DeAnnIso: a tool for online detection and annotation of isomiRs from small RNA sequencing data

Yuanwei Zhang1,2,†, Qiguang Zang2,†, Huan Zhang1,†, Rongjun Ban3, Yifan Yang4, Furhan Iqbal1,5, Ao Li3,6,* and Qinghua Shi1,2,*

1Molecular and Cell Genetics Laboratory, The CAS Key Laboratory of Innate Immunity and Chronic Disease, Hefei National Laboratory for Physical Sciences at Microscale and School of Life Sciences, University of Science and Technology of China, Hefei 230027, China, 2Hefei Institute of Physical Science, China Academy of Science, Hefei 230027, China, 3School of Information Science and Technology, University of Science and Technology of China, Hefei 230027, China, 4Department of statistics, University of Kentucky, Lexington, KY 40536, USA, 5Institute of Pure and Applied Biology, Bahauddin Zakariya University, Multan 60800, Pakistan and 6Research Centers for Biomedical Engineering, University of Science and Technology of China, Hefei 230027, China

Received March 01, 2016; Revised April 24, 2016; Accepted May 05, 2016

ABSTRACT

Small RNA (sRNA) Sequencing technology has revealed that microRNAs (miRNAs) are capable of exhibiting frequent variations from their canonical sequences, generating multiple variants: the isoforms of miRNAs (isomiRs). However, integrated tool to precisely detect and systematically annotate isomiRs from sRNA sequencing data is still in great demand. Here, we present an online tool, DeAnnIso (Detection and Annotation of IsomiRs from sRNA sequencing data). DeAnnIso can detect all the isomiRs in an uploaded sample, and can extract the differentially expressing isomiRs from paired or multiple samples. Once the isomiRs detection is accomplished, detailed annotation information, including isomiRs expression, isomiRs classification, SNPs in miRNAs and tissue specific isomiR expression are provided to users. Furthermore, DeAnnIso provides a comprehensive module of target analysis and enrichment analysis for the selected isomiRs. Taken together, DeAnnIso is convenient for users to screen for isomiRs of their interest and useful for further functional studies. The server is implemented in PHP + Perl + R and available to all users for free at: http://mcg.ustc.edu.cn/bsc/deanniso/ and http://mcg2.ustc.edu.cn/bsc/deanniso/.

INTRODUCTION

MicroRNAs (miRNAs), with average length of 22 nucleotides have emerged as key post-transcriptional regulators of gene expression in many developmental and cellular processes in eukaryotic organisms (1). Recent advances in high-throughput small RNA sequencing technology and algorithmic-based computational approaches reveal the complex and dynamic repertoire of miRNAs—a single miRNA locus can give rise to multiple distinct miRNA isoforms (isomiRs) that differ in their length and sequence composition (2). It has been demonstrated that isomiRs are bona fide miRNA variants, which were originally dismissed as experimental artifacts (3,4). It has also been documented that IsomiRs expression pattern may vary across different tissues, cell types and/or developmental stages (5).

Based on the sequences alignment with precursors, isomiRs can be divided into ‘templated’ and ‘non-templated’ isomiRs. ‘Templated’ isomiRs are most likely derived from variations during miRNA processing by Drosha and/or Dicer enzymes (6). ‘Non-templated’ isomiRs are mostly catalyzed by nucleotidyl transferases possessing 5′–3′ uridylyltransferase and/or adenyltransferase activity, thereby causing an abundance of non-templated nucleotide extensions at 3′ rather than at 5′ ends (7).

On the other hand, nucleotide variations detected in isomiRs may originate from single nucleotide polymorphisms (SNPs) in miRNA genes (8), or they may be resulted from enzymatic modification like A–I editing, that represents the most general type of miRNA editing (9).

Regarding to the specific expression pattern and the heterogeneous nature of isomiRs, they have been documented with differential functionality (2). The 5′ isomiRs have different seed regions as compared to their canonical miRNAs, therefore it can strongly affect the target selection, which is of great functional and evolutionary importance (10,11).
The 3' uridylation and adenylation isomiRs may possess opposing roles in plants by promoting the degradation and stability of isomiRs (12,13). miRNA editing in the seed region may change the targeting profile, for example, A-to-I editing of has-miR-376 within the seed region generates a distinct set of predicted targets including PRPS1, and contributes to tissue-specific regulation of uric-acid levels (14). Studies indicating differential functionality for isomiRs are currently confined to several specific variants, and there are much more isomiRs with biological significance to be fully resolved.

Several tools, such as SeqBuster (15), IsomiRex (16), miRanalyzer (17), isomiRID (18), IsomiRage (19) and CPSS (20) have been designed for the detection of miRNA modifications from sRNA sequencing data. However, these tools mainly focus on the isomiR detection, and do not provide comprehensive annotation information of the detected isomiR. Hence, integrated tools are required to precisely detect and systematically annotate isomiRs from sRNA sequencing data. To meet this demand, we are presenting an online tool, DeAnnIso (Detection and Annotation of IsomiRs) to handle both.

After that, BLAST was used to match precursor sequence information to reference genome. Aligned reads had no more than 3 mismatches (0–3, default is 2) in the first ‘L’ bases (≥5, default is 10) of the left end. This was followed by BLAST application to match these genome mapped reads with precursor sequences downloaded from miRBase (version 21) (21). After that, BLAST was used to match precursor sequence matched reads with mature miRNA downloaded from miRBase (version 21). Default E-value for BLAST was set to 0.01. Finally, all the detected isomiRs were aligned with their canonical miRNAs and the variant nucleotides were then extracted.

**WEB SERVER CONSTRUCTION**

DeAnnIso consists of three modules: (i) detection of isomiRs; (ii) annotation of isomiRs; (iii) target analysis.

**Detection of isomiRs**

Bowtie and BLAST were used for sequence alignment. Initially, the ‘–n’ mode of Bowtie was used for mapping reads into reference genome. Aligned reads had no more than ‘N’ mismatches (0–3, default is 2) in the first ‘L’ bases (≥5, default is 10) of the left end. This was followed by BLAST application to match these genome mapped reads with precursor sequences downloaded from miRBase (version 21) (21).

**Annotation of isomiR**

(i) **IsomiRsexpression.** The expression level of each isomiR detected in a sample was normalized by reads per million (RPM) with the following formula: RPM_{isomiR} = (N_{isomiR}/N_{all}) \times 10^6, where N_{isomiR} was the number of reads mapped to the isomiR and N_{all} was total number of reads mapped in the sample. The expression level of a mature miRNA detected in each sample was normalized with the following formula: RPM_{mature} = \sum RPM_{isomiR} + RPM_{canonical}. For paired samples, the P-values were quantitatively measured by Fisher’s exact test and Bonferroni correction was performed when an adjustment was made to P-values.

(ii) **IsomiRclassification.** IsomiRs were primarily classified into three categories: 5′ isomiRs, 3′ isomiRs and isomiRs with internal modifications (2). Both 5′ and 3′ isomiRs had two subsets: isomiRs with addition (nucleotides adding), and isomiRs with trimming (nucleotides trimming). For isomiRs of ‘Addition’ type, if the isomiRs with the added nucleotides could perfectly align with the precursor miRNA, they were defined as ‘templated (T)’ types, otherwise they were defined as ‘non-templated (NT)’ type. IsomiRs of ‘NT’ type were ultimately divided into A, G, T and C types according to the first added nucleotide (19). IsomiRs with internal modifications were classified into two categories: internal modifications with seed shifting (5′ isomiRs with novel seed sequences that resulted from nucleotides adding or trimming) (22) and internal modifications without seed shifting.

(iii) **SNPs.** The variations found in isomiRs were extracted and matched with the dbSNP (23). Variations that had already been recorded as a genome SNP were displayed with the SNP information on the results page.

(iv) **Tissue specificisomiR expression.** Tissue specific isomiR expression was annotated with an in-house data set that contained 308 919 isomiRs detected from 2727 samples (sRNA sequencing data downloaded from ArrayExpress) of 8 species (Arabidopsis thaliana, Drosophila melanogaster, Danio rerio, Homo sapiens, Mus musculus, Oryza sativa, Solanum lycopersicum and Zea mays).

**Target analysis**

Target analysis consisted of three steps: (i) Prediction of isomiR targets; (ii) Scatter plot analysis of differences on target selection between isomiRs; (iii) Enrichment analysis.

First, miRanda (24) or RNA hybrid (25) was used to predict the target gene for each isomiR. The minimum free energy (MFE) between isomiR and its target gene was also calculated. Second, affected targets were selected based on MFE fold change. For example, isomiR ‘A’ and isomiR ‘B’ could both target gene ‘C’, and if the MFE of ‘AC’ was at...
least twice as that of ‘BC’, the target gene ‘C’ was considered as isomiR-affected gene (target). Scatter plot was used to illustrate the differences on target gene selection, in which isomiR-affected targets were supposed to deviate from the diagonal. Finally, the enrichment analysis was carried out for isomiR-affected targets. The Gene Ontology, Pathway and Protein Domain information used for enrichment analysis were retrieved from DAVID (26). The enrichment \( P \)-values were quantitatively measured by Fisher’s exact test and Bonferroni correction was calculated when an adjustment was made to \( P \)-values.

**UTILITY AND WEB INTERFACE**

**Preparation of files to be uploaded**

Reads files in FASTA format (without adapter sequence) are required as input. Users can use FASTX toolkit ([http://hannonlab.cshl.edu/fastx_toolkit/](http://hannonlab.cshl.edu/fastx_toolkit/)) to convert their data from FASTQ to FASTA format. In order to minimize the size of uploaded files, we recommend users to upload tag files, which will combine the identical reads into unique tags. To further minimize the size of uploading files, compression of the reads or tag file into *tar.gz* format file is recommended. Users can also use the tool (PreprocessFiles) provided by our server to conveniently generate input files from raw sequencing data. Detailed requirements for input files can be found at [http://mcg.ustc.edu.cn/bsc/deanniso/tools.html](http://mcg.ustc.edu.cn/bsc/deanniso/tools.html).

**Data analysis**

Currently, DeAnnIso is able to handle the data from single, paired and group samples, and can complete the assigned job within 1 h, depending on data size, species and selected parameters for analysis. Data from eight species (*Arabidopsis thaliana, Drosophila melanogaster, Danio rerio, Homo sapiens, Mus musculus, Oryza sativa, Solanum lycopersicum* and *Zea mays*) are supported by the current version of DeAnnISO. Bowtie –n mode and BLAST have default analysis parameters. All the analysis parameters can be modified upon request of professional users (Figure 1A). Parameter settings are described in detail in ‘Detection of isomiRs’ section of web server construction. After setting the parameters, users can click the submit button to initiate the isomiR detection job. Status bar will appear at the top of the web page, representing the progress of the job and will refresh automatically. Meanwhile, statistical results of the completed step will be displayed in a graphical form, providing a comprehensive overview of the detection results (Figure 1B).

To perform the differential expression analysis of multiple samples, we have also designed a separate group analysis module, which requires the expression table files of isomiRs from each single case. Users can upload single case expression table file, or alternatively input a single case analysis job ID and the web server will retrieve the corresponding expression table file automatically. We have applied Wilcoxon Rank-sum test to infer the statistically significant differences for a specific experimental group. In all conditions, one isomiR is considered to be differentially expressed with \( P \)-value < 0.01 and fold change > 2 by default.

In group analysis results, the differentially expressing isomiRs between two groups are listed. The annotation information contains the expression value of isomiRs in each sample, the median expression value of the group, fold of median expression and \( P \)-value. All these components are well organized with suitable examples to facilitate users for choosing correct input and to analyze the results.

**Results and description**

Once the isomiR detection job is accomplished, the results page will display the parameters for the detection of isomiRs, and will provide general and detailed results.

**General results.** Preliminary results will be displayed in a graphical form, including: (i) Genome mapping results (Figure 2A); (ii) Length distribution of genome mapped reads (Figure 2B); (iii) Chromosome distribution of genome mapped reads (Figure 2C); (iv) Statistics of miRNA family match results (Figure 2D); (v) Percentage of mature miRNAs that are processed from 5′ or 3′ arms of the hairpin precursor (Figure 2E); (vi) Nucleotide addition/trimming at 5′/3′ ends (Figure 2F); (vii) Comparison of addition/trimming events between sample A and B (Figure 2G); (viii) Templated/non-templated nucleotide(s) addition at 5′ or 3′ ends (Figure 2H); (ix) Comparison of templated/non-templated nucleotide(s) addition at 5′ or 3′ ends between sample A and B (Figure 2I); (x) Non-templated addition type (Figure 2J); (xi) Comparison of non-templated addition between sample A and B (Figure 2K); (xii) Internal modification without seed shifting (Figure 2L); (xiii) Comparison of internal modification without seed shifting between sample A and B (Figure 2M); (xiv) Internal modification with seed shifting (Figure 2N); (xv) Comparison of internal modification with seed shifting between sample A and B (Figure 2O). Each statistical graph can be easily accessed by clicking the items in the ‘Results Index’ box, and can be saved in desired format by clicking the button at the upper right of each graph. The detailed statistical table will be displayed, upon clicking the corresponding part in the graphs.

**Detailed results.** By clicking the ‘Check Detailed Results’ button at the bottom of the page, detailed annotation of isomiRs is displayed in the ‘Detailed results’ page. The annotation information includes: sample source (from which sample the isomiR is detected), canonical miRNA information (miRNA family, stem loop and mature ID), isomiR sequence, classification (as described in the ‘isomiR classification’ section), expression (absolute reads count, ratio – \( P \)-value. All these components are transmitted to the client browser. The list of detailed results
Figure 1. Input page of DeAnnIso.

can be saved by clicking download button on the top of the page.

For more details of the results, users can click the mature ID in the list to view 'detailed information for isomiRs belonging to one mature miRNA' in a new page (Figure 4A). Similar to the 'General results' page, most of the results are displayed in a graphical form, including: (i) Percentage of mature miRNAs that are processed from 5’ or 3’ arms of the hairpin precursor; (ii) Nucleotide addition or trimming at 5’ or 3’ ends; (iii) Comparison of addition/trimming events between sample A and B; (iv) Templated or non-templated nucleotide addition at 5’ or 3’ ends; (v) Comparison of templated/non-templated nucleotide addition at 5’ or 3’ ends between sample A and B; (vi) Non-templated addition type; (vii) Comparison of non-templated addition type between sample A and B; (viii) Internal modification without seed shifting; (ix) Comparison of internal modification without seed shifting between sample A and B; (x) Internal modification with seed shifting and (xi) Comparison of internal modification with seed shifting between sample A and B. The alignment results of all isomiRs belonging to one mature miRNA is available when users click the mature ID (Figure 4B). Users can select or neglect isomiRs of certain types by clicking the ‘+’ buttons on the ‘Alignment’ page (Figure 4B). For each aligned isomiR, RPM values of two samples in different colors are displayed next to the isomiR sequence. Users can click the arrowhead button for more annotation information, including isomiR types and tissue information.

Target analysis. Users can select isomiRs of their interest (range from 2 to 8) on the ‘Alignment’ page for target analysis directly, with desired target prediction tool (miRanda or RNAhybrid) and the threshold (MFE fold change, described in ‘target analysis’ section of web construction) as shown in Figure 4B. By clicking on the ‘start’ button, target analysis will start, with status bar refreshing automatically and showing the progress of job (Supplementary Figure S1A). When the analysis is done, a table will be displayed showing MFE between each isomiR and their predicted targets, and this results can be saved by clicking download button just behind the title of ‘Predicted isomiR targets’ (Supplementary Figure S1B). Next is the scatter plot(s) to illustrate the effects of isomiRs on targets selection (described in ‘target analysis’ section of web construction) (Supplementary Figure S1B).

User can also upload more isomiRs sequences (range from 2 to 50) in *.fasta or .tar.gz file on ‘Target Analysis’ page to perform target analysis separately. In this tab, users should select species for the uploaded samples (Supplementary Figure S1C), while the options are identical to those on the ‘Alignment page’ (Supplementary Figure S1B).

Moreover, functional enrichment analysis for these affected targets is also available to users. By selecting the isomiR sequences (two isomiRs for one analysis), Annotation Categories (Gene Ontology, Pathway, Protein Domains) and subsets of each category from the drop-down box, users can overview the effect of isomiR-affected targets on downstream biological processes, pathways or protein domains. The enrichment fold, P-values (Fisher’s exact test) and Bonferroni adjusted P-values can be optimized to refine the results (Supplementary Figure S1D). Users can obtain the detailed information for each enriched term by clicking the term, and will be redirected to the source providing this annotation information in a new page. All enrichment results can be saved by clicking download button in the footer navigation bar of ‘Enrichment analysis’ page.

Comparison with other isomiR detection tools

Although there are several bioinformatics tools existed to detect isomiRs from sRNA sequencing data with different emphases, DeAnnIso is highly competitive in terms of its broad functionality (Supplementary Table S1).

(i) DeAnnIso provides a comprehensive overview of various visualized statistics results derived from detected isomiRs, which no other web-based tools do. This information is very much important for users to understand the general characteristics within the samples and to discover the global change of certain isomiR types. (ii) Compared to
Figure 2. General results page.
Figure 3. Detailed results page.
Figure 4. Detailed information for isomiRs belonging to hsa-miR-143-3p.
the classification criterion (5' isomiRs, 3' isomiRs and substitutions in the middle) used in most existing tools, DeAnnIso further classify isomiRs into more sophisticated categories to encompass the diversity of isomiRs with sequence and length variations, which enables deeper understanding of isomiRs complexity. (iii) Most of the existing tools can only analyze the differential expressed mature miRNAs or precursors, whereas DeAnnIso can detect both differential expressed miRNAs and isomiRs from paired or group samples, according to the parameters set by the users. (iv) DeAnnIso provides the tissue specific isomiR expression information by an in-house data set mentioned before, which is very convenient for users to search for certain isomiRs expressed in a tissue-or context-dependent manner. (v) DeAnnIso also matches substitutions in isomiRs with dbSNP, which only CPSS can do. (vi) For potential function isomiRs screening, DeAnnIso allows users to set the parameters in the ‘Detailed results’ page to filter out certain isomiRs based on expression or classification. Users can get insights into isomiRs of their interest by selecting them for target prediction and enrichment analysis. All these integrated function modules are customized for users to easily search for isomiRs of biological significance, which no existed tools can fully cover.

**CASE STUDY AND DISCUSSION**

*Spo11* is essential for double strand break (DSB) formation during meiotic chromosome synapsis. Disruption of *Spo11* in mice can cause infertility (27). According to the comparison of our unpublished transcriptome data from the testis of *Spo11*<sup>−/−</sup> and wild type mice, several genes that are directly involved in meiotic recombination (28) had differential expression (Supplementary Table S2). Limited studies are available in literature that have focused on the post-transcriptional regulation of these genes during meiosis. We wondered whether isomiRs were involved in the formation of DSB and tried to use DeAnnIso to analyze sRNA sequencing data from the testis of *Spo11*<sup>−/−</sup> and wild type mice. After uploading the *.fa* files to DeAnnIso server, isomiRs detection was performed automatically and overview of results are available at http://mcg.ustc.edu.cn/bsc/deanniso/status/paired.php?job_id = 699438278459. The most abundant isomiR observed in both samples was the 3' isomiR (Supplementary Figure S2), which in agreement with former studies (15,29–31). We observed that there were more isomiRs with either nucleotide addition at 3' ends or internal modification found in *Spo11*<sup>−/−</sup> mice testis as compared to wild type testis (Supplementary Figure S3).

Several enzymes involved in the biogenesis of different isomiR types have been reported: *Papd4*, *Papd5*, *Mtpap*, *Tut1* and *Papolgc* catalyzed 3' adenlylation of isomiRs; *Zcchc11*, *Zcchc6* catalyzed 3' uridylation of isomiRs; *Adar* catalyzed A to I editing, resulting in A to G substitution (2,5,9). Interestingly, most of these enzymes were highly expressed in *Spo11*<sup>−/−</sup> mice according to transcriptome sequencing data (Supplementary Table S3). Taken together, these results indicated that the isomiRs detected and annotated by DeAnnIso are reliable, and might have a physiological relevance in DSB formation. Therefore, we tried to screen for isomiRs of these three types: 3' addition of A, 3' addition of T and A to G substitution at seed region. By setting the parameters in ‘Detailed results’ page, isomiRs of certain types could be easily found out (Supplementary Figure S4, Supplementary Table S4). Detailed information of isomiRs can be accessed at http://mcg.ustc.edu.cn/bsc/deanniso/rpm_for_mature_paired.php?job_id = 699438278459. The expression of selected isomiRs in *Spo11*<sup>−/−</sup> and wild type mice testis was confirmed experimentally by using stem-loop polymerase chain reaction (PCR) (Figure 5), as a previously report (32).

Since isomiRs with nucleotide substitution at seed region may have more severe influence on target selection (14), we selected iso-miR-30a-5p with A to G substitution for further target analysis. At ‘Alignment’ results page, we selected this isomiR and canonical miR-30a-5p sequences, and performed target analysis...
with default parameters. Detailed results can be accessed at http://mcg.ustc.edu.cn/bse/deanniso/enrichment_status.php?rand_num_enrichment = 4881692432. The scatter plot analysis showed targets selection have changed very substantially between this isomiR and canonical miRNA (Supplementary Figure S5). Several GO terms, related to post-transcriptional regulation of gene expression (GO:0010608) and regulation of RNA splicing (GO:0048024 and GO:0043484) were enriched (Supplementary Table S5). Notably, microtubulin binding (GO:0008017) was enriched. One recent study has revealed that mammalian cells use microtubules in the cytoplasm to promote the mobility of chromatin surrounding DSBs in the nucleus and contributes to DSB repair (33). One target in this GO term, DNA polymerase β (Polb) might be an explanation for downregulated expression of iso-miR-30a-5p rather than the canonical miRNA (Supplementary Table S6). Therefore, higher expression of iso-miR-30a-5p could only be targeted by iso-miR-30a-5p rather than the canonical miRNA (Supplementary Table S6). Therefore, higher expression of iso-miR-30a-5p might be an explanation for downregulated expression of Polb in Spo11−/− mice tests, as compared with wild type mice tests (Supplementary Table S2). Further functional study is needed to illustrate the roles of isomiRs in DSB formation during meiosis.

This case study indicates that DeAnnIso can not only precisely detect isomiRs from sRNA sequencing data, but also provide systematic annotation information for users to search for their interested isomiRs. Furthermore, users can get further understanding of isomiRs’ biological functionality with target analysis and enrichment analysis.

CONCLUSION
In conclusion, we have described the DeAnnIso web server, a web-based tool that is capable of precisely detection and systematic annotation of isomiRs from sRNA sequencing data. With the detailed annotation information, users can conveniently search for isomiRs of interest. Furthermore, the module of target analysis and enrichment analysis help users to reveal isomiRs’ roles in certain biology process.

SUPPLEMENTARY DATA
Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENT
This work was supported by the Joint NSFC-ISF Research Program, jointly funded by the National Natural Science Foundation of China and the Israel Science Foundation. Author contribution: YW Zhang conceived and designed the project. QG Zang and YW Zhang constructed the pipeline. RJ Ban developed the web server. H Zhang performed the experiments, analyzed the data. YW Zhang and H Zhang wrote the paper. YF Yang tested the web server. YF Yang, Furhan Iqbal, A Li and QH Shi revised the manuscript.

FUNDING
National Basic Research Program of China (973) [2012CB944402, 2014CB943101 and 2013CB945502]; National Natural Science Foundation of China [31461143013, 31371519, 31301227, 31501202 and 61571414]; Major Program of Development Foundation of Hefei Center for Physical Science and Technology [2014FXZY003]; Fundamental Research Funds for the Central Universities [WK2340000069]. Funding for open access charge: National Basic Research Program of China (973) [2012CB944402, 2014CB943101 and 2013CB945502]; National Natural Science Foundation of China [31461143013, 31371519, 31301227, 31501202 and 61571414]; Major Program of Development Foundation of Hefei Center for Physical Science and Technology [2014FXZY003]; Fundamental Research Funds for the Central Universities [WK2340000069].

Conflict of interest statement. None declared.

REFERENCES


