BIOINFORMATICS APPLICATIONS NOTE Vol. 29 no. 20 2013, pages 2651–2652
doi:10.1093/bioinformatics/btt457

Gene expression

Advance Access publication August 13, 2013

omiRas: a Web server for differential expression analysis of miRNAs derived from small RNA-Seq data

Sören Müller[1,2,3,†], Lukas Rycak[2,†], Peter Winter[2], Günter Kahl[1], Ina Koch[3] and Björn Rotter[2,*]

[1]Plant Molecular Biology, Institute of Molecular BioSciences, University of Frankfurt am Main, Marie-Curie-Str.9, 60439 Frankfurt am Main, [2]GenXPro GmbH, Frankfurt Innovation Center Biotechnology, Altenhöferallee 3, 60438 Frankfurt am Main and [3]Molecular Bioinformatics Group, Institute of Computer Science, Faculty of Computer Science and Mathematics, Cluster of Excellence, Frankfurt ''Macromolecular Complexes'', Robert-Mayer-Strasse 11-15, 60325 Frankfurt am Main, Germany

Associate Editor: Martin Bishop

ABSTRACT

Summary: Small RNA deep sequencing is widely used to characterize non-coding RNAs (ncRNAs) differentially expressed between two conditions, e.g. healthy and diseased individuals and to reveal insights into molecular mechanisms underlying condition-specific phenotypic traits. The ncRNAome is composed of a multitude of RNAs, such as transfer RNA, small nucleolar RNA and microRNA (miRNA), to name few. Here we present omiRas, a Web server for the annotation, comparison and visualization of interaction networks of ncRNAs derived from next-generation sequencing experiments of two different conditions. The Web tool allows the user to submit raw sequencing data and results are presented as: (i) static annotation results including length distribution, mapping statistics, alignments and quantification tables for each library as well as lists of differentially expressed ncRNAs between conditions and (ii) an interactive network visualization of user-selected miRNAs and their target genes based on the combination of several miRNA–mRNA interaction databases.

Availability and Implementation: The omiRas Web server is implemented in Python, PostgreSQL, R and can be accessed at: http://tools.genxpro.net/omiras/.

Contact: rotter@genxpro.de

Supplementary Information: Supplementary data are available at *Bioinformatics* online.

Received on June 6, 2013; revised on August 1, 2013; accepted on August 2, 2013

1 INTRODUCTION

Non-coding ribonucleic acids (ncRNAs) are functional molecules that do not code for a protein, but play important roles in normal development, physiology and disease (Stefani and Slack, 2008). The ncRNAome is grouped into RNAs of different functions. Among them, microRNAs (miRNAs), which have gained attention due to their implication in the development of diseases like cancer or cardiovascular problems (Lu *et al.*, 2008). MiRNAs can repress gene expression through translational repression or messenger RNA (mRNA) deadenylation and decay. Next-generation sequencing (NGS) offers unbiased insights into the small non-coding transcriptome by small RNA-Sequencing (sRNA-Seq). Several Web services, like DARIO (Fasold *et al.*, 2011), wapRNA (Zhao *et al.*, 2011) or ncPRO-seq (Chen *et al.*, 2012) have been developed for the annotation and quantification of known and novel ncRNAs in a single sRNA-Seq library. miRAnalyzer (Hackenberg *et al.*, 2009) and CPSS (Zhang *et al.*, 2012) allows quantification of known ncRNAs, detection and quantification of novel miRNAs as well as differential expression with functional analysis. Whereas CPSS is focused on experiments without replicates, the new version of miRAnalyzer can handle replicates, but does not allow functional analysis. The sRNA workbench (Stocks *et al.*, 2012) offers a collection of independent tools for initial processing and filtering of sRNA-Seq data as well as annotation and prediction of novel miRNAs and functional analysis. FiRePat (Mohorianu *et al.*, 2012) allows for the construction of heatmaps and correlation networks from multiple samples of sRNA/mRNA datasets. Most tools, except wapRNA and the sRNA toolkit, require local pre-processing of raw sequencing data. None of the tools is specifically designed to detect differential expression between two conditions for various ncRNA classes. Apart from FirePat, no other tool allows for the construction of miRNA–mRNA interaction networks.

Therefore, we developed omiRas—a Web service for the analysis of ncRNA datasets derived from Illumina sRNA-Seq experiments. Starting with raw sequencing data, omiRas offers an efficient way to analyze differential expression of ncRNAs between two groups and to assign functions to differentially expressed miRNAs. MiRNA–mRNA interaction databases allow the user to construct networks of interesting miRNAs and mRNAs to identify miRNAs with implications in the development of differential gene signatures.

2 APPROACH

2.1 Data submission

Sequencing data are submitted in four steps to the omiRas platform: first, give project and condition names, second, the species selection (*Homo sapiens*, *Mus musculus*, *Solanum lycopersicum* or *Arabidopsis thaliana*). Third, upload of all raw FASTQ or FASTA files (compressed or uncompressed) and validity check.

*To whom correspondence should be addressed.
†The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

The file size for each library is limited to 2 GB. In the last step, information about the sequencing adapter has to be provided. Additionally, the user can select whether novel miRNA prediction should be performed.

2.2 Workflow

The omiRas analysis pipeline is implemented in Python 2.7 (a detailed description of the workflow is provided in Supplementary Fig. S1 and Supplementary Methods). omiRas processes up to five projects in parallel; for each project, five CPUs are used for the computation. The workflow starts with adapter removal by a local alignment of the adapter sequence to the 3′-end of the reads accounting for sequencing errors. Subsequently, Q2 basecalls are removed, reads summarized to UniTags and their quantity in each library determined. Singletons are removed from the dataset and the remaining tags are mapped to the genome under consideration with bowtie (Langmead *et al*., 2009) allowing at most two mismatches. The mapping coordinates are overlapped with various models of coding and non-coding RNAs. Tags mapping to exonic regions of protein coding genes are excluded from further analysis. Next, ncRNAs are quantified in each library. For tags mapping to multiple genomic loci, the number of reads corresponding to the tag is normalized with the number of mapping loci. Tags overlapping introns or intergenic regions are used for the prediction of miRNAs not present in miRBase (Kozomara and Griffiths-Jones, 2011) with miRDeep (Friedländer *et al*., 2008) and RNAfold (Gruber *et al*., 2008). Finally, quantification and annotation results are summarized for each library (Fig. 1a). The DESeq package (Anders and Huber, 2010) is used to normalize the data and to independently calculate differential expression between the two conditions for each class of ncRNAs. The interface allows the user to visualize results as heatmaps and correlation plots (Pearson product-moment correlation coefficient). An interaction network of one or more differentially expressed miRNA(s) and the potential targets can be created (Fig. 1b). Targets are determined by the combination of seven target prediction tools (for details, see Supplementary Methods). For an application of omiRas, see Supplementary Application.

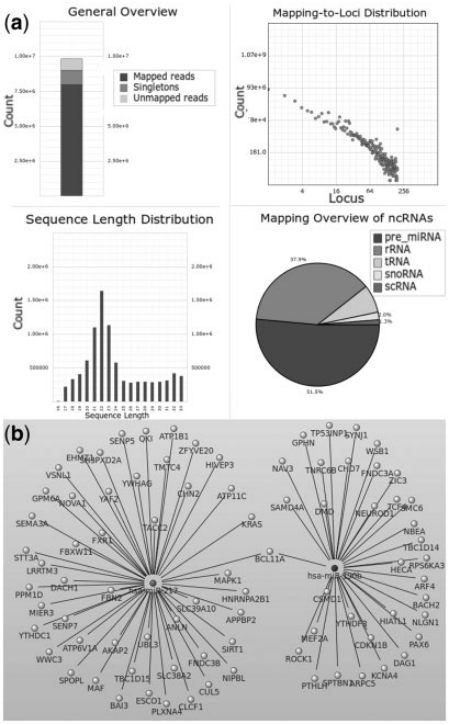

Fig. 1. Exemplary quantification results for a single sRNA-Seq library (**a**) and an interaction network for hsa-miR-217 and 190 b generated by omiRas (**b**)

2.3 Webserver implementation

The platform is implemented in Python 2.7 using the Web framework Django 1.4.5. The system is hosted on a Cherokee Web server 1.2.101. Job queuing is carried out by the asynchronous task queue Celery 3.0.19 that uses the distributed message passing system RabbitMQ 3.0.4. Dynamical plots are generated with the graphics utility canvasXpress (http://canvasxpress.org/).

Funding: This work was supported by the Bundesministerium für Bildung und Forschung (BMBF) [grant number FKZ 0316043].

Conflict of Interest: none declared.

REFERENCES

Anders,S. and Huber,W. (2010) Differential expression analysis for sequence count data. *Genome Biol*., **11**, R106.

Chen,C.-J. *et al*. (2012) ncPRO-seq: a tool for annotation and profiling of ncRNAs in sRNA-seq data. *Bioinformatics*, **28**, 3147–3149.

Fasold,M. *et al*. (2011) DARIO: a ncrna detection and analysis tool for next-generation sequencing experiments. *Nucleic Acids Res*., **39** (**Suppl. 2**), W112–W117.

Friedländer,M.R. *et al*. (2008) Discovering microRNAs from deep sequencing data using mirdeep. *Nat Biotechnol*., **26**, 407–415.

Gruber,A.R. *et al*. (2008) The vienna RNA websuite. *Nucleic Acids Res*., **36** (**Suppl. 2**), W70–W74.

Hackenberg,M. *et al*. (2009) miRanalyzer: a microRNA detection and analysis tool for next-generation sequencing experiments. *Nucleic Acids Res*., **37** (**Suppl. 2**), W68–W76.

Kozomara,A. and Griffiths-Jones,S. (2011) miRBase: integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Res*., **39** (**Suppl. 1**), D152–D157.

Langmead,B. *et al*. (2009) Ultrafast and memory-efficient alignment of short dna sequences to the human genome. *Genome Biol*., **10** (3), R25.

Lu,M. *et al*. (2008) An analysis of human microrna and disease associations. *PLoS One*, **3**, e3420.

Mohorianu,I. *et al*. (2012) Firepat–finding regulatory patterns between sRNAs and genes. *Data Min. Knowl. Discov*., **2**, 273–284.

Stefani,G. and Slack,F.J. (2008) Small non-coding RNAs in animal development. *Nat. Rev. Mol. Cell Biol*., **9**, 219–230.

Stocks,M.B. *et al*. (2012) The UEA sRNA Workbench: a suite of tools for analysing and visualizing next generation sequencing microrna and small RNA datasets. *Bioinformatics*, **28**, 2059–2061.

Zhang,Y. *et al*. (2012) CPSS: a computational platform for the analysis of small RNA deep sequencing data. *Bioinformatics*, **28**, 1925–1927.

Zhao,W. *et al*. (2011) wapRNA: a web-based application for the processing of RNA sequences. *Bioinformatics*, **27**, 3076–3077.