

mirTrans: a resource of transcriptional regulation on microRNAs for human cell lines

Xu Hua^{1,†}, Rongjun Tang^{1,†}, Xiuxiu Xu¹, Zhi Wang¹, Qi Xu¹, Luxiao Chen¹,
Edgar Wingender², Jie Li^{1,3,*}, Chenyu Zhang^{1,*} and Jin Wang^{1,*}

¹The State Key Laboratory of Pharmaceutical Biotechnology and Jiangsu Engineering Research Center for MicroRNA Biology and Biotechnology, NJU Advanced Institute for Life Sciences (NAILS), School of Life Science, Nanjing University, Nanjing 210023, China, ²Institute of Bioinformatics, University Medical Center Goettingen, Goldschmidtstrasse 1, D-37077 Goettingen, Germany and ³Nanjing Shengnongding Biotech. Co. Ltd., Nanjing 211100, China

Received August 15, 2017; Revised September 26, 2017; Editorial Decision October 11, 2017; Accepted October 16, 2017

ABSTRACT

The cell-specific information of transcriptional regulation on microRNAs (miRNAs) is crucial to the precise understanding of gene regulations in various physiological and pathological processes existed in different tissues and cell types. The database, mirTrans, provides comprehensive information about cell-specific transcription of miRNAs including the transcriptional start sites (TSSs) of miRNAs, transcription factor (TF) to miRNA regulations and miRNA promoter sequences. mirTrans also maps the experimental H3K4me3 and DHS (DNase-I hypersensitive site) marks within miRNA promoters and expressed sequence tags (ESTs) within transcribed regions. The current version of database covers 35 259 TSSs and over 2.3 million TF-miRNA regulations for 1513 miRNAs in a total of 54 human cell lines. These cell lines span most of the biological systems, including circulatory system, digestive system and nervous system. Information for both the intragenic miRNAs and intergenic miRNAs is offered. Particularly, the quality of miRNA TSSs and TF-miRNA regulations is evaluated by literature curation. 23 447 TSS records and 2148 TF-miRNA regulations are supported by special experiments as a result of literature curation. EST coverage is also used to evaluate the accuracy of miRNA TSSs. Interface of mirTrans is friendly designed and convenient to make downloads (<http://mcube.nju.edu.cn/jwang/lab/soft/mirtrans/> or <http://120.27.239.192/mirtrans/>).

INTRODUCTION

The microRNA (miRNA) is a subset of non-coding RNAs regarded as ‘small RNAs of big roles’ (1). It prevents or suppresses protein translation by binding to the 3’UTR of mRNA. The target mRNA of miRNA is usually in vast numbers, which enables miRNA to participate in almost all the biological processes. The essential roles of miRNA in gene regulation attract huge research interests, which in turn drives the rapid development of miRNA study. In recent years, miRNAs are also widely recognized as the potential biomarker for clinical diagnoses, treatments and prognoses of various kinds of diseases, especially cancers (2,3). Till 2014, about twenty years after the first miRNA was found (4), 28645 hairpin precursor miRNAs (pre-miRNAs) and 35828 mature miRNAs from 223 species have been recorded in the most comprehensive repository of miRNAs, miR-Base (version 21) (5). For human, 1881 pre-miRNAs and 2588 mature miRNAs were revealed.

A full understanding of miRNA function depends on the knowledge of miRNA regulation network, including the upstream transcriptional regulation and their downstream targets. Compared to the abundance of miRNA expression databases and target databases, the resource for regulating miRNA expression is far behind. The main reason is that the 5’ end of primary miRNA transcript is rapidly sliced by Drosha after transcription so that it is very hard to capture the full-length transcript of miRNA (6,7). This eventually lead to the difficulties in the gain of accurate miRNA transcriptional start sites (TSSs), miRNA promoters and the regulations of transcription factors (TFs) on the miRNA transcription (i.e. TF-miRNA regulations). Several approaches have been conducted by combining high throughput data with appropriate algorithms or literature mining. A few databases about miRNA transcription were

*To whom correspondence should be addressed. Email: jwang@nju.edu.cn
Correspondence may also be addressed to Jie Li. Email: jlee@nju.edu.cn
Correspondence may also be addressed to Chenyu Zhang. Email: cyzhang@nju.edu.cn

†These authors contributed equally to this work as first authors.

built to provide information of either TF-miRNA regulation or miRNA TSSs. ChIPBase (8) provides abundant resources of TF-miRNA regulations and other TF-target regulations by parsing ChIP data. CircuitsDB (9) discovered TF-miRNA regulation circuits for human and mouse. TransmiR (10) offers highly reliable yet somewhat limited number (735 in total) of TF-miRNA regulations with literature curation. However, without accurate annotation of miRNA TSSs and promoters, it's hard to tell whether the parsed TF truly functions in the transcription of the miRNA that is downstream to the binding site of the TF (i.e. TFBS). While miRT (11) provides miRNA TSSs that are collected from a couple of experiments, no TF-miRNA regulations are offered. All of these databases developed earlier did not tackle the problem of cell-specificity of miRNA transcription.

Recent studies have shown that miRNAs expression is distinctly tissue-specific or cell-type specific, which are fundamental to tissue development and function maintenance (12). Increasing number of tissue specific miRNAs are reported to be disease associated (13–16). These miRNAs could be of clinical potentials as biomarkers for the diagnoses, treatments and prognoses of diseases (2,3). So, the knowledge about cell-specific transcription of miRNAs is eagerly demanded to facilitate the investigation of miRNAs in fundamental biological aspects as well as in clinical applications. TSmir (17) offers the regulations of TFs on tissue-specific miRNAs for 12 tissues based on ChIP-seq data. DIANA-miRGen (18) provides information of both miRNA TSSs and TF-miRNA regulations for 428 intergenic miRNAs in nine cell lines including human and mouse.

Here, we introduce mirTrans database to provide comprehensive information about cell-specific transcription of miRNA. mirTrans offers abundant cell-specific data which includes both miRNA TSSs and TF-miRNA regulations, together with some supportive data derived from other public resources [e.g. H3K4me3 peaks and DHS peaks]. Up to now, it contains 35 259 TSSs and 2 340 406 TF-miRNA regulations for 1513 miRNAs in 54 human cell lines. The 54 cell lines belong to 17 tissues which cover most of biological systems, including circulatory system, digestive system, nervous system, reproductive system and respiratory system. So, the database could be applied in quite a variety of biological or medical fields. Information of both intragenic miRNAs and intergenic miRNAs is reported. Specifically, quality evaluation on every miRNA TSS and TF-miRNA regulation is presented by literature evidence and EST evidence. Literature evidence gives the documentary support to miRNA TSS and TF-miRNA regulations, and EST evidence evaluates the accuracy of miRNA TSS by EST coverage. mirTrans also provides a convenient and friendly interface for customers. Flexible query is allowed by either one of the three types of keywords (i.e. miRNA, cell line and TF) or any combinations among them. Genomic distribution graph illustrates the intuitive view of miRNA cell-specific transcription. All the data in mirTrans are offered for free downloads.

RESULTS

Database content

mirTrans generally covers two types of data, i.e. the core data and the supportive data. The core data were produced on our own, which mainly include cell-specific miRNA TSSs and TF-miRNA regulations; the supportive data are derived from other resources, which are the primary materials for either the production or the quality evaluation of core data. Figure 1 illustrates how these data were produced or applied.

Core data

Cell-specific miRNA TSSs. TSS data are produced through a python implementation that is developed on the bases of a method previously published by the authors (19). This implementation allows the input in the form of ‘.bed’, where the TSS-related segments are recorded. The ‘.bed’ file could be obtained when peak-calling tools [e.g. MACS2 (20) and Homer (21)] are used to catch the peak areas in TSS-related sequencing data (e.g. H3K4me3, Pol II and GRO/PRO-seq). With the help of the implementation, it is convenient to integrate multiple types of high-throughput data for the expansion of TSS data. In the future, TSS data in mirTrans will be updated once a certain amount of additional TSS-related data are collected. Up to now, the TSSs in mirTrans are predicted based on H3K4me3 and DHSs (19).

Usually, predictions on miRNA TSSs are focused on intergenic miRNAs, because intragenic miRNAs are considered to be co-transcribed with its host gene. However, more and more evidences have shown that intragenic miRNA is also capable of transcription independent to its host gene (22). Our implementation has a special module to tell whether the transcription is dependent or independent on its host gene, and also returned the right TSSs for every intragenic miRNA in certain cell lines.

In particular, mirTrans provides quality evaluations for all the TSSs, which is quite helpful for customers to make choice. The quality of TSSs is annotated in two ways. One is so called ‘ESTs (expressed sequence tags) evidence’, which evaluates TSS quality with the significance of EST coverage difference around TSS. EST is the expressed segment of transcripts. Theoretically, significant differences of EST coverage between the upstream and the downstream of TSS should be observed. The less the P value of EST evidence is, the higher the TSS quality is. The additional evaluation is termed as ‘literature evidence’, which annotates TSS quality by literature curation. ‘Strong evidence’ means the supports from literatures of traditional molecular experiments, while ‘evidence’ refers to the supports from GRO-Seq, CAGE and other high-throughput transcriptional signals (23–25).

Cell-specific TF-miRNA regulations. TF-miRNA regulation is produced based on the data of miRNA TSS. In general, the promoter that is 1kb upstream to the cell-specific miRNA TSSs is scanned with geneXplain platform (<http://platform.genexplain.com>) in order to obtain the cell-specific transcription factor binding sites (TFBSs). In particular, for the cell lines with DHSs data, TFBSs are scanned

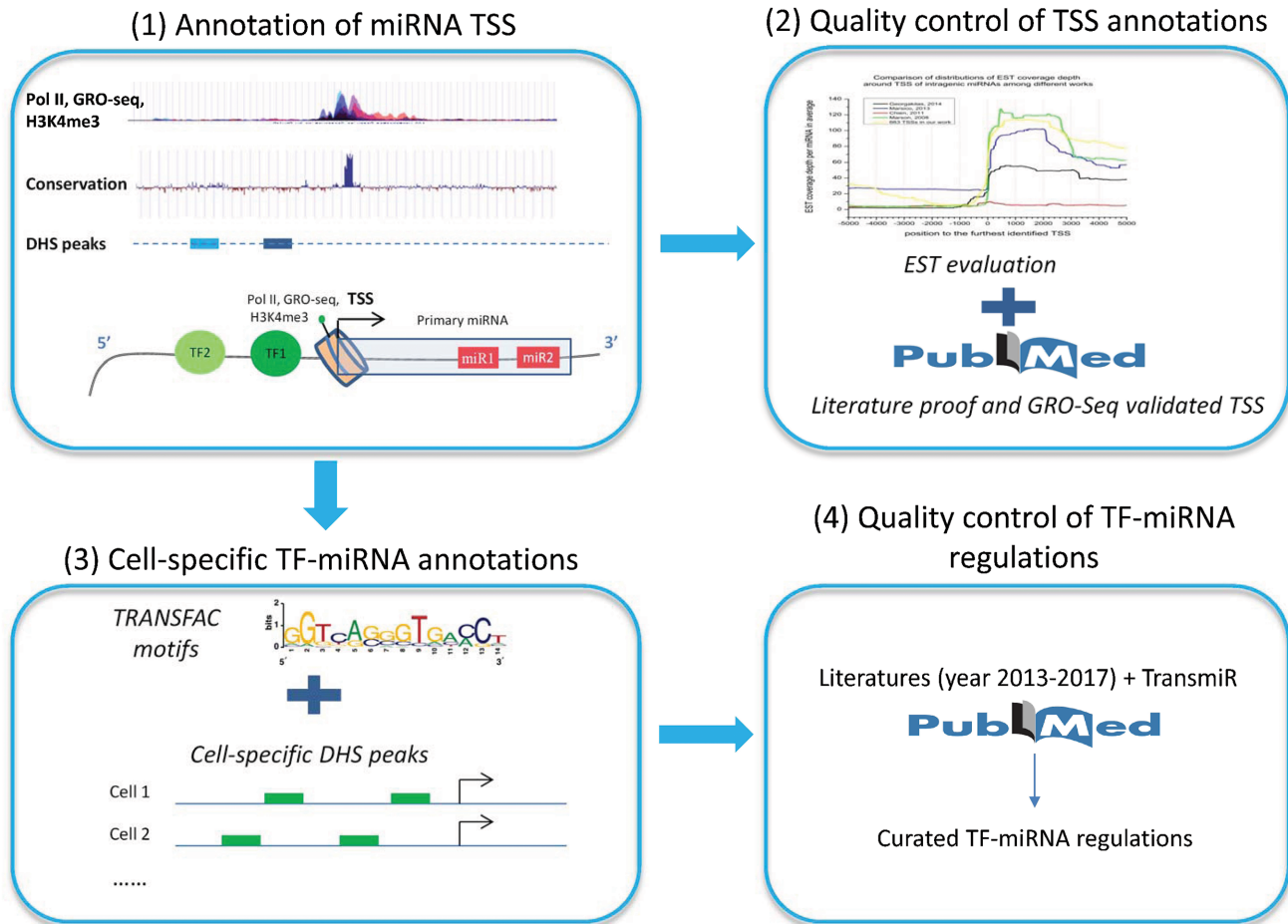


Figure 1. Workflow for the production of core data.

on the DHS segments within the promoter, since DHSs data gives the marks for TFBS regions (26). TF-miRNA regulations are obtained when the related TFs are retrieved from TRANSFAC (27) by TFBSs. Since ChIP-seq data for TF binding is able to provide the experiment-proved TFBSs for certain TFs, it will be employed for the refinement of cell-specific TF-miRNA regulations in the future, as the updating of mirTrans.

Two parameters i.e. the affinity score and the conservation scores, are additionally offered for TFBSs. The affinity score which scales the affinity of TF binding to TFBS is calculated with MATCH (28) provided by TRANSFAC, while the conservation score which measures the conservation of TFBS sequence across species is calculated based on data of UCSC track ‘phastCons100way’ (29,30). Till now, >2.3 million cell specific TF-miRNA regulations are available in mirTrans.

The quality of cell-specific TF-miRNA regulations is also verified with literature evidences. The literature pool was constructed on the basis of TransmiR database (<http://www.cuilab.cn/transmir>) combined with new papers from PubMed during the year 2013–2017. The TF-miRNA regulations of ‘strong evidence’ are supported by promoter-related experiments, e.g. ChIP and luciferase report assays detecting the activity of promoters (31); while the regula-

tions of ‘evidence’ are supported by gene expression experiments or the computational analysis based on expression data (32). The other TF-miRNA regulations recorded in this database were only based on DHS evidence.

Supportive data

So far, H3K4me3 peaks, DHSs peaks, ESTs and pre-miRNAs are the supportive data in mirTrans. H3K4me3 and DHSs data were collected from ENCODE (33), ESTs data were derived from UCSC (34), and pre-miRNAs were derived from miRBase (miRBase version 21). H3K4me3, DHSs, and pre-miRNA data are the source data for the production of cell-specific miRNA TSSs and cell-specific TF-miRNA regulations. ESTs are used to evaluate the quality of TSSs.

Most of the supportive data are graphically illustrated together with miRNA TSSs and TFBSs when miRNA TSSs are retrieved. Theoretically, H3K4me3 peak represents the location of miRNA TSS, DHSs peak represent the region of TFBSs, and EST the position of transcripts. Therefore, the genomic distributions of all these data gives the intuitive illustration of cell-specific transcription of miRNAs.

Table 1. Summary of the contents in mirTrans and other comparable databases/data sets

Database	Cell lines	miRNAs	miRNA TSSs	miRNA TSSs with literature evidence	TF-miRNA regulations	TF-miRNA regulations with literature evidence
mirTrans	54	1513 (546 intergenic / 967 intragenic)	35259	23447	2340406	2148
Liu Q et al. [24] (2017)	27	480 (intergenic)	5839	N/A	1893	9
de Rie D et al. [25] (2017)	84	1357 (371 intergenic / 986 intragenic)	1118	N/A	N/A	N/A
ChIPBase (2016)	N/A	unknown	N/A	N/A	273761	None
DIANA-miRGen (2016)	9	428 (intergenic)	276	None	unknown	None
TSmiR (2014)	12 tissues	116	N/A	N/A	2347	None
CircuitsDB (2013)	N/A	180	N/A	N/A	115 TFs to 180 miRNAs	None
miRT (2012)	N/A	588 (206 intergenic / 382 intragenic)	670	None	N/A	N/A
TransmiR (2009)	N/A	100	N/A	N/A	735	735

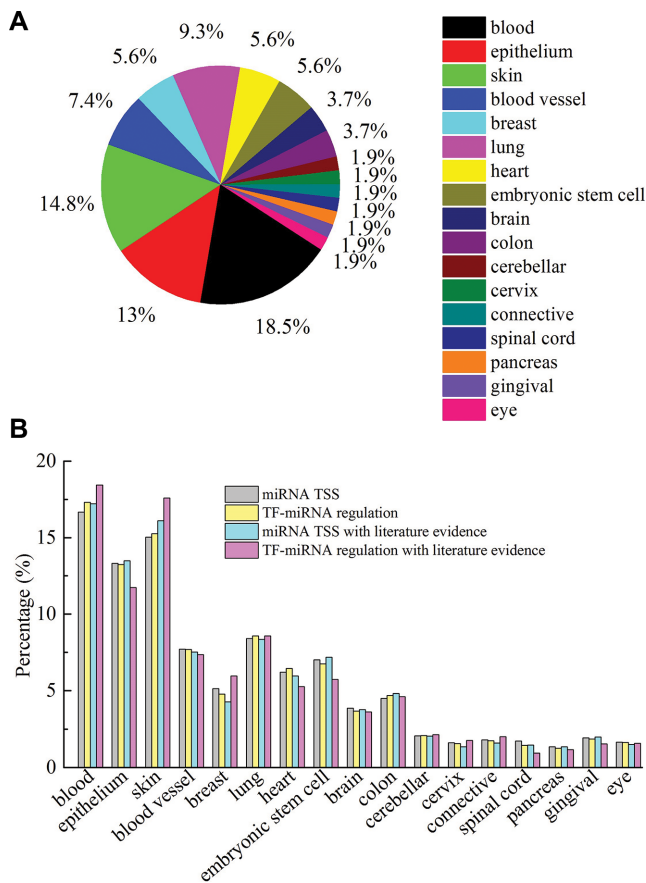


Figure 2. Statistics on the data in mirTrans. (A) Distribution of cell lines in 17 tissues; (B) distribution of miRNA TSSs, TF-miRNA regulations, miRNA TSSs with literature evidence, TF-miRNA regulations with literature evidence in 17 tissues.

Database statistics

Up to now, 35 259 TSSs and 2 340 406 TF-miRNA regulations for 1513 pre-miRNAs (546 intergenic miRNA

and 967 intragenic miRNA) in 54 human cell lines are recorded in mirTrans. Comparing to the previous comparable databases, mirTrans offers the most abundant repository for the cell-specific transcription of miRNAs (see Table 1). Two recent researches reported a large amount of miRNA transcriptional regulation data by using GRO/PRO-seq and CAGE techniques respectively (24,25). The statistics on the results of these two papers are also shown in Table 1. In addition, literature evidence is provided to indicate the data quality. So far, there are all together 23 447 TSS records and 2148 TF-miRNA regulations supported by the specified experimental evidences as described above and which are regarded as the most reliable data. Note that literature evidence will be increasingly accumulated as the update of mirTrans.

The 54 cell lines in mirTrans come from 17 tissues, which cover most of the biological systems including circulatory system, digestive system, nervous system, reproductive system and respiratory system (see Figure 2). About 30% of cell lines are cancer cells. This enables mirTrans a broad range of application in cancer study.

The segment that starts from miRNA TSS to the end of pre-miRNA is annotated as the confident transcript of miRNA. Statistics on EST distribution show that 96.5% (10627/11016) of the confident transcripts are covered by at least one EST. Moreover, the coverage difference of EST between the upstream and the downstream of miRNA TSS is calculated for quality evaluation of TSSs (i.e. EST evidence mentioned above). 78.4% (8640/11016) of TSSs are with significant difference of EST coverage ($P < 0.05$, by *t*-test), which indicates the high-quality of miRNA TSSs in mirTrans.

Database implementation

System facility. All the data in mirTrans are well organized with MySQL. The website is implemented with PHP, JavaScript and HTML. Genomic graphs are illustrated using 'Gviz' package of Bioconductor.

A Search

You can search in miRTrans in various ways by entering an miRNA ID, Cell-line or TF in the search box (case-insensitive). These are as follows:

- miRNA
- Cell-line
- Transcription Factor (TF)

Simple Search

Search by miRNA: (e.g. hsa-mir-17,7a-1 or 6b)

Search by Cell-line: Cell-line:

Search by TF: TF: (e.g. MYC, ah, egr)

Advanced Search

Search by miRNA+Cell-line: miRNA: Cell-line:

Search by miRNA+TF: miRNA: TF:

Search by Cell-line+TF: Cell-line: TF:

Search by miRNA+Cell-line+TF: Cell-line: TF: miRNA:

B Search by miRNA

Information of hsa-mir-17
1. Genomic view for hsa-mir-17

2. Detailed information

miRNA	miRNA location	TSS	Distance from TSS to miRNA	Cell line	TF	EST evidence(p_value)	Literature evidence
hsa-mir-17	chr13:91350605-91350688+	chr13:91347919+	2686	HMEC	ASCL1, BACH1	0	strong evidence
hsa-mir-17	chr13:91350605-91350688+	chr13:91348409+	2196	HFF_Myc	ASCL1, BACH1	0	strong evidence
hsa-mir-17	chr13:91350605-91350688+	chr13:91349171+	1434	CACO2	ASCL1, BACH1	0	strong evidence
hsa-mir-17	chr13:91350605-91350688+	chr13:91349202+	1403	AG04449	ASCL1, BACH1	0	strong evidence
hsa-mir-17	chr13:91350605-91350688+	chr13:91349202+	1403	AG10803	ASCL1, BACH1	0	strong evidence
hsa-mir-17	chr13:91350605-91350688+	chr13:91349202+	1403	H7_hESC_T5	ASCL1, BACH1	0	strong evidence
hsa-mir-17	chr13:91350605-91350688+	chr13:91349202+	1403	HCPePic	ASCL1, BACH1	0	strong evidence
hsa-mir-17	chr13:91350605-91350688+	chr13:91349202+	1403	HECTO	ASCL1, BACH1	0	strong evidence
hsa-mir-17	chr13:91350605-91350688+	chr13:91349202+	1403	HUVEC	ASCL1, BACH1	0	strong evidence
hsa-mir-17	chr13:91350605-91350688+	chr13:91349202+	1403	SAEC	ASCL1, BACH1	0	strong evidence

C Search by Cell-Line

Information of A549
1. Brief Introduction
Description: epithelial cell line derived from a lung carcinoma tissue
Tissue: epithelium_C
More details: [Encode search at UCSC](#)

Number of records: 567

Cell line	miRNA	miRNA location	TSS	Host gene	Host gene location
A549	hsa-let-7a-1	chr9:94175957-94176036+	chr9:94167251+	ENSG00000269946	chr9:94166289-94200627+
A549	hsa-let-7d	chr9:94178834-94178920+	chr9:94167251+	ENSG00000269946	chr9:94166289-94200627+
A549	hsa-let-7f-1	chr9:94176347-94176433+	chr9:94167251+	ENSG00000269946	chr9:94166289-94200627+
A549	hsa-let-7g	chr3:52268278-52268361-	chr3:52277674-	ENSG00000164091	chr3:52254421-52288020-
A549	hsa-let-7i	chr12:62603686-62603769+	chr12:62602384+	ENSG00000257354	chr12:62600752-62622213+
A549	hsa-mir-101-1	chr1:65069404-65069508-	chr1:65066797-	None	None
A549	hsa-mir-103a-1	chr5:168560896-168560973-	chr5:168578819-	ENSG00000120137	chr5:168548495-168579600-
A549	hsa-mir-1178	chr12:119713634-119713724-	chr12:119876980-	ENSG00000122966	chr12:119685790-119877291-
A549	hsa-mir-1180	chr17:19344406-19344574-	chr17:19362158-	ENSG00000108641	chr17:19337554-19378182-
A549	hsa-mir-1183	chr7:21471058-21471146+	chr7:21428650+	ENSG00000105966	chr7:21426034-21514822+
A549	hsa-mir-1199	chr15:14073361-14073479+	chr19:14073089+	ENSG00000141854	chr15:14072536-14075062+
A549	hsa-mir-1200	chr7:36918337-36919432-	chr7:37448891-	ENSG00000155849	chr7:3694361-37445249-
A549	hsa-mir-1204	chr8:12779592-127796028+	chr8:127794918+	ENSG00000249859	chr8:12779453-128101253+
A549	hsa-mir-1225	chr16:2090195-2090284-	chr16:2128611-	ENSG00000008710	chr16:2088710-2135898-
A549	hsa-mir-1226	chr3:47849555-47849629+	chr3:47826529+	ENSG00000132153	chr3:47802909-47850195+
A549	hsa-mir-1227	chr19:2234052-2234149-	chr19:2238623-	ENSG00000104896	chr19:2230084-2237704-
A549	hsa-mir-1228	chr12:57194504-57194576+	chr12:57170862+	ENSG00000123384	chr12:57129493-57213551+
A549	hsa-mir-1229	chr5:179798278-179798346-	chr5:179806118-	ENSG00000161013	chr5:17979597-17980952-
A549	hsa-mir-1233-2	chr15:34528290-34528371-	chr15:34583052-	ENSG00000215252	chr15:34525207-34588503-
A549	hsa-mir-1234	chr8:144400096-14440165-	chr8:144408931-	ENSG00000071894	chr8:14439329-144409349-

D Search by TF

Detailed Information of MYC
Number of records: 8769

TF	miRNA	Cell line	TFBS	Affinity score	Conservation score	Literature evidence	Details
MYC	hsa-let-7a-1	A549	chr9:94166826-94166836+; chr9:94166824-94166834-	0.913.085	0.016.072	strong evidence	Click
MYC	hsa-let-7a-1	AG04449	chr9:94166826-94166836+; chr9:94166824-94166834-	0.913.085	0.016.072	strong evidence	Click
MYC	hsa-let-7a-1	AG04450	chr9:94166826-94166836+; chr9:94166824-94166834-	0.913.085	0.016.072	strong evidence	Click
MYC	hsa-let-7a-1	AG09309	chr9:94166826-94166836+; chr9:94166824-94166834-	0.913.085	0.016.072	strong evidence	Click
MYC	hsa-let-7a-1	AG09319	chr9:94166826-94166836+; chr9:94166824-94166834-	0.913.085	0.016.072	strong evidence	Click
MYC	hsa-let-7a-1	AG10803	chr9:94166826-94166836+; chr9:94166824-94166834-	0.913.085	0.016.072	strong evidence	Click
MYC	hsa-let-7a-1	AoAF	chr9:94166826-94166836+; chr9:94166824-94166834-	0.913.085	0.016.072	strong evidence	Click
MYC	hsa-let-7a-1	BJ	chr9:94166826-94166836+; chr9:94166824-94166834-	0.913.085	0.016.072	strong evidence	Click
MYC	hsa-let-7a-1	CACO2	chr9:94166826-94166836+; chr9:94166824-94166834-	0.913.085	0.016.072	strong evidence	Click
MYC	hsa-let-7a-1	CD14	chr9:94166826-94166836+; chr9:94166824-94166834-	0.913.085	0.016.072	strong evidence	Click
MYC	hsa-let-7a-1	CD20	chr9:94166826-94166836+; chr9:94166824-94166834-	0.913.085	0.016.072	strong evidence	Click
MYC	hsa-let-7a-1	GM05990	chr9:94166826-94166836+; chr9:94166824-94166834-	0.913.085	0.016.072	strong evidence	Click

E

Download

The complete data in miRTrans are organized in the form of the return table when miRNA/cell line/TF is queried. You could download the forms separately or collectively.

China

- miRNA-query table
- Cell-query table
- TF-query table
- All Tables

Abroad

- miRNA-query table
- Cell-query table
- TF-query table
- All Tables

Figure 3. Interface for data query and download. (A–D) Data query; (E) data download. Red square in (B–D) marks the button for customized download or filter option.

Data query. Data retrieving starts from data query. miRTrans offers two ways of queries, i.e. simple search and advanced search (Figure 3A). Simple search allows the query with a single keyword, while advanced search allows the query with multiple keywords. Fuzzy words are supported in both simple and advanced search.

For simple search, there are three entries which respectively relate to three types of keywords, i.e. miRNA, cell line and TF. Different entries go to different forms of return.

- **Simple search by miRNA** returns a graph followed by a table. The graph shows the genomic distribution of pre-miRNA, miRNA TSS miRNA TFBSs, H3K4me3 peak, DHS peak and ESTs (Figure 3B). The table lists the following information that is associated with the input pre-miRNA: genomic location, TSS location, distance from TSS to pre-miRNA, cell line where the TSS are found, TFs that bind to the region of 1kb upstream to the TSS

(i.e. the miRNA promoter), EST evidence and literature evidence for the TSS (Figure 3B).

- **Simple search by cell line** returns a table where all the pre-miRNAs of which TSSs are found in this cell line, are listed. Additional information in the table includes miRNA location, TSS location of the miRNA, host gene of the miRNA and the location of the host gene (Figure 3C). More details can be retrieved when a miRNA or TSS in table is clicked (Figure 3C).
- **Simple search by TF** returns a table where all the TF-miRNA regulations in which the TF participates are listed. Associated information includes the miRNA that the TF regulates, the cell line where the TF-miRNA regulation happens, the TFBS where the TF binds, affinity score that describes the binding strength of the TF to the TFBS, conservation score of the TFBS sequence, and literature evidence (Figure 3D). Detail information about the literature that supports the TF-miRNA regulation can be referred when the links in ‘Details’ are followed (Figure 3D). In particular, a filter option is offered, with which customers are free to retrieve either the whole results or only the literature-supported ones (see the bottom-left button in Figure 3D). Download of the return table is associated with the filter option (see the bottom-right button in Figure 3D).

Advanced search enables a flexible query with any combinations of the three keywords used in simple search (Figure 3A). The results of the advanced search will be returned in the way similar to that of simple search.

Data download. mirTrans provides two ways of data download, i.e. customized download and the complete download. The customized download is associated with the query that the customer submits. Every result of the query can be downloaded by clicking the ‘download’ button at the bottom of the return page (Figure 3B–D). The complete download is specifically offered in the download page (Figure 3E). Customers are allowed to separately download the complete data in the form of the returned table when miRNA/cell line/TF is inquired. A collective download packing all the forms of data is also provided.

Database maintenance

A new version of mirTrans is released every 6 months in regular. The update includes the refresh of supportive data (e.g. H3K4me3, DHS, EST and pre-miRNA), core data (e.g. cell-specific miRNA TSSs and TF-miRNA regulation), EST evidence and literature evidence. The whole updating process is developed in a semi-automatic way. The supportive data are automatically collected from the related database, the core data and EST evidence are automatically produced through the workflow of our own, and the literature curation are manually made after related literatures are collected from PUBMED (<https://www.ncbi.nlm.nih.gov/pubmed>).

CONCLUSION

The wide participations in gene regulation of biological processes and clinical application potentials of miRNAs in-

trigue a great demand on the knowledge of cell-specific regulations of miRNAs. mirTrans offers abundant and high-quality information about cell-specific transcriptions of human miRNAs. Compared with the current databases reported so far, mirTrans specially advances in that, a significant amount of data is provided with sufficient evidence from high throughput data and literature reports manually curated. Firstly, the information for a total of 54 cell lines covering most of the key biological systems ensures the database applicable to a large variety of biological and medical investigations in different tissues or processes; Secondly, nearly all the human miRNAs known so far residing in either intergenic genomic region or intragenic region are included in mirTrans, while many of the other databases only provide intergenic miRNAs which generally account for <50% of the total miRNAs encoded by human genome; Finally, quality evaluation on either miRNA TSSs or TF-miRNA regulations were made by EST data and literature reports. Concerned with these aspects, mirTrans is very helpful for both the experimental biologists and the system biologists, who are interested in gene regulations, pathways and gene networks where miRNAs are involved.

Recent years have shown an ever rapid increase in the genomic and epigenomic data for diverse samples of tissues and cells. Accordingly, mirTrans will be updated regularly by employing new experimental data and information from literature mining. It will also incorporate cell-specific miRNA transcriptional networks by integrating all the regulations between miRNAs and TFs in the future.

ACKNOWLEDGEMENTS

The authors would acknowledge the Center of High Performance Computation of Nanjing University for the support of computational resources.

FUNDING

China Scholarship Council, DZHK (German Center for Cardiovascular Research); National Science Foundation of China [81250044, 31500674]; NJU-Yangzhou Institute of Optoelectronics. Funding for open access charge: National Science Foundation of China [81250044, 31500674].

Conflict of interest statement. None declared.

REFERENCES

1. He, L. and Hannon, G.J. (2004) MicroRNAs: small RNAs with a big role in gene regulation. *Nat. Rev. Genet.*, **5**, 522–531.
2. van Rooij, E. and Olson, E.N. (2012) MicroRNA therapeutics for cardiovascular disease: opportunities and obstacles. *Nat. Rev. Drug Discov.*, **11**, 860–872.
3. Rupaimoole, R. and Slack, F.J. (2017) MicroRNA therapeutics: towards a new era for the management of cancer and other diseases. *Nat. Rev. Drug Discov.*, **16**, 203–222.
4. Lee, R.C., Feinbaum, R.L. and Ambros, V. (1993) The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. *Cell*, **75**, 843–854.
5. Kozomara, A. and Griffiths-Jones, S. (2014) miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res.*, **42**, D68–D73.
6. Houseley, J., LaCava, J. and Tollervey, D. (2006) RNA-quality control by the exosome. *Nat. Rev. Mol. Cell Biol.*, **7**, 529–539.

7. Chen, C.Y. and Shyu, A.B. (2011) Mechanisms of deadenylation-dependent decay. *Wiley Interdiscipl. Rev. RNA*, **2**, 167–183.
8. Zhou, K.R., Liu, S., Sun, W.J., Zheng, L.L., Zhou, H., Yang, J.H. and Qu, L.H. (2017) ChIPBase v2.0: decoding transcriptional regulatory networks of non-coding RNAs and protein-coding genes from ChIP-seq data. *Nucleic Acids Res.*, **45**, D43–D50.
9. Friard, O., Re, A., Taverna, D., De Bortoli, M. and Cora, D. (2010) CircuitsDB: a database of mixed microRNA/transcription factor feed-forward regulatory circuits in human and mouse. *BMC Bioinformatics*, **11**, 435.
10. Wang, J., Lu, M., Qiu, C. and Cui, Q. (2010) TransmiR: a transcription factor-microRNA regulation database. *Nucleic Acids Res.*, **38**, D119–D122.
11. Bhattacharyya, M., Das, M. and Bandyopadhyay, S. (2012) miRT: a database of validated transcription start sites of human microRNAs. *Genomics Proteomics Bioinformatics*, **10**, 310–316.
12. Roux, J., Gonzalez-Porta, M. and Robinson-Rechavi, M. (2012) Comparative analysis of human and mouse expression data illuminates tissue-specific evolutionary patterns of miRNAs. *Nucleic Acids Res.*, **40**, 5890–5900.
13. Esteller, M. (2011) Non-coding RNAs in human disease. *Nat. Rev. Genet.*, **12**, 861–874.
14. Gama-Carvalho, M., Andrade, J. and Bras-Rosario, L. (2014) Regulation of cardiac cell fate by microRNAs: implications for heart regeneration. *Cells*, **3**, 996–1026.
15. Trionfini, P., Benigni, A. and Remuzzi, G. (2015) MicroRNAs in kidney physiology and disease. *Nat. Rev. Nephrol.*, **11**, 23–33.
16. Issler, O. and Chen, A. (2015) Determining the role of microRNAs in psychiatric disorders. *Nat. Rev. Neurosci.*, **16**, 201–212.
17. Guo, Z., Maki, M., Ding, R., Yang, Y., Zhang, B. and Xiong, L. (2014) Genome-wide survey of tissue-specific microRNA and transcription factor regulatory networks in 12 tissues. *Scientific Rep.*, **4**, 5150.
18. Georgakilas, G., Vlachos, I.S., Zagganas, K., Vergoulis, T., Paraskevopoulou, M.D., Kanellos, I., Tsanakas, P., Dellis, D., Fevgas, A., Dalamagas, T. et al. (2016) DIANA-miRGen v3.0: accurate characterization of microRNA promoters and their regulators. *Nucleic Acids Res.*, **44**, D190–D195.
19. Hua, X., Chen, L., Wang, J., Li, J. and Wingender, E. (2016) Identifying cell-specific microRNA transcriptional start sites. *Bioinformatics*, **32**, 2403–2410.
20. Zhang, Y., Liu, T., Meyer, C.A., Eeckhoutte, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W. et al. (2008) Model-based analysis of ChIP-Seq (MACS). *Genome Biol.*, **9**, R137.
21. Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H. and Glass, C.K. (2010) Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell*, **38**, 576–589.
22. Ribas, J., Ni, X., Castanares, M., Liu, M.M., Esopi, D., Yegnasubramanian, S., Rodriguez, R., Mendell, J.T. and Lupold, S.E. (2012) A novel source for miR-21 expression through the alternative polyadenylation of VMP1 gene transcripts. *Nucleic Acids Res.*, **40**, 6821–6833.
23. Georgakilas, G., Vlachos, I.S., Paraskevopoulou, M.D., Yang, P., Zhang, Y., Economides, A.N. and Hatzigeorgiou, A.G. (2014) microTSS: accurate microRNA transcription start site identification reveals a significant number of divergent pri-miRNAs. *Nat. Commun.*, **5**, 5700.
24. Liu, Q., Wang, J., Zhao, Y., Li, C.I., Stengel, K.R., Acharya, P., Johnston, G., Hiebert, S.W. and Shyr, Y. (2017) Identification of active miRNA promoters from nuclear run-on RNA sequencing. *Nucleic Acids Res.*, **45**, e121.
25. de Rie, D., Abugessaisa, I., Alam, T., Arner, E., Arner, P., Ashoor, H., Astrom, G., Babina, M., Bertin, N., Burroughs, A.M. et al. (2017) An integrated expression atlas of miRNAs and their promoters in human and mouse. *Biotechnol.*, **35**, 872–878.
26. Thurman, R.E., Rynes, E., Humbert, R., Vierstra, J., Maurano, M.T., Haugen, E., Sheffield, N.C., Stergachis, A.B., Wang, H., Vernot, B. et al. (2012) The accessible chromatin landscape of the human genome. *Nature*, **489**, 75–82.
27. Matys, V., Fricke, E., Geffers, R., Gossling, E., Haubrock, M., Hehl, R., Hornischer, K., Karas, D., Kel, A.E., Kel-Margoulis, O.V. et al. (2003) TRANSFAC: transcriptional regulation, from patterns to profiles. *Nucleic Acids Res.*, **31**, 374–378.
28. Kel, A.E., Gossling, E., Reuter, I., Cheremushkin, E., Kel-Margoulis, O.V. and Wingender, E. (2003) MATCH: A tool for searching transcription factor binding sites in DNA sequences. *Nucleic Acids Res.*, **31**, 3576–3579.
29. Pollard, K.S., Hubisz, M.J., Rosenbloom, K.R. and Siepel, A. (2010) Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Res.*, **20**, 110–121.
30. Siepel, A., Bejerano, G., Pedersen, J.S., Hinrichs, A.S., Hou, M., Rosenbloom, K., Clawson, H., Spieth, J., Hillier, L.W., Richards, S. et al. (2005) Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res.*, **15**, 1034–1050.
31. Zhang, Y.C., Ye, H., Zeng, Z., Chin, Y.E., Huang, Y.N. and Fu, G.H. (2015) The NF-kappaB p65/miR-23a-27a-24 cluster is a target for leukemia treatment. *Oncotarget*, **6**, 33554–33567.
32. Li, P., Hu, Y., Yi, J., Li, J., Yang, J. and Wang, J. (2015) Identification of potential biomarkers to differentially diagnose solid pseudopapillary tumors and pancreatic malignancies via a gene regulatory network. *J. Transl. Med.*, **13**, 361.
33. Kavanagh, D.H., Dwyer, S., O'Donovan, M.C. and Owen, M.J. (2013) The ENCODE project: implications for psychiatric genetics. *Mol. Psychiatry*, **18**, 540–542.
34. Karolchik, D., Hinrichs, A.S., Furey, T.S., Roskin, K.M., Sugnet, C.W., Haussler, D. and Kent, W.J. (2004) The UCSC Table Browser data retrieval tool. *Nucleic Acids Res.*, **32**, D493–D496.