

Reproducible Research in the era of Next Generation Sequencing: current approaches, examples and future perspectives

Claudia Angelini✉, Dario Righelli, Francesco Russo

Istituto per le Applicazioni del Calcolo "M. Picone", Napoli, Italy

Angelini C *et al.* (2015) *EMBnet.journal* **21**(Suppl A), e808. <http://dx.doi.org/10.14806/ej.21.A.808>

Next Generation Sequencing (NGS) has revolutionised the way of thinking and of performing biomedical research. It is now possible to investigate biological aspects of cell functionalities and to understand previously unexplored disease etiologies by analyzing multi-omic data. Several computational (open-source) tools have been developed to analyze NGS data. Despite the technological advances, the way in which data analysis is performed and described in most of scientific papers does not facilitate the reproducibility of scientific results. Complex analyses are often poorly described and the lack of the technical information impedes the possibility for a researcher to reproduce the results available in literature (Nekrutenko and Taylor, 2012). Therefore, the problem of the "Reproducible Research", here-denoted RR, (Stodden *et al.*, 2014) is emerging as a serious issue for all Life Sciences.

In this work, we first introduce the concept of RR, its main benefits and challenges. Then, we discuss a simple way to develop novel computational tools that implement RR by using R and

Bioconductor packages. In particular, we show how RR can be incorporated within a graphical user-friendly interface and how such tools can automatically generate executable analysis reports. Then, we describe how it is possible to speed up repetitive and computational expensive function calls by using results stored in a cache memory (the latter point is crucial for the analysis of "Big Data" as those collected with NGS experiments). As a concrete working example of our approach, we illustrate the advances we have introduced in RNASeqGUI (Russo and Angelini, 2014).

References

- Nekrutenko A, Taylor J (2012) Next-generation sequencing data interpretation: enhancing reproducibility and accessibility. *Nature Reviews Genetics* **13**, 667-672. <http://dx.doi.org/10.1038/nrg3305>
- Stodden V, Leisch F, Peng RD (2014) Implementing Reproducible Research. In: Chapman & Hall/CRC The R Series.
- Russo F, Angelini C (2014) RNASeqGUI: A GUI for analyzing RNA-seq data. *Bioinformatics* **30**(17), 2514-2516. <http://dx.doi.org/10.1093/bioinformatics/btu308>