miRDeep* manual

2013-01-10

Jiyuan An, John Lai, Melanie Lehman, Colleen Nelson: Australian Prostate Cancer Research Center (APCRC-Q) and Institute of Health and Biomedical Innovation (IHBI), Queensland University of Technology (QUT), Brisbane, Australia
j.an@qut.edu.au
+61 7 3443 7280

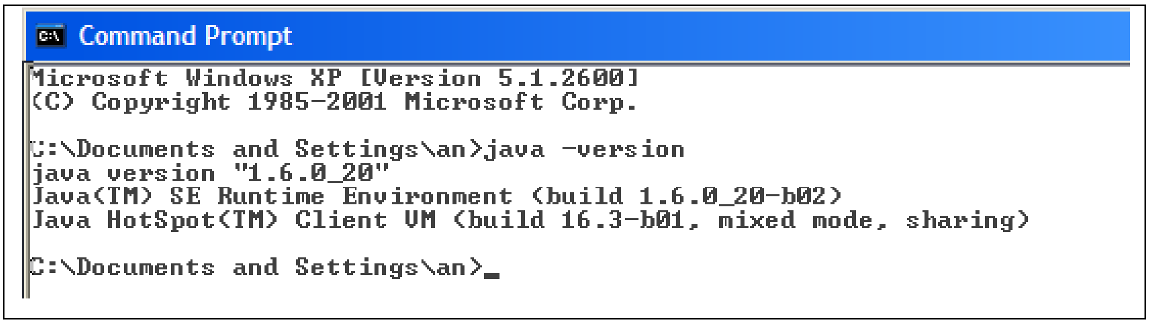
Web page:
http://www.australianprostatecentre.org/research/software/mirdeep-star
https://sourceforge.net/projects/mirdeepstar/

1. Execution with Demo-data

1.1 Install JDK

a. Download JDK 6 (version >= 1.6.0_20) to run miRDeep* from http://jdk6.java.net/.

b. After installing JDK, the java command will run in a command prompt (Windows) as shown in Figure 1 or any other terminal (Linux). If version information is not available from the command prompt, then check whether the java path is in the environment variable.



```
Microsoft Windows XP [Version 5.1.2600]
(C) Copyright 1985-2001 Microsoft Corp.

C:\Documents and Settings\an>java -version
java version "1.6.0_20"
Java(TM) SE Runtime Environment (build 1.6.0_20-b02)
Java HotSpot(TM) Client VM (build 16.3-b01, mixed mode, sharing)

C:\Documents and Settings\an>_
```

Figure 1 - Check whether JDK has been installed in windows.

1.2 Download files

a. Download jar and demo data files from http://www.australianprostatecentre.org/research/software/mirdeep-star.

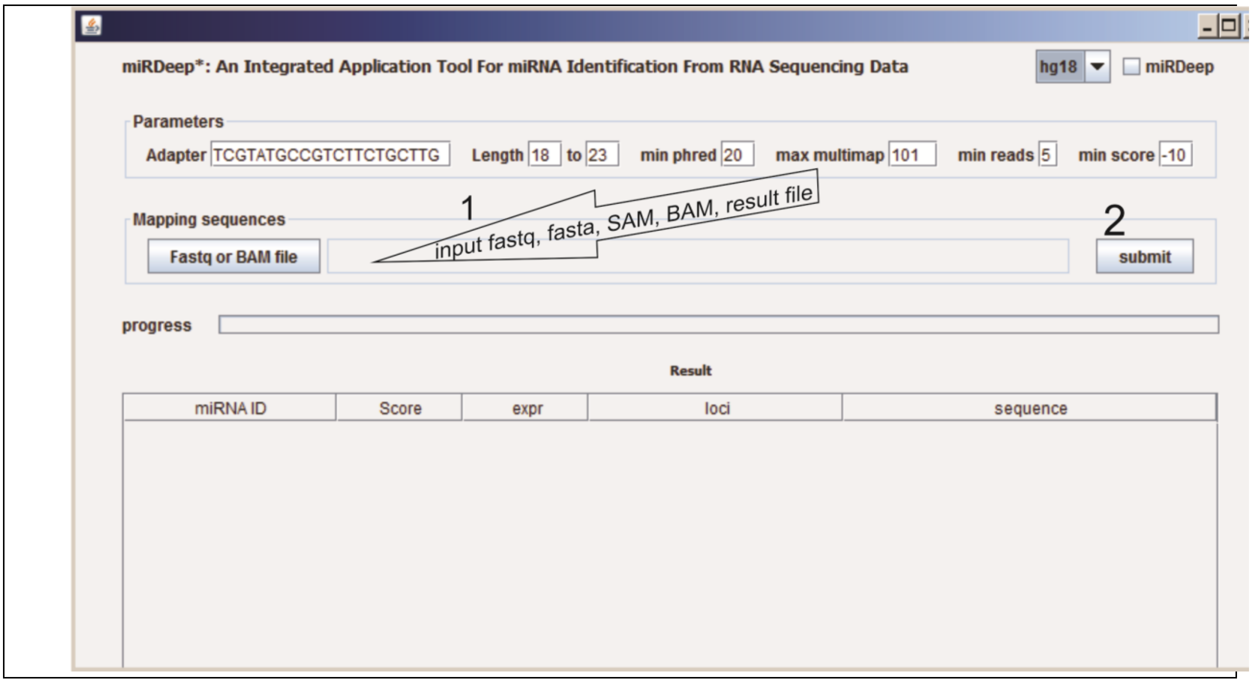
b. There are "miRDeep_star.sh" and "miRDeep_star.bat" files in the unzipped directory. Windows uses the "miRDeep_star.bat" file to run miRDeep*, while "miRDeep_star.sh" is used in Linux.

c. In Linux, type the following command making the file executable.
>chmod +x miRDeep.sh
To execute miRDeep*, type "miRDeep.sh" in command line:
>miRDeep.sh

d. In Windows, double click "miRDeep_star.bat" to run the tool.

A window (as shown in Figure 2) will be displayed after execution of the program.

e. To load raw sequence data in fastQ file, click the “Fastq”, “fasta” or “BAM” file button (labelled 1 in Figure 2). The file selection pop up window will appear.
f. In the unzipped directory, there is a directory called “dat” where you will find a demo data file: “limitDemo.fastq”.
g. miRDeep* accepts the following formats: fastq(“xxx.fastq”), fasta (“xxx.fa”), SAM/BAM(“xxx.sam” or “xxx.bam”) and result (“xxx.result”). Details are shown in the end of this README.
h. To start the whole process of genomic mapping, and identifying miRNAs, select the file and click the submit button (labelled 2 in Figure 2) on the right hand side.



miRDeep*: An Integrated Application Tool For miRNA Identification From RNA Sequencing Data

hg18 ☐ miRDeep

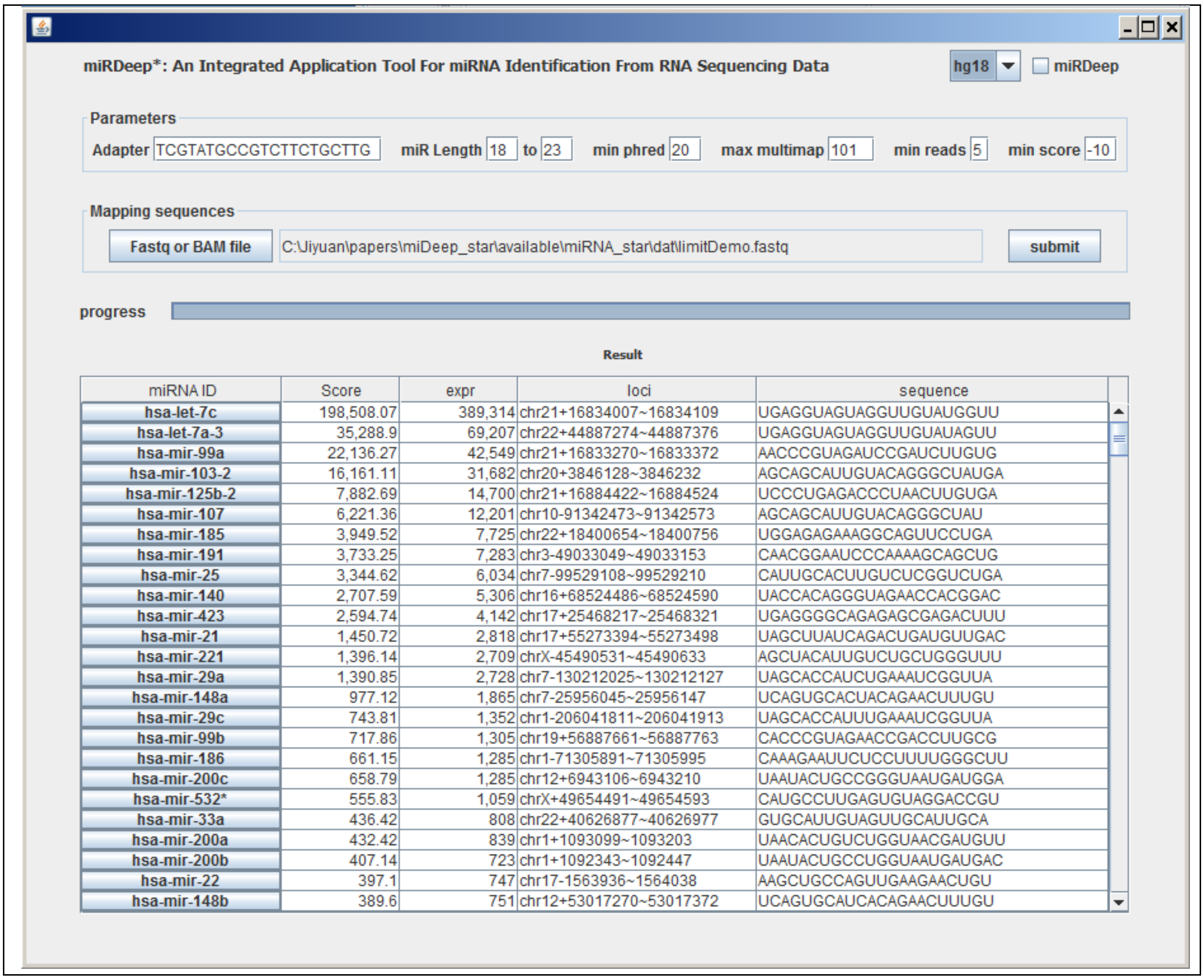
Parameters

Adapter TCGTATGCCGTCTTCTGCTTG Length 18 to 23 min phred 20 max multimap 101 min reads 5 min score -10

Figure 2 screen shot of miRDeep*

1.3 Results

After execution, the identified miRNAs will be shown in a table (as shown in Figure 3). To show the hairpin structure of an identified miRNA, click the miRNA ID in the first column of the table.



miRDeep*: An Integrated Application Tool For miRNA Identification From RNA Sequencing Data

hg18 ☐ miRDeep

Parameters

Adapter TCGTATGCCGTCTTCTGCTTG miR Length 18 to 23 min phred 20 max multimap 101 min reads 5 min score -10

Mapping sequences

C:\Jiyuan\papers\miDeep_star\available\miRNA_star\dat\limitDemo.fastq

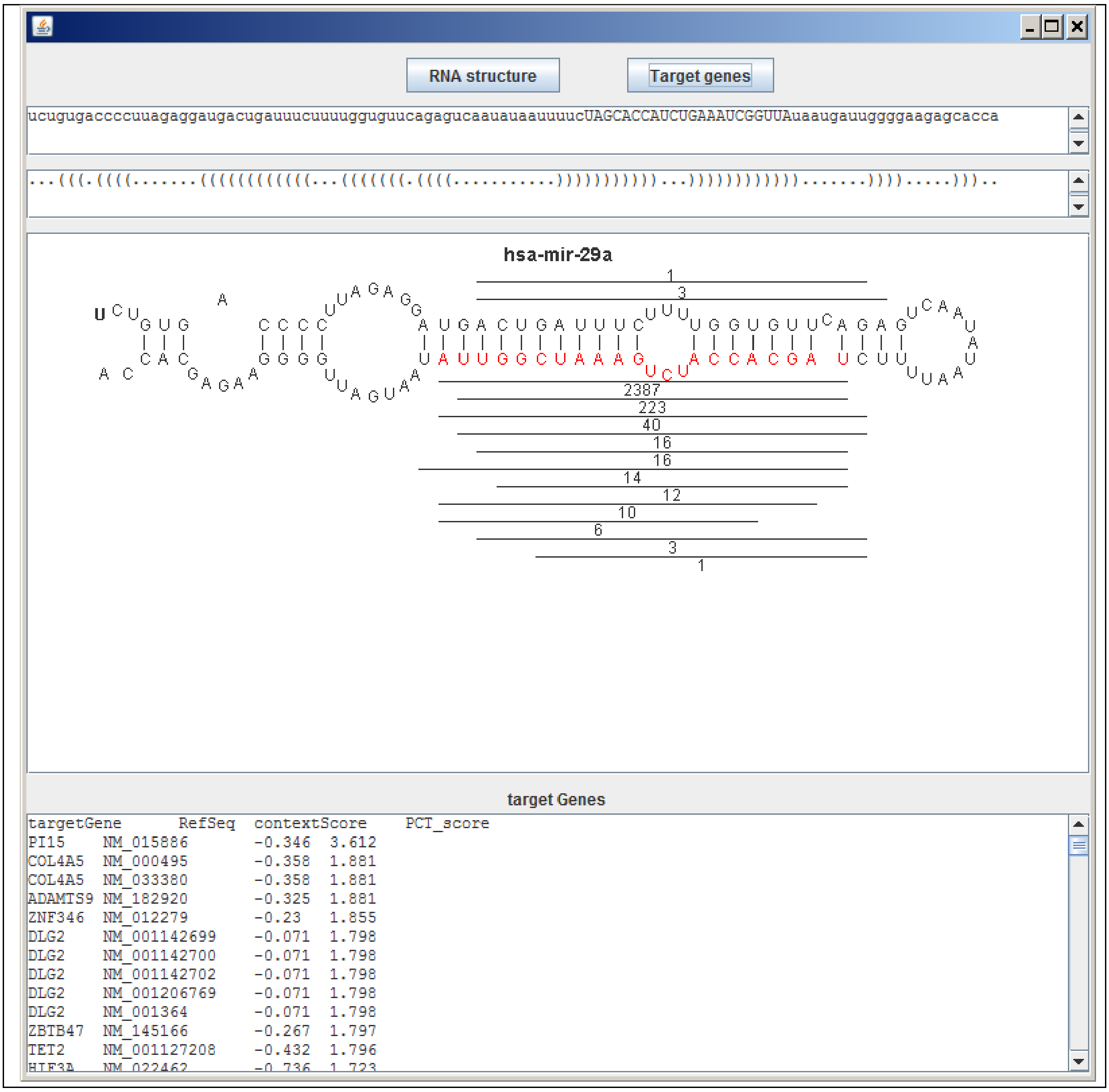
miRNA ID	Score	expr	loci	sequence
hsa-let-7c	198,508.07	389,314	chr21+16834007~16834109	UGAGGUAGUAGGUUGUAUGGUU
hsa-let-7a-3	35,288.9	69,207	chr22+44887274~44887376	UGAGGUAGUAGGUUGUAUAGUU
hsa-mir-99a	22,136.27	42,549	chr21+16833270~16833372	AACCCGUAGAUCCGAUCUUGUG
hsa-mir-103-2	16,161.11	31,682	chr20+3846128~3846232	AGCAGCAUUGUACAGGGCUAUGA
hsa-mir-125b-2	7,882.69	14,700	chr21+16884422~16884524	UCCCUGAGACCCUAACUUGUGA
hsa-mir-107	6,221.36	12,201	chr10-91342473~91342573	AGCAGCAUUGUACAGGGCUAU
hsa-mir-185	3,949.52	7,725	chr22+18400654~18400756	UGGAGAGAAAGGCAGUUCCUGA
hsa-mir-191	3,733.25	7,283	chr3-49033049~49033153	CAACGGAAUCCCAAAAGCAGCUG
hsa-mir-25	3,344.62	6,034	chr7-99529108~99529210	CAUUGCACUUGUCUCGGUCUGA
hsa-mir-140	2,707.59	5,306	chr16+68524486~68524590	UACCACAGGGUAGAACCACGGAC
hsa-mir-423	2,594.74	4,142	chr17+25468217~25468321	UGAGGGGCAGAGAGCGAGACUUU
hsa-mir-21	1,450.72	2,818	chr17+55273394~55273498	UAGCUUAUCAGACUGAUGUUGAC
hsa-mir-221	1,396.14	2,709	chrX-45490531~45490633	AGCUACAUUGUCUGCUGGGUUU
hsa-mir-29a	1,390.85	2,728	chr7-130212025~130212127	UAGCACCAUCUGAAAUCGGUUA
hsa-mir-148a	977.12	1,865	chr7-25956045~25956147	UCAGUGCACUACAGAACUUUGU
hsa-mir-29c	743.81	1,352	chr1-206041811~206041913	UAGCACCAUUUGAAAUCGGUUA
hsa-mir-99b	717.86	1,305	chr19+56887661~56887763	CACCCGUAGAACCGACCUUGCG
hsa-mir-186	661.15	1,285	chr1-71305891~71305995	CAAAGAAUUCUCCUUUUGGGCUU
hsa-mir-200c	658.79	1,285	chr12+6943106~6943210	UAAUACUGCCGGGUAAUGAUGGA
hsa-mir-532*	555.83	1,059	chrX+49654491~49654593	CAUGCCUUGAGUGUAGGACCGU
hsa-mir-33a	436.42	808	chr22+40626877~40626977	GUGCAUUGUAGUUGCAUUGCA
hsa-mir-200a	432.42	839	chr1+1093099~1093203	UAACACUGUCUGGUAACGAUGUU
hsa-mir-200b	407.14	723	chr1+1092343~1092447	UAAUACUGCCUGGUAAUGAUGAC
hsa-mir-22	397.1	747	chr17-1563936~1564038	AAGCUGCCAGUUGAAGAACUGU
hsa-mir-148b	389.6	751	chr12+53017270~53017372	UCAGUGCAUCACAGAACUUUGU

Figure 3 result screen of miRDeep*

In the hairpin structure pop up window (Figure 4), several extra functions are available:

a. User can change the sequence in the top textbox. To show modified sequence: change the sequence in the first textbox, and then click the “RNA structure” button.

b. Click “target genes” to show the target genes for the miRNA. This takes several minutes.



```
ucugugaccccuuagaggaugacugauuucuuuugguguucagagucaauauaauuuucUAGCACCAUCUGAAAUCGGUUAuaaugauuggggaagagcacca
```

```
...(((.((((.......((((((((((((...(((((((.((((...........)))))))))))...))))))))))))).......)))).....)))..
```

```
targetGene      RefSeq   contextScore    PCT_score
PI15    NM_015886       -0.346  3.612
COL4A5  NM_000495       -0.358  1.881
COL4A5  NM_033380       -0.358  1.881
ADAMTS9 NM_182920       -0.325  1.881
ZNF346  NM_012279       -0.23   1.855
DLG2    NM_001142699    -0.071  1.798
DLG2    NM_001142700    -0.071  1.798
DLG2    NM_001142702    -0.071  1.798
DLG2    NM_001206769    -0.071  1.798
DLG2    NM_001364       -0.071  1.798
ZBTB47  NM_145166       -0.267  1.797
TET2    NM_001127208    -0.432  1.796
HIF3A   NM_022462       -0.736  1.723
```

Figure 4 HairPin structure

2. Custom sequence data

2.1. Input file in fastQ format

fastQ has four lines to represent one read.

If your sequence data does not have a fourth line (quality line), simply input an "SSSS..." sequence, with the same length of read, in place of the fourth line.

```
@seq_0
ATGGGTTTGCAGTCCTCGGTTTAAAAAAAAGACGTC
+
SNSNISPSNENSSPINNESNSNIIIOIIONAAAAA<
```

2.2. Input file in fasta format

In the fasta format file, the copy number is appended in the description line within the delimitation table. For example,

```
>t1		234
AGGCGATCACGTAGATTT
```

2.3. Input file in SAM/BAM format

miRDeep* also supports identified miRNAs from aligned sequences. There are several alignment algorithms (such as bowtie, BAW, and soap2). The aligned format should be in SAM or BAM format.

2.4. Show the results identified previously

miRDeep* saves the result file as “XXX.result”. You can load this file in miRDeep* to show the structure.

3. miRDeep* used in other human assemblies or species

3.1 Change human genome assembly hg18 to hg19

download and unzip hg19.zip file from http://sourceforge.net/mirdeepstar. copy hg19 folder into miRDeep_star/genome. Restart miRDeep* and select hg19 from genome combobox in the interface.

3.2 Other species

Download and unzip build_bwt_idx from http://sourceforge.net/mirdeepstar. follow the readme to create index files. Copy created index file folder into miRDeep_star/genome. the new species genome is shown in genome combobox.