Chapter 3
Atlas of miRNAs and Their Promoters in Human and Mouse

Michiel Jan Laurens de Hoon

Keywords Gene expression · MicroRNA biogenesis · Regulatory sites · Noncoding RNAs · Cap Analysis Gene Expression (CAGE) · Transcription start sites · Transcription · Functional annotation · Vertebrate

3.1 Preamble

MicroRNAs (miRNAs) are a class of short (typically 21–24 nucleotides) noncoding RNAs that inhibit the expression of specific genes by binding to complementary target sequences on RNA transcripts, usually located in the 3′ untranslated region (UTR), and repressing translation or inducing RNA degradation. The miRNA biogenesis pathway consists of excision by Drosha of the precursor miRNA (pre-miRNA) from a primary miRNA (pri-miRNA) transcript, followed by processing of the pre-miRNA by Dicer to release the mature miRNA duplex consisting of a guide RNA and a passenger strand RNA. Mature miRNAs have a phosphate group on their 5′ end and a hydroxyl group on their 3′ end; short RNA (sRNA) sequencing libraries for miRNA expression profiling by next-generation sequencing instruments can be produced by ligating adapters to the 5′ and the 3′ end of the mature miRNA, followed by reverse transcription, PCR amplification, and gel purification (Illumina, Inc. 2011).

As pri-miRNAs are long transcripts with a 5′ cap, their expression can be profiled using Cap Analysis Gene Expression (CAGE; Takahashi et al. 2012). In CAGE, transcripts are reverse-transcribed using random priming to include both polyadenylated and non-polyadenylated RNAs, followed by cap-trapping to capture capped transcripts such as mRNAs, long noncoding RNAs, and pri-miRNAs while avoiding ribosomal RNAs. In addition to measuring expression levels, CAGE

M. J. L. de Hoon (✉)
RIKEN Center for Integrative Medical Sciences, Laboratory for Applied Computational Genomics, Yokohama, Kanagawa, Japan
e-mail: michiel.dehoon@riken.jp

I. Abugessaisa, T. Kasukawa (eds.), *Practical Guide to Life Science Databases*,
https://doi.org/10.1007/978-981-16-5812-9_3

identifies the exact 5′ end of the profiled transcript and therefore its transcription start site and promoter region.

In the fifth edition (FANTOM5) of the FANTOM (Functional Annotation of the Mammalian Genome) project (http://fantom.gsc.riken.jp/), RNA samples from human and mouse, mostly from primary cells, were subjected to CAGE profiling to create an expression atlas of transcription initiation at single-nucleotide resolution (Forrest et al. 2014). A subset of 422 human and 78 mouse RNA samples in FANTOM5 were selected for sRNA library production and sequencing to produce a complementary atlas of miRNA expression in human and mouse (De Rie et al. 2017). By making use of the CAGE data in FANTOM5, for each miRNA the associated pri-miRNA and its promoter was identified. Importantly, each short RNA library in the FANTOM5 collection had a matching CAGE library produced from the same RNA sample. As the expression levels of mature miRNAs observed by sRNA sequencing were correlated to the expression levels of the corresponding pri-miRNAs in the matching CAGE library, pri-miRNA CAGE expression levels could be used as a proxy for the expression level of mature miRNAs, allowing the miRNA expression atlas to be extended to the 1829 human and 1029 mouse CAGE libraries included in FANTOM5 (De Rie et al. 2017).

The FANTOM5 expression atlas of miRNAs and their promoters (http://fantom.gsc.riken.jp/5/suppl/De_Rie_et_al_2017/) thus created is a comprehensive resource of miRNAs in human and mouse, their expression levels in primary cells, tissues, and cell lines, as well as their promoters and associated CAGE expression levels. Using this atlas, the expression pattern across samples of miRNAs can be evaluated as an indication of the cell types in which the miRNA is biologically most relevant. Additionally, sequence analysis of the promoter region around the transcription start site of the identified pri-miRNA enabled an analysis of how the cell type specific expression patterns of each miRNA are encoded in the regulatory control region of the corresponding pri-miRNA (De Rie et al. 2017).

Target users of this atlas are scientists focusing on specific miRNAs or specific cell types, as well as system biologists interested in a global analysis of the cellular regulatory network and the role of miRNAs therein.

3.2 Database Content

Table 3.1 shows an overview of the sRNA data in FANTOM5. Most samples are from human, and most human samples were derived from primary cells. As described previously (De Rie et al. 2017), all sRNA sequencing data were generated using the same protocol to prepare barcoded Illumina TruSeq Small RNA libraries (Illumina, Inc. 2011) and the same sequencing protocol on the Illumina HiSeq2000 sequencer, allowing direct comparison of the expression level of each miRNA between different samples. Similarly, CAGE sequencing data were generated using the same library preparation and sequencing protocol (Kanamori-Katayama et al. 2011) as described in the corresponding publications (Forrest et al. 2014; Arner

Table 3.1 Overview of sRNA libraries in FANTOM5

Organism	Origin	RNA source	Size	# of samples
Human	Primary cells	Total RNA	Short	286
	Primary cells, ES, iPS	Nuclear RNA	Short	16
	Primary cells, ES, iPS	Cytoplasmic RNA	Short	9
	Primary cells, ES, iPS	Nuclear RNA	Long	9
	Primary cells, ES, iPS	Cytoplasmic RNA	Long	9
	Tissue	Total RNA	Short	6
	Time course	Total RNA	Short	87
	Total: 422			
Mouse	Primary cells	Total RNA	Short	1
	Whole body	Total RNA	Short	14
	Time course	Total RNA	Short	27
	Primary cells, ES, iPS	Nuclear RNA	Short	9
	Primary cells, ES, iPS	Cytoplasmic RNA	Short	9
	Primary cells, ES, iPS	Nuclear RNA	Long	9
	Primary cells, ES, iPS	Cytoplasmic RNA	Long	9
	Total: 78			
Rat	Primary cells	Total RNA	Short	3
	Whole body	Total RNA	Short	3
	Total: 6			
Dog	Primary cells	Total RNA	Short	3
	Whole body	Total RNA	Short	3
	Total: 6			
Chicken	Primary cells	Total RNA	Short	3
	Whole body	Total RNA	Short	2
	Total: 5			

RNAs with sizes between 15 and 40 or 50 nucleotides ("short") or between 80 and 280 bp ("long") were selected for sequencing (Fort et al. 2014; De Rie et al. 2017). The sRNA libraries for rat, dog, and chicken samples are unpublished and have not yet been integrated in the database
ES embryonic stem cells, *iPS* induced pluripotent stem cells

et al. 2015). Detailed information on each sample is provided in the FANTOM5 Semantic catalog of Samples, Transcription initiation And Regulators (SSTAR; http://fantom.gsc.riken.jp/5/sstar; Abugessaisa et al. 2016).

To generate a miRNA expression table, sequence reads were mapped using bwa (Li and Durbin 2009) to genome assembly hg19 for human and mm9 for mouse and assigned to miRNAs previously annotated in miRBase release 21 (Kozomara and Griffiths-Jones 2014) or to candidate novel miRNAs identified using miRDeep2 (Friedländer et al. 2012) based on genomic overlap. Expression values were converted to c.p.m. (counts-per-million) by normalizing against the total miRNA expression in each sample (De Rie et al. 2017). The FANTOM5 CAGE data was used to identify the pri-miRNA transcript associated with each miRNA by applying a computational pipeline followed by manual curation (De Rie et al. 2017) and to create an expression table for all identified pri-miRNAs. Cell ontology enrichment

analysis of the human sRNA and CAGE data in FANTOM5 was performed by evaluating the statistical significance of expression enrichment or depletion of each miRNA or pri-miRNA in cell ontology clusters of primary cell types, retaining the three most enriched and depleted cell ontology terms as a systematic annotation of cell type specific expression.

3.3 Database Architecture

CAGE expression data are referred to by the RNA sample ID (a 4- or 5-digit number) of the RNA sample from which the CAGE library was produced. Each RNA sample number is associated with a FANTOM5 sample ontology ID (FF ontology ID) for human and mouse. Short RNA expression data are provided per RNA sample identified by concatenating the sRNA library ID (of the form SRhi*nnnnn*, where *nnnnn* is a 5-digit number), a 6-nucleotide barcode, and the RNA sample number. In a few cases, RNA samples from the same cellular origin were pooled before sRNA library construction; in such cases, the RNA sample numbers and FF ontology IDs are concatenated by a + sign.

Figure 3.1 shows a schematic view of the data stored in the FANTOM5 miRNA atlas. Each table in this schema corresponds to one flat file available at the FANTOM5 miRNA atlas website. At the core is the miRNA promoter annotation table, which associates each pre-miRNA with the corresponding mature miRNA as well as the promoter of the predicted pri-miRNA. Each pre-miRNA is identified by its pre-miRNA ID (i.e. the pre-miRNA name in miRBase) and its miRBase accession number, as well as by its genomic coordinates on genome assembly hg19 (for human) or mm9 (for mouse). Likewise, the mature miRNA is identified by its miRBase miRNA ID and miRBase accession number. Candidate novel miRNAs are indicated by their number in the accompanying publication (De Rie et al. 2017). The promoter of the predicted pri-miRNA is specified by its CAGE peak ID in FANTOM5 (Forrest et al. 2014), as well as by the genomic coordinate of the transcription start site.

The sRNA expression table provides the expression level, normalized to counts-per-million, of each mature miRNA (identified by the miRNA ID) in the FANTOM5 sRNA samples. A table of sRNA library descriptions shows the RNA sample (specified by the FANTOM5 sample ontology ID and corresponding sample

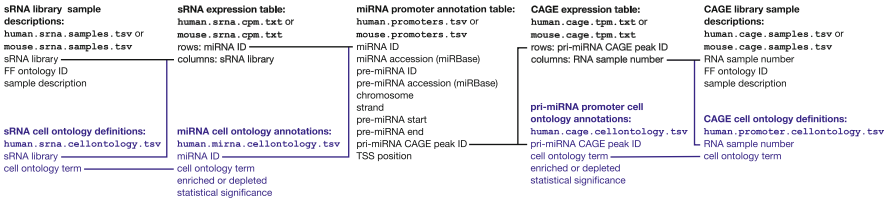


Fig. 3.1 Database architecture of the FANTOM5 miRNA atlas

Table 3.2 Accession numbers for raw sequencing data at DDBJ

Accession number	Data description
DRA000991	CAGE data, human and mouse (Forrest et al. 2014)
DRA001101	sRNA data, human (Andersson et al. 2014)
DRA002711	CAGE data, mouse (Arner et al. 2015)
DRA002747	CAGE data, human (Arner et al. 2015)
DRA002748	CAGE data, mouse (Arner et al. 2015)
DRA003804	sRNA data, human (De Rie et al. 2017)
DRA003807	sRNA data, mouse (De Rie et al. 2017)
DRA000914	sRNA data, human and mouse; CAGE data, human and mouse (Fort et al. 2014)

description) from which the sRNA library was produced. The sRNA cell ontology definitions table lists the sRNA libraries associated with each of the cell ontology terms; the miRNA cell ontology annotation table shows the three most enriched and depleted cell ontology terms for each miRNA, together with the statistical significance found.

The CAGE expression table provides the expression level, normalized to tags-per-million (t.p.m.), of all CAGE peaks associated with predicted pri-miRNAs for each CAGE library in FANTOM5, specified by their RNA sample number. The table of CAGE library descriptions shows the FANTOM5 sample ontology number and sample name for each CAGE library. The CAGE cell ontology definitions table shows the CAGE libraries (referenced by their RNA sample number) associated with each cell ontology term, while the pri-miRNA promoter cell ontology annotations show the three most enriched and depleted cell ontology terms for each pri-miRNA (identified by the CAGE peak ID of its associated promoter), together with the statistical significance found.

Raw sequencing data are available from the DNA Data Bank of Japan (DDBJ; https://www.ddbj.nig.ac.jp) as shown in Table 3.2. Sequence alignments to the human and mouse genome are available as part of the FANTOM5 data files (http://fantom.gsc.riken.jp/5/datafiles/). Expression tables as well as promoter, cell ontology, and sample annotations can be downloaded directly from the FANTOM5 miRNA atlas website, as described below.

3.4 Using the FANTOM5 miRNA Atlas Interactively

The FANTOM5 miRNA atlas is available at http://fantom.gsc.riken.jp/5/suppl/De_Rie_et_al_2017/. The landing page (Fig. 3.2) provides links to the miRNA expression viewer including novel miRNAs (Fig. 3.2Ⓑ) or excluding them (Fig. 3.2Ⓐ) for faster loading. The landing page also provides a link to an interactive miRNA

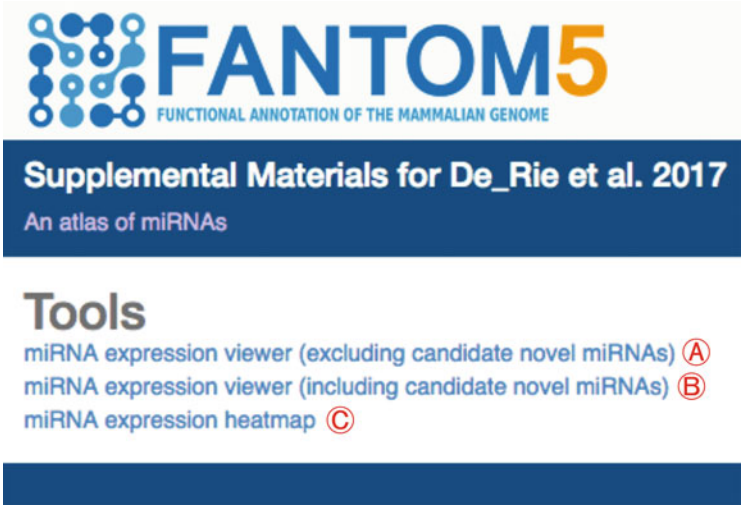


Fig. 3.2 Landing page of the FANTOM5 miRNA atlas at http://fantom.gsc.riken.jp/5/suppl/De_Rie_et_al_2017/

expression heatmap (Fig. 3.2Ⓒ), showing the expression profile of mature miRNAs in the robust set in the human primary cells.

3.4.1 Using the miRNA Expression Viewer

The miRNA expression viewer (Fig. 3.3) visualizes the miRNA expression data and annotation files shown in Fig. 3.1. At the top of the miRNA expression viewer, the user can select to access either the human or the mouse data (Fig. 3.3Ⓐ). The three panels below show, from left to right, the miRNA or sample list (Fig. 3.3Ⓑ), the expression chart, cell ontology analysis, and annotation data (Fig. 3.3Ⓒ), and the expression table (Fig. 3.3Ⓓ).

In the left panel, the user can select to list miRNAs, the sRNA samples, or the CAGE samples (Fig. 3.4Ⓐ). Selecting miRNAs will show a list of all mature miRNAs (both guide RNA and passenger strand RNA), the pre-miRNA from which they originate, and the associated pri-miRNA promoter (Fig. 3.4). Paralogous miRNAs will be listed once for each instance on the genome. The miRNAs can be sorted alphabetically by mature miRNA name (Fig. 3.4Ⓑ), pre-miRNA name (Fig. 3.4Ⓒ), or promoter name (Fig. 3.4Ⓓ) by clicking on the corresponding label. To search for miRNAs, the name of the mature miRNA, of the pre-miRNA, or of the promoter of the pri-miRNA can be entered in the boxes below the label (Fig. 3.4Ⓑ–Ⓓ). Selecting a miRNA from the list will highlight all instances of the mature miRNA (Fig. 3.4Ⓔ), the pre-miRNA associated with the selected miRNA (Fig. 3.4Ⓕ), and the promoter of the associated pri-miRNA both for the guide RNA

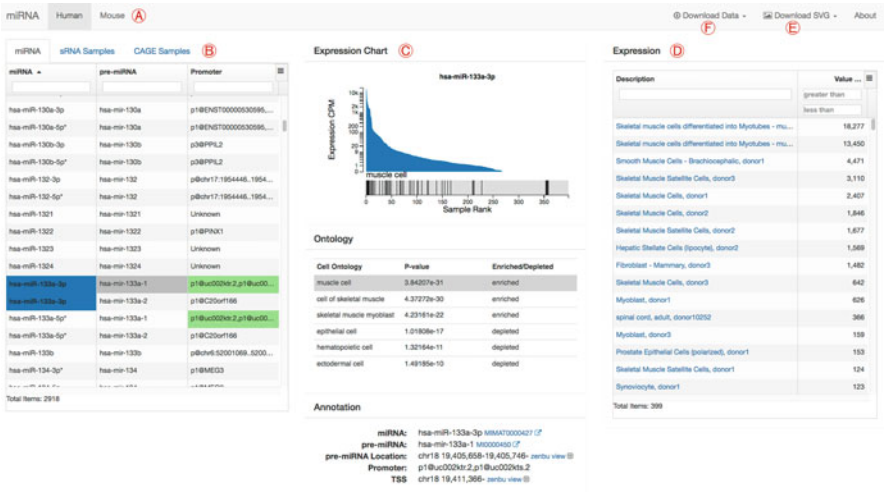


Fig. 3.3 FANTOM5 miRNA expression viewer showing human pre-miRNA hsa-mir-133a-1 with its associated mature miRNA hsa-miR-133a-5p (guide) and pri-miRNA promoter p1@uc002ktr.2, p1@uc002kts.2

(Fig. 3.4Ⓕ) and for the passenger strand RNA (Fig. 3.4Ⓕ). The promoter is also highlighted for any other miRNAs originating from the same pri-miRNA (Fig. 3.4Ⓘ). In addition to the columns shown, the pre-miRNA ID (Fig. 3.5Ⓐ) and the miRNA ID as defined by miRBase (Fig. 3.5Ⓑ) can be included in this table by clicking on the options button (Fig. 3.4Ⓙ) to open the options menu (Fig. 3.5).

The center panel (Fig. 3.6) shows the expression chart (Fig. 3.6Ⓐ) for the miRNA selected in the left panel, with the expression in counts-per-million (c.p.m.) on the vertical axis on a logarithmic scale, and the samples sorted by the expression of the miRNA on the horizontal axis. The expression chart can be downloaded as an editable vector image file in the Scalable Vector Graphics (SVG) format by clicking on "Download SVG" (Fig. 3.3Ⓔ).

For guide strand mature miRNAs, below the expression chart the cell ontology panel (Fig. 3.6Ⓑ) shows a table with the cell ontology clusters in which expression of the miRNA is most enriched or depleted, with the statistical significance shown as the P-value. Selecting a cell ontology term from this table will indicate the expression rank of the associated samples on the sample rank bar of the expression chart (Fig. 3.6Ⓒ) as a visual representation of the expression enrichment or depletion of the miRNA in the selected cell ontology cluster samples. Figures 3.6 and 3.7 show the examples of hsa-miR-16-5p and hsa-miR-100-5p with enriched and depleted, respectively, expression in leukocytes. Further below, the annotation panel (Fig. 3.6Ⓓ) provides links to the mature miRNA and the pre-miRNA in the miRBase (Kozomara and Griffiths-Jones 2014) database (Fig. 3.6Ⓔ), the genomic coordinates of the pre-miRNA, the FANTOM5 name of the promoter associated with the pri-miRNA, the coordinates of the transcription start site (TSS), and links to

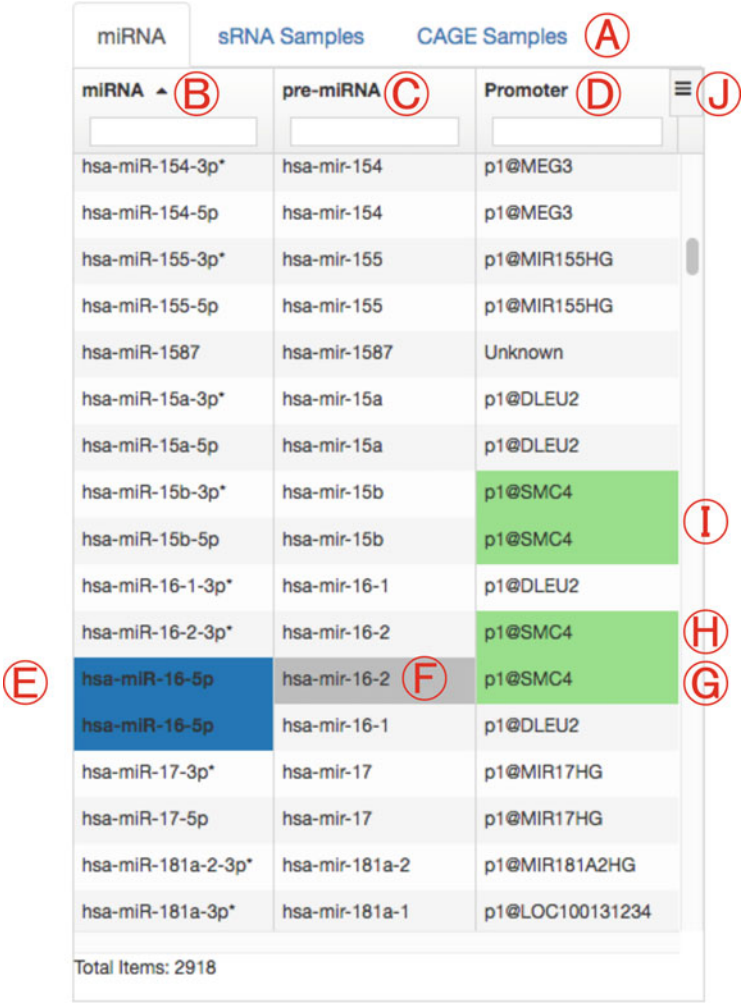


Fig. 3.4 Left panel of the miRNA expression viewer, showing the list of miRNAs

the pre-miRNA and TSS region in the ZENBU (Severin et al. 2014) genome browser (Fig. 3.6Ⓕ).

The right panel (Fig. 3.8) shows an expression table with the expression level in c.p.m. of the selected miRNA in each of the FANTOM5 samples. By default, samples are sorted by expression level of the miRNA. The samples can be sorted alphabetically by clicking on the description label (Fig. 3.8Ⓐ) and sorted by increasing or decreasing expression level by clicking on the value label (Fig. 3.8Ⓑ). Samples with miRNA expression levels greater than or less than a user-specified value can be selected by entering the desired maximum and minimum values in the boxes below the value label (Fig. 3.8Ⓒ). Clicking on the option button

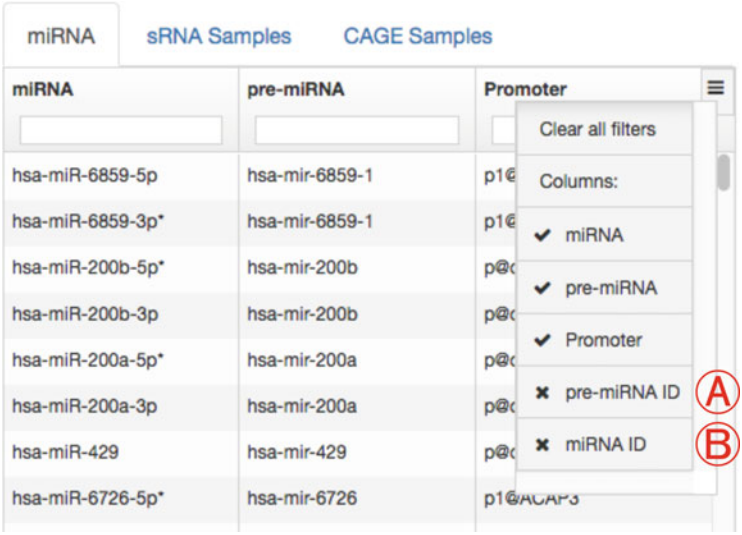


Fig. 3.5 Options menu for the left panel of the miRNA expression viewer, in which the columns to be shown in the list of miRNAs can be selected

(Fig. 3.8Ⓓ) to the right of the value label will display open the options menu displaying a list of columns to be shown in this panel (Fig. 3.9), which allows including the rank (Fig. 3.9Ⓐ) and name (Fig. 3.9Ⓑ) of each sample as additional columns in the expression table. Here, the expression ranks are numbered starting from 0, and the name consists of the sRNA library, barcode, and RNA sample number concatenated by periods. This panel also allows exporting the expression table as a downloadable file of comma-separated values (csv) (Fig. 3.9Ⓒ).

Selecting sRNA samples or CAGE samples in the left panel (Fig. 3.4Ⓐ). The sample ID (Fig. 3.11Ⓐ) and SSTAR ID (Fig. 3.11Ⓑ) can be shown as additional columns by selecting them in the options menu (Fig. 3.11) accessible by clicking the options button (Fig. 3.10Ⓐ). Choosing one of the samples will show the expression levels of miRNAs in the expression chart in the central panel, together with the sample annotation information (Fig. 3.12); the expression table in the right panel will show the expression levels numerically for each mature miRNA in c.p.m. (Fig. 3.13). Clicking the options button (Fig. 3.13Ⓐ) will display the options menu from which the miRNA expression rank (Fig. 3.14Ⓐ) and miRBase name (Fig. 3.14Ⓑ) can be shown as additional columns. The options menu also allows exporting the data visible in the expression table as a downloadable file with comma-separated values (csv) (Fig. 3.14Ⓒ).

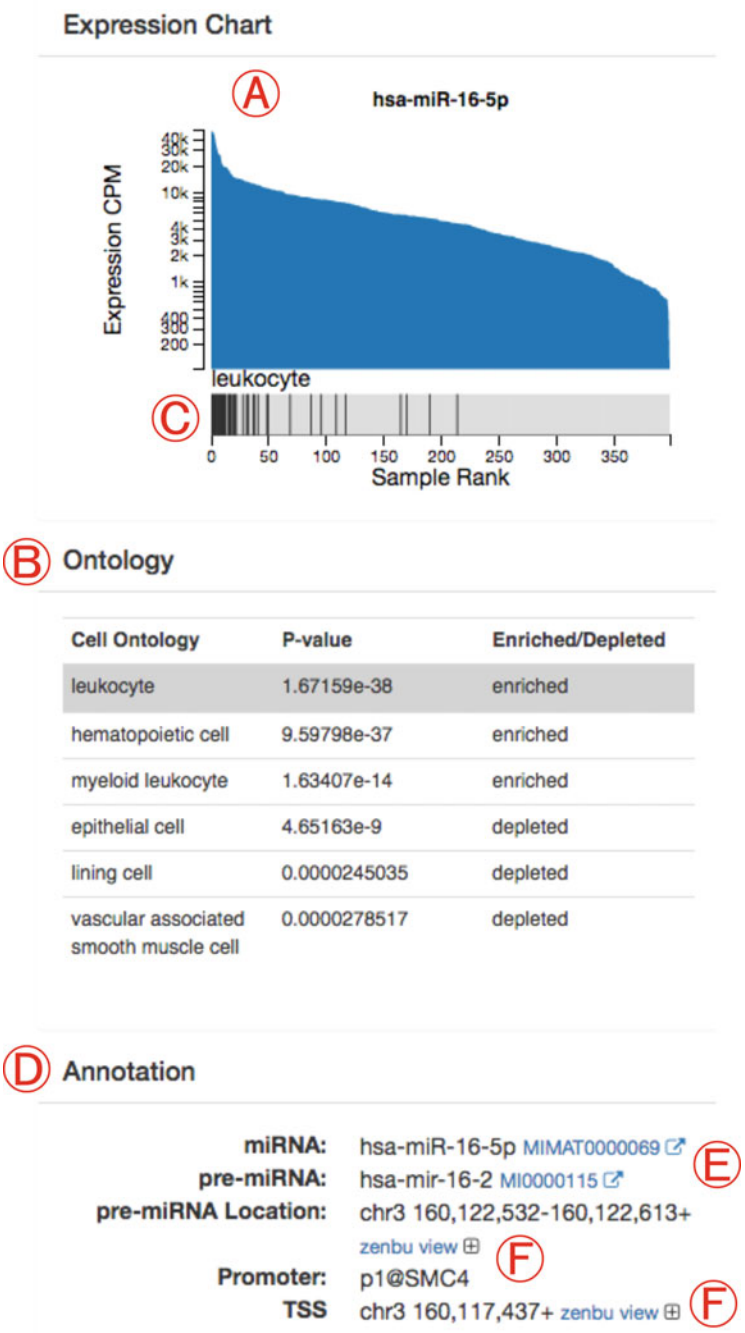
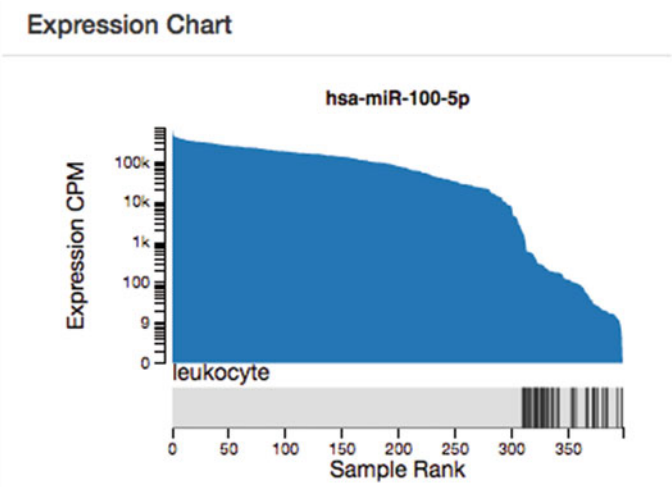


Fig. 3.6 Central panel of the miRNA expression viewer, showing the expression chart and sample rank, the cell ontology results, and the annotation data of the miRNA, pre-miRNA, and promoter. For the selected miRNA, hsa-miR-16-5p, expression is enriched in leukocytes



Ontology

Cell Ontology	P-value	Enriched/Depleted
muscle cell	6.56765e-15	enriched
smooth muscle cell	1.22534e-10	enriched
vascular associated smooth muscle cell	3.83387e-10	enriched
leukocyte	1.82418e-56	depleted
hematopoietic cell	2.61462e-38	depleted
myeloid leukocyte	4.445e-24	depleted

Fig. 3.7 Expression chart, sample rank, and cell ontology results for miRNA hsa-miR-100-5p, for which expression is depleted in leukocytes

3.4.2 Using the Interactive miRNA Expression Heatmap

The interactive miRNA expression heatmap can be accessed by clicking on the link on the landing page of the FANTOM5 miRNA atlas (Fig. 3.2©). The heatmap (Fig. 3.15) shows the expression level of the 735 annotated mature miRNAs (guide strand only) in the robust set (De Rie et al. 2017) as rows, in 118 primary cell types in human, after averaging over donors, as columns. One cell type ("Fibroblast—Pulmonary Artery") was dropped from the full set of samples (Table 3.1) as the corresponding sRNA library had fewer than 100,000 reads. Cell types were grouped

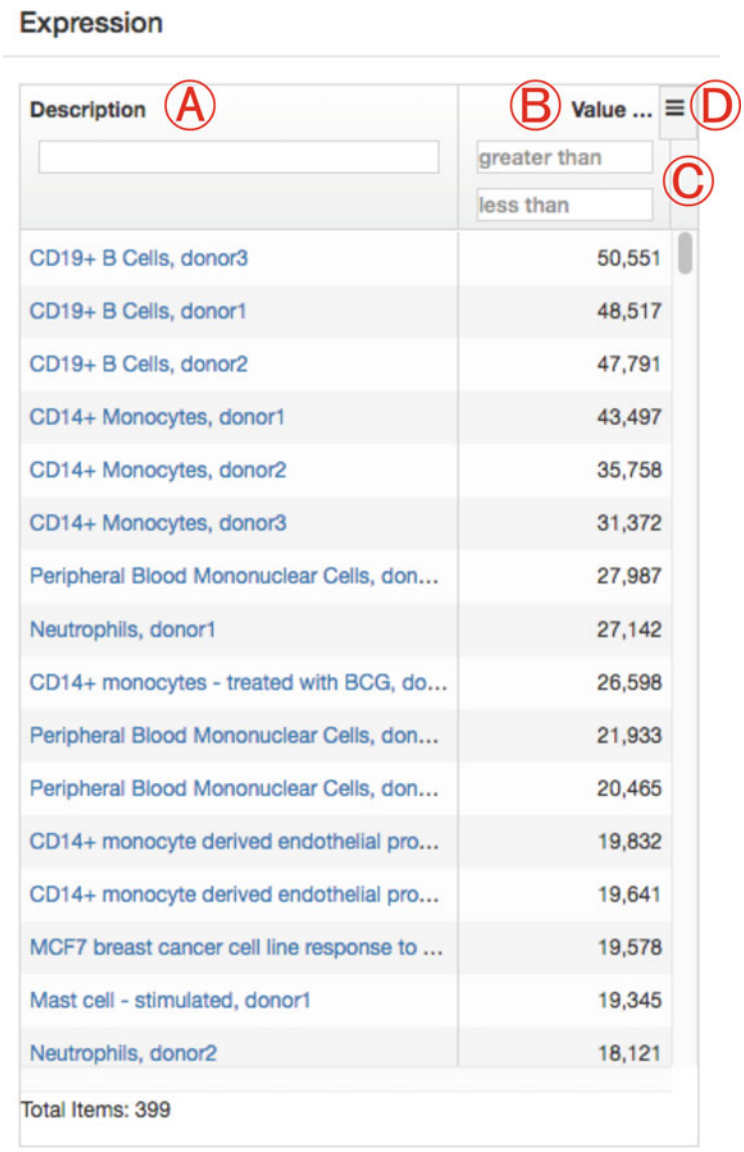


Fig. 3.8 Expression table for miRNA hsa-miR-16-5p, which is highly expressed in CD19+ B cells, CD14+ monocytes, and other leukocytes

by category (Fig. 3.15Ⓐ) as indicated by the color bar below the cell type name (Fig. 3.15Ⓑ).

To sort the heatmap, both miRNAs and cell types were clustered using pairwise centroid-linkage hierarchical clustering with the Pearson correlation as the similarity measure (De Hoon et al. 2004) after normalizing the expression of each miRNA to

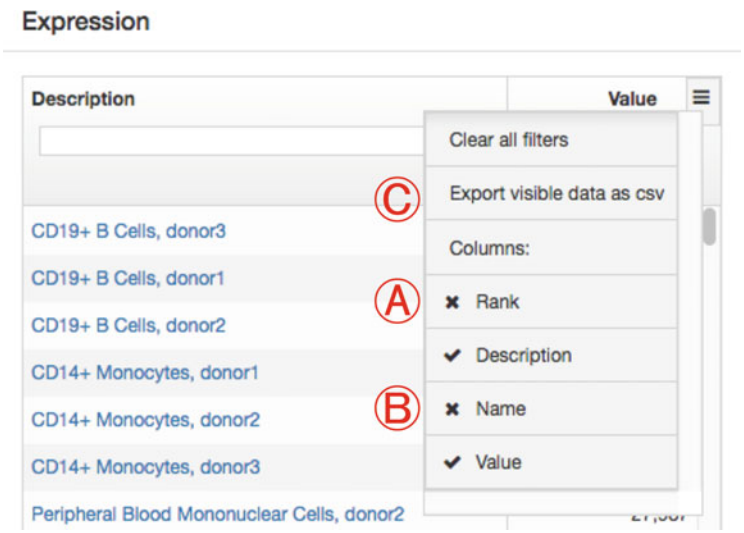


Fig. 3.9 Options menu for the right panel of the miRNA expression viewer, in which the columns to be shown in the expression table can be selected, and data can be exported as a downloadable file in the csv (comma-separated values) format

Z-scores by subtracting the mean and dividing by the standard deviation across samples. The hierarchical clustering tree itself is not displayed. Each cell in the heatmap is colored based on the calculated Z-score of the expression of the miRNA in the cell type (Fig. 3.15©). Hovering with the mouse over a cell in the heatmap will show the mature miRNA name, the cell type, the category to which the cell type belongs, and the Z-score of the miRNA expression level in the cell type (Fig. 3.16). Clicking on a cell or on a miRNA name will redirect the browser to the miRNA expression viewer for the corresponding miRNA. The expression data shown in the heatmap can be downloaded as a tab-delimited file by clicking on "Download Data" (Fig. 3.15Ⓓ).

3.5 Database Access and Mining Methods

A flat file for each table shown in Fig. 3.1 can be downloaded from the FANTOM5 miRNA atlas website by clicking on the download button (Fig. 3.3Ⓕ).

The miRNA promoter annotation table is provided as the tab-delimited files `human.promoters.tsv` and `mouse.promoters.tsv` for human and mouse, respectively. This file contains two lines for each pre-miRNA, corresponding to the guide strand and the passenger strand of the mature miRNA. Each line shows the name and miRBase ID of the pre-miRNA, its chromosome, strand, and genomic coordinates, the name and miRBase ID of the guide RNA or passenger strand RNA of the associated mature miRNA, the short description of the FANTOM5 CAGE

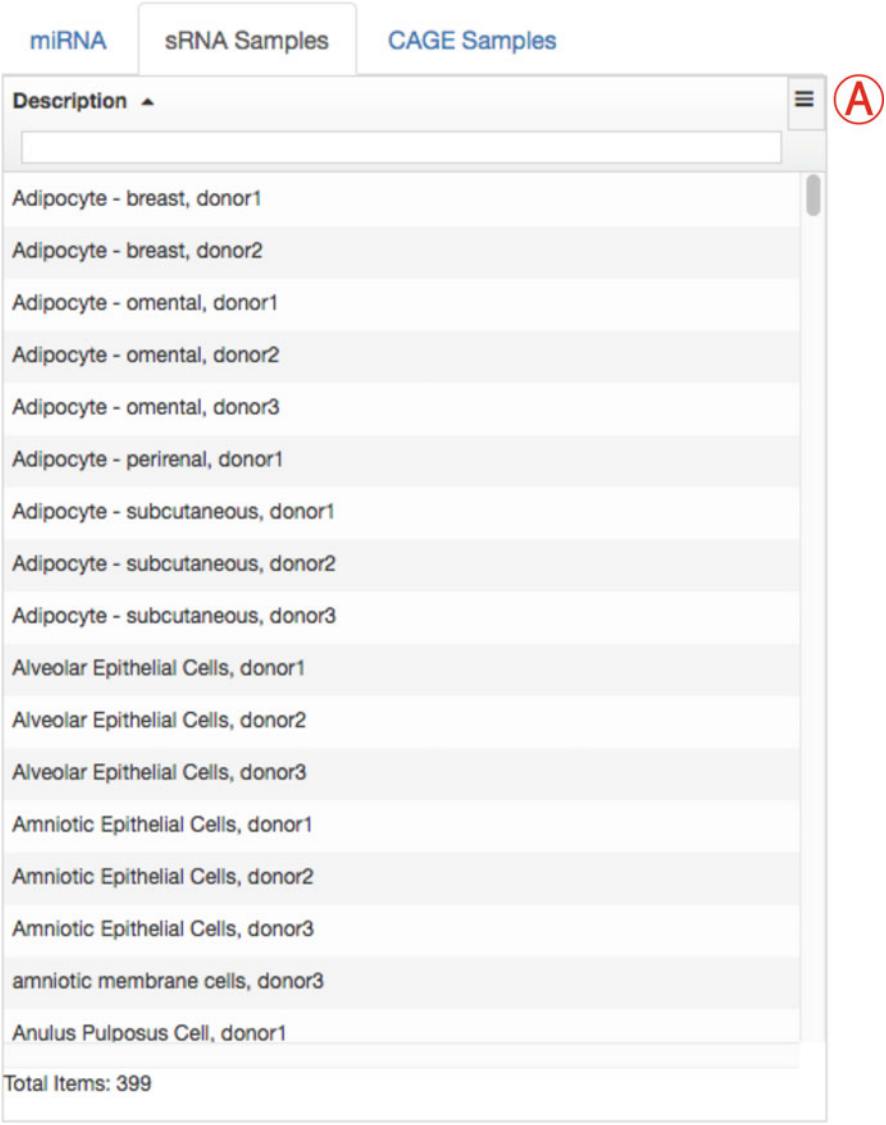


Fig. 3.10 Left panel of the miRNA expression viewer, showing the list of sRNA samples

peak of the promoter associated with the pri-miRNA, and the transcription start site of the pri-miRNA.

The sRNA expression tables, `human.srna.cpm.txt` for human and `mouse.srna.cpm.txt` for mouse, are tab-delimited files with the expression of both the guide and passenger strand of each miRNA, identified by its miRBase ID, in each sRNA library. Expression values are normalized to counts-per-million (c.p.m.) in

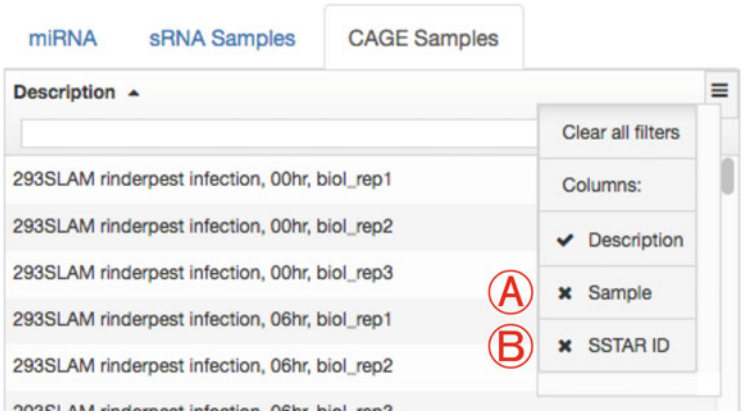


Fig. 3.11 Left panel of the miRNA expression viewer, showing the list of CAGE samples, together with the options menu in which columns to be included can be selected

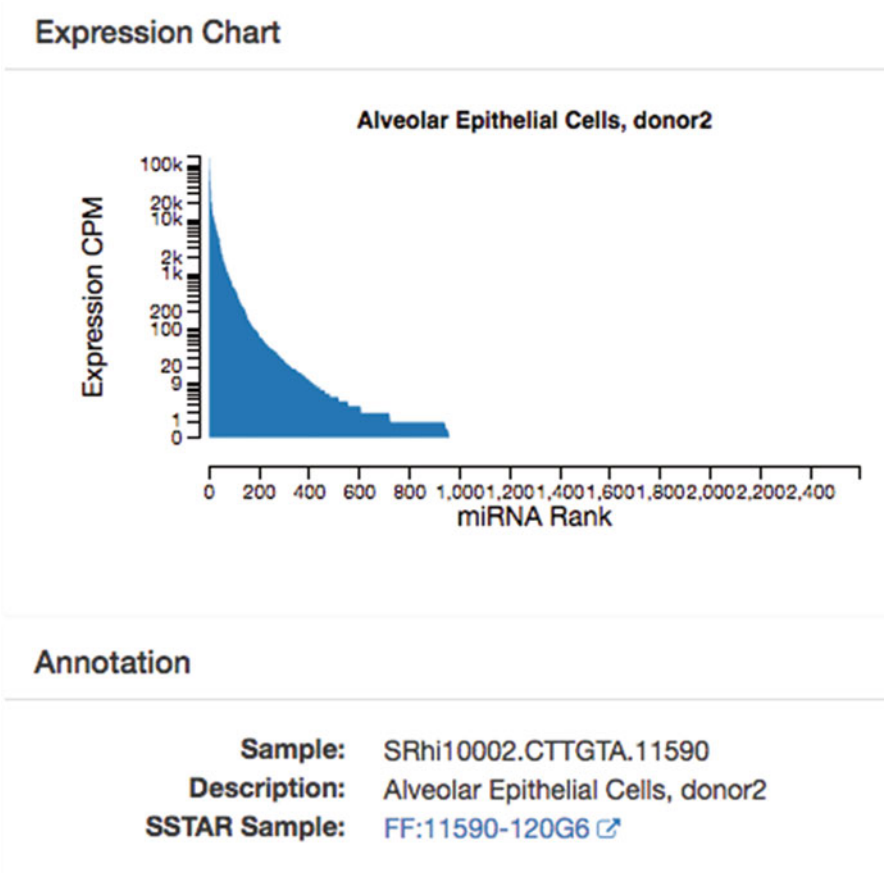


Fig. 3.12 Expression chart for miRNA expression in "alveolar epithelial cells, donor 2" as measured by sRNA sequencing, together with the annotation data of this sample

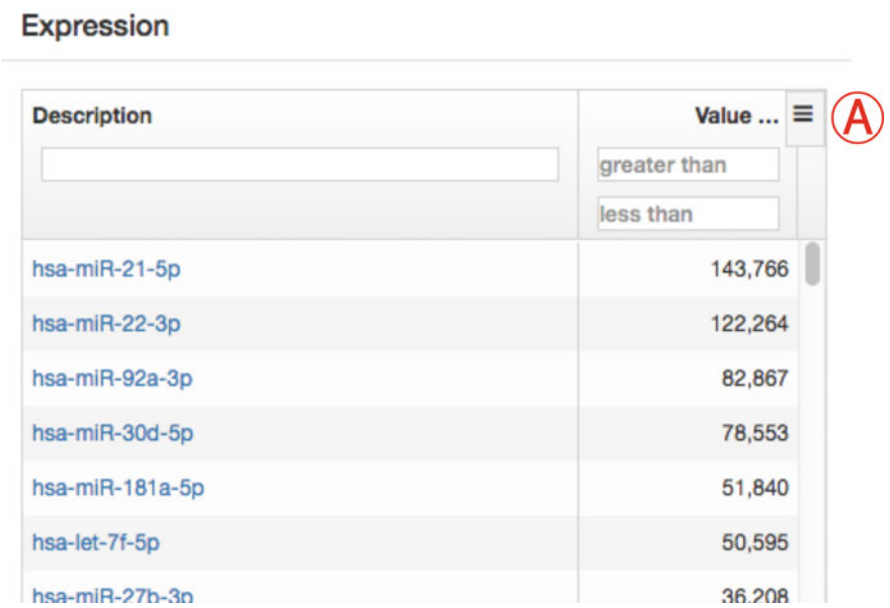


Fig. 3.13 Right panel of the miRNA expression viewer, showing the expression table for miRNA expression in sample "alveolar epithelial cells, donor 2" as measured by sRNA sequencing

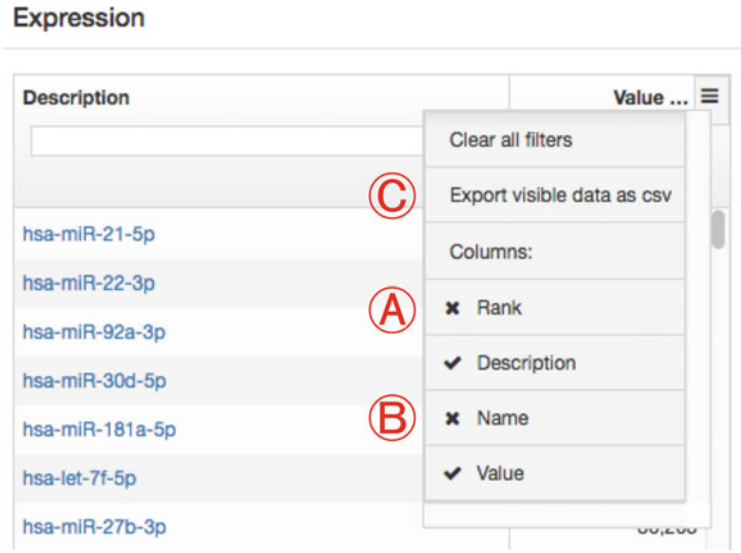


Fig. 3.14 Options menu for the right panel of the miRNA expression viewer, in which the columns to be shown in the expression table can be selected, and data can be exported as a downloadable file in the csv (comma-separated values) format

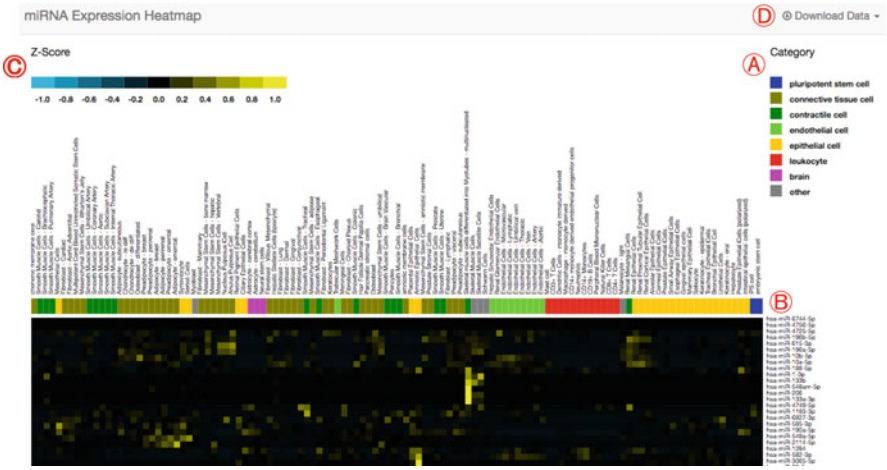


Fig. 3.15 Interactive heatmap showing the expression of mature miRNAs (guide RNA) in the robust set (De Rie et al. 2017), normalized to Z-scores, in human primary cell types after averaging over donors

each library separately. The sRNA library IDs are associated with an FF ontology ID and sample description in the files `human.srna.samples.tsv` and `mouse.srna.samples.tsv`.

The CAGE expression tables, `human.cage.tpm.txt` and `mouse.cage.tpm.txt` for human and mouse, respectively, are tab-delimited files with the CAGE expression level of the FANTOM5 CAGE peaks associated with pri-miRNAs as shown in the miRNA promoter annotation table. Each column corresponds to one CAGE library, as identified by its associated RNA sample number, and is normalized to tags-per-million (t.p.m.). The RNA sample numbers are associated with an FF ontology ID and sample description in the files `human.cage.samples.tsv` and `mouse.cage.samples.tsv`.

The miRNA cell ontology annotations based on the sRNA expression patterns across cell types are available in the file `human.mirna.cellontology.tsv`, listing for each miRNA (identified by miRBase ID) the three cell ontology clusters in which the expression of the miRNA is most enriched, and the three cell ontology clusters in which the expression is most depleted. The sRNA cell ontology definitions are provided in the file `human.srna.cellontology.tsv`, listing the sRNA samples associated with each cell ontology cluster. Similarly, the file `human.promoter.cellontology.tsv` lists for each FANTOM5 CAGE peak associated with a pri-miRNA the three cell ontology clusters in which the CAGE expression levels of the pri-miRNA are most enriched, and the three cell ontology clusters in which the expression is most depleted. The CAGE cell ontology definitions are provided in the file `human.cage.cellontology.tsv`, listing the RNA sample numbers associated with each cell ontology cluster.

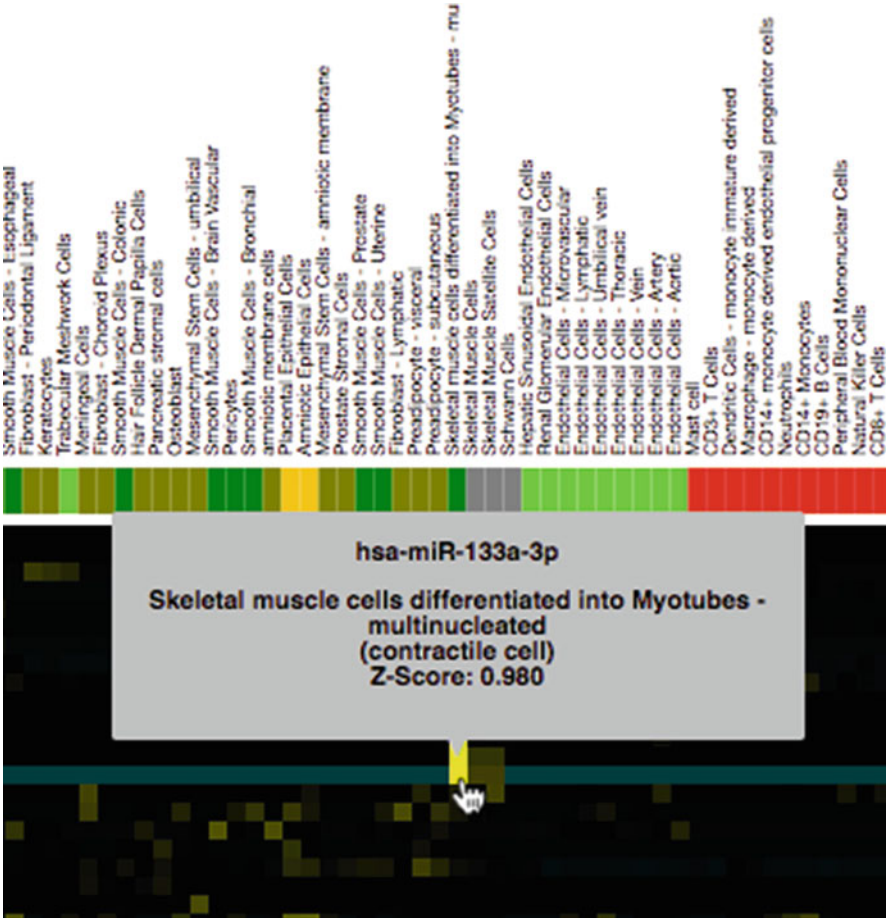


Fig. 3.16 Popup shown when hovering the mouse over a particular cell in the interactive heatmap, with the mature miRNA name, sample description, cell type category, and Z-score

3.6 Summary and Future Development of the Database

The FANTOM5 expression atlas of miRNAs and their promoters provides a basis for a detailed analysis of the transcriptional regulation of miRNAs and their role in defining cell types. The atlas will be extended in the near future sRNA sequencing data for rat, dog, and chicken (Table 3.1), together with promoter annotations for miRNAs in these three species, opening the door to cross-species comparisons of miRNA expression and regulation.

Acknowledgements **Funding**: FANTOM5 was made possible by the following grants: Research Grant for RIKEN Omics Science Center from MEXT to Yoshihide Hayashizaki; Grant of the Innovative Cell Biology by Innovative Technology (Cell Innovation Program) from the MEXT to

Yoshihide Hayashizaki; Research Grant from MEXT to the RIKEN Center for Life Science Technologies; Research Grant to RIKEN Preventive Medicine and Diagnosis Innovation Program from MEXT to Yoshihide Hayashizaki.

References

Abugessaisa I, Shimoji H, Sahin S et al (2016) FANTOM5 transcriptome catalog of cellular states based on Semantic MediaWiki. Database (Oxford) 2016:baw105

Andersson R, Gebhard C, Miguel-Escalada I (2014) An atlas of active enhancers across human cell types and tissues. Nature 507(7493):455–461

Arner E, Daub CO, Vitting-Seerup K et al (2015) Transcribed enhancers lead waves of coordinated transcription in transitioning mammalian cells. Science 347(6225):1010–1014

De Hoon MJL, Imoto S, Nolan J, Miyano S (2004) Open source clustering software. Bioinformatics 20(9):1453–1454

De Rie D, Abugessaisa I, Alam T et al (2017) An integrated expression atlas of miRNAs and their promoters in human and mouse. Nat Biotechnol 35(9):872–878

Forrest ARR, Kawaji H, Rehli M et al (2014) A promoter-level mammalian expression atlas. Nature 507(7493):462–470

Fort A, Hashimoto K, Yamada D et al (2014) Deep transcriptome profiling of mammalian stem cells supports a regulatory role for retrotransposons in pluripotency maintenance. Nat Genet 46(6):558–566

Friedländer MR, Mackowiak SD, Li N et al (2012) miRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades. Nucleic Acids Res 40(1):37–52

Illumina, Inc. (2011) TruSeq® Small RNA Sample Preparation Guide, Catalog # RS-930-1012, Part # 15004197 Rev. C, March 2011. San Diego, California

Kanamori-Katayama M, Itoh M, Kawaji H et al (2011) Unamplified cap analysis of gene expression on a single-molecule sequencer. Genome Res 21(7):1150–1159

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

Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25(14):1754–1760

Severin J, Lizio M, Harshbarger J et al (2014) Interactive visualization and analysis of large-scale sequencing datasets using ZENBU. Nat Biotechnol 32(3):217–219

Takahashi H, Lassmann T, Murata M, Carninci P (2012) 5′ end-centered expression profiling using cap-analysis gene expression and next-generation sequencing. Nat Protoc 7(3):542–561