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

Yuanwei Zhang[1,2,†], Qiguang Zang[3,†], Huan Zhang[1,†], Rongjun Ban[3], Yifan Yang[4], Furhan Iqbal[1,5], Ao Li[3,6,*] and Qinghua Shi[1,2,*]

[1]Molecular 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, [2]Hefei Institute of Physical Science, China Academy of Science, Hefei 230027, China, [3]School of Information Science and Technology, University of Science and Technology of China, Hefei 230027, China, [4]Department of statistics, University of Kentucky, Lexington, KY 40536, USA, [5]Institute of Pure and Applied Biology, Bahauddin Zakariya University, Multan 60800, Pakistan and [6]Research Centers for Biomedical Engineering, University of Science and Technology of China, Hefei 230027, China

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′ uridyltransferase 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).

*To whom correspondence should be addressed. Tel: +86 551 63600344; Fax: +86 551 63600344; Email: aoli@ustc.edu.cn
Correspondence may also be addressed to Qinghua Shi. Tel: +86 551 63600344; Fax: +86 551 63600344; Email: qshi@ustc.edu.cn
†These authors contributed equally to this work as the first authors.

The 3′ uridylation and adenlylation 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 from sRNA sequencing data). It is implemented on our previously published algorithm (20), and integrates a systematic annotation pipeline. Upon submission of the *.fa files, DeAnnIso can detect isomiRs in each uploaded sample and can extract the isomiRs having with differential expression between paired samples. Once the isomiRs detection is accomplished, DeAnnIso will provide additional systematic annotation information to users, including: (i) IsomiRs expression; (ii) IsomiRs classification; (iii) SNPs in isomiRs and (iv) Tissue specific isomiR expression. Furthermore, DeAnnIso has the module of isomiR target prediction and enrichment analysis to help users further investigate the functionality of isomiRs.

In order to test the practicability of our tool, we apply DeAnnIso to analyze the sRNA sequencing data generated from the testis of *Spo11*$^{-/-}$ and wild type mice. We find iso-miR-30a-5p with A to G substitution in seed region, specifically targeting *Polb*, contrary to its canonical miRNA. This observation is probably related to the reproductive phenotype of *Spo11*$^{-/-}$ mice. Our result indicates that DeAnnIso can not only precisely detect isomiRs from sRNA sequencing data, but also provide systematic annotation information and target analysis that help users in searching isomiRs of potential biological significance.

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 ($\geq$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 miR-Base (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.

Annotation of isomiR

(i) ***IsomiRs expression***. The expression level of each isomiR detected in a sample was normalized by reads per million (RPM) with the following formula: $\text{RPM}_{\text{isomiR}} = (N_{\text{isomiR}}/N_{\text{all}}) \times 10^6$, 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: $\text{RPM}_{\text{mature}} = \sum \text{RPM}_{\text{isomiR}} + \text{RPM}_{\text{canonical}}$. 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) ***IsomiRs classification***. 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 specific isomiR 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/) 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.

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 requirement 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 – calculated by $RPM_{isomiR}/RPM_{mature}$, RPM of isomiR and RPM of mature), difference test for an isomiR or mature miRNA (fold change, *P*-value and Bonferroni adjusted *P*-value) (Figure 3A and B). Users can easily filter isomiRs of their interests by setting the parameters on the top of the page (Figure 3C). For a quick search for specific isomiRs, users can input the miRNA family name, stem loop, mature ID or sequence in the 'search' box above the list. In order to reduce the course of users' computing resources, the filter process is performed on server and the results are transmitted to the client browser. The list of detailed results

A

B

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 tar-

gets, 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
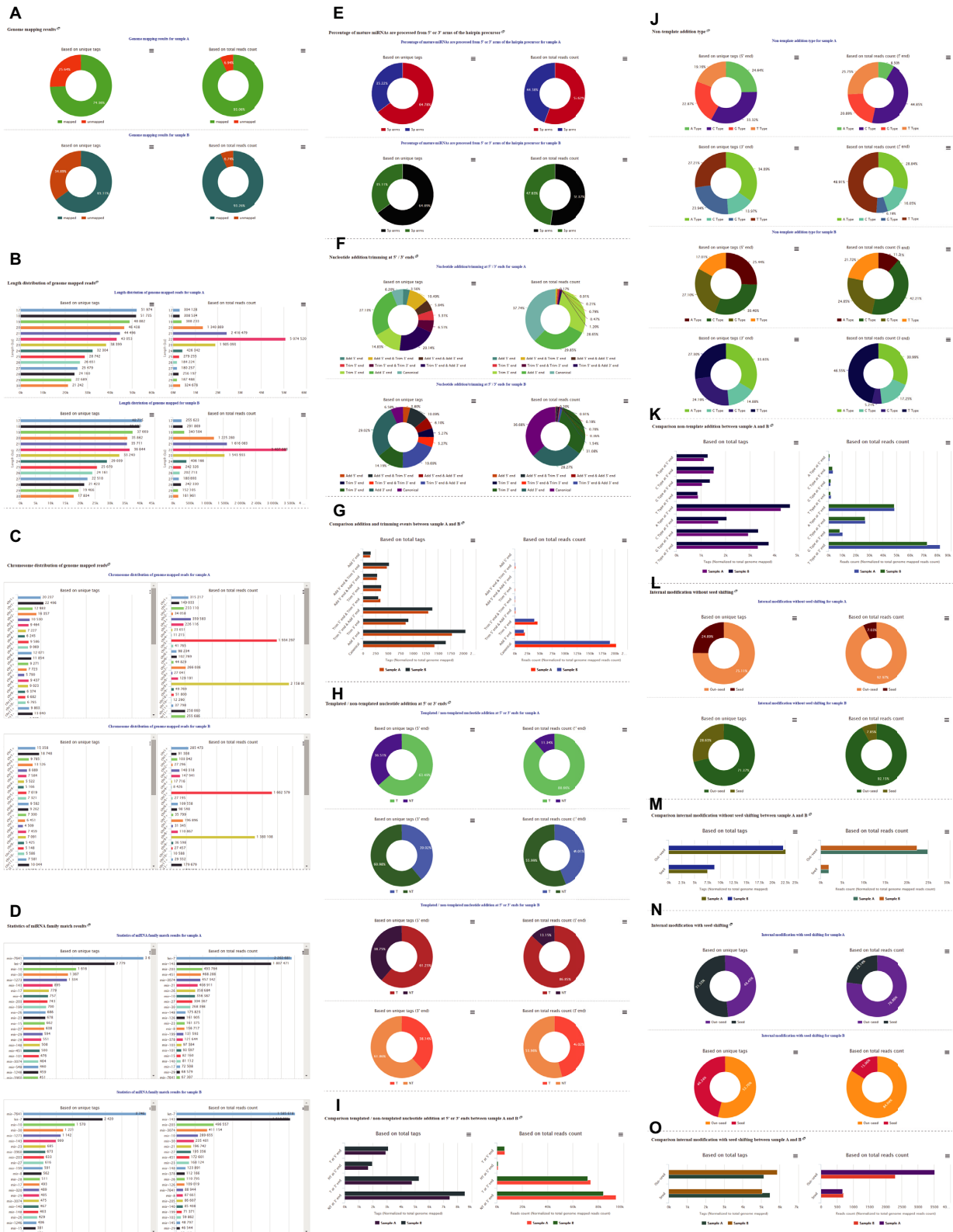
Figure 2. General results page.

A

Show [10 ▼] entries

Search: [　　　　　　]

Sample source	Canonical			Sequence	Classfication										
	miRNA family	Stem loop	Mature		5' end				3' end				Internal		
					Add/Trim/N	Nucleotides	T/NT	Type	Add/Trim/N	Nucleotides	T/NT	Type	Seed/Out-seed/N	Site	
AB	mir-143	hsa-mir-143	hsa-miR-143-3p	TGAGATGAAGCACTGTAGCT	N	-	-	-	Trim	C	-	-	Out-seed	0	C->T
AB	let-7	hsa-let-7a-3	hsa-let-7a-5p	TGAGGTAGTAGGTTGTATAGTT	N	-	-	-	N	-	-	-	N	-	-
AB	let-7	hsa-let-7c	hsa-let-7c-5p	TGAGGTAGTAGGTTGTATGGTT	N	-	-	-	N	-	-	-	N	-	-
AB	mir-378	hsa-mir-378a	hsa-miR-378a-3p	ACTGGACTTGGAGTCAGAAGGC	N	-	-	-	N	-	-	-	N	-	-
AB	let-7	hsa-let-7b	hsa-let-7b-5p	TGAGGTAGTAGGTTGTGTGGTT	N	-	-	-	N	-	-	-	N	-	-
AB	mir-30	hsa-mir-30d	hsa-miR-30d-5p	TGTAAACATCCCCGACTGGAAGC	N	-	-	-	Add	C	T	C	Out-seed	0	G->C
AB	mir-30	hsa-mir-30a	hsa-miR-30a-5p	TGTAAACATCCTCGACTGGAAGC	N	-	-	-	Add	C	T	C	Out-seed	0	G->C
AB	mir-143	hsa-mir-143	hsa-miR-143-3p	TGAGATGAAGCACTGTAGC	N	-	-	-	Trim	TC	-	-	N	-	-
AB	let-7	hsa-let-7b	hsa-let-7b-5p	TGAGGTAGTAGGTTGTGTGGT	N	-	-	-	Trim	T	-	-	N	-	-
AB	mir-23	hsa-mir-23a	hsa-miR-23a-3p	ATCACATTGCCAGGGATTTCC	N	-	-	-	N	-	-	-	N	-	-

Showing 1 to 10 of 1,484 entries

Previous [1] 2 3 4 5 … 149 [Next]

B

Expression								Difference test				
Sample A				Sample B				For isomiR				
Reads count	Ratio	RPM of isomiR	RPM of mature miRNA	Reads count	Ratio	RPM of isomiR	RPM of mature miRNA	Fold change	p value	Bonferroni	Fold change	
630404	0.34917517305	46434.1638961	132982.432544	730121	0.481520587859	71101.2261079	147659.784235	1.53122658280162	0	0	1.11038808796302	
575740	0.622677424239	42407.734598	68105.4635148	523583	0.684145880945	50987.977704	74527.9320159	1.20232731569644	0	0	1.24643556988276	8.175
75180	0.595141027366	5537.59246722	9304.67269536	88534	0.64598367055	8621.68866836	13346.6046611	1.55693809528094	0	0	1.43441121457389	
59152	0.487952155084	4357.00544854	8929.16529448	56964	0.508984336605	5547.31372472	10898.790642	1.27319412156833	0	0	1.2206018305472	
58103	0.358096822903	4279.73842941	11951.3443172	82014	0.382606505066	7986.75282318	20874.5871213	1.8661777944861	0	0	1.74662797123938	
41147	0.503007261436	3030.79698389	6025.35433631	42694	0.605588652482	4157.66119239	6865.48728307	1.37180458291641	0	0	1.13943295279719	1.133
40037	0.607587829122	2949.03683972	4853.67990992	49700	0.693378721504	4839.92507757	6980.20421952	1.64118840849319	0	0	1.43810613107287	
38125	0.0211170986741	2808.20314995	132982.432544	37394	0.0246616394575	3641.53236118	147659.784235	1.29674819332447	4.90437654258499e-279	2.1281561131239e-274	1.11038808796302	
37587	0.231653878155	2768.57526025	11951.3443172	50815	0.237058911343	4948.50689772	20874.5871213	1.78738391864232	0	0	1.74662797123938	
35264	0.351946665071	2597.4682198	7380.2893381	35373	0.355986957309	3444.72172573	9676.53914	1.32618435886155	6.12927546703005e-308	2.65967650340835e-303	1.3111458471069	

Showing 1 to 10 of 1,484 entries

Previous [1] 2 3 4 5 … 149 [Next]

C

Detailed results ⊕ ⧉

Sample source	Sample AB ▼	Reads count (≥) [3]	Ratio (≥) [0]
Difference test	IsomiR	⦿ p value ○ Bonferroni (<) [0.1]	Fold change [≥ ▼] [1]
	mature miRNA	⦿ p value ○ Bonferroni (<) [0.1]	Fold change [≥ ▼] [1]
Type	5' end [Add ▼] Nucleotides [All selected ▼]	3' end [Add ▼] Nucleotides [All selected ▼]	
	Internal [All ▼] Substitution [All selected ▼]		

[Filter]

Figure 3. Detailed results page.

A

Detailed information for isomiRs belonging to **hsa-miR-143-3p**

miRNA family	mir-143
Stem loop	hsa-mir-143
Mature	hsa-miR-143-3p Related mature: hsa-miR-143-5p (Click to mature name to view alignment results)

Show 10 entries   Search:

Sample source	Sequence	5' end				3' end				Internal		RPM	Fold change
		Trim/Add/N	Nucleotides	T/NT	Type	Trim/Add/N	Nucleotides	T/NT	Type	Seed/Out-seed/N	Site		
A	GGAGATGAAGCACTGTAAA	N	-	-	-	Trim	TC	-	-	Out-seed	1|18|19	0.0736577875396 0	-
A	TGAGATGAAGCACTGAGAAAA	N	-	-	-	N	-	-	-	Out-seed	16|17|18|19|20|21	0.0736577875396 0	-
A	GGGATGAAGCACTGTAGCTCA	Trim	T	-	-	Add	A	T	A	Seed	3	0.0736577875396 0	-
A	GATGAAGCACTCTAGCTC	Trim	TGA	-	-	N	-	-	-	Out-seed	12	0.0736577875396 0	-
A	TGAGAAGAAGCACTGTAAA	N	-	-	-	Trim	TC	-	-	Seed	6|18|19	0.147315575079 0	-
A	TGAGATGAAGCACTGTGC	N	-	-	-	Trim	CTC	-	-	Out-seed	17|18	0.0736577875396 0	-
A	TGAGATGAAGCACTGGAGAAA	N	-	-	-	N	-	-	-	Out-seed	16|19|20|21	0.0736577875396 0	-
A	TGAGATGAAGCACTGAAGAA	N	-	-	-	Trim	C	-	-	Out-seed	16|19|20	0.147315575079 0	-
A	AGATGAAGCACTGGAGCTCT	Trim	TG	-	-	Add	T	NT	T	Out-seed	14	0.0736577875396 0	-
A	TGTGATGAAGCACTGTAA	N	-	-	-	Trim	CTC	-	-	Seed	3|18	0.147315575079 0	-

Showing 1 to 10 of 1,158 entries   Previous **1** 2 3 4 5 ... 116 Next

---

B

Alignment

➕ hsa-miR-143-5p ➕ hsa-miR-143-3p

5' end   Add (➕ A type ➕ C type ➕ G type ➕ T type) ➕ Trim   ➕ Internal   ➕ Canonical
3' end   Add (➕ A type ➕ C type ➕ G type ➕ T type) ➕ Trim

🔘 miRanda ⚪ RNAhybrid   Minimum free energy (MFE) fold change: 2   [Start] Could not select more than 8 sequences ✕

Sample A: ACGT   Sample B: ACGT   Sample AB: ACGT

```
>>hsa-mir-143
GCGCAGCGCCCTGTCTCCCAGCCTGAGGTGCAGTGCTGCATCTCTGGTCAGTTGGGAGTCTGAGATGAAGCACTGTAGCTCAGGAAGAGAGAAGTTGTTCTGCAGC hsa-mir-143
(((((((.((.....((((.(..(((((((.(((((((((((.((((((.(((((.....))).)) .)))))))).))))))))).))))))..).)))).....)).)))).)) -51.7999992371
*************************GGTGCAGTGCTGCATCTCTGGT*********************************************** hsa-miR-143-5p
*******************************************TGAGATGAAGCACTGTAGCTC*********************** hsa-miR-143-3p
```

☐	----------------------CAGAGGTGCAGTGCTGCATCTCTGG----------------------------------	--- 0	| 0.194765596683	➤
☐	-----------------GGTGCACTGCTGCATCTCTGG---	--- 0.0736577875396	| 0	➤
☐	--------------GGTGCAGTGCTGCATCTCTGG----------------------------	--- 98.1121730027	| 126.500255046	➤
☐	---------------- **More Information** ✕	--- 0.0736577875396	| 0	➤
☐	------------------	--- 0.0736577875396	| 0	➤
☐	------------------ **IsomiR Form:** 0 |T->G	--- 0	| 0.0973827983415	➤
☐	------------------ **Type:** |SNP	--- 0	| 0.0973827983415	➤
☐	------------------ **Tissue:** embryo|cervix|testis|kidney|whole	--- 0.147315575079	| 0.0973827983415	➤
☐	------------------ body|blastocyst|tonsil|liver|bladder|endometrium|brain|fibroblast|co...	--- 0	| 0.0973827983415	➤
☐	------------------	--- 0.368288937698	| 0.194765596683	➤
☐	------------------	--- 0.0736577875396	| 0	➤
☐	------------------ [Close]	--- 0.147315575079	| 0	➤
☐	------------------	--- 0.147315575079	| 0.194765596683	➤

Figure 4. Detailed information for isomiRs belonging to a mature miRNA.
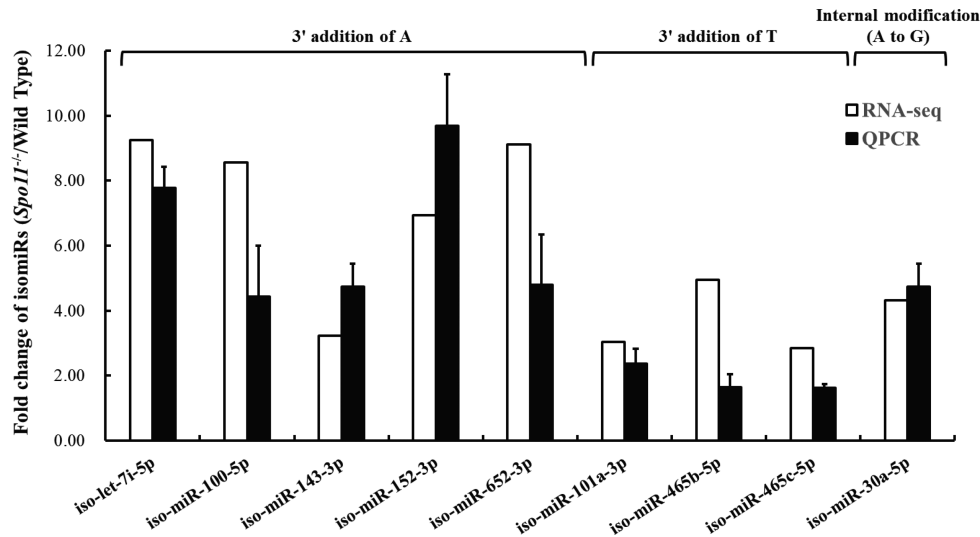
Figure 5. The expression of selected isomiRs in *Spo11*$^{-/-}$ and wild type mice testis was confirmed by stem-loop polymerase chain reaction (PCR).

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 analysis 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 convienient 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 classfication. Users can get insights into isomiRs of their interest by selecting them for target prediction and enrichment analysis. All theses 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*$^{-/-}$ 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*$^{-/-}$ 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 is 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*$^{-/-}$ 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 *Papolg* catalyzed 3′ adenylation of isomiRs; *Zcchc11*, *Zcchc6* catalyzed 3′ uridylation of isomiRs; *Adar* catalyzed A to I editing, resulting in A to G subsition (2,5,9). Interestingly, most of these enzymes were highly expressed in *Spo11*$^{-/-}$ mice testis 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*$^{-/-}$ 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 at seed region 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/bsc/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*) is critical for the repair of SPO11-induced DSBs (34). According to target analysis results, *Polb* 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 testis, as compared with wild type mice testis (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 systematica 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

FUNDING

REFERENCES

1. Krol,J., Loedige,I. and Filipowicz,W. (2010) The widespread regulation of microRNA biogenesis, function and decay. *Nat. Rev. Genet.*, **11**, 597–610.
2. Neilsen,C.T., Goodall,G.J. and Bracken,C.P. (2012) IsomiRs–the overlooked repertoire in the dynamic microRNAome. *Trends Genet.*, **28**, 544–549.
3. Sdassi,N., Silveri,L., Laubier,J., Tilly,G., Costa,J., Layani,S., Vilotte,J.L. and Le Provost,F. (2009) Identification and characterization of new miRNAs cloned from normal mouse mammary gland. *BMC Genomics*, **10**, 149.
4. Reese,T.A., Xia,J., Johnson,L.S., Zhou,X., Zhang,W. and Virgin,H.W. (2010) Identification of novel microRNA-like molecules generated from herpesvirus and host tRNA transcripts. *J Virol.*, **84**, 10344–10353.
5. BurroughsWyman,S.K., Knouf,E.C., Parkin,R.K., Fritz,B.R., Lin,D.W., Dennis,L.M., Krouse,M.A., Webster,P.J. and Tewari,M. (2011) Post-transcriptional generation of miRNA variants by multiple nucleotidyl transferases contributes to miRNA transcriptome complexity. *Genome Res.*, **21**, 1450–1461.
6. Kuchenbauer,F., Morin,R.D., Argiropoulos,B., Petriv,O.I., Griffith,M., Heuser,M., Yung,E., Piper,J., Delaney,A., Prabhu,A.L. *et al.* (2008) In-depth characterization of the microRNA transcriptome in a leukemia progression model. *Genome Res.*, **18**, 1787–1797.
7. Martin,G. and Keller,W. (2007) RNA-specific ribonucleotidyl transferases. *RNA*, **13**, 1834–1849.
8. Ryan,B.M., Robles,A.I. and Harris,C.C. (2010) Genetic variation in microRNA networks: the implications for cancer research. *Nat. Rev. Cancer*, **10**, 389–402.
9. Nishikura,K. (2010) Functions and regulation of RNA editing by ADAR deaminases. *Annu. Rev. Biochem.*, **79**, 321–349.
10. Tan,G.C., Chan,E., Molnar,A., Sarkar,R., Alexieva,D., Isa,I.M., Robinson,S., Zhang,S., Ellis,P., Langford,C.F. *et al.* (2014) 5′ isomiR variation is of functional and evolutionary importance. *Nucleic Acids Res.*, **42**, 9424–9435.
11. Tan,G.C. and Dibb,N. (2015) IsomiRs have functional importance. *Malays. J. Pathol.*, **37**, 73–81.
12. Ibrahim,F., Rymarquis,L.A., Kim,E.J., Becker,J., Balassa,E., Green,P.J. and Cerutti,H. (2010) Uridylation of mature miRNAs and siRNAs by the MUT68 nucleotidyltransferase promotes their degradation in Chlamydomonas. *Proc. Natl. Acad. Sci. U.S.A.*, **107**, 3906–3911.
13. Katoh,T., Sakaguchi,Y., Miyauchi,K., Suzuki,T., Kashiwabara,S., Baba,T. and Suzuki,T. (2009) Selective stabilization of mammalian microRNAs by 3 ' adenylation mediated by the cytoplasmic poly(A) polymerase GLD-2. *Genes Dev.*, **23**, 433–438.
14. Kawahara,Y., Zinshteyn,B., Sethupathy,P., Iizasa,H., Hatzigeorgiou,A.G. and Nishikura,K. (2007) Redirection of silencing targets by adenosine-to-inosine editing of miRNAs. *Science*, **315**, 1137–1140.
15. Pantano,L., Estivill,X. and Marti,E. (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.

16. Sablok,G., Milev,I., Minkov,G., Minkov,I., Varotto,C., Yahubyan,G. and Baev,V. (2013) isomiRex: web-based identification of microRNAs, isomiR variations and differential expression using next-generation sequencing datasets. *FEBS Lett.*, **587**, 2629–2634.

17. Hackenberg,M., Sturm,M., Langenberger,D., Falcon-Perez,J.M. and Aransay,A.M. (2009) miRanalyzer: a microRNA detection and analysis tool for next-generation sequencing experiments. *Nucleic Acids Res.*, **37**, W68–W76.

18. de Oliveira,L.F., Christoff,A.P. and Margis,R. (2013) isomiRID: a framework to identify microRNA isoforms. *Bioinformatics*, **29**, 2521–2523.

19. Muller,H., Marzi,M.J. and Nicassio,F. (2014) IsomiRage: From functional classification to differential expression of miRNA isoforms. *Front Bioeng. Biotechnol.*, **2**, 38.

20. Zhang,Y., Xu,B., Yang,Y., Ban,R., Zhang,H., Jiang,X., Cooke,H.J., Xue,Y. and Shi,Q. (2012) CPSS: a computational platform for the analysis of small RNA deep sequencing data. *Bioinformatics*, **28**, 1925–1927.

21. Kozomara,A. and Griffiths-Jones,S. (2014) miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res.*, **42**, D68–D73.

22. Guo,L., Zhao,Y., Yang,S., Zhang,H. and Chen,F. (2014) A genome-wide screen for non-template nucleotides and isomiR repertoires in miRNAs indicates dynamic and versatile microRNAome. *Mol. Biol. Rep.*, **41**, 6649–6658.

23. Smigielski,E.M., Sirotkin,K., Ward,M. and Sherry,S.T. (2000) dbSNP: a database of single nucleotide polymorphisms. *Nucleic Acids Res.*, **28**, 352–355.

24. Betel,D., Wilson,M., Gabow,A., Marks,D.S. and Sander,C. (2008) The microRNA.org resource: targets and expression. *Nucleic Acids Res.*, **36**, D149–D153.

25. Kruger,J. and Rehmsmeier,M. (2006) RNAhybrid: microRNA target prediction easy, fast and flexible. *Nucleic Acids Res.*, **34**, W451–W454.

26. Huang,D.W., Sherman,B.T. and Lempicki,R.A. (2009) Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.*, **37**, 1–13.

27. Romanienko,P.J. and Camerini-Otero,R.D. (2000) The mouse Spo11 gene is required for meiotic chromosome synapsis. *Mol. Cell*, **6**, 975–987.

28. Baudat,F., Imai,Y. and de Massy,B. (2013) Meiotic recombination in mammals: localization and regulation. *Nat. Rev. Genet.*, **14**, 794–806.

29. Vickers,K.C., Sethupathy,P., Baran-Gale,J. and Remaley,A.T. (2013) Complexity of microRNA function and the role of isomiRs in lipid homeostasis. *J. Lipid Res.*, **54**, 1182–1191.

30. Vaz,C., Ahmad,H.M., Bharti,R., Pandey,P., Kumar,L., Kulshreshtha,R. and Bhattacharya,A. (2013) Analysis of the microRNA transcriptome and expression of different isomiRs in human peripheral blood mononuclear cells. *BMC Res. Notes*, **6**, 390.

31. McGahon,M.K., Yarham,J.M., Daly,A., Guduric-Fuchs,J., Ferguson,L.J., Simpson,D.A. and Collins,A. (2013) Distinctive profile of IsomiR expression and novel microRNAs in rat heart left ventricle. *PLoS One*, **8**, e65809.

32. Schamberger,A. and Orban,T.I. (2014) 3′ IsomiR species and DNA contamination influence reliable quantification of microRNAs by stem-loop quantitative PCR. *PLoS One*, **9**, e106315.

33. Lottersberger,F., Karssemeijer,R.A., Dimitrova,N. and de Lange,T. (2015) 53BP1 and the LINC Complex Promote Microtubule-Dependent DSB Mobility and DNA Repair. *Cell*, **163**, 880–893.

34. Kidane,D., Jonason,A.S., Gorton,T.S., Mihaylov,I., Pan,J., Keeney,S., de Rooij,D.G., Ashley,T., Keh,A., Liu,Y. *et al.* (2010) DNA polymerase beta is critical for mouse meiotic synapsis. *EMBO J.*, **29**, 410–423.