

DirectRMDb: a database of post-transcriptional RNA modifications unveiled from direct RNA sequencing technology

Yuxin Zhang^{1,2,5,†}, Jie Jiang^{2,5,†}, Jiongming Ma^{1,2,5,†}, Zhen Wei^{2,6,*}, Yue Wang^{3,7}, Bowen Song^{3,5}, Jia Meng^{2,4,5}, Guifang Jia⁸, João Pedro de Magalhães⁶, Daniel J. Rigen⁵, Daiyun Hang^{2,7,*} and Kunqi Chen^{1,9,*}

¹Key Laboratory of Ministry of Education for Gastrointestinal Cancer, School of Basic Medical Sciences, Fujian Medical University, Fuzhou, Fujian 350004, China, ²Department of Biological Sciences, Xi'anJiaotong-Liverpool University, Suzhou, Jiangsu 215123, China, ³Department of Mathematical Sciences, Xi'anJiaotong-Liverpool University, Suzhou, Jiangsu 215123, China, ⁴AI University Research Centre, Xi'an Jiaotong-Liverpool University, Suzhou, Jiangsu 215123, China, ⁵Institute of Systems, Molecular and Integrative Biology, Xi'anJiaotong-Liverpool University, Suzhou, Jiangsu 215123, China, ⁶Institute of Life Course and Medical Sciences, Xi'anJiaotong-Liverpool University, Suzhou, Jiangsu 215123, China, ⁷Department of Computer Science, University of Liverpool, L69 7ZB, Liverpool, UK, ⁸Synthetic and Functional Biomolecules Center, Beijing National Laboratory for Molecular Sciences, Key Laboratory of Bioorganic Chemistry and Molecular Engineering of Ministry of Education, College of Chemistry and Molecular Engineering, Peking University, Beijing, China and ⁹Fujian Key Laboratory of Tumor Microbiology, Department of Medical Microbiology, School of Basic Medical Sciences, Fujian Medical University, Fuzhou, Fujian 350004, China

Received August 07, 2022; Revised October 20, 2022; Editorial Decision October 22, 2022; Accepted October 25, 2022

ABSTRACT

With advanced technologies to map RNA modifications, our understanding of them has been revolutionized, and they are seen to be far more widespread and important than previously thought. Current next-generation sequencing (NGS)-based modification profiling methods are blind to RNA modifications and thus require selective chemical treatment or antibody immunoprecipitation methods for particular modification types. They also face the problem of short read length, isoform ambiguities, biases and artifacts. Direct RNA sequencing (DRS) technologies, commercialized by Oxford Nanopore Technologies (ONT), enable the direct interrogation of any given modification present in individual transcripts and promise to address the limitations of previous NGS-based methods. Here, we present the first ONT-based database of quantitative RNA modification profiles, DirectRMDb, which includes 16 types of modification and a total of 904,712 modification sites in 25 species identified from 39 independent

studies. In addition to standard functions adopted by existing databases, such as gene annotations and post-transcriptional association analysis, we provide a fresh view of RNA modifications, which enables exploration of the epitranscriptome in an isoform-specific manner. The DirectRMDb database is freely available at: <http://www.rnamd.org/directRMDb/>.

INTRODUCTION

Conceptually similar to DNA modifications, RNA molecules undergo chemical modifications as well. The first RNA chemical modifications were documented in the 1950s in tRNAs and rRNAs (1). To date, >170 different modification types have been described, including N6-methyladenosine (m⁶A), pseudouridine (Ψ), N5-methylcytosine (m⁵C), N1-methyladenosine (m¹A), methylation of 2'-O in the four nucleotides (i.e. Am, Tm, Cm and Gm) and N7-methylguanosine (m⁷G) (2–7). With recent advanced technologies to map these RNA modifications, our understanding of them has been revolutionized, and they are now understood to be far more widespread and important than previously thought. Systematic studies

*To whom correspondence should be addressed. Tel: +86 0591 22862299; Email: kunqi.chen@fjmu.edu.cn

Correspondence may also be addressed to Zhen Wei. Email: zhen.wei01@xjtlu.edu.cn

Correspondence may also be addressed to Daiyun Hang. Email: daiyun.huang@liverpool.ac.uk

†Contributed equally to this work.

of this post-transcriptional regulatory layer have revealed its profound roles in shaping cellular processes, modulating disease risks, and governing cellular fate (8–11). For instance, m⁶A, one of the most prevalent RNA modifications, is proven to regulate cardiac gene expression, cell growth, stress response, stabilize junctional RNA, etc (12–15). Pseudouridine, the first discovered post-transcriptional modification (16), has recently been implicated in tumor development, maintenance, and progression (17).

RNA-seq has become a popular choice for analyzing complex epitranscriptomics. However, next-generation sequencing (NGS) platforms are typically blind to nucleotide modifications and thus need specific protocols to highlight RNA modifications on the molecules. These typically involve three strategies: (i) antibody immunoprecipitation, which specifically recognizes the modified bases with antibodies; (ii) enzyme-digestion, where RNAs are digested with modification-sensitive enzymes; (iii) chemical treatment, using chemical compounds that selectively react with the modified ribonucleotide of interest. Example of immunoprecipitation methods include m⁶A-seq (3), PA-m⁶A-seq (18), m⁶A-CLIP-seq (19), miCLIP (20), m⁶A-LAIC-Seq (21), m⁶ACE-Seq (22). Mazter-seq (23), m⁶A-REF-Seq (24), and DART-seq (25) are enzyme digestion-based methods that quantify m⁶A modification with single-nucleotide resolution. Pseudo-seq (6) and AlkAniline-seq (26) are typical chemical-based detection methods. These methods are similar in that they enrich fragments harboring modified ribonucleotides, followed by high-throughput sequencing and bioinformatics analysis to detect these changes.

Although these methods provide invaluable information, they are limited by the availability of high-quality antibodies and the lack of practical chemical reactivities towards a particular RNA modification (27). Thus, only a few of the over 170 known modification types can be accurately and effectively profiled. When selective antibodies or chemical treatments are available, the RNA modification to be studied should be chosen beforehand, and customized protocols must be designed for the chosen type, limiting our ability to characterize the epitranscriptome in a systematic and flexible manner (28). Also, these methods require multiple ligation and amplification steps during the library preparation, introducing undesired biases and artifacts (29,30). Finally, with respect to the sequencing itself, NGS platforms face the problem of short read length. Mapping modifications on highly repetitive splicing isoforms and characterizing the co-occurrence of distant modifications in the same transcripts remain challenges (27). Thus, most of existing NGS-based methods have the isoform-ambiguity issue and they report only genome-based coordinate of RNA modifications.

The continuing discoveries of novel classes of RNA modifications in various organisms call for more sensitive, plastic, and convenient modification profiling methods. A promising alternative to NGS technologies is the direct RNA sequencing (DRS) platform developed by Oxford Nanopore Technologies (ONT) (31). Each nucleotide will cause distinct ionic current signals as it passes through a sensitive channel. This platform infers the RNA sequence by deconvoluting the serial electric signal event when the

molecule is threading through the sensitive protein channel (32). Natural modifications along the molecule can result in characteristic signals that suggest both the position and identity of modifications (33). Theoretically, direct RNA sequencing allows the real-time and simultaneous detection of any given modification in the native RNA molecule. Additionally, nanopore sequencing offers ultra-long reads that can cover the entire length of the RNA molecule, which benefits the study of RNA modifications on splicing isoforms (34).

ONT sequencing platforms have yielded robust data of reasonable quality, and several pilot studies have detected RNA modifications from the data. For example, EpiNano (m⁶A) (35), ELIGOS (27), DRUMMER (36) and the work of Parker *et al.* (37) screen RNA modifications by examining the sequencing error profiles. Another body of work, such as xPore (38), m⁶Anet (m⁶A) (39), MINES (m⁶A) (34), nanom⁶A (m⁶A) (40), and nanoPsu (pseudouridine) (41) utilized the variation in current signal intensities. These tools were confirmed to have high accuracy for modification detection with single-nucleotide resolution.

To date, various comprehensive databases of RNA modification sites reported by NGS approaches are publicly available, including MODOMICS (42), RMBase (43), REPIC (44), m⁶A-atlas (45), m⁵C-atlas (46), MeT-DB (47), RMVar (48) and M6A2Target (49), which have together provided invaluable information to help decipher the complexities of epitranscriptomes. However, due to the low sensitivity and detection chemistry of NGS-based approaches, a huge proportion of modified sites have not been detected, and the landscape of RNA modifications on the transcriptome is yet to be well-studied (50). To address this gap, we have developed DirectRMDb, the first comprehensive database of RNA modification sites derived from direct RNA sequencing data. In this study, a collection of 16 quantitative modification profiles among 25 species and various cell types or tissues under different conditions were integrated from direct RNA sequencing samples. Data from other studies or techniques were also collected to validate the collected sites. A significant advantage of DirectRMDb is that it provides isoform-level information, including isoform-specific distributions of RNA modifications, isoform expression levels and secondary structure. We constructed a user-friendly web interface for the query, visualization, and sharing of the modification profiles and their association with transcriptional and post-transcriptional regulatory machinery (i.e. RNA binding proteins, miRNAs, splicing events), as well as their potential involvement in pathogenesis. As the first DRS-based database, DirectRMDb is expected to provide new insight into the complex epitranscriptome (Figure 1).

MATERIALS AND METHODS

Collection of candidate modification sites

125 direct RNA sequencing samples for 25 species, including 44 FAST5 and 81 FASTQ samples, were collected from 39 independent studies in the Gene Expression Omnibus (GEO) database (Supplementary Table S1). Eight modification detection tools, namely nanom⁶A (40), MINES (34), xPore (38), m⁶Anet (39), DRUMMER (36), ELIGOS

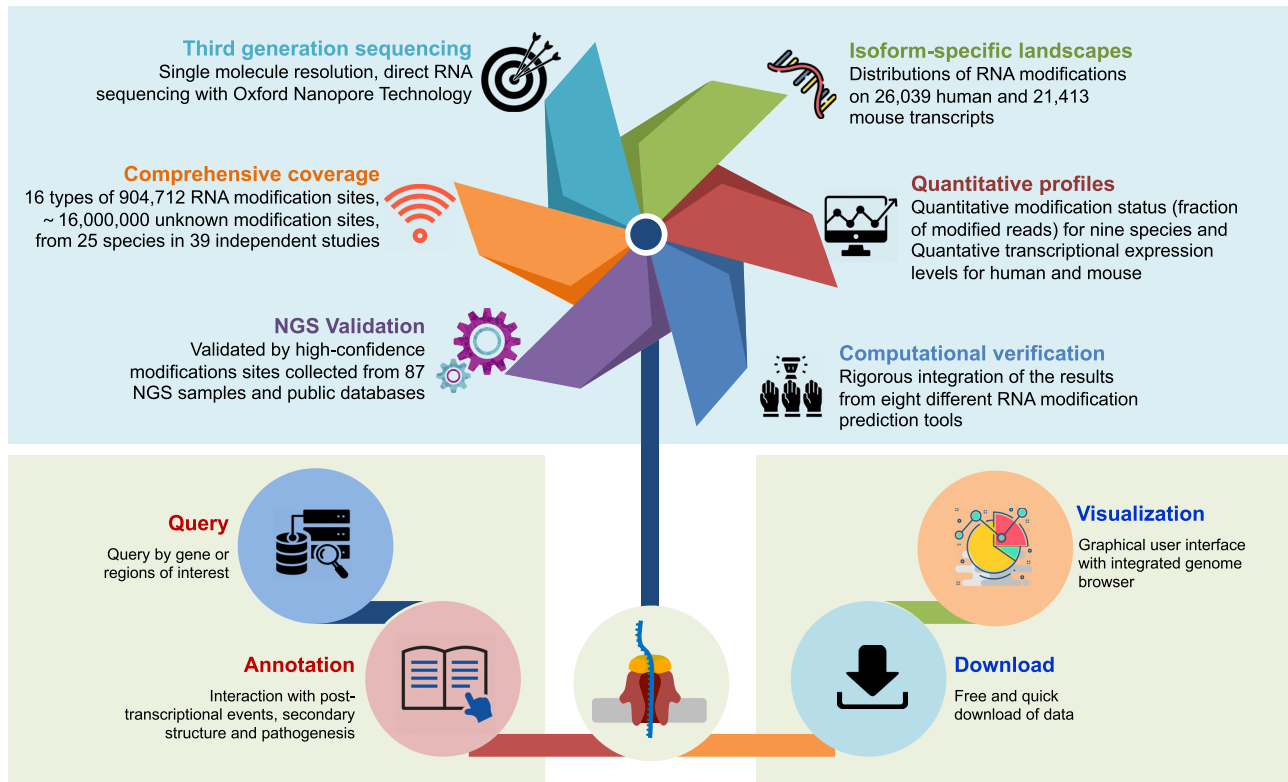


Figure 1. The overall design of DirectRMDb. DirectRMDb is the first comprehensive database that integrates quantitative modification profiles determined by direct RNA sequencing. For quality assurance, eight different software tools for mining RNA modifications were rigorously integrated, and additional next-generation sequencing samples were collected for validation. DirectRMDb provides an isoform-specific view of modification sites, including their distribution on individual transcripts and the secondary structure predicted from the RNA primary sequences. The potential involvement of reported sites in pathogenesis and their potential interactions with post-transcriptional machinery can also be queried.

(27), the work of Parker *et al.* (37), and Nanopsu (41) were used to infer possible modification sites from samples (Table 1). It is worth noting that although direct RNA sequencing allows the detection of RNA modifications with an isoform-level resolution, some tools (e.g. MINES) still rely on genome-level features and thus cannot distinguish between transcripts. Supplementary Figure S1 shows the general workflow for candidate site collection. As the colors indicate, the eight tools can be categorized into three classes in terms of their required input and thus the different pipelines of data pre-processing: (i) Tombo-based (i.e. nanom6A and MINES): the raw data (FAST5) was re-squiggled (i.e. a new assignment from current signal level data to the reference sequences was defined) with either transcriptome or genome reference using Tombo ‘re-squiggle’ function. Specifically for MINES, the Tombo ‘de novo modification detection’ function was used to infer non-canonical bases from the re-squiggled current signals. This Tombo output was provided to nanom6A and MINES as input and candidate m⁶A sites returned. (ii) Nanopolish-based (i.e. xPore and m6Anet): the Nanopolish (51) ‘event-align’ function was used to map the signal events extracted from the raw FAST5 sample to the reference transcriptome. m6Anet and xPore then analyzed the output TXT files to identify possible modifications. It is notable that xPore is a comparative method, which requires modification-free sam-

ples as control. (iii) BAM-based (i.e. DRUMMER, ELIGOS, the work of Parker *et al.*, and NanoPsu): FAST5 samples were base called into FASTQ format with Guppy before alignment. Base-called reads, as well as downloaded FASTQ samples, were aligned to either reference genome or transcriptome using Minimap2 (52) with -ax map-on settings. The resulting SAM files were transformed into BAM files, sorted and indexed with Samtools (53). ELIGOS and NanoPsu examine the error distribution profiles from the alignment file directly, while DRUMMER and Simpson’s work requires control samples to perform the modification detection. Samples were analyzed by some or all of the eight tools depending on their data format (i.e. FASTQ or FAST5), the availability of control samples, and authentic reference sequences. References used for each species are summarized in Supplementary Table S2.

The landscape of RNA modifications on transcripts

Nanom6A, xPore, m6Anet, and DRUMMER detect bulk-level RNA modifications by examining either error distribution profiles or current signals distributions along transcripts. Nanom6A maps detected sites to the reference genome and present the results with genome coordinates. Therefore, only xPore, m6Anet and DRUMMER were used to predict modifications in individual transcripts. The work-

Table 1. brief description and comparison of modification calling tools

	Modification	Input	Isoform-level?	Control sample?	Algorithm
nanoPsu	Ψ	BAM	No	No	Sequencing error
ELIGOS	Mixed		No	No	
the work of Parker <i>et al.</i>	/		No	Yes	
Drummer			Yes	Yes	Current signal
xpore		Nanopolish output	Yes	Yes	
m6Anet	m ⁶ A		Yes	No	
MINES		Tombo output	No	No	
Nanom6A			No	No	

Note: ‘/’ means that detected modification types depend on the modification-free samples. For example, if an m⁶A-free sample is used as a control, reported sites are expected to be m⁶A methylation.

flow to run the three tools is shown in Supplementary Figure S1. To compare the isoform-level modification patterns and for the simplicity of results presentation, we converted the transcripts’ coordinates to genomic coordinates while keeping the isoform-level information.

Integration of results and validation by NGS methods

The collection of sites reported by each software could contain a significant proportion of false-positive sites. To ensure reliability, we collated the results from different samples and tools and then applied strict filtration criteria to generate reliable modification profiles for each species. To screen high-confident m⁶A sites, we searched for its known consensus DRACH motifs (where D denotes A, G or U, R denotes A or G, and H denotes A, C or U) from the candidate sites and only sites reported in multiple cases (i.e. reported by more than one tools or conditions) were kept. Candidate uridines suggested by both ELIGOS and NanoPsu (probability > 0.8) were considered pseudouridine sites.

ELIGOS identified a set of putative modification sites without characterizing their modification type. Therefore, NGS technologies were used to label these unknown types of candidate modification sites. Specifically, 10 modification profiles for human (i.e. m¹A, methylation of 2’-O in the four nucleotides, m⁷G, AtoI, m⁶Am and m⁵U, m⁵C), two modification profiles for mouse (i.e. m¹A and m⁵C), three modification profiles (i.e. f⁵C, dihydrouridine and ac⁴c) for yeast, and two modification profiles (i.e. m6A and Y) for *Arabidopsis thaliana* were collected from public resources, including m6A-atlas, m5C-atlas, RMDisease, MODOMICS, and supplementary data of published works. To ensure reliability and save space, unlabeled ELIGOS results were excluded from the final proposed profiles but can be downloaded from DirectRMDDB.

To further validate our results, we collected high-confident modification sites and modification-enriched peaks derived from next generation sequencing samples (Supplementary table S3). Additionally, multiple modification profiles reported by LC-MS techniques were downloaded from MODOMICS and RMBase. Cross-validation was performed between candidate and NGS/LC-MS-derived sites. Sites confirmed by other techniques were clearly labeled. We also compared our results with sites published by other ONT-based modification detection studies (38,54). Overlapped sites were indicated as well.

Secondary structure prediction

RNA plays a vital role in the cell, not only as an intermediate product for the transmission of genetic information, but also as a functional element. Single-stranded-RNA molecules can fold into specific and stable structures. It is known that there is a strong association between their functions and structures (55). The three-dimensional structure of RNA molecules can only be determined by X-ray crystallography, nuclear magnetic resonance, and other laborious and high-cost methods (56). Therefore, we present the secondary structure of isoforms, which is easier to predict computationally. We use RNAfold (57), a widely used RNA secondary structure prediction software, with default parameter settings, to infer the structure from the RNA primary sequences. The landscapes of RNA modifications on each isoform under different conditions were annotated and highlighted on the predicted structure. For a better view, ultra-long reads (>2001nt) were cut into 2001nt fragments that contained modified bases.

Quantitative profiles of putative modification sites

44 FAST5 samples from nine species were collected to quantify the modification status of high-confident modification sites under different cell lines/tissues and conditions. The Tombo ‘de novo modification detection’ function was used to investigate non-canonical (i.e. modified) bases within individual reads and the fractions of modified reads aligned to each genomic position were output with Tombo ‘text output’ command. The modification fraction is used to quantify the modification status of reported sites. In addition to modification status, the transcripts’ expression profiles were also estimated from the BAM file with transcriptome reference using nanocount (58), isoform expression level calculation software designed for direct RNA sequencing data.

Basic annotation for modification sites

Gene annotation files were downloaded from Gencode (Human and Mouse) (59) and NCBI (60) (other species). Those high-confidence sites (genome-wide) were annotated by collected gene annotations and were classified into different gene types and genomic regions using ChIPseeker (61). In addition to basic gene annotation, the potential interactions between modifications and splicing events, miRNA as well as RNA binding proteins (RBPs) were included in

Table 2. The data statistics for DirectRMDB

Species	RNA modifications						Total
	m ⁶ A	Ψ	2'-O-Me	m ⁵ C	m ¹ A	Other	
Human	195 871	134 834	1506	26 033	2979	3803	365 026
Mouse	186 175	45 397	/	970	693	/	233 235
Yeast	148	59	/	/	/	19	226
22 other species	203 973	102 241	/	/	/	11	306 225

Note: The numbers in the table indicate the total count of each modification type. In human, 'Other' refers to m⁷G, m⁵U, m⁶Am, and AtoI modifications, while in yeast, 'Other' refers to ac⁴c, D, Y and f⁵C. Please refer to Supplementary Table S4 for more details.

human and mouse. miRNA target sites, RBP binding, and other events information were obtained from starBase (62), POSTAR (63) and the UCSC genome browser database (64), respectively. Since nanopore sensor protein takes a k-mer (4–6nt) as input each time, the presence of non-canonical bases could cause misleading signals thus influencing the deconvolution of adjacent bases. Therefore, for each site, we indicated the presence of other modifications within 5bp upstream and downstream as a warning of false positives.

Potential involvement of individual modification sites in pathogenesis

It is known that RNA modifications are closely related to the progression of diseases. To investigate the contribution of individual modification sites in disease development, we analyzed their positional relationship with potentially disease-associated genetic mutations. Sites that exactly overlapped with mutations were indicated. Collection of single nucleotides polymorphisms (SNPs), including both common variations and clinical mutations for human and mouse, was downloaded from dbSNP (65).

Database and web interface implementation

MySQL was used to store and manage the metadata. Hypertext Markup Language (HTML), Cascading style sheets (CSS) and Hypertext Preprocessor (PHP) were used to build the web interface. Genome browser JBrowse2 (66) was used to provide an integrated view of reference sequences, modification site information, related RBP binding, splicing, miRNA binding event, and associated SNPs.

RESULTS

The eight modification detection tools, applied to 125 direct RNA sequencing samples, suggested more than 16,000,000 candidate modification sites. By manually integrating, evaluating, and filtering the results, a total of 904 712 sites of 16 chemical modifications, namely m⁶A, Ψ, m¹A, m⁶Am, 2'-O-Me, m⁵U, m⁷G, m⁵C, D, f⁵C, Y and ac⁴c, across 25 species, including *Homo sapiens*, *Mus musculus*, *Arabidopsis thaliana*, *Sus scrofa* and *Escherichia coli*, were confidently identified (Table 2 and Supplementary Table S4). Among these proposed sites, 149 353 human sites and 91 910 mouse sites were further confirmed by other techniques (i.e. NGS techniques and LC-MS) (Supplementary Figure S2). The landscapes of RNA modifications on human and mouse

transcriptomes were evaluated. 225 041 sites in 26 039 distinct human transcripts and 228 558 sites in 21 413 mouse transcripts, corresponding respectively to 88 230 and 112 820 bases on the human and mouse genome, were found. We also predicted the secondary structure of isoforms and calculated their expression levels under specific cell lines and conditions.

Quantitative modification profiles (i.e. the fraction of modified reads) for nine species under 44 different cell lines/tissues and conditions were calculated. Gene annotation of 20 species was successfully performed, while annotation of the *Bipolaris sorokiniana*, *Candida nivariensis*, *Chikungunya virus* and *influenza A virus* failed since no feasible annotation file is available for these species. Since non-canonical bases can influence the deconvolution of adjacent nucleotides due to the nanopore sequencing chemistry, we evaluated the interaction between reported modification sites. A total of 105 581 sites were screened as consecutive modifications (i.e. has other modifications within 5 bp up and downstream). For human and mouse, we also investigated the interaction between RNA modification and RNA binding proteins, splicing sites, and miRNA targets. For human, 171 RNA binding proteins, 826 miRNAs, and 101 587 splicing events are suggested to be associated with respectively 275 956, 54 390 and 108 738 modification sites. For mouse, we identified 39 RNA binding proteins, 905 miRNA and 79 010 splicing events that are related to RNA modifications. Also, 80 614 human sites and 5636 mouse modification sites are documented SNP sites, suggesting their potential involvement in disease development.

Comprehensive atlas of various types of RNA modifications

We constructed DirectRMDB, the first database that integrates direct RNA sequencing data to explore post-transcriptional modifications of RNAs. A user-friendly web interface was provided to search, browse, visualize and download the 16 types of high confidently collected modification sites and their potential relationships with miRNA targets, RBPs, splicing events, and pathogenesis. Jbrowser2 was integrated for interactive exploration of individual sites or regions of interest. We also provided isoform-level information, including the landscape of RNA modifications on individual transcripts, annotated secondary structures and transcripts expression levels under particular cell lines, tissues, and conditions. The DirectRMDB database is freely available at: <http://www.rnamd.org/directRMDB/>, and has a mirror at: www.xjtlu.edu.cn/biologicalsciences/directRMDB.

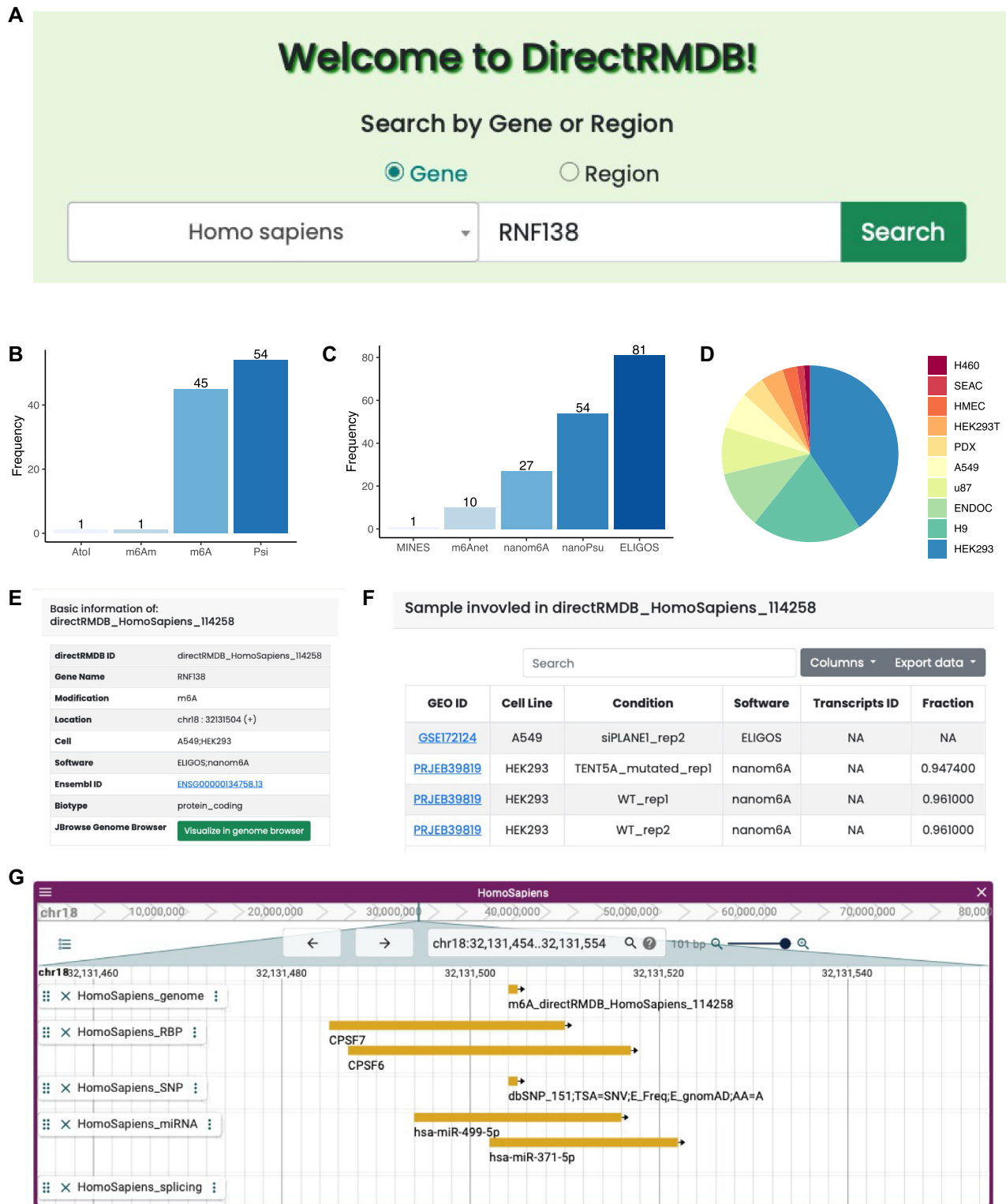


Figure 2. Case study on protein-coding gene RNF138. (A) Searching by gene name. (B) 101 sites of four modification types on the RNF gene. (C) Number of modifications detected by different software. (D) Pie chart of the number of modifications detected in each cell line. (E) Basic information of the example site with ID of 'directRMDb_HomoSapiens_114258'. (F) Conditions involved in the example site and fraction of modified reads under different conditions. (G) Genome browser view of the example site and its relationships with RNA binding proteins, miRNAs and pathogenesis.

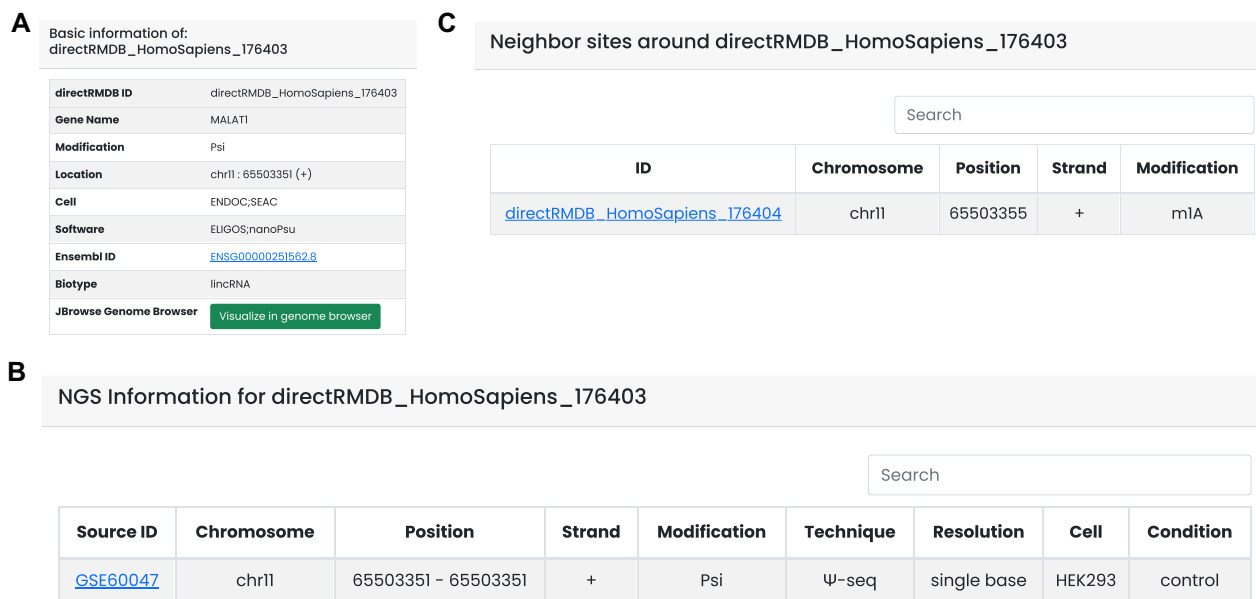


Figure 3. Case study on lncRNA MALAT1. (A) Basic information of the example site with ID of 'directRMDb.HomoSapiens.176403'. (B) Information of other techniques or studies that confirmed the example site. (C) Information of site located within the 11bp region centered by the example site.

Case study on protein-coding gene RNF138

Ring finger protein 138 (RNF138) is a ubiquitin ligase belonging to the E3 ligase family, which harbors a ring finger protein domain, three zinc-finger-like domains, and a ubiquitin-interacting motif (67,68). It promotes cell survival via counteracting apoptotic signaling or directly influencing genome stability. Emerging evidence has linked the RNF138 protein with tumorigenesis, neurodegenerative disorders, and chronic inflammatory conditions (69,70). By searching through the *Homo sapiens* repository from directRMDb with the gene name RNF138 (Figure 2A), 101 entries, with one AtoI, one m⁶Am, 45 m⁶A and 54 Ψ sites, detected by MINES, m6Anet, Nanopsu, nanom6A, and ELIGOS from 7 different cell lines were returned (Figure 2C and D). Among the results, ELIGOS screened the majority of them (i.e. one AtoI, one m⁶Am, 25 m⁶A and 54 Ψ sites), suggesting its high sensitivity in screening modified bases. MINES only contributes to one m6A site, which can be explained by its strict evaluation criteria. Entries without available RBP binding, miRNA, splicing site, SNP, and transcriptomic information or not confirmed by NGS methods can be removed by clicking the corresponding buttons in the top filters box. Users can also retrieve sites of specific modification types from certain cell lines, tools or RNA types (e.g. mRNA, rRNA and tRNA) of interest. Detailed information on individual sites can be acquired by clicking the site ID. Taking 'directRMDb.HomoSapiens_114258' as an example, from the basic information returned, it is an m⁶A site reported by ELIGOS and nanom6A from four different samples under two cell lines (Figure 2E). Figure 2F shows the fraction of modified reads on different samples. Since no transcript information is available for these samples, the transcripts ID column is filled with NAs (for Not Available). In terms of interaction with regulatory events, this m⁶A site is expected to associate with two RNA binding

proteins and two miRNAs. Also, this modification might play an important role in pathogenesis since it is a reported SNP site, where the adenosine base is mutated to guanosine in certain cases (Figure 2G).

Case study on lncRNA MALAT1

Metastasis-associated lung adenocarcinoma transcript 1 (MALAT1), a long non-coding RNA (lncRNA) that has been confirmed to influence cancer development and metastasis (71). 73 modification sites from the DirectRMDb homo sapiens sets, including m⁶A, m⁵C, pseudouridine, and 2'-O-Me, were found on MALAT1 transcripts. Similarly, detailed information for individual modification sites, including interaction with RNA binding proteins, miRNAs, and other sites, can be acquired by clicking the site ID. Taking `directRMDb_HomoSapiens.176403` as an example (Figure 3A). It is a pseudouridine sites screened by ELIGOS and nanoPsu from human ENDOC and SEAC cell lines. Also, a previous study (GSE60047) based on Ψ -seq also found pseudouridine modification at this position (Figure 3B). Figure 3C shows that an m¹A site is located 4 bp downstream of the examples site. Since the nanopore sensor protein can interact with ~5nt regions simultaneously, the presence of another non-canonical bases nearby may cause misleading signals and thus influence the analysis. Although the example site is reported in multiple cases and was supported by other techniques, it is possible that the site is false positive as the result of adjacent misleading signals.

Case study: isoform level exploration of RNA modifications

TXNDC12 (chr1:52020131..52056171, GRCh38.p14 assembly) and KTI12 (chr1:52042103..52033810) are two genes that share common regions on chromosome 1. For an RNA modification site located within the shared regions,

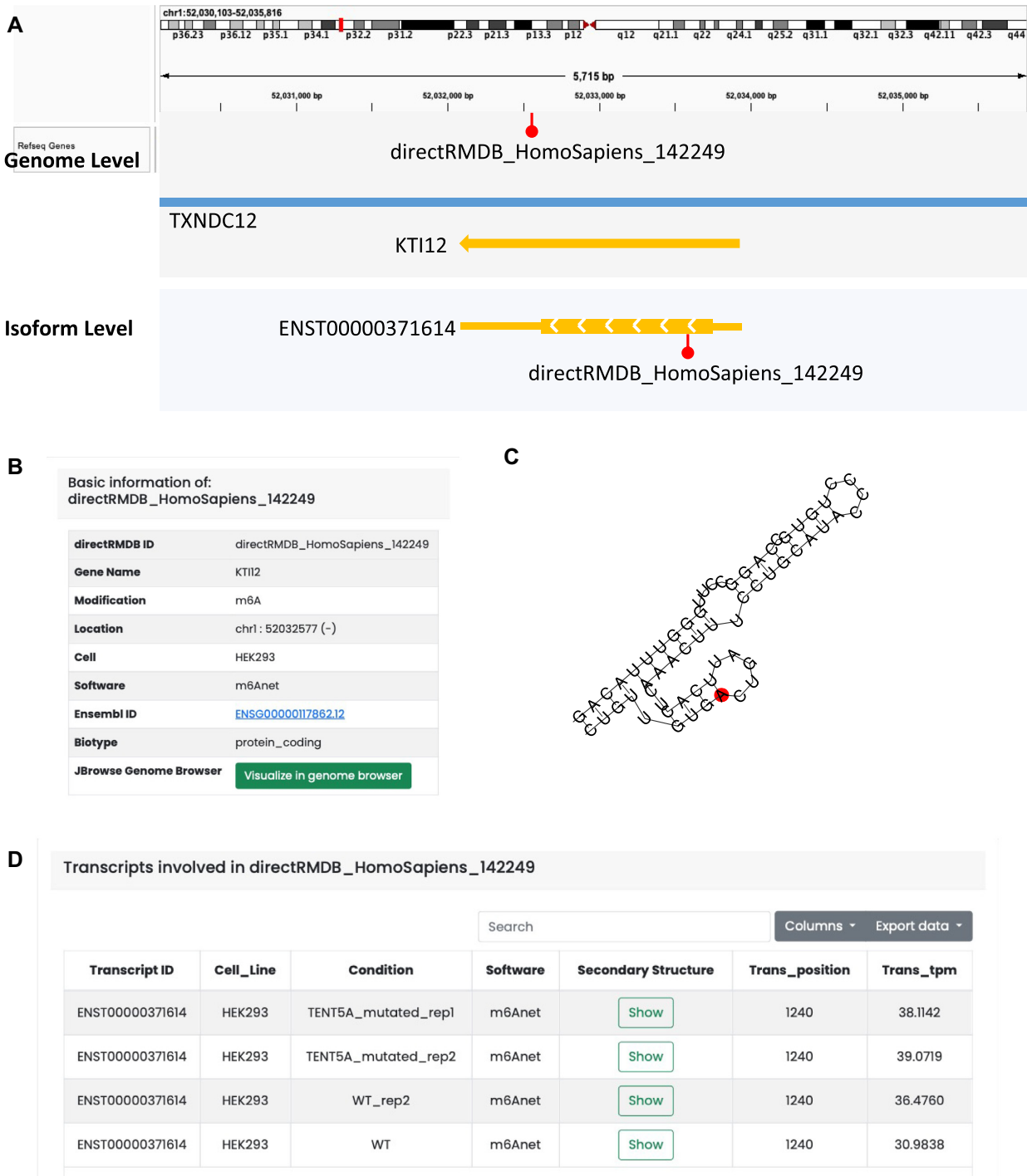


Figure 4. Case study: isoform level exploration of RNA modifications. (A) Genome and isoform level views of RNA modifications. (B) Basic information of the example site with ID of ‘directRMDb_HomoSapiens_142249’. (C) Screenshot of predicted secondary structure of ENST00000371614. The example modification site is highlighted (red). (D) Transcripts information, including transcriptomic coordinates, expression levels.

it can be difficult with NGS epitranscriptomics profiling methods to decide which genes or transcripts it belongs to. Fortunately, direct RNA sequencing technologies offer a solution to this isoform ambiguity problem thanks to the longer reads. By more precisely aligning reads to transcriptome references, modifications can be confidently located in an isoform-specific manner (Figure 4A). Here, the m6A site

with ID of ‘directRMDb_HomoSapiens_142249’ is taken as an example (Figure 4B). From a genome-wide view, it is located on the shared region of TXNDC12 and KTI12 and was wrongly assigned to TXNDC12 by ChipSeeker in a random manner. On the directRMDb details page, in contrast, we can see that m6Anet unambiguously assigned this site to ENST00000371614, an isoform for the KTI12 gene, under

four different conditions. Also, expression levels of KTI12 isoforms under the four conditions are displayed (Figure 4C). The picture of the predicted RNA secondary structure with highlighted modified bases can be queried by clicking the ‘show’ button on the ‘secondary structure’ column (Figure 4D).

Conclusion

Maps of various RNA modifications have been constructed by coupling antibody immunoprecipitation or chemical probing with high-throughput sequencing. However, customized protocols are required for each RNA modification type, thus limiting our ability to characterize the plasticity of the whole epitranscriptomics systematically and in an unbiased fashion. Fortunately, the development of direct RNA sequencing platforms enables the mapping of diverse RNA modification types simultaneously and detection of any given modification present in native RNA molecules. With the rapid accumulation of direct RNA sequencing data and designed ONT tools, we constructed DirectRMDb, the first database of multiple RNA modifications unveiled by direct RNA sequencing technologies. By taking advantage of direct RNA sequencing technologies, DirectRMDb offers several novel features compared with existing epitranscriptomics databases: (i) since ONT direct sequencing generate ultra-long reads and is less vulnerable to isoform ambiguity, we confidently presented isoform-specific distribution of RNA modification sites. (ii) we provided transcript expression levels under different conditions. (iii) we integrated novel modifications sites that have not been detected by NGS-methods. Also, a user-friendly graphical interface integrated with a genome browser was constructed to facilitate the query, visualization, and analysis of this novel, fine-grained epitranscriptomics data. Due to the nature of ONT direct sequencing, the results might contain some false positive site. Therefore, we clearly indicated the tools, samples, and other techniques (i.e. NGS techniques and LC-MS) that support each site. Users could filter, select and use sites based on their understanding and knowledge. Overall, DirectRMDb provides a fresh view of the epitranscriptome. We will continue to update and improve the database by integrating the latest sequencing data and advanced tools to ensure that it remains a valuable resource for the research community.

DATA AVAILABILITY

No new data were generated or analysed in support of this research. The DirectRMDb database is freely available at: <http://www.rnamd.org/directRMDb/>, and has a mirror at: www.xjtlu.edu.cn/biologicalsciences/directRMDb.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

FUNDING

National Natural Science Foundation of China [32100519 and 31671373]; XJTLU Key Program Special Fund [KSF-E-51 and KSF-P-02]; Scientific Research Foundation for

Advanced Talents of Fujian Medical University [XR-CZX202109]. Funding for open access charge: Scientific Research Foundation for Advanced Talents of Fujian Medical University [XR-CZX202109].

Conflict of interest statement. None declared.

REFERENCES

- Kemp, J.W. and Allen, F.W. (1958) Ribonucleic acids from pancreas which contain new components. *Biochim. Biophys. Acta*, **28**, 51–58.
- Boccalletto, P., Machnicka, M.A., Purta, E., Piatkowski, P., Baginski, B., Wirecki, T.K., de Crécy-Lagard, V., Ross, R., Limbach, P.A., Kotter, A. *et al.* (2018) MODOMICS: a database of RNA modification pathways. 2017 update. *Nucleic Acids Res.*, **46**, D303–D307.
- Dominissini, D., Moshitch-Moshkovitz, S., Schwartz, S., Salmon-Divon, M., Ungar, L., Osenberg, S., Cesarkas, K., Jacob-Hirsch, J., Amariglio, N., Kupiec, M. *et al.* (2012) Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. *Nature*, **485**, 201–206.
- Li, X., Xiong, X., Zhang, M., Wang, K., Chen, Y., Zhou, J., Mao, Y., Lv, J., Yi, D., Chen, X.W. *et al.* (2017) Base-Resolution mapping reveals distinct m(1)A methylome in Nuclear- and Mitochondrial-Encoded transcripts. *Mol. Cell*, **68**, 993–1005.
- Hussain, S., Aleksic, J., Blanco, S., Dietmann, S. and Frye, M. (2013) Characterizing 5-methylcytosine in the mammalian epitranscriptome. *Genome Biol.*, **14**, 215.
- Carlile, T.M., Rojas-Duran, M.F., Zinshteyn, B., Shin, H., Bartoli, K.M. and Gilbert, W.V. (2014) Pseudouridine profiling reveals regulated mRNA pseudouridylation in yeast and human cells. *Nature*, **515**, 143–146.
- Marchand, V., Blanloelil-Oillo, F., Helm, M. and Motorin, Y. (2016) Illumina-based ribomethseq approach for mapping of 2'-O-Me residues in RNA. *Nucleic Acids Res.*, **44**, e135.
- Schwartz, S., Agarwala, S.D., Mumbach, M.R., Jovanovic, M., Mertins, P., Shishkin, A., Tabach, Y., Mikkelsen, T.S., Satija, R., Ruvkun, G. *et al.* (2013) High-resolution mapping reveals a conserved, widespread, dynamic mRNA methylation program in yeast meiosis. *Cell*, **155**, 1409–1421.
- Torres, A.G., Batlle, E. and Ribas de Pouplana, L. (2014) Role of tRNA modifications in human diseases. *Trends Mol. Med.*, **20**, 306–314.
- Batista, P.J., Molinie, B., Wang, J., Qu, K., Zhang, J., Li, L., Bouley, D.M., Lujan, E., Haddad, B., Daneshvar, K. *et al.* (2014) m(6)A RNA modification controls cell fate transition in mammalian embryonic stem cells. *Cell Stem Cell*, **15**, 707–719.
- Hausmann, I.U., Bodi, Z., Sanchez-Moran, E., Mongan, N.P., Archer, N., Fray, R.G. and Soller, M. (2016) m(6)A potentiates *sxl* alternative pre-mRNA splicing for robust drosophila sex determination. *Nature*, **540**, 301–304.
- Chen, J., Zhang, Y.C., Huang, C., Shen, H., Sun, B., Cheng, X., Zhang, Y.J., Yang, Y.G., Shu, Q., Yang, Y. *et al.* (2019) m(6)A regulates neurogenesis and neuronal development by modulating histone methyltransferase *ezh2*. *Genomics Proteomics Bioinformatics*, **17**, 154–168.
- Engel, M., Eggert, C., Kaplick, P.M., Eder, M., Roh, S., Tietze, L., Namendorf, C., Arloth, J., Weber, P., Rex-Haffner, M. *et al.* (2018) The role of m(6)A/m-RNA methylation in stress response regulation. *Neuron*, **99**, 389–403.
- Kmieczyk, V., Riechert, E., Kalinski, L., Boileau, E., Malovrh, E., Malone, B., Gorska, A., Hofmann, C., Varma, E., Jürgensen, L. *et al.* (2019) m(6)A-mRNA methylation regulates cardiac gene expression and cellular growth. *Life Sci. Alliance*, **2**, e201800233.
- Liu, B., Merriman, D.K., Choi, S.H., Schumacher, M.A., Plangger, R., Kreutz, C., Horner, S.M., Meyer, K.D. and Al-Hashimi, H.M. (2018) A potentially abundant junctional RNA motif stabilized by m(6)A and mg(2). *Nat. Commun.*, **9**, 2761.
- Cohn, W.E. (1960) Pseudouridine, a carbon-carbon linked ribonucleoside in ribonucleic acids: isolation, structure, and chemical characteristics. *J. Biol. Chem.*, **235**, 1488–1498.
- Nombela, P., Miguel-López, B. and Blanco, S. (2021) The role of m(6)A, m(5)C and Ψ RNA modifications in cancer: novel therapeutic opportunities. *Mol. Cancer*, **20**, 18.

18. Chen, K., Lu, Z., Wang, X., Fu, Y., Luo, G.Z., Liu, N., Han, D., Domissini, D., Dai, Q., Pan, T. *et al.* (2015) High-resolution N(6)-methyladenosine (m(6)A) map using photo-crosslinking-assisted m(6)A sequencing. *Angew. Chem. Int. Ed Engl.*, **54**, 1587–1590.
19. Ke, S., Alemu, E.A., Mertens, C., Gantman, E.C., Fak, J.J., Mele, A., Haripal, B., Zucker-Scharff, I., Moore, M.J., Park, C.Y. *et al.* (2015) A majority of m6A residues are in the last exons, allowing the potential for 3' UTR regulation. *Genes Dev.*, **29**, 2037–2053.
20. Linder, B., Grozhik, A.V., Olarerin-George, A.O., Meydan, C., Mason, C.E. and Jaffrey, S.R. (2015) Single-nucleotide-resolution mapping of m6A and m6Am throughout the transcriptome. *Nat. Methods*, **12**, 767–772.
21. Molinie, B., Wang, J., Lim, K.S., Hillebrand, R., Lu, Z.X., Van Wittenberghe, N., Howard, B.D., Daneshvar, K., Mullen, A.C., Dedon, P. *et al.* (2016) m(6)A-LAIC-seq reveals the census and complexity of the m(6)A epitranscriptome. *Nat. Methods*, **13**, 692–698.
22. Koh, C.W.Q., Goh, Y.T. and Goh, W.S.S. (2019) Atlas of quantitative single-base-resolution N(6)-methyl-adenine methylomes. *Nat. Commun.*, **10**, 5636.
23. Garcia-Campos, M.A., Edelheit, S., Toth, U., Safra, M., Shachar, R., Viukov, S., Winkler, R., Nir, R., Lasman, L., Brandis, A. *et al.* (2019) Deciphering the “m(6)A code” via antibody-independent quantitative profiling. *Cell*, **178**, 731–747.
24. Zhang, Z., Chen, L.Q., Zhao, Y.L., Yang, C.G., Roundtree, I.A., Zhang, Z., Ren, J., Xie, W., He, C. and Luo, G.Z. (2019) Single-base mapping of m(6)A by an antibody-independent method. *Sci. Adv.*, **5**, eaax0250.
25. Meyer, K.D. (2019) DART-seq: an antibody-free method for global m(6)A detection. *Nat. Methods*, **16**, 1275–1280.
26. Marchand, V., Ayadi, L., Ernst, F.G.M., Hertler, J., Bourguignon-Igel, V., Galvanin, A., Kotter, A., Helm, M., Lafontaine, D.L.J. and Motorin, Y. (2018) AlkAniline-Seq: profiling of m(7)G and m(3)C RNA modifications at single nucleotide resolution. *Angew. Chem. Int. Ed Engl.*, **57**, 16785–16790.
27. Jenjaroenpun, P., Wongsurawat, T., Wadley, T.D., Wassenaar, T.M., Liu, J., Dai, Q., Wanchai, V., Akeel, N.S., Jamshidi-Parsian, A., Franco, A.T. *et al.* (2021) Decoding the epitranscriptional landscape from native RNA sequences. *Nucleic Acids Res.*, **49**, e7.
28. Begik, O., Lucas, M.C., Prysacz, L.P., Ramirez, J.M., Medina, R., Milenkovic, I., Cruciani, S., Liu, H., Vieira, H.G.S., Sas-Chen, A. *et al.* (2021) Quantitative profiling of pseudouridylation dynamics in native RNAs with nanopore sequencing. *Nat. Biotechnol.*, **39**, 1278–1291.
29. Carrara, M., Beccuti, M., Lazzarato, F., Cavallo, F., Cordero, F., Donatelli, S. and Calogero, R.A. (2013) State-of-the-art fusion-finder algorithms sensitivity and specificity. *Biomed. Res. Int.*, **2013**, 340620.
30. Hansen, K.D., Brenner, S.E. and Dudoit, S. (2010) Biases in illumina transcriptome sequencing caused by random hexamer priming. *Nucleic Acids Res.*, **38**, e131.
31. Garalde, D.R., Snell, E.A., Jachimowicz, D., Sipos, B., Lloyd, J.H., Bruce, M., Pantic, N., Admassu, T., James, P., Warland, A. *et al.* (2018) Highly parallel direct RNA sequencing on an array of nanopores. *Nat. Methods*, **15**, 201–206.
32. Lu, H., Giordano, F. and Ning, Z. (2016) Oxford nanopore MinION sequencing and genome assembly. *Genomics Proteomics Bioinformatics*, **14**, 265–279.
33. McIntyre, A.B.R., Alexander, N., Grigorev, K., Bezdan, D., Sichtig, H., Chiu, C.Y. and Mason, C.E. (2019) Single-molecule sequencing detection of N6-methyladenine in microbial reference materials. *Nat. Commun.*, **10**, 579.
34. Lorenz, D.A., Sathe, S., Einstein, J.M. and Yeo, G.W. (2020) Direct RNA sequencing enables m(6)A detection in endogenous transcript isoforms at base-specific resolution. *RNA*, **26**, 19–28.
35. Liu, H., Begik, O., Lucas, M.C., Ramirez, J.M., Mason, C.E., Wiener, D., Schwartz, S., Mattick, J.S., Smith, M.A. and Novoa, E.M. (2019) Accurate detection of m(6)A RNA modifications in native RNA sequences. *Nat. Commun.*, **10**, 4079.
36. Price, A.M., Hayer, K.E., McIntyre, A.B.R., Gokhale, N.S., Abebe, J.S., Della Fera, A.N., Mason, C.E., Horner, S.M., Wilson, A.C., Depledge, D.P. *et al.* (2020) Direct RNA sequencing reveals m(6)A modifications on adenovirus RNA are necessary for efficient splicing. *Nat. Commun.*, **11**, 6016.
37. Parker, M.T., Knop, K., Sherwood, A.V., Schurch, N.J., Mackinnon, K., Gould, P.D., Hall, A.J., Barton, G.J. and Simpson, G.G. (2020) Nanopore direct RNA sequencing maps the complexity of arabidopsis mRNA processing and m(6)A modification. *Elife*, **9**, e49658.
38. Pratanwanich, P.N., Yao, F., Chen, Y., Koh, C.W.Q., Wan, Y.K., Hendra, C., Poon, P., Goh, Y.T., Yap, P.M.L., Chooi, J.Y. *et al.* (2021) Identification of differential RNA modifications from nanopore direct RNA sequencing with xPore. *Nat. Biotechnol.*, **39**, 1394–1402.
39. Hendra, C., Pratanwanich, P.N., Wan, Y.K., Goh, W.S.S., Thiery, A. and Göke, J. (2021) Detection of m6A from direct RNA sequencing using a multiple instance learning framework. bioRxiv doi: <https://doi.org/10.1101/2021.09.20.461055>, 22 September 2021, preprint: not peer reviewed.
40. Gao, Y., Liu, X., Wu, B., Wang, H., Xi, F., Kohlen, M.V., Reddy, A.S.N. and Gu, L. (2021) Quantitative profiling of N(6)-methyladenosine at single-base resolution in stem-differentiating xylem of populus trichocarpa using nanopore direct RNA sequencing. *Genome Biol.*, **22**, 22.
41. Huang, S., Zhang, W., Katanski, C.D., Dersh, D., Dai, Q., Lolans, K., Yewdell, J., Eren, A.M. and Pan, T. (2021) Interferon inducible pseudouridine modification in human mRNA by quantitative nanopore profiling. *Genome Biol.*, **22**, 330.
42. Boccaletto, P., Stefaniak, F., Ray, A., Cappannini, A., Mukherjee, S., Purta, E., Kurkowska, M., Shirvanizadeh, N., Destefanis, E., Groza, P. *et al.* (2022) MODOMICS: a database of RNA modification pathways. 2021 update. *Nucleic Acids Res.*, **50**, D231–D235.
43. Xuan, J.J., Sun, W.J., Lin, P.H., Zhou, K.R., Liu, S., Zhang, L.L., Qu, L.H. and Yang, J.H. (2018) RMBase v2.0: deciphering the map of RNA modifications from epitranscriptome sequencing data. *Nucleic Acids Res.*, **46**, D327–D334.
44. Liu, S., Zhu, A., He, C. and Chen, M. (2020) REPIC: a database for exploring the N(6)-methyladenosine methylome. *Genome Biol.*, **21**, 100.
45. Tang, Y., Chen, K., Song, B., Ma, J., Wu, X., Xu, Q., Wei, Z., Su, J., Liu, G., Rong, R. *et al.* (2021) m6A-Atlas: a comprehensive knowledgebase for unraveling the N6-methyladenosine (m6A) epitranscriptome. *Nucleic Acids Res.*, **49**, D134–D143.
46. Ma, J., Song, B., Wei, Z., Huang, D., Zhang, Y., Su, J., de Magalhaes, J.P., Rigden, D.J., Meng, J. and Chen, K. (2022) m5C-Atlas: a comprehensive database for decoding and annotating the 5-methylcytosine (m5C) epitranscriptome. *Nucleic Acids Res.*, **50**, D196–D203.
47. Liu, H., Wang, H., Wei, Z., Zhang, S., Hua, G., Zhang, S.W., Zhang, L., Gao, S.J., Meng, J., Chen, X. *et al.* (2018) MeT-DB V2.0: elucidating context-specific functions of N6-methyl-adenosine methyltranscriptome. *Nucleic Acids Res.*, **46**, D281–D287.
48. Luo, X., Li, H., Liang, J., Zhao, Q., Xie, Y., Ren, J. and Zuo, Z. (2021) RMVar: an updated database of functional variants involved in RNA modifications. *Nucleic Acids Res.*, **49**, D1405–D1412.
49. Deng, S., Zhang, H., Zhu, K., Li, X., Ye, Y., Li, R., Liu, X., Lin, D., Zuo, Z. and Zheng, J. (2021) M6A2Target: a comprehensive database for targets of m6A writers, erasers and readers. *Brief Bioinform.*, **22**, bbaa055.
50. Jonkhout, N., Tran, J., Smith, M.A., Schonrock, N., Mattick, J.S. and Novoa, E.M. (2017) The RNA modification landscape in human disease. *RNA*, **23**, 1754–1769.
51. Simpson, J.T., Workman, R.E., Zuzarte, P.C., David, M., Dursi, L.J. and Timp, W. (2017) Detecting DNA cytosine methylation using nanopore sequencing. *Nat. Methods*, **14**, 407–410.
52. Li, H. (2018) Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*, **34**, 3094–3100.
53. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G. and Durbin, R. (2009) The sequence alignment/map format and SAMtools. *Bioinformatics*, **25**, 2078–2079.
54. Piechotta, M., Naarmann-de Vries, I.S., Wang, Q., Altmüller, J. and Dieterich, C. (2022) RNA modification mapping with JACUSA2. *Genome Biol.*, **23**, 115.
55. McCaskill, J.S. (1990) The equilibrium partition function and base pair binding probabilities for RNA secondary structure. *Biopolymers*, **29**, 1105–1119.
56. Andronescu, M., Bereg, V., Hoos, H.H. and Condon, A. (2008) RNA STRAND: the RNA secondary structure and statistical analysis database. *BMC Bioinf.*, **9**, 340.

57. Denman,R.B. (1993) Using RNAFOLD to predict the activity of small catalytic RNAs. *BioTechniques*, **15**, 1090–1095.
58. Gleeson,J., Leger,A., Prawer,Y.D.J., Lane,T.A., Harrison,P.J., Haerty,W. and Clark,M.B. (2022) Accurate expression quantification from nanopore direct RNA sequencing with nanocount. *Nucleic Acids Res.*, **50**, e19.
59. Frankish,A., Diekhans,M., Jungreis,I., Lagarde,J., Loveland,J.E., Mudge,J.M., Sisu,C., Wright,J.C., Armstrong,J., Barnes,I. *et al.* (2021) gencode 2021. *Nucleic Acids Res.*, **49**, D916–D923.
60. Sayers,E.W., Beck,J., Bolton,E.E., Bourexis,D., Brister,J.R., Canese,K., Comeau,D.C., Funk,K., Kim,S., Klimke,W. *et al.* (2021) Database resources of the national center for biotechnology information. *Nucleic Acids Res.*, **49**, D10–D17.
61. Yu,G., Wang,L.G. and He,Q.Y. (2015) CHIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics*, **31**, 2382–2383.
62. Li,J.H., Liu,S., Zhou,H., Qu,L.H. and Yang,J.H. (2014) starBase v2.0: decoding miRNA-ceRNA, miRNA-ncRNA and protein-RNA interaction networks from large-scale CLIP-Seq data. *Nucleic Acids Res.*, **42**, D92–D97.
63. Hu,B., Yang,Y.T., Huang,Y., Zhu,Y. and Lu,Z.J. (2017) POSTAR: a platform for exploring post-transcriptional regulation coordinated by RNA-binding proteins. *Nucleic Acids Res.*, **45**, D104–D114.
64. Lee,B.T., Barber,G.P., Benet-Pagès,A., Casper,J., Clawson,H., Diekhans,M., Fischer,C., Gonzalez,J.N., Hinrichs,AngieS., Lee,ChristopherM. *et al.* (2021) The UCSC genome browser database: 2022 update. *Nucleic Acids Res.*, **50**, D1115–D1122.
65. Sherry,S.T., Ward,M.H., Kholodov,M., Baker,J., Phan,L., Smigielski,E.M. and Sirotkin,K. (2001) dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.*, **29**, 308–311.
66. Skinner,M.E., Uzilov,A.V., Stein,L.D., Mungall,C.J. and Holmes,I.H. (2009) JBrowse: a next-generation genome browser. *Genome Res.*, **19**, 1630–1638.
67. Ismail,I.H., Gagné,J.P., Genois,M.M., Strickfaden,H., McDonald,D., Xu,Z., Poirier,G.G., Masson,J.Y. and Hendzel,M.J. (2015) The RNF138 E3 ligase displaces ku to promote DNA end resection and regulate DNA repair pathway choice. *Nat. Cell Biol.*, **17**, 1446–1457.
68. Schmidt,C.K., Galanty,Y., Sezaniecka-Clift,M., Coates,J., Jhujh,S., Demir,M., Cornwell,M., Beli,P. and Jackson,S.P. (2015) Systematic E2 screening reveals a UBE2D-RNF138-CtIP axis promoting DNA repair. *Nat. Cell Biol.*, **17**, 1458–1470.
69. Lee,K., Byun,K., Hong,W., Chuang,H.Y., Paek,C.G., Bayarsaikhan,E., Paek,S.H., Kim,H., Shin,H.Y., Ideker,T. *et al.* (2013) Proteome-wide discovery of mislocated proteins in cancer. *Genome Res.*, **23**, 1283–1294.
70. Long,P., Samnakay,P., Jenner,P. and Rose,S. (2012) A yeast two-hybrid screen reveals that osteopontin associates with MAP1A and MAP1B in addition to other proteins linked to microtubule stability, apoptosis and protein degradation in the human brain. *Eur. J. Neurosci.*, **36**, 2733–2742.
71. Li,Z.X., Zhu,Q.N., Zhang,H.B., Hu,Y., Wang,G. and Zhu,Y.S. (2018) MALAT1: a potential biomarker in cancer. *Cancer Manag Res*, **10**, 6757–6768.