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

METHODS

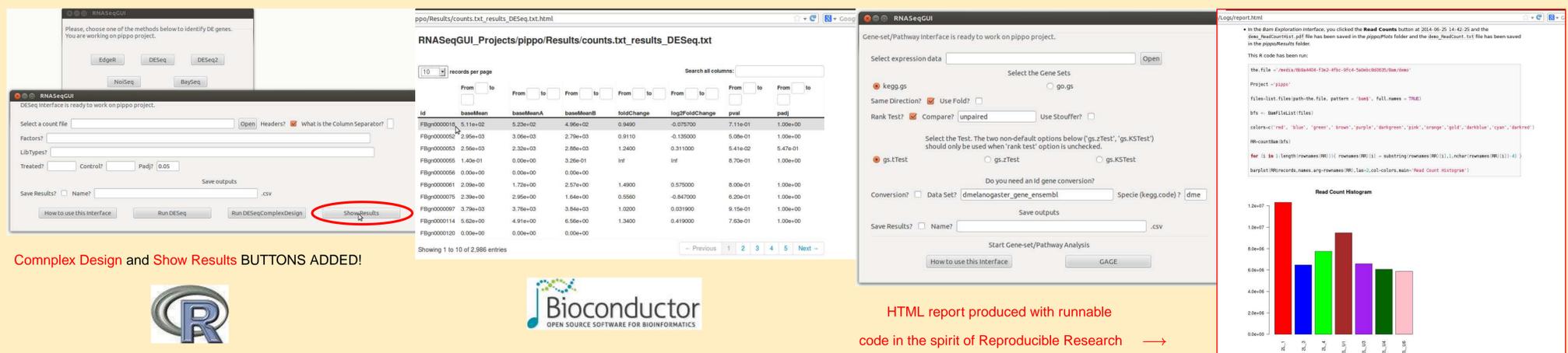
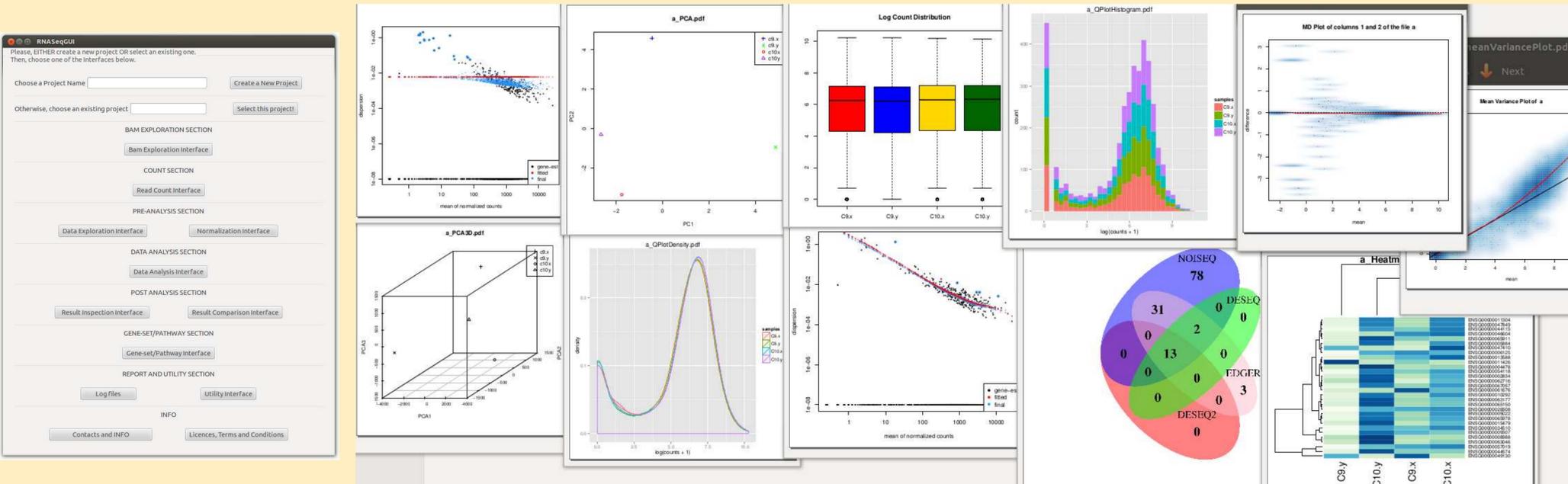
RNASeqGUI is implemented in **R**. It requires the **RGTK2** graphical library [1, 2] to run and uses **BiocParallel** to speed up the computations. A preliminary version of RNASeqGUI was published [5] on **Bioinformatics**.

Here, we present some new features, such as:

- **Reactive Data Exploration** via a web browser,
- **Reproducible Research**,
- **Complex Design Analysis**,
- **Pathview/GAGE analysis**,
- **Utility Interface**,
- **FeatureCounts** a new alternative method included in the **Read Count Interface** to count reads,
- **Venn Diagrams DE 4 sets**.

STRUCTURE OF THE GUI AND USAGE

RNASeqGUI is divided into 7 main sections. Each section is dedicated to a particular step of the data analysis process.



Complex Design and Show Results BUTTONS ADDED!

HTML report produced with runnable code in the spirit of Reproducible Research

RESULTS

We used the dataset published by [4]. 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 [4] 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.

REFERENCES

- [1] Lawrence M and Duncan TL (2010) RGtk2: A Graphical User Interface Toolkit for R. *Journal of Statistical Software* 37(8).
- [2] Lawrence M and Verzani J. *Programming Graphical User Interfaces in R*. CRC The R Series. Chapman & Hall, June 2012.
- [3] 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*.
- [4] 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.
- [5] F. Russo and C. Angelini. RNASeqGUI: A GUI for analysing RNA-seq data. *Bioinformatics* (2014) doi: 10.1093/bioinformatics/btu308

ACKNOWLEDGEMENTS

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



www.combolab.it



EUROPEAN COOPERATION IN SCIENCE AND TECHNOLOGY