Bioinformatics, 32(13), 2016, 2069–2071
doi: 10.1093/bioinformatics/btw070
Advance Access Publication Date: 2 March 2016
Applications Note

Data and text mining

IsomiR Bank: a research resource for tracking IsomiRs

Yuanwei Zhang[1,†], Qiguang Zang[2,†], Bo Xu[3,†], Wei Zheng[1,†], Rongjun Ban[2], Huan Zhang[1], Yifan Yang[4], Qiaomei Hao[1], Furhan Iqbal[1], Ao Li[2,*] and Qinghua Shi[1,5,*]

[1]Molecular and Cell Genetics Laboratory, The CAS Key Laboratory of Innate Immunity and Chronic Diseases, Hefei National Laboratory for Physical Sciences at Microscale, School of Life Sciences, CAS Center for Excellence in Molecular Cell Science, University of Science and Technology of China, Collaborative Innovation Center of Genetics and Development, Collaborative Innovation Center for Cancer Medicine, Hefei 230027, Anhui, China, [2]School of Information Science and Biology, Centers for Biomedical Engineering, University of Science and Technology of China, Hefei 230027, China, [3]Reproductive Medical Centre of Anhui, Provincial Hospital Affiliated to Anhui Medical University, Hefei 230001, China, [4]Department of Statistics, University of Kentucky, Lexington, KY 40536, USA and [5]Hefei Institute of Physical Science, China Academy of Science, Hefei 230027, China

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

Associate Editor: Jonathan Wren

Received on November 30, 2015; revised on January 31, 2016; accepted on February 1, 2016

Abstract

Summary: Next-Generation Sequencing (NGS) technology has revealed that microRNAs (miRNAs) are capable of exhibiting frequent differences from their corresponding mature reference sequences, generating multiple variants: the isoforms of miRNAs (isomiRs). These isomiRs mainly originate via the imprecise and alternative cleavage during the pre-miRNA processing and post-transcriptional modifications that influence miRNA stability, their sub-cellular localization and target selection. Although several tools for the identification of isomiR have been reported, no bioinformatics resource dedicated to gather isomiRs from public NGS data and to provide functional analysis of these isomiRs is available to date. Thus, a free online database, IsomiR Bank has been created to integrate isomiRs detected by our previously published algorithm CPSS. In total, 2727 samples (Small RNA NGS data downloaded from ArrayExpress) from eight species (*Arabidopsis thaliana*, *Drosophila melanogaster*, *Danio rerio*, *Homo sapiens*, *Mus musculus*, *Oryza sativa*, *Solanum lycopersicum* and *Zea mays*) are analyzed. At present, 308 919 isomiRs from 4706 mature miRNAs are collected into IsomiR Bank. In addition, IsomiR Bank provides target prediction and enrichment analysis to evaluate the effects of isomiRs on target selection.
Availability and implementation: IsomiR Bank is implemented in PHP/PERL + MySQL + R format and can be freely accessed at http://mcg.ustc.edu.cn/bsc/isomir/
Contacts: aoli@ustc.edu.cn or qshi@ustc.edu.cn
Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 Introduction

MicroRNAs (miRNAs), with average length of 22 nucleotides, are ubiquitously expressed and regulate many essential biological processes mainly via post-transcriptional silencing of genes through mRNA decay and/or translational repression (Bartel, 2004). During this process, the primary miRNA transcripts (pri-miRNAs) are mainly cleaved by complexes of RNAses III (Drosha and DGCR8) and give rise to one or more precursor miRNAs (pre-miRNAs) hairpins. Following Dicer processing, the hairpins release short double-stranded RNA. One of the resultant strand (defined as mature miRNA) binds to the protein Argonaut 2 (Ago2), and gets incorporated into the RNA Induced Silencing Complexes (RISCs). It guides RISCs to the 3′UTR of mRNAs by base pairing with fully or partially complementary sequences (Griffiths-Jones *et al.*, 2006). Previously, miRNAs were only annotated as the canonical sequences. However, extensive application of high-throughput sequencing technology to detect expression profile of miRNA has revealed that miRNAs can frequently exhibit differences from their corresponding canonical mature sequences, generating multiple variants known as isoforms of miRNAs (isomiRs) (Guo and Lu, 2010; Li *et al.*, 2012; Llorens *et al.*, 2013; Vaz *et al.*, 2013; Wang *et al.*, 2008). Generally, the reference miRBase miRNA (having the canonical sequence) often differs from the most abundant isomiR (Ameres and Zamore, 2013; Loher *et al.*, 2014). There are two main processes that lead to the production of isomiRs. These isomiRs mainly originate via the imprecise and alternative cleavage (not completely random) during the pre-miRNA processing and post-transcriptional modifications that have influence on miRNA stability, their sub-cellular localization and target selection (Ameres and Zamore, 2013; Loher *et al.*, 2014). On the other hand, miRNA editing is the process that generates isomiRs by post-transcriptional enzymatic editing of the miRNA genome (Ameres and Zamore, 2013; Neilsen *et al.*, 2012). It has also been reported recently that isomiRs can significantly affect the half-life of miRNA, as well as their sub-cellular localization and their target specification (Ameres and Zamore, 2013; Cloonan *et al.*, 2011; Li *et al.*, 2012; Wang *et al.*, 2008).

Here, we have established a database, IsomiR Bank, to support and facilitate the ongoing isomiRs research. Our database contains 308 919 isomiRs detected from 2727 samples (Small RNA NGS data downloaded from ArrayExpress database) of eight species (*Arabidopsis thaliana*, *Drosophila melanogaster*, *Danio rerio*, *Homo sapiens*, *Mus musculus*, *Oryza sativa*, *Solanum lycopersicum* and *Zea mays*) with manually curated metadata for each sample (organism, sex, tissue, development stage, age, experiment condition, treatment and sequencing platform). IsomiR Bank also has a user-friendly web interface allowing researchers to quickly find and navigate the isomiRs of their interest and related annotation along with the functional analysis that can help in evaluating the effects of isomiR on target selection and downstream pathways. Our database is not only a collection of isomiRs from NGS data but also a tool to find out the candidate functional isomiRs for further experimental studies.

2 Content and construction

The general process of data collection, annotation and model development for IsomiR Bank are illustrated in Figure 1 (The detail methods for IsomiR Bank are provided in Supplementary Information). IsomiR Bank is developed in a user friendly mode, and includes a search engine to find the isomiRs of interest. The search option provides an interface for querying IsomiR Bank either with the sequence of isomiRs or by the name of miRNA/miRNA family or by the sources of isomiRs [e.g. tissue]. For example, if an isomiR sequence of GTAAAGCAAGATAACCGAAAGT is input to the search engine (Supplementary Fig. S1A), the search results will appear in a tabular format containing species information in which this isomiR has been reported along with isomiR Sequence, Family Accession (the miRNA family accession number provided by miRBase for this searched isomiR's canonical miRNA), Family Name, Mature miRNA Name (the canonical miRNA of searched isomiR), Tissue origin, information of reads per million (RPM) of the isoform, Development stage, Genotype and Sex (Supplementary Fig. S1A). If users click on an interested entry in the isomiR Sequence, the database will display a summary of that isomiR, which is subdivided in four parts as: (i) IsomiR Source; (ii) Detailed Sample Information; (iii) IsomiR Form; (iv) Related Canonical miRNA Information; (v) IsomiR Analysis (Supplementary Fig. S1A). In order to predict the influence of isomiRs on targeted gene selection and downstream pathway function, IsomiR Bank provides target prediction for the isomiRs by miRanda (www.microrna.org) and then enrichment analysis for these predicted isomiR targets. When the analysis is finished, the third tab [Enrichment Analysis] will provide figures to illustrate the effects of all these isoforms on targets selection (Supplementary Fig. S1B). Moreover, functional enrichment analysis for these affected targets will also be provided to users. By selecting the Annotation Categories, users can overview the effect of their selected isomiRs on downstream biological processes, pathway and enriched protein domains (Supplementary Fig. S2). In addition to searching by a specific isomiR sequence or its name, all entries of IsomiR Bank can be browsed by the name of organism, tissue, miRNA families or by experimental conditions (Supplementary Fig. S3). In this way, users can browse the isomiR in a particular species in which the isomiR was identified and reported, or they can browse isomiR in each miRNA family (The examples for search and browse option and documentation for IsomiR Bank is provided in Supplementary information).

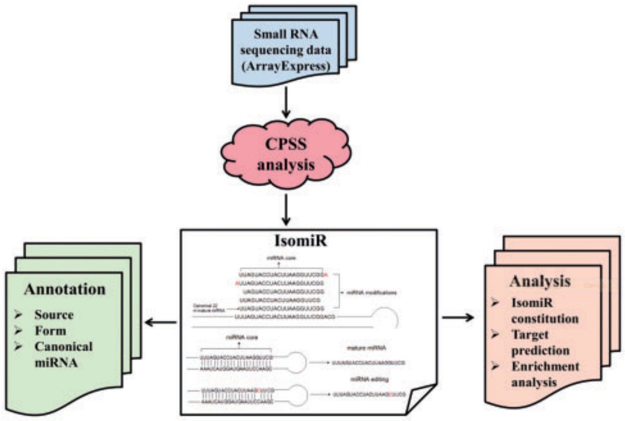

Fig. 1. IsomiR Bank database scheme

3 Discussion

Until now, a number of tools, such as MODOMICS, SeqBuster, IsomiRex, MiRGator v3.0 and CPSS have been generated and are in use for the detection of RNA modifications from small RNA deep sequencing data (Cho *et al.*, 2013; Dunin-Horkawicz *et al.*, 2006; Pantano *et al.*, 2010; Sablok *et al.*, 2013; Zhang *et al.*, 2012). However, these tools only focus on isomiRs detection and are unable to provide the functional annotation for these isomiRs. A public resource for recording the detected isomiRs with functional annotations is highly in demand. Hence, we have established a database, IsomiR Bank, which is a comprehensive and consistently annotated resource of miRNA isoforms for eight species. We have collected 2727 samples for isomiR analysis (Supplementary Fig. S4A). Among them, 977 samples were from *Homo sapiens*, 207 were from *Mus musculus* and 1543 were from the other six species. Presently, IsomiR Bank contains 308 919 isomiRs of 4706 mature miRNAs from 813 miRNA families (The mean number of isomiRs for each mature miRNA is 65.64, range from 1 to 1850, and other statistical analysis for isomiRs in IsomiR Bank are provided in Supplementary Information).

We analyzed a number of isomiRs shared between any two species among the eight species in the database. Interestingly, many isomiRs are conserved in animals or plants, but a specific isomiR of animals could not be found in plants and vice versa (Supplementary Table S1). This observation is consistent with previous studies showing few miRNAs were known to be structurally or functionally conserved between plants and animals (Arteaga-Vazquez *et al.*, 2006; Millar and Waterhouse, 2005). Moreover, we observed that more conserved isomiRs could be found within two species having closer genetic/evolutionary relationship (Supplementary Table S1). These observations suggest that the isomiRs are also conserved like canonical miRNAs during evolutionary process and they might be involved in regulation of same biological processes across different species (Ameres and Zamore, 2013; Cloonan *et al.*, 2011; Li *et al.*, 2012; Wang *et al.*, 2008).

Funding

This work was supported by the National Natural Science Foundation of China (31301227, 31501199, 31571555 and 61571414), the National Basic Research Program (2013CB947900?2012CB944402, 2014CB943101, and 2013CB945502) of China (973), Major Program of Development Foundation of Hefei Center for Physical Science and Technology (2014FXZY003), and the Fundamental Research Funds for the Central Universities (WK2340000069).

Conflict of Interest: none declared.

References

Ameres,S.L. and Zamore,P.D. (2013) Diversifying microRNA sequence and function. *Nat. Rev. Mol. Cell Biol.*, **14**, 475–488.

Arteaga-Vazquez,M. *et al.* (2006) A family of microRNAs present in plants and animals. *Plant Cell*, **18**, 3355–3369.

Bartel,D.P. (2004) MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell*, **116**, 281–297.

Cloonan. *et al.* (2011) MicroRNAs and their isomiRs function cooperatively to target common biological pathways. *Genome Biol.*, **12**, R126.

Cho,S. *et al.* (2013) MiRGator v3.0: a microRNA portal for deep sequencing, expression profiling and mRNA targeting. *Nucleic Acids Res.*, **41**, D252–D257.

Dunin-Horkawicz,S. *et al.* (2006) MODOMICS: a database of RNA modification pathways. *Nucleic Acids Res.*, **34**, D145–D149.

Griffiths-Jones,S. *et al.* (2006) miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Res.*, **34**, D140–D144.

Guo,L. and Lu,Z. (2010) Global expression analysis of miRNA gene cluster and family based on isomiRs from deep sequencing data. *Comput. Biol. Chem.*, **34**, 165–171.

Li,S.C. *et al.* (2012) MicroRNA 3′ end nucleotide modification patterns and arm selection preference in liver tissues. *BMC Syst. Biol.*, **6**, S2–S14.

Llorens,F. *et al.* (2013) A highly expressed miR-101 isomiR is a functional silencing small RNA. *BMC Genomics*, **14**, 104.

Loher,P. *et al.* (2014) IsomiR expression profiles in human lymphoblastoid cell lines exhibit population and gender dependencies. *Oncotarget*, **5**, 8790–8802.

Millar,A.A. and Waterhouse,P.M. (2005) Plant and animal microRNAs: similarities and differences. *Funct. Integr. Genomics*, **5**, 129–135.

Neilsen,C.T. *et al.* (2012) IsomiRs - the overlooked repertoire in the dynamic microRNAome. *Trends Genet.*, **28**, 544–549.

Pantano,L. *et al.* (2010) SeqBuster, a bioinformatic tool for the processing and analysis of small RNAs datasets, reveals ubiquitous miRNA modifications in human embryonic cells. *Nucleic Acids Res.*, **38**, e34.

Sablok,G. *et al.* (2013) isomiRex: web-based identification of microRNAs, isomiR variations and differential expression using next-generation sequencing datasets. *FEBS Lett.*, **587**, 2629–2634.

Vaz,C. *et al.* (2013) Analysis of the microRNA transcriptome and expression of different isomiRs in human peripheral blood mononuclear cells. *BMC Res. Notes*, **6**, 390.

Wang,E.T. *et al.* (2008) Alternative isoform regulation in human tissue transcriptomes. *Nature*, **456**, 470–476.

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