D356–D364 *Nucleic Acids Research, 2022, Vol. 50, Database issue* *Published online 13 October 2021*
https://doi.org/10.1093/nar/gkab917

scAPAatlas: an atlas of alternative polyadenylation across cell types in human and mouse

Xiaoxiao Yang[1,2,†], Yang Tong[1,2,†], Gerui Liu[1,2,†], Jiapei Yuan[3] and Yang Yang[1,2,*]

[1]The Province and Ministry Co-sponsored Collaborative Innovation Center for Medical Epigenetics, Tianjin Key Laboratory of Inflammation Biology, Tianjin Key Laboratory of Medical Epigenetics, Department of Pharmacology, School of Basic Medical Sciences, Tianjin Medical University, Tianjin 300070, China, [2]Department of Bioinformatics, School of Basic Medical Sciences, Tianjin Medical University, Tianjin 300070, China and [3]State Key Laboratory of Experimental Hematology, National Clinical Research Center for Blood Diseases, Institute of Hematology & Blood Diseases Hospital, Chinese Academy of Medical Sciences & Peking Union Medical College, Tianjin 300020, China

Received August 14, 2021; Revised September 15, 2021; Editorial Decision September 22, 2021; Accepted September 25, 2021

ABSTRACT

Alternative polyadenylation (APA) has been widely recognized as a crucial step during the post-transcriptional regulation of eukaryotic genes. Recent studies have demonstrated that APA exerts key regulatory roles in many biological processes and often occurs in a tissue- and cell-type-specific manner. However, to our knowledge, there is no database incorporating information about APA at the cell-type level. Single-cell RNA-seq is a rapidly evolving and powerful tool that enable APA analysis at the cell-type level. Here, we present a comprehensive resource, scAPAatlas (http://www.bioailab.com:3838/scAPAatlas), for exploring APA across different cell types, and interpreting potential biological functions. Based on the curated scRNA-seq data from 24 human and 25 mouse normal tissues, we systematically identified cell-type-specific APA events for different cell types and examined the correlations between APA and gene expression level. We also estimated the crosstalk between cell-type-specific APA events and microRNAs or RNA-binding proteins. A user-friendly web interface has been constructed to support browsing, searching and visualizing multi-layer information of cell-type-specific APA events. Overall, scAPAatlas, incorporating a rich resource for exploration of APA at the cell-type level, will greatly help researchers chart cell type with APA and elucidate the biological functions of APA.

INTRODUCTION

Alternative polyadenylation (APA) is emerging as an important regulatory mechanism that contributes to transcriptome complexity and sophisticated dynamics of gene regulation across eukaryotic species (1,2). Most eukaryotic genes harbor multiple polyadenylation (polyA) sites, leading to generation of distinct APA isoforms with different coding regions or 3′ untranslated regions (3′UTRs) (3,4). Since the 3′UTRs contain key RNA regulatory elements interacting with regulatory RNA-binding proteins (RBPs) or microRNAs (miRNAs), APA could modulate RBPs or miRNAs targetability, and ultimately regulating the RNA cellular localization, translation efficiency and degradation rates (5–8). Increasing evidences have revealed that APA plays vital roles in diverse biological processes, such as cell proliferation, differentiation and tumorigenesis (2,9–11). APA often occurs in a tissue-specific manner as documented in previous studies (12,13). Furthermore, recent studies have detailed the existence of cell-type-specific APA preference within the same tissue (14–16). For example, the long isoform of *Itsn1* is restricted to neurons while the short isoform is expressed in astrocytes and microglia of the adult mouse brain (14).

Several databases have been developed to characterize genome-wide APA in different species. For instance, early databases could annotate limited polyA sites based on expression sequence tags, such as polyA_DB_2 and PACdb (17,18). The advent of next-generation sequencing technology provides an unprecedented opportunity to detect genome-wide APA events. Several databases, such as PolyA_DB_3, PolyAsite and APADB (19–21), have curated genome-wide polyA sites in different species using 3′ sequencing datasets. Recently, several databases, such as TC3A and APAatlas (22,23), were constructed to not only annotate but also quantify APA events in different tissues

*To whom correspondence should be addressed. Tel: +86 18511896924; Email: yy@tmu.edu.cn
†The authors wish it to be known that, in their opinion, the first three authors should be regarded as Joint First Authors.

based on large-scale RNA-seq datasets. However, to our knowledge, no database has been developed to systematically explore APA across different cell types. The recent development of single-cell RNA-seq methods, such as Smart-seq, 10x Genomics and CEL-seq2, has provided opportunities for studying APA in different cell types (24–26). Several bioinformatics methods, such as scAPA, have been developed to identify polyA sites and quantify APA using 3′-tag based scRNA-seq data (27).

Here, we aim to profile the APA landscape of different cell types and identify cell-type-specific APA events for each cell type from a certain tissue. We first collected and uniformly processed about 150TB 10x Genomics scRNA-seq data from 49 studies, incorporating 305 scRNA-seq datasets of 24 human and 25 mouse normal tissues (Supplementary Table S1). We then applied the scAPA method to these datasets to identify polyA sites and quantify the relative polyA site usage (27). Thus, we could identify cell-type-specific APA events for different cell types in a certain tissue. Moreover, we also examined the correlations between polyA site usage and the corresponding gene expression level to explore the effects of polyA site choice on gene expression. In addition, we predicted the putative binding sites of miRNAs and RBPs on the alternative regions regulated by cell-type-specific APA events. By designing an interactive user interface, we constructed a versatile database, scAPAatlas, which allows users to browse, search and visualize cell-type-specific APA events and other information. To our knowledge, scAPAatlas is the first comprehensive resource for exploring APA at the cell-type level, which could benefit researchers in understanding biological functions of APA. ScAPAatlas is freely available at http://www.bioailab.com:3838/scAPAatlas or http://47.100.223.144:3838/scAPAatlas.

MATERIALS AND METHODS

scRNA-seq data collection and processing

The scAPAatlas was designed for users to explore APA at the cell-type level in human and mouse. We manually collected recently published 10x Genomics scRNA-seq data from the Gene Expression Omnibus (GEO), Sequence Read Archive (SRA), ArrayExpression and NCBI BIOPROJECT (Supplementary Table S1 and Supplementary Figure S1). Several datasets are publicly available with alignment results, so we could directly download the aligned reads in BAM format. For the other datasets, we downloaded the sequencing reads in FASTQ format. Then we processed and aligned the sequencing reads to the reference genome using Cell Ranger (version 5.0.1) *CellRanger count* pipeline with default parameters to generate the BAM files containing the cell barcodes and unique molecular identifiers (UMIs) (26). The reference genome sequence of human (hg38) and mouse (mm10) were downloaded from Ensembl (28). The gene expression matrices generated by Cell Ranger were further filtered and normalized using Seurat package (29).

As the cell-type annotation for each single cell is required for downstream APA analysis at the cell-type level, we downloaded the cell-type annotation information directly if it is available in the original publications. If the cell-type annotation information was not available, we employed the Seurat package to perform cell clustering and assign each single cell to the corresponding cell cluster with the standard pipeline described in Seurat implementation (29). Each cell-type cluster was annotated using marker genes provided in the original publications. In addition, we applied t-distributed Stochastic Neighbor Embedding (t-SNE) algorithm to visualize the cell clustering result and count the number of single cells per cell type. The marker gene expression pattern was visualized using dot-plot heatmaps in which the color intensity represents the average level of expression and the size of dots represents the percentage of cells within each cell-type cluster expressing the marker genes.

Alternative polyadenylation analysis

To systematically explore the cell-type-specific APA regulation, we applied the scAPA pipeline to analyze the compiled 10x Genomics scRNA-seq datasets (27). For each dataset, we first employed the Drop-seq tool (version 2.4.0) to filter the BAM files using *FilterBAM* function, and then removed PCR duplicates using the UMI-tools *dedup* function (version 1.1.1) (30,31). We next employed the scAPA pipeline to identify APA peaks with default procedures described in the scAPA implementation using Homer *findPeaks* (32). For each tissue, we could identify the APA peaks in BED format. Then, we merged all APA peaks from different tissues using bedtools *merge* to generate APA peaks annotation for human and mouse, respectively (33).

In further, we applied the scAPA pipeline to profile the APA landscape based on the defined APA peaks. At first, we utilized the Drop-seq tool *FilterBamByTag* and SAMtools *merge* to merge all processed reads from single cells assigned to the same cell-type cluster to generate an individual BAM file for each cell type based on the cell-type annotation (30,34). We next used featureCounts to count reads aligned to defined APA peak regions for each cell type to generate an APA peak counts matrix for each scRNA-seq dataset (35). In each dataset, the resulting count matrix was normalized into the expression level measured in counts per million (CPM), and the APA peaks were further filtered with the procedures implemented in scAPA pipeline. To quantify the relative usage of polyA sites, we calculated the relative expression level of an APA isoform over the total expression level of all APA isoforms in a gene, defined by a metric called PolyA site Usage (PAU):

$$PAU_{ig} = \frac{C_{ig}}{\sum_i^n C_{ig}}$$

Where g is a given gene, C_{ig} is the CPM value of APA isoform i in gene g and n is the number of APA isoforms of the gene.

In addition, to visualize the reads coverage of scRNA-seq data for each cell type, we took the BAM file as input and generated coverage tracks in bedGraph format for different cell types as output. The coverage is calculated using bedtools *genomecov* and normalized in counts per million. Then, we converted the bedGraph files to bigWig files using the *bedGraphToBigWig* utility from UCSC Genome Browser.

Identification of cell-type-specific APA events

In each dataset, only genes with counts >10 in a cell type were considered for analysis. Here, we defined that an APA event is cell-type-specific in a certain cell type, when the PAU of the APA event in the cell type is significantly higher than in other cell types. Chi-squared test was used to calculate the *P*-value of differential usage of polyA sites across different cell types. The resulting *P*-values were adjusted for multiple comparisons into *Q*-values by the Benjamini–Hochberg procedure to control the false discovery rate (FDR). For a specific gene, the PAUs of the polyA site i in all k cell types are P_{i1}, P_{i2}, P_{i3}, …, P_{ik}. Then, to decide whether the PAU of the polyA site i in cell type j is higher than in other cell types, we calculate the differences between $P_{ij}$and other PAU respectively to get a set of PAU difference ($P_{ij} - \mathrm{P_{i1}}$, $P_{ij} - P_{i2}$, $P_{ij} - P_{i3}$, …, $P_{ij} - P_{ik}$). Here, we refer to the minimal value of the set of PAU difference ($P_{ij} - \mathrm{P_{i1}}$, $P_{ij} - P_{i2}$, $P_{ij} - P_{i3}$, …, $P_{ij} - P_{ik}$) as the difference of PAU. Thus, if the difference of PAU > 0.2 and *Q*-value < 0.05, the APA event with polyA site i is defined as a cell-type-specific event in cell type j (Supplementary Figure S2).

Correlation analysis between APA and gene expression level

To identify the APA event whose PAU is significantly correlated with its gene expression level, we used Spearman's rank correlation to calculate the correlation coefficient (R_s) between PAU and corresponding gene expression level. The significant correlations between PAU and gene expression level were defined by the absolute value of $R_s > 0.3$ and *P*-value < 0.05. In total, 2444 and 2223 APA events significantly correlated with their gene expression level were identified in human and mouse, respectively.

miRNA-binding sites prediction analysis

The miRNA-binding sites on 3′UTR in human and mouse were predicted using TargetScan (Release 7.2) (36). Then we could obtain the names of miRNA family, genomic coordinates of miRNA-binding sites and context++ score percentiles. To identify miRNA-binding sites potentially altered by cell-type-specific APA events, we took the proximal and distal polyA sites to define the alternative regions regulated by cell-type-specific APA events. Then, intersecting the miRNA-binding sites with the defined alternative regions using bedtools *intersect* function, we could identify miRNA binding sites potentially altered by cell-type-specific APA events. The resulting miRNA-binding sites were prepared in BED format for displaying using JBrowse in scAPAatlas (37).

RBP-binding sites prediction analysis

We downloaded computationally predicted RBP-binding sites from the MotifMap-RNA database (http://motifmap-rna.ics.uci.edu), and obtained the RBP names and genomic coordinates of RBP-binding sites (38). We then identified RBP-binding sites potentially altered by cell-type-specific APA events through intersecting the RBP-binding sites with previously defined alternative regions regulated by the cell-type-specific APA events. In addition, we downloaded the binding peaks of 122 RBPs identified by eCLIP-seq experiments in HepG2 and K549 human cell lines from the ENCODE data portal (https://www.encodeproject.org/) (39). Then we could label each predicted RNA-binding site in human with eCLIP-seq support evidence by intersecting them with the eCLIP-seq peaks. The resulting RBP-binding sites were prepared in BED format for displaying using JBrowse in scAPAatlas.

RESULTS

The APA landscape across different cell types in human and mouse tissues

To construct a comprehensive database for exploring APA at the cell-type level, about 150TB 10x Genomics scRNA-seq data were manually curated from publications through literature searching (Supplementary Table S1). These publications were obtained by searching for key words, including the name of tissue and 'scRNA-seq' or '10x Genomics', from the PubMed database (https://pubmed.ncbi.nlm.nih.gov). We manually collected scRNA-seq datasets of human and mouse normal tissues from 49 studies (Supplementary Table S1). Specially, the human data contain 194 scRNA-seq datasets of 24 diverse human normal tissues, involving a total of 385 964 single cells, while 111 scRNA-seq datasets, covering 25 mouse normal tissues, were included in the mouse data with 428 225 single cells in total (Figure 1A; Supplementary Figure S1 and Supplementary Table S1). Furthermore, the cell-type annotation for each tissue was obtained from original publications or estimated based on known marker genes. Finally, 299 human cell types and 286 mouse cell types were annotated in the compiled scRNA-seq dataset.

Applying the scAPA method to these datasets, we identified 38 687 APA events in human, while 35 375 APA events were identified in mouse (Figure 1B). To demonstrate the reliability of the identified polyA sites, we computed the genomic distributions of the identified polyA sites and compared them with the annotated polyA sites in polyA_DB3 (19). The results showed that the identified polyA sites are congruent with annotated polyA sites (Supplementary Figures S3–6). In addition, we have conducted motif enrichment analysis on the identified polyA sites in introns and intergenic regions. The results showed that the canonical polyA motif (AAUAAA) is top significantly enriched for each tissue (Supplementary Figures S7 and S8). The APA events could be classified into two types according to the genomic positions of the polyA sites, including tandem 3′UTR-APAs and upstream regions APAs (UR-APAs). We have classified the identified APA events into two different types and computed the prevalence of tandem 3′UTR-APAs and UR-APAs in each tissue (Supplementary Figures S9 and S10).

Through profiling the APA landscape, we could identify cell-type-specific APA events for each cell type in a certain tissue (Figure 1B; Supplementary Figure S11A and Supplementary Tables S2 and S3). The correlations between APA and corresponding gene expression level were also estimated in human and mouse, respectively (Supplementary Figure S11B). Additionally, potential miRNA-binding sites and RBP-binding sites on alternative regions regulated

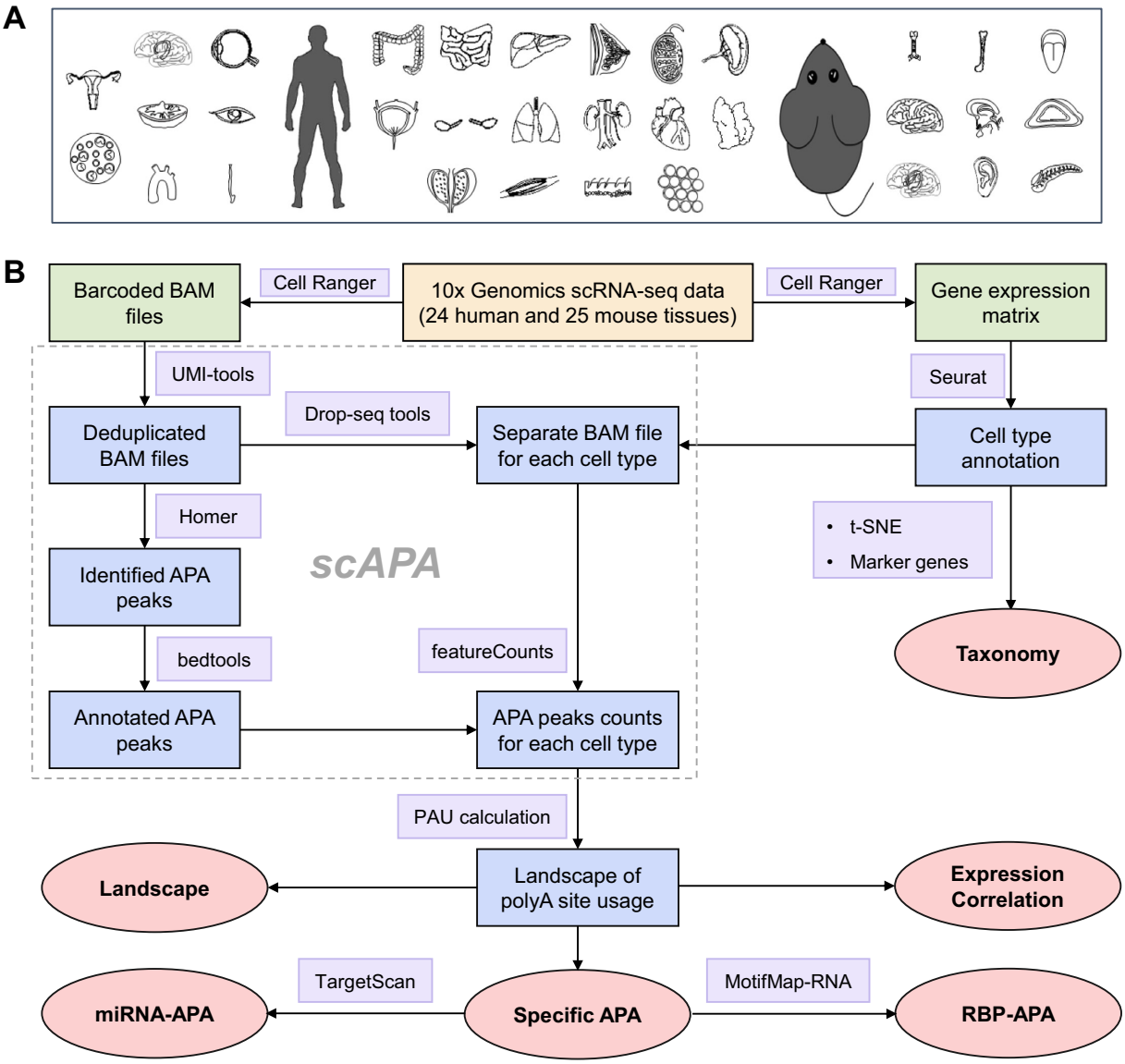

Figure 1. Framework to construct the scAPAatlas database. (**A**) The scAPAatlas database incorporates scRNA-seq data of 24 human and 25 mouse normal tissues. (**B**) The flowchart depicting the construction pipeline of the scAPAatlas database.

by cell-type-specific APA events were identified to connect miRNAs or RBPs to APA events (Figure 1B).

Web design and interface

Based on the above results, we constructed a versatile database with user-friendly interface, scAPAatlas, for browsing, searching and visualizing multi-layer information of cell-type-specific events in different human and mouse tissues. This database could also be used to explore the correlations between APA and gene expression, crosstalk between APA and miRNAs or RBPs (Figure 2). In scAPAatlas, we designed six function modules, including (i) 'Taxonomy' module, (ii) 'APA landscape' module, (iii) 'Specific APA' module, (iv) 'Expression Correlation' module, (v) 'miRNA-APA' module and (vi) 'RBP-APA' module (Figure 2). In addition, we designed a 'Home' page to give a brief description of this database and summary statistics of each function module. Besides, scAPAatlas also provides a 'Download' page for researchers downloading all data in batches. The 'Help' page was designed to provide sufficient guidelines so that first-time users could easily access and obtain information from the database. The 'About' page provides a detailed description of the database, including the introduction to scAPAatlas, the pipeline of database construction and the data resources used.

Function modules

The 'Taxonomy' module provides the taxonomy of cell types in each tissue (Figures 2 and 3A). Users could choose a species and a tissue to search within. The scAPAatlas will return a 2D representation of various cell types based on the t-SNE mapping of single cells. In addition, A dot-plot

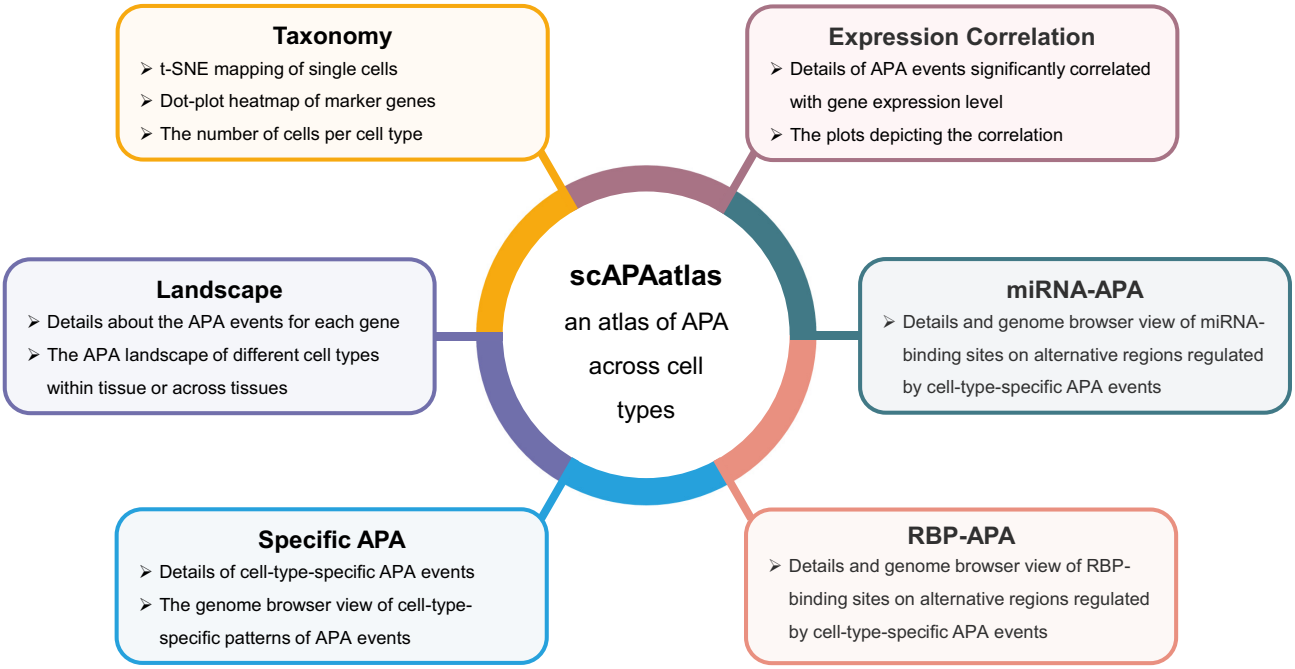

Figure 2. The overall design of function modules in scAPAatlas. The scAPAatlas database provides six function modules: (i) 'Taxonomy' module; (ii) 'Landscape' module; (iii) 'Specific APA' module; (iv) 'Expression Correlation' module; (v) 'miRNA-APA' module; (vi) 'RBP-APA' module.

heatmap depicting the expression level and distributions of representative marker genes for each cell type is provided in the 'Marker Gene' panel. To further detail the cellular taxonomy in each tissue, users could select the 'Summary' panel to retrieve a table with the number of single cells per cell type. For example, we can choose 'Human' for species and 'Heart' for tissue to retrieve the cell taxonomy of heart comprising nine different cell types, including endothelial cells, pericytes, fibroblasts, atrial cardiomyocytes, ventricular cardiomyocytes, lymphoid cells, myeloid cells, neuronal cells and smooth muscle cells (Figure 3A).

The 'Landscape' module enables users to query the APA landscape in different cell types within tissue or across tissues (Figures 2 and 3B). Users could choose a species and enter a gene symbol to query the APA landscape in different cell types within a certain tissue by selecting the tissue from the drop-down list. Details about the polyA sites with gene symbol, APA ID and APA position will be displayed in a table. For each polyA site, a bar chart is provided to depict the landscape of APA in different cell types within the tissue. When users select 'All' from the drop-down list of 'Tissue' box, the bar chart will be rebuilt to display the APA landscape in different cell types across tissues. For example, the APA landscape of the gene *CLOCK* in human heart could be retrieved by entering the gene symbol and selecting 'Human' for species and 'Heart' for tissue (Figure 3B).

The 'Specific APA' module enables users to interrogate cell-type-specific APA events identified for each cell type (Figures 2 and 3C). Users could choose a species and a tissue to search within. The scAPAatlas returns a table with the cell type, gene symbol, gene position, specific APA ID, specific APA position, PAU, difference of PAU and q-value (Figure 3C). When users choose a different cell type, the table will be rebuilt to display the query results. Besides, users could also query cell-type-specific APA events for a specific gene by entering the gene symbol. In addition, when users click the 'Display' button in the 'JBrowse' column for a cell-type-specific APA event, an easy-to-use genome browser implemented in a modal window will be displayed. The reads coverage of each cell type will be displayed as a track in the genome browser with annotation tracks of gene models and identified polyA sites. Thus, users could navigate easily to particular genomic coordinates to view the cell-type-specific pattern of APA events (Figure 3C). This module could benefit researchers in charting cell types with APA and understanding APA at the cell-type level. For instance, by browsing and searching cell-type-specific APA events in human heart, we observed that the gene *CLOCK* which plays an important role in regulation of circadian rhythms exhibits a cell-type-specific APA pattern in human heart. The atrial cardiomyocytes predominantly use the distal polyA site of *CLOCK* to express the long isoform, while other cell types express both the long and short APA isoforms (Figure 3C).

In the 'Expression Correlation' module, the genes whose APA events are significantly correlated with their gene expression level are provided (Figures 2 and 3D). When users choose a species, a table containing gene symbol, APA ID, APA position, the Spearman's correlation coefficient (Rs), P-value of each significantly expression-correlated APA events will be displayed (Figure 3D). Users could search for a gene by entering a gene symbol in the search box. Besides, clicking the 'Plot' button in the table opens a gene-wise scatter plot that displays the correlation between PAU and gene expression level, accompanied by their Spearman's correla-

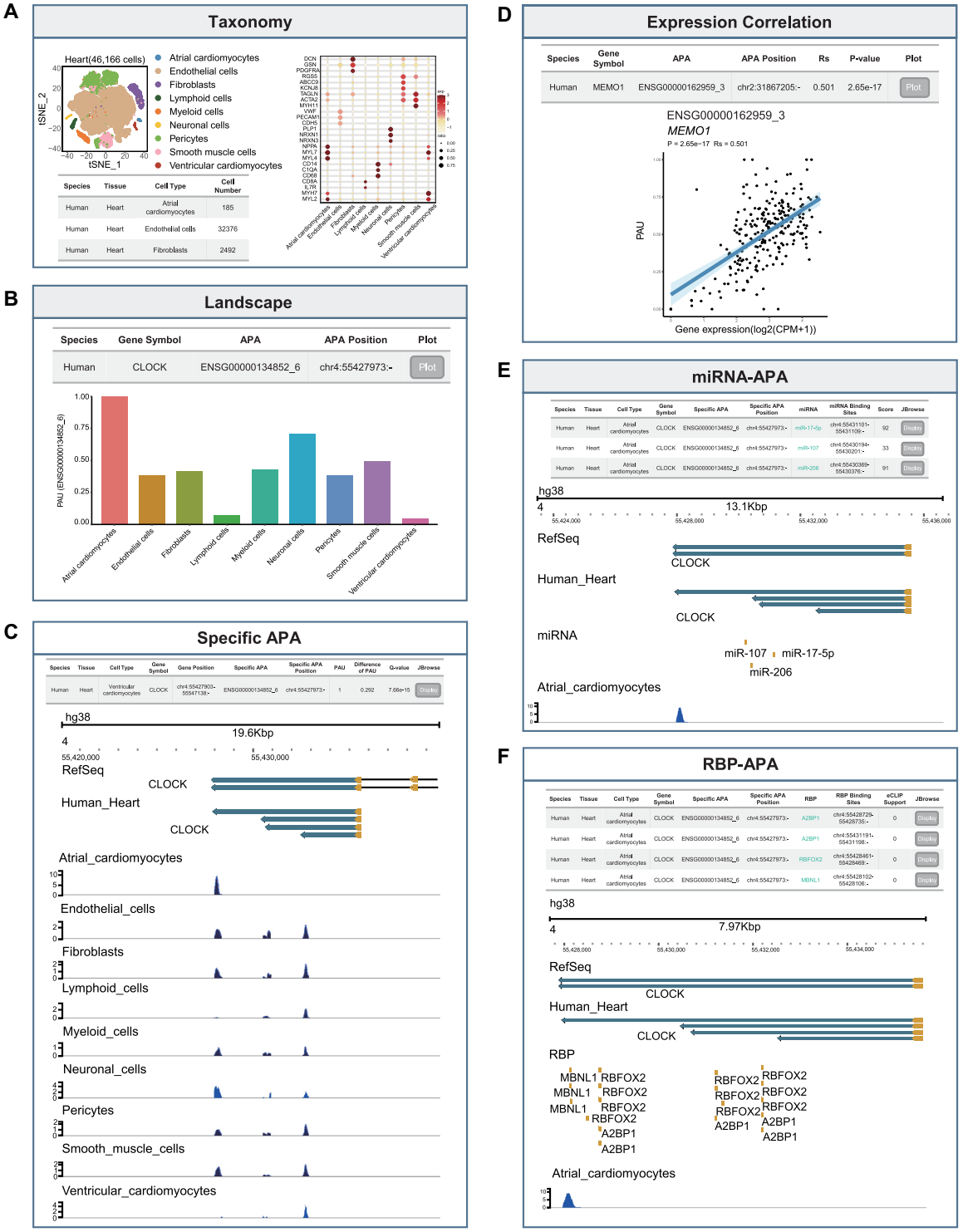

Figure 3. The schematic features in each function module of scAPAatlas. (**A**) In the 'Taxonomy' module, users can choose 'Human' for species and 'Heart' for tissue to visualize the t-SNE mapping of 9 different cell types and the dot-plot heatmap of representative marker genes. Additionally, the table with the number of single cells per cell type is also provided. (**B**) Example of the 'Landscape' module showing the APA landscape of the gene *CLOCK* in human heart. The bar chart showing the APA landscape in different cell types. (**C**) Example of the 'Specific APA' module showing the cell-type-specific APA events of *CLOCK* in atrial cardiomyocytes of heart. The table provides the detailed information of the APA events and the genome browser view showing cell-type-specific patterns of APA events across cell types in heart. (**D**) Example of the 'Expression Correlation' module showing that the polyA site (ENSG00000162959_3) usage of *MEMO1* is positively correlated with its gene expression. (**E**) Example of the 'miRNA-APA' module showing miRNA-binding sites on alternative regions regulated by cell-type-specific APA event of *CLOCK*. The table provides the detailed information of miR-17–5p, miR-107 and miR-206 binding sites on *CLOCK*, which also be visualized in genome browser view. (**F**) Example of the 'RBP-APA' module showing RBP-binding sites on alternative regions regulated by cell-type-specific APA event of *CLOCK*. The table provides the detailed information of A2BP1, RBFOX2 and MBNL1 binding sites on *CLOCK*, which also be visualized in genome browser view.

tion coefficients and *P*-value (Figure 3D). The correlations between APA and gene expression will help researchers recognize the effects of APA on gene expression. As an example, the PAU of the short isoform of *MEMO1* is positively correlated with its gene expression level, which may be in line with previous studies that 3′UTR shortening through APA may upregulated its gene expression level by escaping post-transcriptional repression (40) (Figure 3D).

The 'miRNA-APA' module allows users to browse, search and visualize miRNA-binding sites on alternative regions regulated by cell-type-specific APA events (Figures 2 and 3E). Details with the cell type, gene symbol, specific APA ID, specific APA position, miRNA name, miRNA-binding sites and context++ score percentile estimated from TargetScan will be displayed in a table (Figure 3E). To search results of a certain cell type, users can choose a cell type from the drop-down list. Entering a gene symbol in the search box will return the result of a certain gene. A modal window with the genome browser is embedded for each row to display the crosstalk between miRNAs and APA events, so that users could click the 'Display' button to bring it to front (Figure 3E). In addition, users could click the miRNA name to open a webpage from miRBase describing detail information of the miRNA, including miRNA sequence, related scientific literature and word cloud representing the functional roles of the miRNA. This module could facilitate a deep understanding of biological functions of APA at the cell-type level. As an illustration, several miRNA-binding sites were discovered on the alternative regions regulated by APA events of *CLOCK* gene, which are cell-type-specific in atrial cardiomyocytes of heart (Figure 3E). Among these miRNAs, some miRNAs such as miR-17–5p, miR-107 and miR-206 were reported previously to be involved in regulation of circadian rhythms through interacting with *CLOCK* gene (41–43).

Moreover, the 'RBP-APA' module provides RNA-binding sites on alternative regions regulated by cell-type-specific APA events (Figures 2 and 3F). As in 'miRNA-APA' module, users could choose a species and a tissue to retrieve a table with the cell type, gene symbol, specific APA ID, specific APA position, RBP and RBP-binding sites (Figure 3F). Specially, additional eCLIP-seq data support evidence is denoted for each RBP-binding site in the table for human data. The select box for cell type and search box for gene symbol were designed for users to search data of a certain cell type or a certain gene, respectively. Users could click the 'Display' button to open a modal window with the genome browser to visualize the crosstalk between RBP and APA events (Figure 3F). Moreover, each RBP name provides a link to the webpage from CISBP-RNA, which enables users to conveniently get the detailed information of the RBP, including RBP-binding motif, RNA-recognition motif (RRM) and orthologs. This module could help researchers decipher cell-type-specific APA regulation. For example, several RBP-binding sites were discovered on the alternative regions regulated by cell-type-specific APA events of *CLOCK* gene (Figure 3F). Among these RBPs, some RBPs such as A2BP1, RBFOX2 and MBNL1 are reported previously to be implicated in regulation of alternative polyadenylation (44,45).

Download of tables and figures

All the tables generated in the modules of 'Taxonomy', 'Landscape', 'Specific APA', 'Expression Correlation', 'miRNA-APA' and 'RBP-APA' can be downloaded by clicking the 'Download' button below the table. Moreover, the 't-SNE' plot of cell-type clustering and the dotplot heatmap of representative marker genes in each tissue of 'Taxonomy' module could be downloaded by clicking the 'Download' sign on the bottom left corner of the corresponding plot. The bar charts depicting the APA landscape could be downloaded by clicking the camera icon on the top right corner. Additionally, on the 'Specific APA', 'miRNA-APA' and 'RBP-APA' page, the genome browser view incorporating genome tracks of alignment data, miRNA-binding sites and RBP-binding sites, could be downloaded by using the built-in plugin 'Export SVG' of JBrowse. In addition, users could download all data in batches from the 'Download' page. Users can freely utilize the plots and explore the tables they downloaded from scAPAatlas in their studies.

DISCUSSION AND FUTURE DIRECTIONS

Here, we present a user-friendly database, scAPAatlas, for exploring APA at the cell-type level in diverse human and mouse tissues. Through collecting and processing 305 scRNA-seq datasets from 49 studies, scAPAatlas provides the APA landscape in different cell types. Cell-type-specific APA events were estimated for each cell type in a certain tissue, which provides an additional layer of information in charting cell identity and help researchers find out potential functional APA events. The scAPAatlas also examined the correlations between APA and gene expression level, benefiting researchers recognizing those APA events which affect gene expression level. In addition, putative miRNA-binding and RBP-binding sites on alternative regions regulated by APA events are also estimated for each cell-type-specific APA event, which could give a hint of the underlying mechanisms of post-transcriptional regulation.

As advances in high-throughput sequencing technology, scRNA-seq will be applied to more tissues and more species, we will continue to collect new incoming scRNA-seq data. Besides, we will further integrate other types of functional genomic data, such as RNA modification data and RNA degradome data, with the APA landscape. We will keep maintaining scAPAatlas to ensure it remains a valuable resource for the research community.

DATA AVAILABILITY

ScAPAatlas is publicly and freely available at http://www.bioailab.com:3838/scAPAatlas or http://47.100.223.144:3838/scAPAatlas. The tables and figures in scAPAatlas could be freely downloaded.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We gratefully acknowledge the technical support by the IT center of Tianjin Medical University.

FUNDING

National Natural Science Foundation of China [32100534]; Talent Excellence Program from Tianjin Medical University (to Y. Y). Funding for open access charge: National Natural Science Foundation of China [32100534]; Talent Excellence Program from Tianjin Medical University.

Conflict of interest statement. None declared.

REFERENCES

1. Tian,B. and Manley,J.L. (2017) Alternative polyadenylation of mRNA precursors. *Nat. Rev. Mol. Cell Biol.*, **18**, 18–30.
2. Gruber,A.J. and Zavolan,M. (2019) Alternative cleavage and polyadenylation in health and disease. *Nat. Rev. Genet.*, **20**, 599–614.
3. Hoque,M., Ji,Z., Zheng,D., Luo,W., Li,W., You,B., Park,J.Y., Yehia,G. and Tian,B. (2013) Analysis of alternative cleavage and polyadenylation by 3' region extraction and deep sequencing. *Nat. Methods*, **10**, 133–139.
4. Tian,B., Hu,J., Zhang,H. and Lutz,C.S. (2005) A large-scale analysis of mRNA polyadenylation of human and mouse genes. *Nucleic Acids Res.*, **33**, 201–212.
5. Berkovits,B.D. and Mayr,C. (2015) Alternative 3' UTRs act as scaffolds to regulate membrane protein localization. *Nature*, **522**, 363–367.
6. Di Giammartino,DC., Nishida,K. and Manley,J.L. (2011) Mechanisms and consequences of alternative polyadenylation. *Mol. Cell*, **43**, 853–866.
7. Fabian,M.R., Sonenberg,N. and Filipowicz,W. (2010) Regulation of mRNA translation and stability by microRNAs. *Annu. Rev. Biochem.*, **79**, 351–379.
8. Blazie,S.M., Geissel,H.C., Wilky,H., Joshi,R., Newbern,J. and Mangone,M. (2017) Alternative polyadenylation directs tissue-specific miRNA targeting in caenorhabditis elegans somatic tissues. *Genetics*, **206**, 757–774.
9. Brumbaugh,J., Di Stefano,B., Wang,X., Borkent,M., Forouzmand,E., Clowers,K.J., Ji,F., Schwarz,B.A., Kalocsay,M., Elledge,S.J. *et al.* (2018) Nudt21 controls cell fate by connecting alternative polyadenylation to chromatin signaling. *Cell*, **172**, 106–120.
10. Masamha,C.P., Xia,Z., Yang,J., Albrecht,T.R., Li,M., Shyu,A.-B., Li,W. and Wagner,E.J. (2014) CFIm25 links alternative polyadenylation to glioblastoma tumour suppression. *Nature*, **510**, 412–416.
11. Grassi,E., Santoro,R., Umbach,A., Grosso,A., Oliviero,S., Neri,F., Conti,L., Ala,U., Provero,P., DiCunto,F. *et al.* (2018) Choice of alternative polyadenylation sites, mediated by the RNA-Binding protein elavl3, plays a role in differentiation of inhibitory neuronal progenitors. *Front Cell Neurosci.*, **12**, 518.
12. Lianoglou,S., Garg,V., Yang,J.L., Leslie,C.S. and Mayr,C. (2013) Ubiquitously transcribed genes use alternative polyadenylation to achieve tissue-specific expression. *Genes Dev.*, **27**, 2380–2396.
13. Zhang,H., Lee,J.Y. and Tian,B. (2005) Biased alternative polyadenylation in human tissues. *Genome Biol.*, **6**, R100.
14. Hwang,H.-W., Saito,Y., Park,C.Y., Blachère,N.E., Tajima,Y., Fak,J.J., Zucker-Scharff,I. and Darnell,R.B. (2017) cTag-PAPERCLIP reveals alternative polyadenylation promotes cell-type specific protein diversity and shifts araf isoforms with microglia activation. *Neuron*, **95**, 1334–1349.
15. Singh,I., Lee,S.-H., Sperling,A.S., Samur,M.K., Tai,Y.-T., Fulciniti,M., Munshi,N.C., Mayr,C. and Leslie,C.S. (2018) Widespread intronic polyadenylation diversifies immune cell transcriptomes. *Nat. Commun.*, **9**, 1716.
16. Yang,Y., Paul,A., Bach,T.N., Huang,Z.J. and Zhang,M.Q. (2021) Single-cell alternative polyadenylation analysis delineates GABAergic neuron types. *BMC Biol.*, **19**, 144.
17. Lee,J.Y., Yeh,I., Park,J.Y. and Tian,B. (2007) PolyA_DB 2: mRNA polyadenylation sites in vertebrate genes. *Nucleic Acids Res.*, **35**, D165–D168.
18. Brockman,J.M., Singh,P., Liu,D., Quinlan,S., Salisbury,J. and Graber,J.H. (2005) PACdb: PolyA cleavage site and 3′-UTR database. *Bioinformatics*, **21**, 3691–3693.
19. Wang,R., Nambiar,R., Zheng,D. and Tian,B. (2018) PolyA_DB 3 catalogs cleavage and polyadenylation sites identified by deep sequencing in multiple genomes. *Nucleic Acids Res.*, **46**, D315–D319.
20. Herrmann,C.J., Schmidt,R., Kanitz,A., Artimo,P., Gruber,A.J. and Zavolan,M. (2020) PolyASite 2.0: a consolidated atlas of polyadenylation sites from 3′ end sequencing. *Nucleic Acids Res.*, **48**, D174–D179.
21. Müller,S., Rycak,L., Afonso-Grunz,F., Winter,P., Zawada,A.M., Damrath,E., Scheider,J., Schmäh,J., Koch,I., Kahl,G. *et al.* (2014) APADB: a database for alternative polyadenylation and microRNA regulation events. *Database*, **2014**, bau076.
22. Feng,X., Li,L., Wagner,E.J. and Li,W. (2018) TC3A: the cancer 3′ UTR atlas. *Nucleic Acids Res.*, **46**, D1027–D1030.
23. Hong,W., Ruan,H., Zhang,Z., Ye,Y., Liu,Y., Li,S., Jing,Y., Zhang,H., Diao,L., Liang,H. *et al.* (2020) APAatlas: decoding alternative polyadenylation across human tissues. *Nucleic Acids Res.*, **48**, D34–D39.
24. Hashimshony,T., Senderovich,N., Avital,G., Klochendler,A., de Leeuw,Y., Anavy,L., Gennert,D., Li,S., Livak,K.J., Rozenblatt-Rosen,O. *et al.* (2016) CEL-Seq2: sensitive highly-multiplexed single-cell RNA-Seq. *Genome Biol.*, **17**, 77.
25. Picelli,S., Faridani,O.R., Björklund,Å.K., Winberg,G., Sagasser,S. and Sandberg,R. (2014) Full-length RNA-seq from single cells using Smart-seq2. *Nat. Protoc.*, **9**, 171–181.
26. Zheng,G.X.Y., Terry,J.M., Belgrader,P., Ryvkin,P., Bent,Z.W., Wilson,R., Ziraldo,S.B., Wheeler,T.D., McDermott,G.P., Zhu,J. *et al.* (2017) Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.*, **8**, 14049.
27. Shulman,E.D. and Elkon,R. (2019) Cell-type-specific analysis of alternative polyadenylation using single-cell transcriptomics data. *Nucleic Acids Res.*, **47**, 10027–10039.
28. Yates,A.D., Achuthan,P., Akanni,W., Allen,J., Allen,J., Alvarez-Jarreta,J., Amode,M.R., Armean,I.M., Azov,A.G., Bennett,R. *et al.* (2020) Ensembl 2020. *Nucleic Acids Res.*, **48**, D682–D688.
29. Satija,R., Farrell,J.A., Gennert,D., Schier,A.F. and Regev,A. (2015) Spatial reconstruction of single-cell gene expression data. *Nat. Biotechnol.*, **33**, 495–502.
30. Macosko,EZ., Basu,A., Satija,R., Nemesh,J., Shekhar,K., Goldman,M., Tirosh,I., Bialas,AR., Kamitaki,N., Martersteck,EM. *et al.* (2015) Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell*, **161**, 1202–1214.
31. Smith,T., Heger,A. and Sudbery,I. (2017) UMI-tools: modeling sequencing errors in Unique Molecular Identifiers to improve quantification accuracy. *Genome Res.*, **27**, 491–499.
32. 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.
33. Quinlan,A.R. (2014) BEDTools: The Swiss-Army tool for genome feature analysis. *Curr. Protoc. Bioinform.*, **47**, 11.12.11–11.12.34.
34. Li,H., Handsaker,B., Wysoker,A., Fennell,T., Ruan,J., Homer,N., Marth,G., Abecasis,G., Durbin,R. and Genome Project Data Processing, S. (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, **25**, 2078–2079.
35. Liao,Y., Smyth,G.K. and Shi,W. (2014) featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*, **30**, 923–930.
36. Garcia,D.M., Baek,D., Shin,C., Bell,G.W., Grimson,A. and Bartel,D.P. (2011) Weak seed-pairing stability and high target-site abundance decrease the proficiency of lsy-6 and other microRNAs. *Nat. Struct. Mol. Biol.*, **18**, 1139–1146.
37. Buels,R., Yao,E., Diesh,C.M., Hayes,R.D., Munoz-Torres,M., Helt,G., Goodstein,D.M., Elsik,C.G., Lewis,S.E., Stein,L. *et al.* (2016) JBrowse: a dynamic web platform for genome visualization and analysis. *Genome Biol.*, **17**, 66.
38. Liu,Y., Sun,S., Bredy,T., Wood,M., Spitale,R.C. and Baldi,P. (2017) MotifMap-RNA: a genome-wide map of RBP binding sites. *Bioinformatics*, **33**, 2029–2031.
39. Van Nostrand,E.L., Pratt,G.A., Shishkin,A.A., Gelboin-Burkhart,C., Fang,M.Y., Sundararaman,B., Blue,S.M., Nguyen,T.B., Surka,C., Elkins,K. *et al.* (2016) Robust transcriptome-wide discovery of

RNA-binding protein binding sites with enhanced CLIP (eCLIP). *Nat. Methods*, **13**, 508–514.
40. Sandberg,R., Neilson,J.R., Sarma,A., Sharp,P.A. and Burge,C.B. (2008) Proliferating cells express mRNAs with shortened 3' untranslated regions and fewer microRNA target sites. *Science*, **320**, 1643–1647.
41. Gao,Q., Zhou,L., Yang,S.-Y. and Cao,J.-M. (2016) A novel role of microRNA 17-5p in the modulation of circadian rhythm. *Sci. Rep.*, **6**, 30070.
42. Daimiel-Ruiz,L., Klett-Mingo,M., Konstantinidou,V., Micó,V., Aranda,J.F., García,B., Martínez-Botas,J., Dávalos,A., Fernández-Hernando,C. and Ordovás,J.M. (2015) Dietary lipids modulate the expression of miR-107, an miRNA that regulates the circadian system. *Mol. Nutr. Food Res.*, **59**, 552–565.
43. Zhou,W., Li,Y., Wang,X., Wu,L. and Wang,Y. (2011) MiR-206-mediated dynamic mechanism of the mammalian circadian clock. *BMC Syst. Biol.*, **5**, 141.
44. Chen,P.-F., Hsiao,J.S., Sirois,C.L. and Chamberlain,S.J. (2016) RBFOX1 and RBFOX2 are dispensable in iPSCs and iPSC-derived neurons and do not contribute to neural-specific paternal UBE3A silencing. *Sci. Rep.*, **6**, 25368.
45. Batra,R., Charizanis,K., Manchanda,M., Mohan,A., Li,M., Finn,DJ., Goodwin,M., Zhang,C., Sobczak,K., Thornton,CA. *et al.* (2014) Loss of MBNL leads to disruption of developmentally regulated alternative polyadenylation in RNA-Mediated disease. *Mol. Cell*, **56**, 311–322.