

Supplementary information

Small RNA-seq: The RNA 5'-end adapter ligation problem and how to circumvent it

Lodoe Lama^{1,2,6}, Jose Cobo^{1,2}, Diego Buenaventura³ and Kevin Ryan*^{1,2,4,5}

¹Department of Chemistry and Biochemistry, The City College of New York, New York, NY 10031, USA

²Biochemistry, ³Biology and ⁴Chemistry Ph.D. Programs, The City University of New York Graduate Center, 365 Fifth Avenue, New York, NY 10016, USA

⁵Corresponding author: Department of Chemistry and Biochemistry, The City College of New York, 160 Convent Avenue, New York, NY 10031. Tel.: +1 212 650 8312. E-mail address: kryan@ccny.cuny.edu

⁶Current address: Rockefeller University

Table of contents:

	Page
Figure S1 5' adapter ligation attempts	1
Supplemental Discussion on TS2126 Rnl1/CircLigase bias experiments.	2
Supplemental Methods for Figures S2 through S4.	2
Table S1	3
Figure S2	4
Figure S3	5
Figure S4	6
Application of Coligo-seq to 5'p RNA (HeLa miRNA)	7
Coligo-seq procedure for 5' phosphorylated RNA (HeLa miRNA).	7
Figure S5	8
Table S2	8
Supplemental References	16

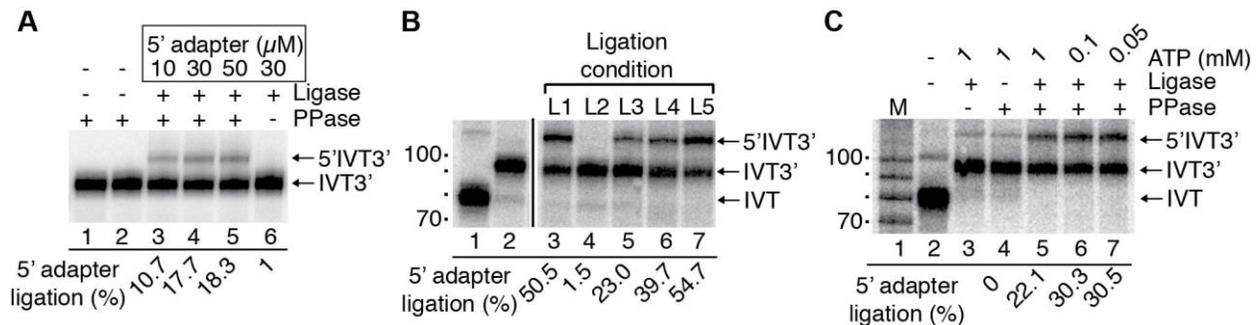


Figure S1. Examples of 5' adapter ligation optimization attempts using coligo 122 transcript as test case. (A) 5' adapter ligation to gel-purified 3' adapter-ligated transcripts (denoted by IVT3') using the optimized 3' adapter ligation conditions for 20% DMSO and 15% PEG-8000, but with varying concentrations of the 5' adapter. The desired ligation product is denoted 5'IVT3'. In the Lane 6 control reaction, the 5'ppp transcripts were not converted to 5'p before the ligation reaction. PPase, 5' polyphosphatase. (B) 5' ligation efficiency under various

ligation reaction conditions for 5' polyphosphatase treated coligo 122 transcripts: (L1) using the optimized 3' adapter ligation conditions; (L2) as in L1 but no PEG; (L3) as in L1 but no DMSO; (L4) as in L1 but incubated for 16 h instead of 2 h; and (L5) as in L1 but with 0.1 mM ATP instead of 1 mM ATP. (C) 5' adapter ligation of 3' adapter ligated transcripts as a function of ATP concentration. Optimized 3' adapter ligation conditions were used except ATP concentration was varied. IVT, Pol III in vitro transcript; IVT3', 3' adapter-ligated transcripts; 5'IVT3', fully adapter-ligated transcripts; PPase, RNA 5' polyphosphatase; PEG, polyethylene glycol 8000; DMSO, Dimethyl sulfoxide; M, RNA decade marker. 5' adapter ligation percentage is defined as $[5'IVT3' \text{ transcripts}/(IVT3' \text{ transcripts} + 5'IVT3' \text{ transcripts})] \times 100$.

Supplemental Discussion on TS2126 Rnl1 (CircLigase) bias assay.

The problem with low quality reads in the first TS2126 Rnl1/CircLigase circularization bias experiment. The TS2126 Rnl1/CircLigase bias study described in Figure 3 produced sequencing results with an unusually high percentage of low quality reads (phred score <20). The data in Figure 3 came from the ~3% of reads having a phred score greater than 30, whereas 94% of the reads had a phred score <20. We surmised that the unusually short size of the DNA library (133 bp end-to-end), resulting from the small size of the randomized region (14-bp) either reduced the efficiency of flow cell cluster generation or interfered with location of the cluster coordinates (Fig. S2B)[1]. To test this possible cause, and to assess the validity of the Figure 3 data, the experiment was repeated from the beginning, but this time the cDNA was lengthened to 171-bp using two successive PCR reactions (Fig. S2C). The P5/P7 flow cell sequences were then added by GeneWiz, who in this experiment conducted the MiSeq Illumina sequencing. The ligation reaction and 171-bp intermediate cDNA library gel image are shown in Figure S3. The new library preparation method produced 94% good quality reads (phred >30), showing that the amplicon length had indeed caused the large percentage of low quality reads.

As shown in Figure S4, the results from the repeat experiment were closely similar to those shown in Figure 3, further supporting our conclusions on TS2126 Rnl1/CircLigase. As in the first experiment, there was a slight under-representation in C+G content in the 7-nt on either side of the ligation site, but otherwise no significant preference in nucleotide identity (Fig. S4A,B). Importantly, all 16 dinucleotides were found to have been ligated, and the percentage of reads ranged only from 4 to 8%, while 6.25% is the percentage for completely unbiased ligation. Within this range, AA, TT, TA and TG were found in the top five in both experiments, and CC, GT, GG and GC were found in the bottom five in both experiments. These data indicate that there are reproducible differences in ligation sequence preference, but these differences are small and all possible combinations of 5' and 3' nucleotides undergo efficient ligation.

Supplemental Methods for Figures S2 through S4.

Circularization of LBprimer was performed as described in the main text; 0.25 μ M pre-circularized LBprimer oligo was circularized using 3 μ M TS2126 Rnl 1 for 90 minutes at 60 $^{\circ}$ C followed by extraction and ethanol precipitation. The circular product was separated by 12 % DPAGE, excised, extracted and precipitated. The purified circular product was then treated with 1 mg/ml RNase A (1 h, 37 $^{\circ}$ C) followed by extraction and ethanol precipitation. The precipitated product was used as a template for PCR 1. All PCR reaction conditions were identical to those described in the main text. The product was run on 2.5% agarose gel electrophoresis, extracted

using Genscript gel elution buffer. The 106 bp amplicon produced by PCR 1 was amplified using forward and reverse primers that contained partial adapter sequences that serve as a constant region for the addition of Illumina technical sequences. PCR 2 was performed at 12, 16, and 21 cycles, products from all the cycles were pooled together and run on 2.5% agarose gel electrophoresis. The 171 bp product was excised, extracted using Genscript gel elution buffer and sent for Illumina sequencing by GeneWiz (South Plainfield, NJ).

Bioinformatics analysis was done using the bioinformatic tools available at usegalaxy.org[2]. The fastq files produced from the sequencing run were uploaded to usegalaxy.org. The reads were filtered to retain only reads having a phred quality score above 30, and 94% of the reads were retained. The retained reads were statistically analyzed to determine the percent nucleotide distribution at each randomized position using the “FASTQ summary statistics” tool[3]. The reads were trimmed to retain only the randomized region using the “trim sequences tool”. The sequences were then collapsed and the sequence logo was generated using the sequence logo generator WebLogo 3[4]. To analyze the ligation site dinucleotides the sequences were further trimmed to retain only the n7-N7 positions. The dinucleotide sequences were analyzed by the “FASTQ summary statistics” tool. High throughput sequencing result files are available at the National Center for Biotechnology Information database, BioSample accession SAMN09534545.

Table S1. Oligonucleotide sequences used in TS2126 Rnl1 bias experiments.

PCR 1	
Forward Primer	TTGCCTAGCAGTAGCTATTTAGTGCAGACGTGTGCTCTTCCGATCT
Reverse Primer	TTCCTTAGCAGAGCTGTGGAGTTCAGAGTTCTACAGTCCGACGA
PCR 2	
Forward Primer	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTTGCCTAGCAGTAGCTATTTAGTG
Reverse Primer	GACTGGAGTTCAGACGTGTGCTCTTCCGATCTTTCCTTAGCAGAGCTGTGGAG
Partial Adapters	
Forward	ACACTCTTTCCCTACACGACGCTCTTCCGATCT
Reverse	GACTGGAGTTCAGACGTGTGCTCTTCCGATCT
Illumina technical sequences	
P5	AATGATACGGCGACCACCGA
P7	CAAGCAGAAGACGGCATACGAGAT

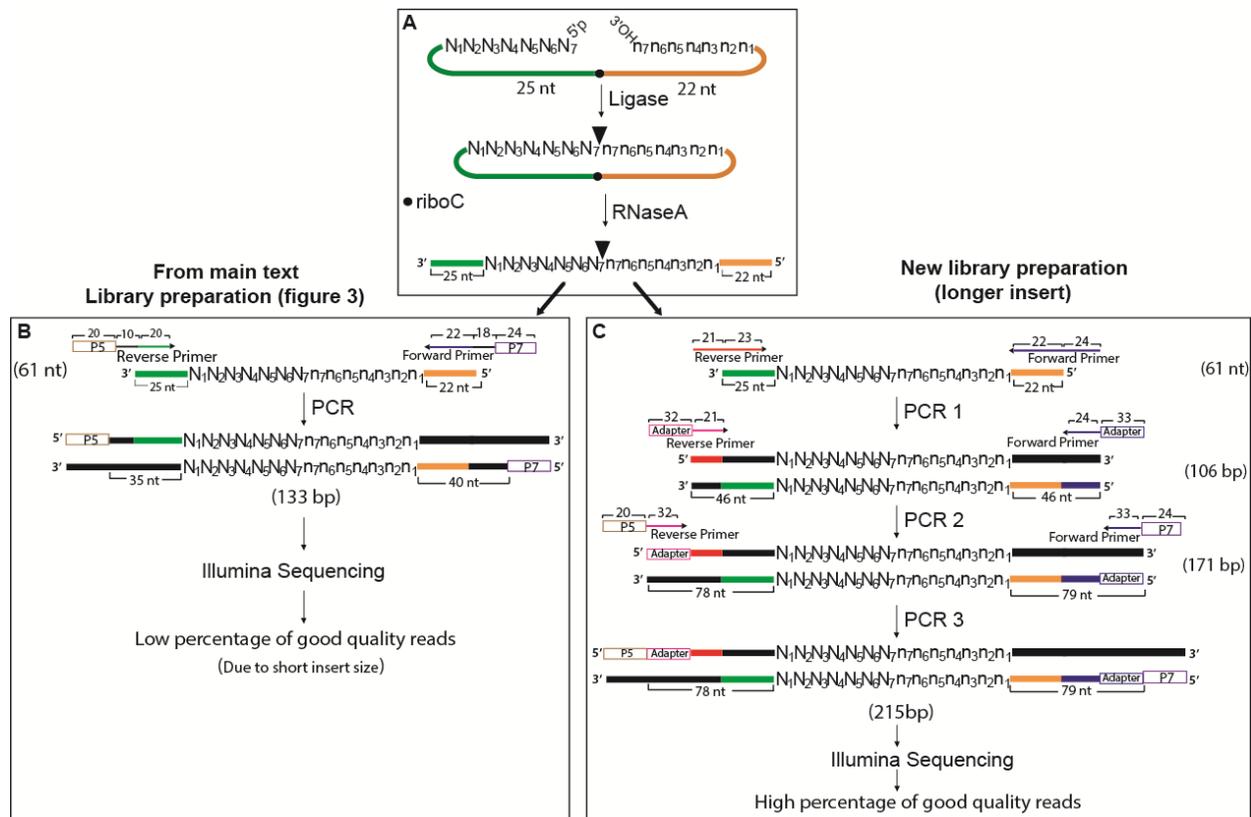


Figure S2. TS2126 Rnl1 (CircLigase) circularization bias experiments and sequencing fragment size. Library preparation schematic for high throughput sequencing. (A) The re-linearized LBprimer construct was produced as described in the main text and Figure 3. The re-linearized LBprimer construct consists of a 14-nucleotide random region flanked by a 22 nt constant region at 5' end and a 25 nt constant region at the 3' end. (B) Library preparation schematic described in Figure 3A. The PCR fragment/amplicon produced by this method has an 89 bp insert region (133 bp total size). Most reads obtained from this experiment were of low quality; the 3% with phred score greater than 30 were used for analysis in Figure 3 of the main text. (C) In order to increase the percentage of good quality reads the library preparation protocol was modified to produce a longer insert size. The original re-linearized LBprimer was first amplified by PCR with primers that included filler sequences to increase the size of the construct to 106 bp (PCR 1). A second PCR was performed to add partial adapter sequences that serve as a constant region for the addition of Illumina technical sequences (PCR 2,) this step also serves to increase the size of the construct to 171 bp. The 171 bp amplicon was further lengthened by GeneWiz when a third PCR was done to introduce the Illumina P7 and P5 technical sequences (PCR 3). The final amplicon was brought to 215 bp containing a 171 bp insert suitable for sequencing using the MiSeq 300 cycle kit. This library preparation method produced 94% good quality reads having a phred score over 30.

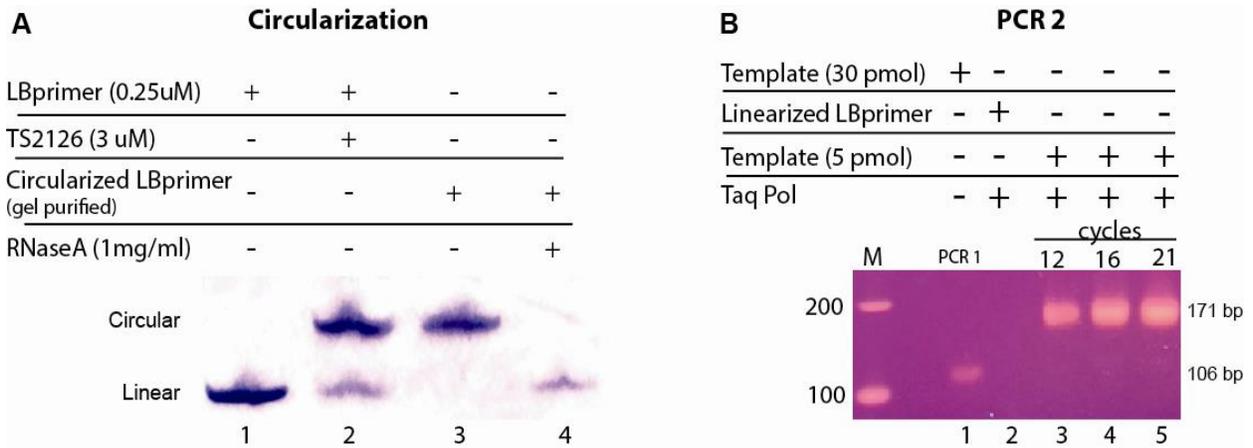


Figure S3. TS2126 Rnl1 circularization bias experiment; library preparation. (A) Circularization of LBprimer using 3 μ M TS2126 Rnl1 as described in Figure 3. Pre-circularized LBprimer oligo 0.25 μ M (lane 1) was circularized using 3 μ M TS2126 Rnl1 (lane 2). The oligo was gel purified (lane 3) and re-linearized using 1 mg/ml RNase A (lane 4). Stains-All gel stain. (B) Second PCR using forward and reverse primers that contained partial adapter sequences that serve as a constant region for the addition of Illumina technical sequences (PCR 2). The 106 bp amplicon produced by PCR 1 (lane 1) was amplified using forward and reverse primers that contained partial adapter sequences. PCR 2 was performed at 12, 16, and 21 cycles (lanes 3, 4, and 5) products from all the cycles were pooled together and sent for sequencing at GeneWiz. Ethidium bromide gel staining.

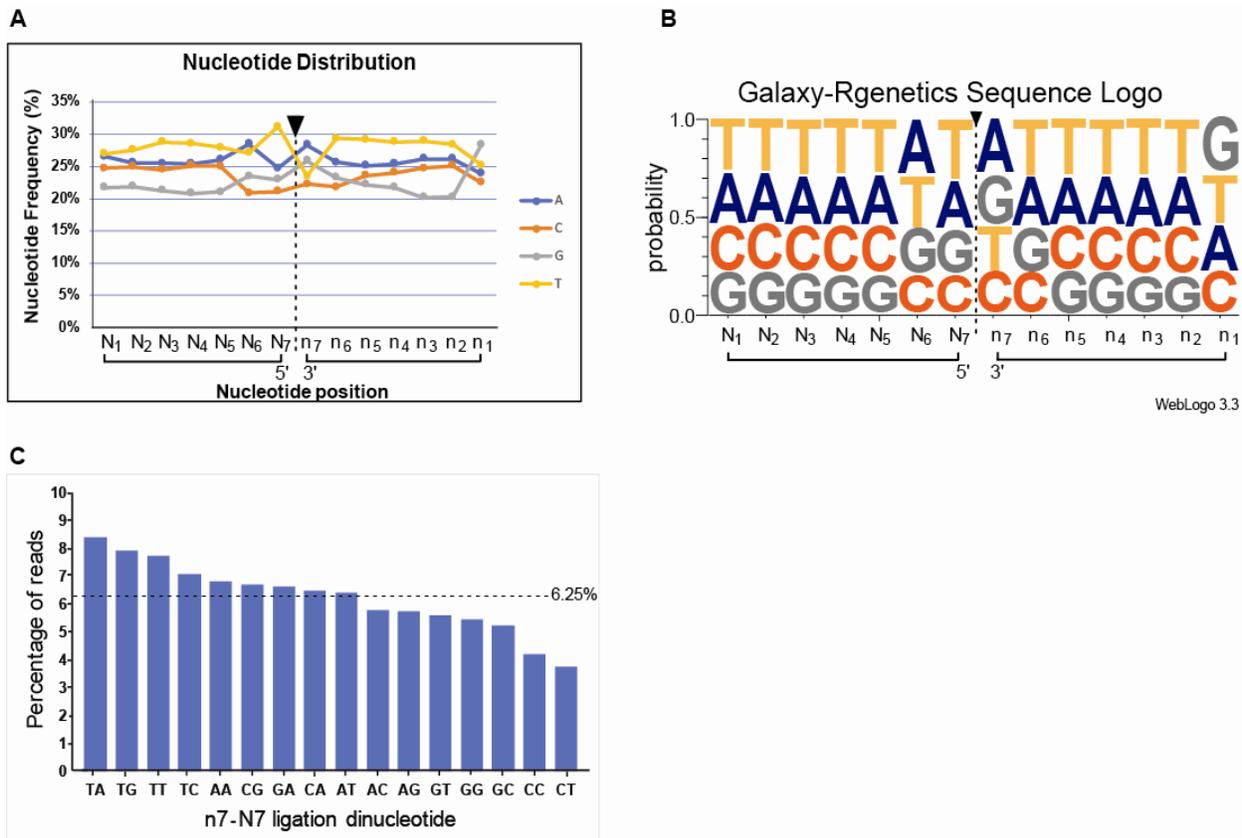


Figure S4. TS2126 Rnl1 circularization bias analysis results. (A) Nucleotide distribution at each randomized position of the circularized LBprimer at its 5' and 3' end. (B) The sequence logo shows the probability of a nucleotide being present at each randomized position. The sequence logo was generated using the sequence logo generator WebLogo 3[4] freely available at usegalaxy.org. (C) The dinucleotide distribution shows that all possible 5' and 3' nucleotides are circle-ligated by TS2126 Rnl1. The results from this experiment confirm the results obtained from the original experiment shown in Figure 3 of the main text.

Application of Coligo-seq to 5'p RNA (e.g. HeLa miRNA). Coligo-seq will work on RNA with any 5' modification that is inert to the ligase used to add the 3' adapter (e.g. triphosphate). However, some cellular small RNAs contain a 5' phosphate, which as noted in the Results section will lead to ligation side-reactions. Coligo-seq can be adapted to 5' monophosphorylated RNA by using a pre-adenylylated 3' adapter (Fig. 6), which is also commonly used in kit-based small RNA-seq methods. We followed the procedure outlined in Fig. 6 to make a HeLa cell miRNA cDNA library and compared it to a library made from the other half of the same RNA sample using a standard RNA-seq protocol. A summary of the Illumina sequencing results are shown in Figure S5. NCBI Biosample accession codes are given below. A spreadsheet of the miRNAs found by the two methods can be found in Table S2.

Coligo-seq procedure for 5' phosphorylated RNA (e.g. HeLa miRNA). Total RNA was isolated from a 10 cm plate of confluent HeLa cells using 2 ml of Trizol reagent (Invitrogen). Four μg of the total RNA was used for 3' adapter ligation in a 10 μl final reaction volume containing 10 μM pre-adenylylated adapter Adapter4, 200 U of truncated T4 Rnl2(1-249)K227Q (available from NEB), 20% PEG-8000, 32 units RNasin RNase inhibitor (Promega), and trace amounts of 5' ^{32}P end-labeled small RNA size markers M1 (19 nt) and M2 (24 nt) (sequences shown in Table 1). The reaction was incubated at 16 $^{\circ}\text{C}$ for 17 h. The products were PCI extracted, ethanol precipitated, and purified by 12% DPAGE gel. The labeled, 3'-ligated size-marker products were located via XAR film exposure, and region between and including the markers was excised, extracted, and precipitated. Half of the material isolated was used to continue the Coligo-seq protocol: The size-fractionated 3' adapter-ligated small RNAs were used as a template for cDNA strand synthesis using 5' phosphorylated RTprimer2 spiked with a trace amount of 5' ^{32}P end-labeled RTprimer2. The products were PCI extracted, ethanol precipitated, and resolved by 10 % DPAGE. The 5' radiolabeled cDNA products from the 19-24 nt RNA were located using Kodak Biomax XAR film exposure, excised, extracted, and precipitated. The cDNA was then circularized and re-linearized according to the Coligo-seq procedure. The products were PCR amplified using primers Primer2F and Primer2R4. The PCR products were treated with PmeI restriction enzyme digestion to remove PCR products coming from the RNA size markers. The PCR products were purified on a 2.5 % agarose gel. The bands were visualized using ethidium bromide, excised, and extracted using Quickclean II Gel extraction kit (Genscript). The gel-purified PCR products were sequenced on an Illumina Miseq instrument at the Personalized Genomic Medicine Laboratory, Department of Pathology and Cell Biology, Columbia University Medical Center, New York, NY. The raw sequencing data are available on the NCBI website under Biosample accession number SAMN10416069. The resulting reads were analyzed and mapped onto reference human genome (hg19) using NextGene software (Summarized in Fig. S5). The percentage of miRNA reads for a given miRNA sequence was calculated as (Number of reads for a given miRNA sequence/Total miRNA reads)*100. The reads mapping to known miRNA genes were used for comparative analysis, shown in Supplementary Information spreadsheet Table S2.

Standard RNA-seq procedure for 5' phosphorylated RNA. The other half of the size-fractionated, 3' adapter ligated small RNAs was used for 5' adapter ligation using 5' RNA adapter Adapter5 following a standard RNA-seq procedure[5] and the products were extracted, resolved on 10% DPAGE gel. The 5' adapter-ligated products were excised, extracted, and precipitated as described above. This template was reverse transcribed with Superscript III RT (Invitrogen) and primer RTprimer3 followed by PCR amplification using PCR primers Primer2F

and Primer2R3. The PCR products were digested with restriction enzyme PmeI to remove PCR products coming from the RNA size markers. The resulting PCR products were resolved on a 2.5 % agarose gel, as described above for the Coligo-seq PCR products, and continued with the same procedure described above through Illumina sequencing and analysis. The raw sequencing data is available on the NCBI website under Biosample accession number SAMN10416079.

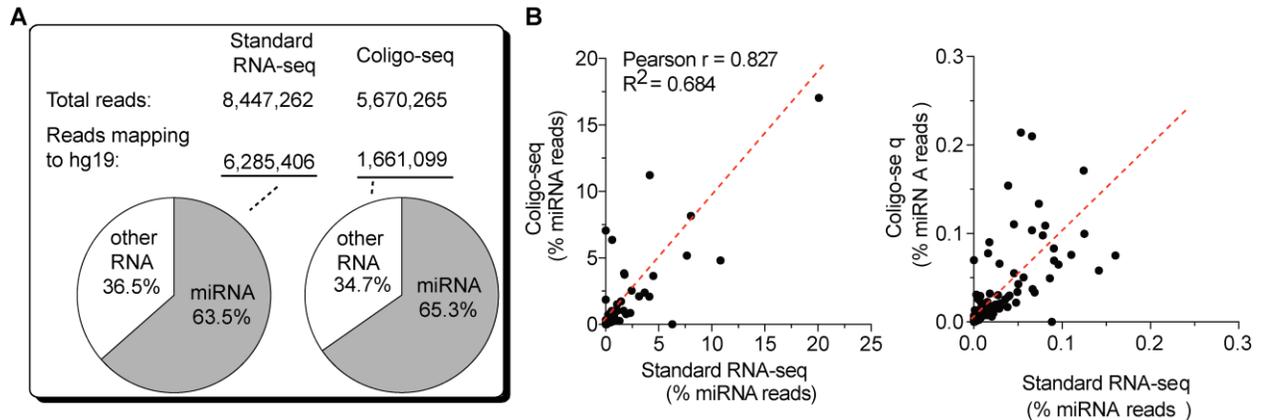


Figure S5. Application of Coligo-seq to HeLa cell miRNA. A sample of total RNA from HeLa cell culture was split in half and a cDNA library was made from each half using the standard RNA-seq and the Coligo-seq procedures. **A.** The number and percentages of total seq reads, and reads mapping to known human miRNAs. **B.** Plots of percent reads (two different scales of same data) found by the two library construction methods. Each dot represents one human miRNA found in both cDNA libraries. Red line denotes slope = 1. The raw sequence data are available on the NCBI website under the following Biosample accession codes: Standard RNA-seq method, SAMN10416079; Coligo-seq method, SAMN10416069. A searchable Excel spreadsheet of miRNAs found can be found in Supplementary Table S2.

Table S2. Comparison of miRNA reads obtained using standard RNA seq method (Method1) versus Coligo-seq method (Method 2).

Chromosome number	Chromosome Position Start	Chromosome Position End	Gene	Length	Read Counts (Method 1)	% Read Counts (Method 1)	Read Counts (Method 2)	% Read counts (Method 2)
reads not found in Method 2					NCBI raw data set:		NCBI raw data set:	
reads not found in Method 1					SAMN10416079		SAMN10416069	
17	57918634	57918655	MIR21; +	22	437515	20.09123634	106423	17.03591175
13	92003615	92003636	MIR17HG; +	22	235648	10.82125107	30082	4.815446824

13	92003326	92003348	MIR17HG; +	23	175083	8.040030473	50959	8.157381647
12	62997471	62997492	MIRLET7I; +	22	166845	7.661731204	32342	5.177221633
13	92002872	92002894	MIR17HG; +	23	136920	6.287537753	0	0
22	46509571	46509592	MIRLET7BH G; +	22	97805	4.491328001	22728	3.638237997
13	92003193	92003215	MIR17HG; +	23	90396	4.151097449	70149	11.22926598
6	72113298	72113319	MIR30A; -	22	89638	4.11628914	13043	2.087888868
22	46508632	46508653	MIRLET7BH G; +	22	79932	3.670577474	14890	2.383551732
17	1617208	1617229	MIR22HG; -	22	68825	3.160530133	13126	2.101175288
9	21512157	21512177	MIR31HG; -	21	53361	2.4504039	15908	2.546510474
11	122022983	122023004	MIR100HG; -	22	50643	2.325589939	5391	0.86297699
8	135817162	135817183	MIR30D; -	22	42181	1.937004309	4856	0.777335609
19	13947409	13947429	MIR23A; -	21	38932	1.787806163	23344	3.736845644
13	92003010	92003032	MIR17HG; +	23	37781	1.734950802	23952	3.834172672
8	22102488	22102509	MIR320A; -	22	37041	1.700969076	6377	1.020813257
9	96938635	96938656	MIRLET7F1; +	22	31571	1.449779831	10777	1.725153594
19	52196521	52196544	MIR125A; +	24	28520	1.309674092	1654	0.264767936
X	45605608	45605630	MIR221; -	23	23866	1.095956588	6378	1.020973334
8	135812813	135812834	MIR30B; -	22	23398	1.074465442	9381	1.501685614
13	92002909	92002930	MIR17HG; +	22	17609	0.808627317	4123	0.659998911
17	46657266	46657288	MIR10A; -	23	16065	0.737724905	1498	0.239795869
9	96941123	96941144	MIRLET7D; +	22	13681	0.628248642	5706	0.913401356
13	92003499	92003521	MIR17HG; +	23	13383	0.614564109	39705	6.355871157
X	45606442	45606462	MIR222; -	21	12562	0.576862761	6268	1.003364826
1	207975210	207975231	MIR29C; -	22	12012	0.551606073	3624	0.580120314
7	25989542	25989563	MIR148A; -	22	11739	0.539069571	1668	0.267009019
1	9211794	9211815	MIR34A; -	22	10564	0.485112101	5448	0.872101399
19	52195871	52195892	MIR99B; +	22	10501	0.482219062	1012	0.161998278
7	129410287	129410310	MIR182; -	24	8977	0.412235075	1343	0.21498388
12	54385546	54385567	MIR196A2; +	22	7324	0.336327246	3026	0.484394059
17	8090266	8090287	MIR4521; +	22	7055	0.323974429	1473	0.235793936
19	52196046	52196067	MIRLET7E; +	22	6358	0.291967317	2094	0.33520197
7	129414579	129414601	MIR96; -	23	5444	0.249995293	4495	0.719547685
2	177015057	177015079	MIR10B; +	23	5435	0.249582002	614	0.098287493
12	62997527	62997548	MIRLET7I; +	22	5082	0.2333718	1127	0.180407173
7	129414807	129414828	MIR183; -	22	4545	0.208712088	846	0.135425438
1	155165028	155165049	MIR92B; +	22	3491	0.160311089	470	0.075236354
11	64658876	64658897	MIR194-2; -	22	3083	0.141575218	363	0.058108078

6	72113257	72113278	MIR30A; -	22	2725	0.12513541	622	0.099568111
13	92003362	92003383	MIR17HG; +	22	2710	0.124446592	1069	0.171122686
9	28888923	28888943	MIR873; -	21	2404	0.110394689	474	0.075876664
19	13947301	13947322	MIR27A; -	22	2089	0.095929494	406	0.064991404
9	96941177	96941198	MIRLET7D; +	22	1979	0.090878157	433	0.069313492
22	22007643	22007664	MIR130B; +	22	1971	0.090510787	520	0.083240222
	568112	568133	MIR210HG; -	22	1925	0.088398409	0	0
17	57918672	57918692	MIR21; +	21	1877	0.086194189	308	0.049303824
X	133680710	133680731	MIR424; -	22	1761	0.080867324	680	0.108852598
6	72086706	72086728	MIR30C2; -	23	1699	0.078020206	612	0.097967338
19	13947140	13947161	MIR24-2; -	22	1601	0.073519924	835	0.133664587
11	64658675	64658695	MIR192; -	21	1498	0.068790035	208	0.033296089
1	110141562	110141583	MIR197; +	22	1449	0.066539893	232	0.037137945
17	29887069	29887090	MIR193A; +	22	1437	0.065988838	647	0.103570045
X	133303713	133303735	MIR19B2; -	23	1433	0.065805153	1311	0.209861405
17	1953223	1953244	MIR132; -	22	1225	0.056253533	315	0.050424365
9	21512120	21512141	MIR31HG; -	22	1163	0.053406415	1338	0.214183493
X	139006319	139006340	MIR505; -	22	1096	0.050329692	268	0.04290073
19	13985539	13985560	MIR181C; +	22	1076	0.049411267	211	0.033776321
1	155164987	155165008	MIR92B; +	22	1044	0.047941787	137	0.021930597
13	92003578	92003600	MIR17HG; +	23	996	0.045737566	345	0.055226686
9	28863673	28863694	MIR876; -	22	989	0.045416118	690	0.110453371
X	133303574	133303595	MIR92A2; -	22	879	0.04036478	186	0.029774387
19	13985724	13985746	MIR181D; +	23	850	0.039033064	179	0.028653846
13	92003461	92003483	MIR17HG; +	23	846	0.038849379	963	0.154154487
12	54385583	54385604	MIR196A2; +	22	830	0.038114639	105	0.016808122
17	29887035	29887056	MIR193A; +	22	775	0.03558897	160	0.025612376
8	135817122	135817143	MIR30D; -	22	664	0.030491711	132	0.02113021
19	52195909	52195930	MIR99B; +	22	644	0.029573286	112	0.017928663
13	92003051	92003073	MIR17HG; +	23	635	0.029159995	411	0.065791791
19	46522219	46522240	MIR769; +	22	629	0.028884467	94	0.015047271
19	4770712	4770734	MIR7-3HG; +	23	617	0.028333412	127	0.020329823
19	13947444	13947465	MIR23A; -	22	598	0.027460908	190	0.030414696
14	104583805	104583826	MIR203; +	22	531	0.024384185	122	0.019529437
14	104583806	104583827	MIR203; +	22	531	0.024384185	122	0.019529437
1	207975837	207975858	MIR29B2; -	22	522	0.023970893	120	0.019209282
6	72086667	72086688	MIR30C2; -	22	483	0.022179964	62	0.009924796
17	72744802	72744822	MIR3615; +	21	442	0.020297193	38	0.006082939
X	8095021	8095042	MIR651; +	22	429	0.019700217	98	0.01568758

22	22007314	22007336	MIR301B; +	23	392	0.018001131	200	0.03201547
19	13947261	13947281	MIR27A; -	21	385	0.017679682	563	0.090123548
22	22007605	22007625	MIR130B; +	21	355	0.016302044	47	0.007523635
17	1617246	1617267	MIR22HG; -	22	353	0.016210202	485	0.077637514
9	96938295	96938315	MIRLET7A1; +	21	342	0.015705068	90	0.014406961
12	7073318	7073339	MIR141; +	22	336	0.01542954	141	0.022570906
X	45605649	45605670	MIR221; -	22	329	0.015108092	66	0.010565105
12	7072905	7072927	MIR200C; +	23	320	0.0146948	103	0.016487967
10	104196277	104196298	MIR146B; +	22	299	0.013730454	48	0.007683713
20	57392681	57392702	MIR296; -	22	286	0.013133478	54	0.008644177
1	9211752	9211773	MIR34A; -	22	257	0.011801762	104	0.016648044
X	139006355	139006376	MIR505; -	22	230	0.010561888	46	0.007363558
8	135812775	135812796	MIR30B; -	22	223	0.010240439	57	0.009124409
X	133674423	133674444	MIR450A1; -	22	215	0.009873069	87	0.013926729
17	6921298	6921318	MIR497HG; -	21	214	0.009827148	96	0.015367426
17	6920986	6921006	MIR497HG; -	21	212	0.009735305	89	0.014246884
19	52196559	52196580	MIR125A; +	22	192	0.00881688	57	0.009124409
9	28863634	28863655	MIR876; -	22	191	0.008770959	119	0.019049205
11	93466890	93466911	MIR1304; -	22	181	0.008311747	76	0.012165879
X	133303880	133303902	MIR20B; -	23	179	0.008219904	32	0.005122475
19	46522258	46522280	MIR769; +	23	169	0.007760692	35	0.005602707
1	207975248	207975269	MIR29C; -	22	159	0.007301479	183	0.029294155
11	122017231	122017252	MIR100HG; -	22	156	0.007163715	41	0.006563171
2	56216155	56216176	MIR216A; -	22	148	0.006796345	46	0.007363558
14	100774213	100774234	MIR345; +	22	143	0.006566739	153	0.024491834
1	25350043	25350064	MIR4425; +	22	140	0.006428975	22	0.003521702
12	95702221	95702242	MIR331; +	22	140	0.006428975	22	0.003521702
13	92003158	92003179	MIR17HG; +	22	139	0.006383054	74	0.011845724
11	93466857	93466878	MIR1304; -	22	137	0.006291211	16	0.002561238
9	28888887	28888908	MIR873; -	22	133	0.006107526	19	0.00304147
4	24521875	24521898	MIR573; -	24	133	0.006107526	23	0.003681779
X	133675394	133675415	MIR542; -	22	132	0.006061605	25	0.004001934
1	207975795	207975817	MIR29B2; -	23	111	0.005097259	80	0.012806188
19	52196091	52196112	MIRLET7E; +	22	96	0.00440844	11	0.001760851
X	133304274	133304296	MIR106A; -	23	96	0.00440844	164	0.026252685
1	65524160	65524181	MIR101-1; -	22	92	0.004224755	82	0.013126343
X	45606479	45606500	MIR222; -	22	92	0.004224755	183	0.029294155
11	122017276	122017297	MIR100HG; -	22	84	0.003857385	11649	1.864741043
12	95702256	95702276	MIR331; +	21	80	0.0036737	26	0.004162011

3	195426331	195426352	MIR570; +	22	75	0.003444094	30	0.00480232
11	122022948	122022969	MIR100HG; -	22	69	0.003168566	56	0.008964332
21	42539491	42539513	MIR3197; +	23	63	0.002893039	11	0.001760851
20	57392716	57392736	MIR296; -	21	63	0.002893039	193	0.030894928
22	31556048	31556069	MIR3928; -	22	62	0.002847118	19	0.00304147
2	56210155	56210177	MIR217; -	23	61	0.002801196	11	0.001760851
X	133674261	133674282	MIR450B; -	22	56	0.00257159	18	0.002881392
7	25989580	25989601	MIR148A; -	22	51	0.002341984	15	0.00240116
5	148808541	148808561	MIR143HG; +	21	51	0.002341984	18	0.002881392
19	13985577	13985598	MIR181C; +	22	49	0.002250141	9	0.001440696
1	94312436	94312455	MIR760; +	20	49	0.002250141	23	0.003681779
1	117214409	117214430	MIR320B1; +	22	47	0.002158299	22	0.003521702
9	96938244	96938265	MIRLET7A1; +	22	47	0.002158299	2056	0.32911903
17	1953584	1953604	MIR212; -	21	42	0.001928693	16	0.002561238
7	129410246	129410266	MIR182; -	21	41	0.001882771	25	0.004001934
7	129414768	129414789	MIR183; -	22	40	0.00183685	12	0.001920928
17	29902458	29902479	MIR365-2; +	22	35	0.001607244	7	0.001120541
17	1953622	1953644	MIR212; -	23	35	0.001607244	8	0.001280619
17	46657226	46657247	MIR10A; -	22	35	0.001607244	10	0.001600773
10	106028099	106028120	MIR4482; -	22	31	0.001423559	11	0.001760851
X	133675430	133675452	MIR542; -	23	31	0.001423559	11	0.001760851
16	2320752	2320773	MIR3677; +	22	28	0.001285795	9	0.001440696
X	133680674	133680694	MIR424; -	21	28	0.001285795	13	0.002081006
10	106028135	106028156	MIR4482; -	22	27	0.001239874	14	0.002241083
11	64658840	64658861	MIR194-2; -	22	25	0.001148031	10	0.001600773
17	46709894	46709915	MIR196A1; -	22	24	0.00110211	12	0.001920928
5	148810224	148810246	MIR143HG; +	23	23	0.001056189	47	0.007523635
X	63005935	63005955	MIR1468; -	21	22	0.001010268	3	0.000480232
X	133304114	133304136	MIR18B; -	23	22	0.001010268	84	0.013446497
19	18392939	18392961	MIR3188; +	23	21	0.000964346	2	0.000320155
16	2320716	2320737	MIR3677; +	22	20	0.000918425	8	0.001280619
1	110141523	110141545	MIR197; +	23	19	0.000872504	12	0.001920928
11	64658631	64658652	MIR192; -	22	19	0.000872504	13	0.002081006
7	129414537	129414558	MIR96; -	22	19	0.000872504	15	0.00240116
5	148810262	148810283	MIR143HG; +	22	18	0.000826583	1	0.000160077
2	56227899	56227920	MIR216B; -	22	18	0.000826583	2	0.000320155

1	1103296	1103317	MIR200A; +	22	17	0.000780661	7	0.001120541
21	9826237	9826260	MIR3687; +	24	15	0.000688819	17	0.002721315
16	2321807	2321827	MIR940; +	21	14	0.000642898	13	0.002081006
22	46486970	46486991	MIRLET7BH G; +	22	14	0.000642898	13	0.002081006
1	1102540	1102561	MIR200B; +	22	13	0.000596976	2	0.000320155
22	46508680	46508700	MIRLET7BH G; +	21	13	0.000596976	3	0.000480232
2	177015096	177015117	MIR10B; +	22	13	0.000596976	5	0.000800387
1	65524125	65524145	MIR101-1; -	21	13	0.000596976	8	0.001280619
10	132760897	132760921	MIR378C; -	25	13	0.000596976	37	0.005922862
5	170813682	170813702	MIR3912; -	21	12	0.000551055	1	0.000160077
14	104583765	104583787	MIR203; +	23	12	0.000551055	11	0.001760851
12	69978524	69978545	MIR3913-1; -	22	11	0.000505134	1	0.000160077
12	69978524	69978545	MIR3913-1; -	22	11	0.000505134	1	0.000160077
16	14995416	14995437	MIR3179-1; +	22	11	0.000505134	2	0.000320155
22	46486939	46486960	MIRLET7BH G; +	22	11	0.000505134	12	0.001920928
14	102026699	102026720	MIR1247; -	22	10	0.000459213	1	0.000160077
2	219866370	219866391	MIR375; -	22	10	0.000459213	4	0.000640309
15	90549996	90550018	MIR3174; +	23	10	0.000459213	4	0.000640309
9	131154911	131154932	MIR219-2; -	22	10	0.000459213	8	0.001280619
9	131154913	131154934	MIR219-2; -	22	10	0.000459213	8	0.001280619
22	46509504	46509527	MIRLET7BH G; +	24	9	0.000413291	1	0.000160077
1	154166189	154166209	MIR190B; -	21	9	0.000413291	4	0.000640309
9	131154955	131154977	MIR219-2; -	23	9	0.000413291	5	0.000800387
9	131154955	131154975	MIR219-2; -	21	9	0.000413291	5	0.000800387
22	46509625	46509646	MIRLET7BH G; +	22	8	0.00036737	2	0.000320155
5	148808507	148808528	MIR143HG; +	22	8	0.00036737	3	0.000480232
17	6920947	6920968	MIR497HG; -	22	8	0.00036737	3	0.000480232
9	96938691	96938712	MIRLET7F1; +	22	8	0.00036737	11	0.001760851
5	180649577	180649599	MIR4638; -	23	7	0.000321449	1	0.000160077
7	93346249	93346270	MIR4652; +	22	7	0.000321449	1	0.000160077
22	46509464	46509484	MIRLET7BH G; +	21	7	0.000321449	1	0.000160077
3	50712561	50712584	MIR4787; +	24	7	0.000321449	3	0.000480232

17	6921257	6921278	MIR497HG; -	22	7	0.000321449	8	0.001280619
4	183090496	183090517	MIR1305; +	22	7	0.000321449	9	0.001440696
8	27559207	27559228	MIR3622B; -	22	6	0.000275528	0	0
8	27559207	27559227	MIR3622B; -	21	6	0.000275528	0	0
8	27559243	27559262	MIR3622B; -	20	6	0.000275528	0	0
8	27559243	27559264	MIR3622B; -	22	6	0.000275528	0	0
16	15005138	15005159	MIR3180-1; +	22	6	0.000275528	1	0.000160077
17	36858523	36858544	MIR4734; -	22	6	0.000275528	2	0.000320155
15	89155087	89155110	MIR7-2; +	24	6	0.000275528	15	0.00240116
15	89155087	89155109	MIR7-2; +	23	6	0.000275528	15	0.00240116
7	93346289	93346310	MIR4652; +	22	5	0.000229606	0	0
8	8905963	8905985	MIR4660; +	23	5	0.000229606		0
6	2854274	2854295	MIR4645; -	22	5	0.000229606	1	0.000160077
2	240882440	240882461	MIR4786; -	22	4	0.000183685	1	0.000160077
6	167411318	167411339	MIR3939; -	22	4	0.000183685	1	0.000160077
19	50004089	50004110	MIR150; -	22	4	0.000183685	1	0.000160077
1	205417489	205417511	MIR135B; -	23	4	0.000183685	2	0.000320155
6	27115408	27115432	MIR3143; +	25	4	0.000183685	3	0.000480232
17	8090552	8090571	MIR3676; +	20	4	0.000183685	21	0.003361624
2	56216116	56216137	MIR216A; -	22	3	0.000137764	0	0
2	87929283	87929304	MIR4435-1; +	22	3	0.000137764	0	0
9	94398559	94398578	MIR3910-1; +	20	3	0.000137764	0	0
11	111384176	111384198	MIR34C; +	23	3	0.000137764	0	0
6	2854313	2854331	MIR4645; -	19	3	0.000137764	1	0.000160077
9	94398595	94398614	MIR3910-1; +	20	3	0.000137764	3	0.000480232
2	66585389	66585410	MIR4778; -	22	3	0.000137764	7	0.001120541
6	18572075	18572096	MIR548A1; +	22	3	0.000137764	10	0.001600773
2	134884701	134884723	MIR3679; +	23	2	9.18425E-05	0	0
2	134884739	134884760	MIR3679; +	22	2	9.18425E-05	0	0
6	36590257	36590278	MIR3925; -	22	2	9.18425E-05	0	0
6	126805792	126805812	MIR588; +	21	2	9.18425E-05	0	0
12	95703704	95703725	MIR3685; +	22	2	9.18425E-05	0	0
14	101512305	101512326	MIR381; +	22	2	9.18425E-05	0	0
15	89155125	89155146	MIR7-2; +	22	2	9.18425E-05	0	0
15	89155127	89155148	MIR7-2; +	22	2	9.18425E-05	0	0
17	25620948	25620968	MIR4522; -	21	2	9.18425E-05	0	0

18	33514051	33514073	MIR3929; -	23	2	9.18425E-05	0	0
20	60528677	60528697	MIR1257; -	21	2	9.18425E-05	0	0
2	240882480	240882501	MIR4786; -	22	2	9.18425E-05	1	0.000160077
6	33175632	33175652	MIR219-1; +	21	2	9.18425E-05	1	0.000160077
11	43581225	43581244	MIR670; +	20	2	9.18425E-05	1	0.000160077
13	115039360	115039381	MIR4502; +	22	2	9.18425E-05	1	0.000160077
14	100774249	100774270	MIR345; +	22	2	9.18425E-05	2	0.000320155
17	29902345	29902366	MIR4725; +	22	2	9.18425E-05	2	0.000320155
17	1953259	1953280	MIR132; -	22	2	9.18425E-05	43	0.006883326
1	1103258	1103279	MIR200A; +	22	1	4.59213E-05	0	0
1	1104435	1104456	MIR429; +	22	1	4.59213E-05	0	0
1	5922771	5922792	MIR4689; -	22	1	4.59213E-05	0	0
1	249120617	249120638	MIR3124; +	22	1	4.59213E-05	0	0
2	12339264	12339285	MIR3681; +	22	1	4.59213E-05	0	0
3	103242929	103242950	MIR548AB; -	22	1	4.59213E-05	0	0
5	1309429	1309450	MIR4457; -	22	1	4.59213E-05	0	0
5	89312458	89312478	MIR3660; -	21	1	4.59213E-05	0	0
6	36590223	36590243	MIR3925; -	21	1	4.59213E-05	0	0
8	124228071	124228094	MIR4663; -	24	1	4.59213E-05	0	0
9	140732886	140732908	MIR602; +	23	1	4.59213E-05	0	0
10	104196313	104196334	MIR146B; +	22	1	4.59213E-05	0	0
10	121718034	121718056	MIR4682; +	23	1	4.59213E-05	0	0
11	74110346	74110367	MIR548AL; +	22	1	4.59213E-05	0	0
12	69978560	69978581	MIR3913-1; -	22	1	4.59213E-05	0	0
12	69978560	69978581	MIR3913-1; -	22	1	4.59213E-05	0	0
12	124020960	124020981	MIR3908; +	22	1	4.59213E-05	0	0
14	105144060	105144077	MIR4710; -	18	1	4.59213E-05	0	0
16	2324662	2324682	MIR4717; +	21	1	4.59213E-05	0	0
17	27188601	27188622	MIR144; -	22	1	4.59213E-05	0	0
19	54264402	54264424	MIR516A2; +	23	1	4.59213E-05	0	0
19	54291185	54291207	MIR372; +	23	1	4.59213E-05	0	0
20	59053177	59053198	MIR4533; +	22	1	4.59213E-05	0	0
20	61809866	61809887	MIR124-3; +	22	1	4.59213E-05	0	0
22	38243739	38243760	MIR659; -	22	1	4.59213E-05	0	0
22	41488549	41488565	MIR1281; +	17	1	4.59213E-05	0	0
X	133674594	133674615	MIR450A2; -	22	1	4.59213E-05	0	0
X	149396251	149396272	MIR2114; +	22	1	4.59213E-05	0	0
2	66585429	66585450	MIR4778; -	22	1	4.59213E-05	1	0.000160077
5	159912379	159912400	MIR146A; +	22	1	4.59213E-05	1	0.000160077

12	7072866	7072887	MIR200C; +	22	1	4.59213E-05	1	0.000160077
14	101515947	101515968	MIR655; +	22	1	4.59213E-05	1	0.000160077
14	101520653	101520674	MIR382; +	22	1	4.59213E-05	1	0.000160077
15	89911263	89911285	MIR9-3; +	23	1	4.59213E-05	1	0.000160077
17	29421413	29421434	MIR4733; -	22	1	4.59213E-05	1	0.000160077
17	77681025	77681049	MIR4739; -	25	1	4.59213E-05	1	0.000160077
X	133674559	133674580	MIR450A2; -	22	1	4.59213E-05	1	0.000160077
17	29421376	29421397	MIR4733; -	22	1	4.59213E-05	2	0.000320155
X	133674224	133674245	MIR450B; -	22	1	4.59213E-05	2	0.000320155
17	27188421	27188442	MIR451; -	22	1	4.59213E-05	3	0.000480232
21	9825859	9825879	MIR3648; +	21	1	4.59213E-05	4	0.000640309
1	247365318	247365343	MIR3916; -	26	0	0	1	0.000160077
3	87275349	87275370	MIR4795; -	22	0	0	1	0.000160077
4	5925006	5925025	MIR378D1; -	20	0	0	1	0.000160077
8	10524498	10524514	MIR4286; +	17	0	0	1	0.000160077
10	59064249	59064270	MIR3924; -	22	0	0	1	0.000160077
10	118927206	118927228	MIR3663; -	23	0	0	1	0.000160077
15	89151352	89151372	MIR1179; +	21	0	0	1	0.000160077
16	30886622	30886639	MIR4519; -	18	0	0	1	0.000160077
16	56892439	56892461	MIR138-2; +	23	0	0	1	0.000160077
17	29902300	29902320	MIR4725; +	21	0	0	1	0.000160077
18	56118319	56118340	MIR122; +	22	0	0	1	0.000160077
18	56118320	56118341	MIR122; +	22	0	0	1	0.000160077
22	42319275	42319295	MIR378I; -	21	0	0	1	0.000160077
1	55691329	55691349	MIR4422; +	21	0	0	2	0.000320155
2	177465752	177465770	MIR1246; -	19	0	0	2	0.000320155
3	195426296	195426317	MIR570; +	22	0	0	2	0.000320155
12	7073276	7073297	MIR141; +	22	0	0	2	0.000320155
17	27717686	27717706	MIR4523; +	21	0	0	2	0.000320155
13	41301964	41301982	MIR320D1; -	19	0	0	5	0.000800387
11	568112	568133	MIR210HG; -	22	0	0	437	0.069953802
13	92002872	92002894	MIR17HG; +	23	0	0	44078	7.055889406

References

1. Mitra A, Skrzypczak M, Ginalski K, Rowicka M. Strategies for achieving high sequencing accuracy for low diversity samples and avoiding sample bleeding using illumina platform. Oudejans C, editor. PLoS ONE. Public Library of Science; 2015;10: e0120520. doi:10.1371/journal.pone.0120520

2. Afgan E, Baker D, van den Beek M, Blankenberg D, Bouvier D, Cech M, et al. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. *Nucleic Acids Res.* 2016;44: W3–W10. doi:10.1093/nar/gkw343
3. Blankenberg D, Gordon A, Kuster Von G, Coraor N, Taylor J, Nekrutenko A, et al. Manipulation of FASTQ data with Galaxy. *Bioinformatics.* 2010;26: 1783–1785. doi:10.1093/bioinformatics/btq281
4. Crooks GE, Hon G, Chandonia J-M, Brenner SE. WebLogo: a sequence logo generator. *Genome Res. Cold Spring Harbor Lab;* 2004;14: 1188–1190. doi:10.1101/gr.849004
5. Hafner M, Renwick N, Farazi TA, Mihailović A, Pena JTG, Tuschl T. Barcoded cDNA library preparation for small RNA profiling by next-generation sequencing. *Methods.* 2012;58: 164–170. doi:10.1016/j.ymeth.2012.07.030