iMir: an innovative and complete pipeline for smallRNA-Seq data analysis

Giorgio Giurato[1], Antonio Rinaldi[1], Adnan Hashim[1], Giovanni Nassa[1], Maria Ravo[1], Francesca Rizzo[1], Roberta Tarallo[1], Angela Cordella[2], Giovanna Marchese[2], Domenico Memoli[1], Alessandro Weisz[1] ✉

[1]Department of Medicine and Surgery, Laboratory of Molecular Medicine and Genomics, University of Salerno, Baronissi, Italy

[2]Genomix4Life Srl, Spin-Off of the University of Salerno, Baronissi, Italy

Giurato G *et al.* (2015) *EMBnet.journal* **21**(Suppl A), e810. http://dx.doi.org/10.14806/ej.21.A.810

Next-generation sequencing allows researchers to gage the depth and variation of small non-coding RNA populations, comprising miRNAs, piRNAs, tRNAs and other regulatory small transcripts. The accurate analysis of smallRNA-Seq data remain a non-trivial computational problem, requiring implementation of multiple statistical and bioinformatics tools. Here we present iMir (Giurato *et al.*, 2013), a modular pipeline for comprehensive analysis of smallRNA-Seq data, comprising specific tools for adapter trimming, quality filtering, differential expression analysis, biological target prediction and other useful options by integrating multiple open source modules and resources in an automated workflow (Figure 1).

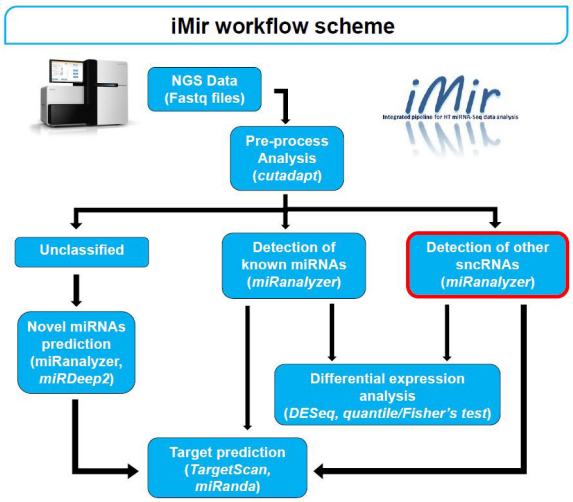

Figure 1. The pipeline accepts NGS data as input and then proceeds automatically to perform several independent analysis, most of them can be selected or excluded according to the user's needs.

iMir is based on reliable, flexible and fully automated workflow, allowing to rapidly and efficiently analyze high-throughput smallRNA-Seq data, such as those produced by the most recent high-performance next generation sequencers. This pipeline allowed us to investigate piRNA expression patterns in rat liver and their modulation during regenerative proliferation (Rizzo *et al.*, 2014) and to identify >100 human piRNAs in breast cancer, some of which showing significant differences in expression in mammary epithelial compared to cancer cells or in normal respect to cancerous mammary tissues (Hashim *et al.*, 2014), and in endometrial hyperplasia and cancer (Ravo *et al.*, 2015).

Acknowledgements

Work supported by: Italian Ministry of Health (Grant Young Researcher GR-2011–02350476 to M.R.), Italian Ministry for Education, University and Research (Grants PRIN 2010LC747T to A.W. and FIRB RBFR12W5V5 _ 003 to R.T.), Italian Association for Cancer Research (Grants IG 13176 to A.W.), National Research Council Flagship Project Interomics. G.N. is supported by a 'Mario e Valeria Rindi' fellowship of the Italian Foundation for Cancer Research, A.R. is a PhD student of the Research Doctorate 'Molecular and Translational Oncology and Innovative Medical-Surgical Technologies', University of Catanzaro 'Magna Graecia'.

References

Giurato G, De Filippo MR, Rinaldi A, Hashim A, Nassa G, *et al.* (2013) iMir: an integrated pipeline for high-throughput analysis of small non-coding RNA data obtained by smallRNA-Seq. *BMC Bioinformatics* **14**:362. http://dx.doi.org/10.1186/1471-2105-14-362

Hashim A, Rizzo F, Marchese G, Ravo M, Tarallo R, *et al.* (2014) RNA sequencing identifies specific PIWI-interacting small non-coding RNA expression patterns in breast cancer. *Oncotarget* **5**(20), 9901-9910.

Ravo M, Cordella A, Rinaldi A, Bruno G, Alexandrova E *et al.* (2015) Small non coding RNA deregulation in endometrial carcinogenesis. *Oncotarget* (Epub ahead of print).

Rizzo F, Hashim A, Marchese G, Ravo M, Tarallo R *et al.* (2014) Timed regulation of P-element-induced wimpy testis-interacting RNA expression during rat liver regeneration. *Hepatology* **60**(3), 798-806. http://dx.doi.org/10.1002/hep.27267