RMBase v2.0: deciphering the map of RNA modifications from epitranscriptome sequencing data

Jia-Jia Xuan[1,2], Wen-Ju Sun[1,2], Peng-Hui Lin[1,2], Ke-Ren Zhou[1,2], Shun Liu[1,2], Ling-Ling Zheng[1,2], Liang-Hu Qu[1,2,*] and Jian-Hua Yang[1,2,*]

[1]Key Laboratory of Gene Engineering of the Ministry of Education, Sun Yat-sen University, Guangzhou 510275, PR China and [2]State Key Laboratory for Biocontrol, Sun Yat-sen University, Guangzhou 510275, PR China

ABSTRACT

More than 100 distinct chemical modifications to RNA have been characterized so far. However, the prevalence, mechanisms and functions of various RNA modifications remain largely unknown. To provide transcriptome-wide landscapes of RNA modifications, we developed the RMBase v2.0 (http://rna.sysu.edu.cn/rmbase/), which is a comprehensive database that integrates epitranscriptome sequencing data for the exploration of post-transcriptional modifications of RNAs and their relationships with miRNA binding events, disease-related single-nucleotide polymorphisms (SNPs) and RNA-binding proteins (RBPs). RMBase v2.0 was expanded with ∼600 datasets and ∼1 397 000 modification sites from 47 studies among 13 species, which represents an approximately 10-fold expansion when compared with the previous release. It contains ∼1 373 000 N6-methyladenosines (m^{6}A), ∼5400 N1-methyladenosines (m^{1}A), ∼9600 pseudouridine (Ψ) modifications, ∼1000 5-methylcytosine (m^{5}C) modifications, ∼5100 2′-O-methylations (2′-O-Me), and ∼2800 modifications of other modification types. Moreover, we built a new module called 'Motif' that provides the visualized logos and position weight matrices (PWMs) of the modification motifs. We also constructed a novel module termed 'modRBP' to study the relationships between RNA modifications and RBPs. Additionally, we developed a novel web-based tool named 'modMetagene' to plot the metagenes of RNA modification along a transcript model. This database will help researchers investigate the potential functions and mechanisms of RNA modifications.

INTRODUCTION

More than 100 different types of RNA modifications have been characterized across all living organisms (1–5). RNA modifications occur in diverse RNA molecules, including mRNAs, tRNAs, rRNAs, lncRNAs and snoRNAs (1–8). Increasing numbers of studies have demonstrated that RNA modifications play important roles in RNA splicing, protein localization and translation, stem cell pluripotency and human diseases (1–12).

To determine the global landscape of RNA modifications, many studies have recently developed transcriptome-wide sequencing technologies (e.g, Pseudo-seq, Ψ-seq, CeU-seq, Aza-IP, MeRIP-seq, m^{6}A-seq, miCLIP, m^{6}A-CLIP, RiboMeth-seq, Nm-seq and m^{1}A-seq) to identify distinct epitranscriptomic marks (4,5,9,12–19). These new methods have helped researchers to identify the genomic locations of RNA modifications and reveal the distinct distributions of various modification types (e.g. Ψ, m^{6}A, m^{5}C, 2′-O-Me and m^{1}A) throughout the transcriptome. Moreover, in combination with other emerging technologies and tools (e.g. CLIP-seq) (20), researchers have used these technologies to identify novel RNA-modifying enzymes and their targets, and have revealed the spatial-temporal dynamics of distinct RNA modifications under different physiological and pathological conditions (1–12). Although these sequencing technologies have provided comprehensive profiling of valuable data for functional epitranscriptomic investigations, the integration of these large-scale data sets for the exploration of the prevalence, mechanisms and functions of various modifications remains a daunting challenge.

In this study, we developed RMBase v2.0 to perform a large-scale integration of RNA modification sites derived from high-throughput epitranscriptome sequencing data that covered 13 species including humans, mice, zebrafish, yeast, etc. (Figure 1, Table 1). RMBase provides web interfaces that display the maps of RNA modifications for various cell types. Moreover, by integrating miRNA targets, RNA-binding protein (RBP) binding sites, single-nucleotide variations (SNVs) and genome-wide association
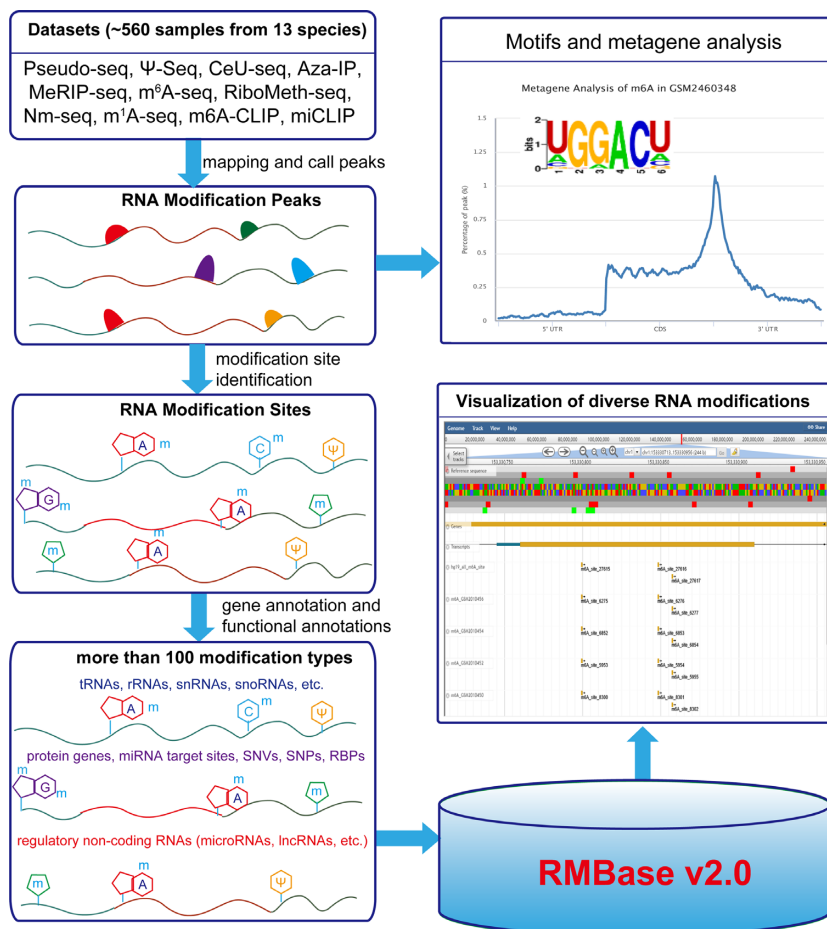
Figure 1. The scheme of the RMBase v2.0 workflow. RMBase v2.0 provides the comprehensive transcriptome-wide landscape of more than 100 types of RNA modifications. All results generated by RMBase v2.0 are deposited in MySQL database and displayed in the visual browser and web pages.

study (GWAS) data, RMBase can be used to explore the relationships between these data and RNA modification sites. Additionally, we also developed novel web-based tools that include the 'modMetagene', 'Motif' module and genome browser to visualize metagene profiles, logos of modification motifs and various types of genomic features. RMBase v2.0 is expected to help researchers investigate the potential functions and mechanisms of RNA modifications.

MATERIALS AND METHODS

Integration of the public epitranscriptome sequencing data sets and genome sets

We manually collected high-throughput Pseudo-seq, Ψ-seq, CeU-seq, Aza-IP, MeRIP-seq, m6A-seq, m1A-seq, miCLIP, m6A-CLIP, RiboMeth-seq and Nm-seq data from the Gene Expression Omnibus (GEO) and Sequence Read Archive (SRA) databases (21). The barcodes and 3′-adapters of the raw sequencing data were clipped using cutadapt software (22). All trimmed reads were aligned to the genomes using hisat2 (23) and the mapping results were then converted into BAM format for display in the genome browser and peak calling using samtools (24). Other known RNA modification sites were integrated and curated as described in our RMBase v1.0 (25). The genome sequences and annotations

of all 13 species were downloaded from the UCSC genome browser (26), GENCODE (27) and Ensembl databases (28) (Supplementary Table S1).

Identification and annotation of modification sites

The m6A modification peaks were called with the exome-Peak program (29) with strict criteria (false discovery rate (FDR) <0.05, P-value <0.01 and fold change (FC) >2). To locate the m6A modification sites in a genome, we predicted the exact m6A positions from the m6A-seq or MeRIP-Seq peaks by searching for consensus RRACH motifs (where R denotes A or G and H denotes A, C or U) (18,30) among the 13 species. Similarly, the m1A modification sites of four species were identified from the m1A-seq peaks by searching for the GAAGAAG motif (14,19). We performed *de novo* motif identifications of the m6A and m1A peak data by using the HOMER software (31) to obtain their position weight matrices (PWMs) and accurate motif regions. We used these PWMs to score the m6A and m1A modification sites.

We assigned all modification sites to multiple types of genes, including tRNAs, rRNAs, Mt-tRNAs, Mt-rRNAs, scRNAs, snRNAs, snoRNAs, microRNAs (miR-NAs), lincRNAs, misc_RNAs, protein-coding genes, pro-

Table 1. The modification site statistics in RMBase v2.0. The statistical data indicating the number of each RNA modification type for 13 species. m^6A is N6-methyladenosine methylation, m^1A is N1-methyladenosine methylation, m^5C is 5-methylcytosine methylation, Ψ is pseudouridine modification and $2'$-O-Me is $2'$-O-methylation and 'other types' contains diverse rare modification types

Species	m^6A	m^1A	m^5C	Ψ	$2'$-O-Me	Other types
Human	477 452	2574	680	4128	4795	525
Mouse	490 704	1052	97	3320	59	435
Rhesus	38 838	/	/	/	/	/
Chimpanzee	38 369	/	/	/	/	/
Rat	60 769	/	/	/	/	/
Pig	121 409	/	/	/	/	/
Zebrafish	43 027	/	/	/	/	/
S. cerevisiae	67 671	1220	211	2122	242	1864
Fly	6798	/	/	/	/	/
A. thaliana	20331	/	/	/	/	/
S. pombe	/	565	/	/	/	/
E. coli	2173	/	/	/	/	/
P. aetuginosa	5814	/	/	/	/	/

cessed_transcripts, pseudogenes and gene regions covering CDS, 3′ UTR, 5′ UTR, intron, exon and intergenic region.

Association analysis of the RBP and miRNA binding sites with the RNA modifications

The RBP-RNA and miRNA-target interactions that were supported by the CLIP-seq data were downloaded from our starBase platform for humans and mice (32,33) (Supplementary Table S2). To study the relationships between the RBPs and RNA modifications, we intersected their binding sites with curated and identified RNA modification sites to obtain regulatory pairs. We also annotated the RBPs that were putatively affected by the m^6A modifications as readers, writers and erasers to facilitate the queries and use by researchers. A similar workflow was applied to the interactions between the miRNA targets and RNA modifications.

Identification of disease-related SNVs and SNPs associated with RNA modification sites

To investigate the relationships of disease-related SNVs and SNPs with RNA modifications, we collected cancer somatic mutations from the COSMIC database (34) and previous publications (35,36). As described in our previous study (37), the human disease/trait-associated SNPs were curated from published GWAS data provided by the NHGRI GWAS Catalog (38), Johnson and O'Donnell (39), dbGAP (40) and GAD (41), as well as the SNPs in linkage disequilibrium (LD, $r^2 \geq 0.5$; Supplementary Table S2) with reported disease-related loci that were selected in at least one of the four populations genotype data (CEU, CHB, JPT and YRI) from the HapMap project (release 28) (42). All SNVs and SNPs were intersected with the RNA modification regions that were extended an additional 10nt in both the 5′- and 3′-directions for each modification site to identify the SNVs and SNPs that might interact with the RNA modifications. The modification regions were defined as described in our previous RMBase (25).

Construction of a web-based function tool and the RMBase genome browser

We developed 'modMetagene' to present a metagene plot of the RNA modifications along a transcript model from uploaded user data. We also improved the web-based tool 'modAnnotation' to identify and annotate the modification sites that were based on all of the RNA modification sites recorded in our database.

We used JBrowse (43), which is a fast and embeddable genome browser that was built completely with JavaScript and HTML5, to construct the improved RMBase Genome Browser that is used to integrate and display reference sequences, RNA modification sites, protein-coding genes, transcripts, modification sequencing data and aligned sequencing reads.

Database and web interface implementation

All data sets were processed and stored in a MySQL Database Management System. The database query and user interface were developed using PHP and JavaScript. The query result table is based on jQueryUI and DataTables and is a highly flexible tool for sorting and filtering search results.

DATABASE CONTENT AND WEB INTERFACE

Comprehensive atlas of various types of RNA modifications

By manually collecting and identifying RNA modification sites from high-throughput epitranscriptome sequencing data, we obtained a total collection of 1,397,244 RNA modification sites, of which covered more than 100 types of chemical modifications among the 13 species (Table 1) including ~1 373 000, ~5400, ~9600, ~1000 and ~5100 m^6A, m^1A, Ψ, m^5C and $2'$-O-Me with sites, respectively. To study the distributions of RNA modifications on the transcript products, we mapped their sites onto the genomic coordinates of the genes with comprehensive annotated information such as gene types and regions. Our results demonstrated consistencies with previous studies in that the m^6A, Ψ and $2'$-O-Me modifications tended to occur in protein-coding genes, whereas the m^5C modifications often appeared among tRNA molecules.

Exploring the associations between RNA modifications and RBPs

To investigate whether RNA modifications influence the function of RBPs or vice versa, the relationships between CLIP-Seq experimentally supported RBP binding sites (32,33) and all RNA modification sites from RMBase were constructed. We located all of the identified modification sites in the binding regions of the RBPs and found that thousands of RNA modifications were associated with RBPs. In human cells, RMBase characterized regulatory relationships between ∼30 RNA modification types that included m^{1}A, m^{6}A, 2′-*O*-Me, m^{5}C and 120 RBPs (Supplementary Table S3). In RMBase, many RBPs bound to thousands of RNA modification sites, such as AGO, and m^{6}A writers (e.g. METTL3 and METTL14) and readers (e.g. YTHDF1 and YTHDF2; Supplementary Table S3).

Selection of RNA modifications associated with disease-related SNPs or SNVs

Some RNA modifications have also been found to be correlated with a variety of human diseases and cancers, such as leukemia, glioblastoma and colon cancer (44,45). To investigate the clinical connections between RNA modifications and SNPs as well as SNVs, we systematically drew the distributions of the SNP and SNV coordinates around the RNA modification sites. We identified 192 283 and 1862 sites of RNA modifications assembled around disease-associated SNVs and SNPs, which included somatic point mutations, such as substitutions and deletions. Additionally, we obtained an overview that at least 10 RNA modification types were potentially correlated with dozens of diseases or cancers, such as melanoma (MEL), lung adenocarcinoma (LUAD) and uterine corpus endometrial carcinoma (UCEC) (Supplementary Table S4).

Web interfaces for web-based modules that were developed to explore different types of RNA modification sites

In RMBase v2.0, we provide multiple user-friendly web interfaces to help users explore RNA modification sites that have been identified from high-throughput epitranscriptome sequencing data. In most of the modules deposited on the web interface, the users are allowed to choose their favored organism to browse the corresponding relationships between RNA modifications and multiple interacting factors, such as RBPs, microRNA targets, SNPs and SNVs.

The interface provides users with six basic web-based modules (m^{6}A, m^{1}A, m^{5}C, 2′-*O*-Me, Pseudouridine/Ψ and otherType) to quickly retrieve the RNA modification sites from various modification types. In the result pages of these modules, users can browse genomic coordinates, associated genes, located gene regions, sequence contexts, motif scores and supporting evidence of RNA modifications. The motif scores provided by RMBase in m^{6}A and m^{1}A web pages will improve the accuracies of analyses of modification prediction.

To display the consensus sequence preferences of the modification sites of m^{6}A and m^{1}A under different cell conditions, we provided the 'Motif' module to visualize the PWMs and the distribution patterns of the transcripts (Figure 2). The identified PWMs can be evaluated in many aspects, including *P*-values, percentages of targets and backgrounds.

The constructed clinical connections between RNA modifications and SNPs as well as SNVs were placed in the 'modVar' module. In the result pages, RNA modification sites were connected to SNPs or SNVs with information about disease types, mutation types, mutation regions and the relative distances between these features.

Additionally, the identified regulatory relationships between RNA modifications and post-transcriptional regulation-associated factors, including RBPs and miRNAs were deposited in 'modRBP' and 'modMirTar'. In these modules, users can inspect the RBPs or miRNAs of interest in terms of whether they have interactions with any RNA modifications.

Web-based analysis tools developed for epitranscriptome sequencing data

To facilitate the research on RNA modifications, we developed a web-based tool suite termed 'modTool' that consists of 'modMetagene' and 'modAnnotation' by employing bin strategies and gene annotation files. The tool suite allows users to upload their identified RNA modification sites, which are formatted into the BED6 format and take up less than 2MB disk volume. After a few seconds to minutes of running, the 'modMetagene' sub-module displays the analysis results with metagene plots, and the 'modAnnotation' sub-module was displays its results with a data table that is similar to that of the m^{6}A web-based module. Additionally, users can export or download the analysis results for use in downstream analyses.

RMBase genome browser

To facilitate the comparative analyses of the RNA modification sites and exploration epitranscriptome sequencing datasets, we developed an improved genome browser in RMBase v2.0. We provide users with an integrated view of various genomic features including genomic coordinates and genome-wide experiment densities of RNA modification sites, and information about genome and gene annotations downloaded from GECODE or Ensembl. For example, Figure 3 illustrates the visualization of the '2-*O*-Me_site_3709' modification site located within rRNA.

EXAMPLE APPLICATIONS

M^{6}A modifications occurred along long-noncoding RNAs

While m^{6}A modification is well characterized in mRNA, there are fewer reports on m^{6}A in long-noncoding RNAs (lncRNAs). Thus, we filtered the 'lincRNA' in the 'modGene' web page, and found that 15 lincRNAs were associated with >50 RNA modification sites (Supplementary Figure S1). We were able to recapitulate the highly methylated lincRNA XIST that has been reported in a recent study (46). The results page sorted by the number of modification sites revealed that the XIST contains 89 m^{6}A modification sites. Additionally, a recent study (46) revealed that m^{6}A RNA
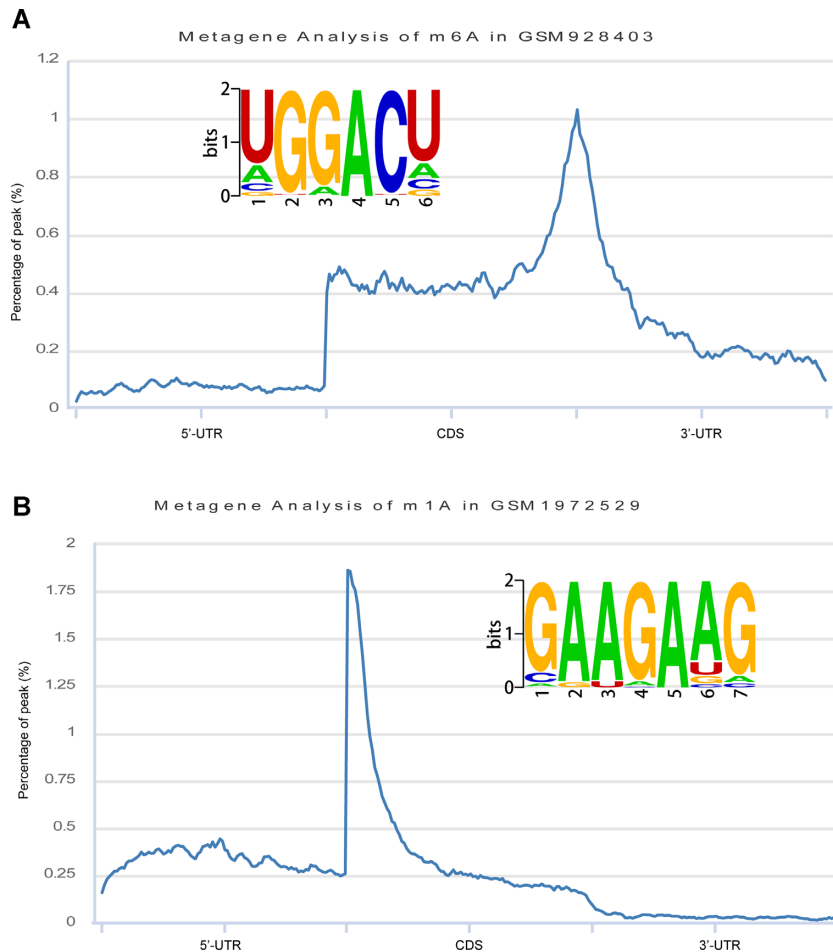
Figure 2. *De novo* motif logos identified and metagene analyses of RNA modifications. (**A**) The m[6]A modification motif and metagene. (**B**) The m[1]A modification motif and metagene.

methylation promotes XIST-mediated transcriptional repression. Thus, we speculate that other highly methylated lincRNAs identified by RMBase may play important roles in various biological processes.

M[6]A modification sites bound by RNA-binding proteins

The function of m[6]A is heavily dependent on m[6]A associated proteins that are characterized as readers, writers and erasers. Consequently, we used the 'modRBP' web page to explore the connection between m[6]A modification sites and related RNA binding proteins (RBPs). We found that many RBPs, as well as known m[6]A-related RBPs (e.g. YTHDC1, YTHDF1, METTL3/14, WTAP and ALKBH5), interacted with a large number of m[6]A sites (Supplementary Table S3). Interestingly, consistent with a recent report about FMR1 as a sequence-context-dependent m[6]A reader (47,48), we also discovered FMR1 bound to 74,460 m[6]A modification sites (Supplementary Figure S2).

DISCUSSION AND CONCLUSIONS

By integrating and analyzing numerous high-throughput epitranscriptome sequencing data and collecting data from public resources, RMBase v2.0 provides a gallery of RNA modification marks that cover more than 100 types of RNA modifications on transcript products and revealed that RNA modifications are involved in complex post-transcription regulatory networks.

In comparison to other databases and our previous release version (RMBase v1.0) (25), the advances and improvements of RMBase v2.0 are listed as follows: (i) RMBase v2.0 expanded the previous version by up to 10-fold with ∼600 datasets from 47 studies in 13 species. (ii) N1-methyladenosine (m[1]A) is one of the most prevalent post-transcriptional modifications in ncRNAs and mRNAs. We first integrated ∼5000 m[1]A modification sites identified from m[1]A-seq data into RMBase v2.0. (iii) RMBase v2.0 integrated over 4000 2′-*O*-Me sites identified from Nm-seq data (16) and provides the first map of the 2′-*O*-Me sites in mRNAs and other regulatory ncRNAs. (iv) We performed *de novo* motif identifications of modification sites of m[6]A and m[1]A and identified thousands of motif matrices and sequences. We also constructed a module called 'Motif' that presents *de novo* identified PWMs and visualized logos of modification motifs. (v) We built a web-based tool called 'modMetagene' for plotting the metagenes of RNA modification sites uploaded by the user. (vi) We added disease-related SNV data to our database and built a dis-
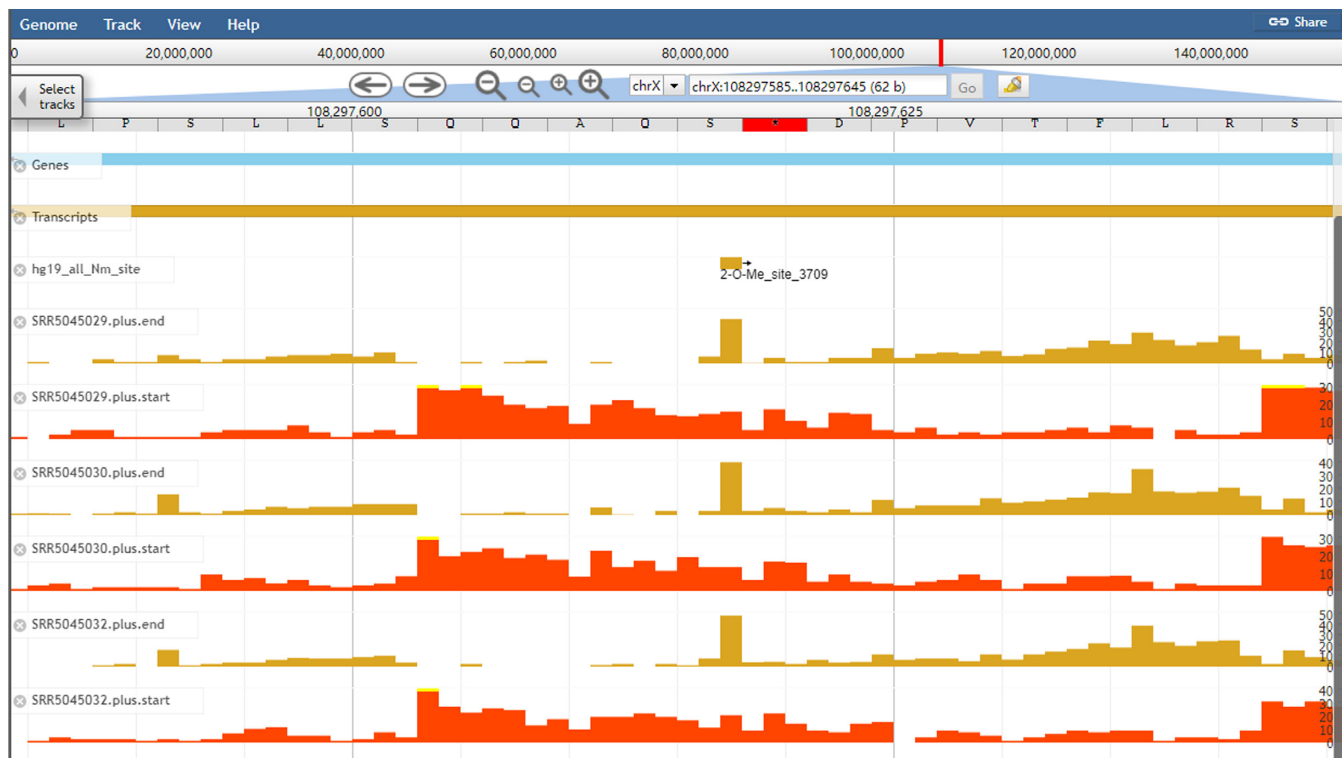
Figure 3. An example of the display of RNA modification sites and epitranscriptome data in the improved RMBase genome browser. Visualization of the genomic context of a '2-*O*-Me_site_3709' modification site located within rRNA using the RMBase browser.

ease link between RNA modification sites and SNVs sites. (vii) RMBase v2.0 includes a newly developed functional module called 'modRBP', which displays the relationships between post-transcriptional RNA modification sites and RNA-binding proteins (RBPs).

RMBase v2.0 allows for the global investigation of more than 100 RNA modification types and reveals extensive and complex post-transcriptional modifications of RNA (Figure 1, Table 1). RMBase v2.0 provides a variety of interfaces and graphic visualizations to facilitate analyses of the massive modification sites in normal tissues and cancer cells. Overall, RMBase v2.0 provides researchers with a comprehensive and powerful platform to discover potential functional roles of RNA modifications hidden in these data.

FUTURE DIRECTIONS

Recent advances in high-throughput epitranscriptome sequencing technology have produced large amounts of single-nucleotide-resolution modification sequencing data. We developed an automatic pipeline that is used to map, annotate, analyze and merge all high-throughput epitranscriptome sequencing data sets, and integrate these data into our local MySQL database. RMBase will continue to improve the computer server performance for storing and analyzing these new incoming data. We also developed new tools to decode the maps of RNA modifications from epitranscriptome sequencing data. We will maintain RMBase to ensure that it remains a useful resource for the research community.

AVAILABILITY

RMBase v2.0 is freely available at http://rna.sysu.edu.cn/rmbase/. All of the data files can be downloaded and used in accordance with the GNU Public License and the licenses of the primary data sources.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR online.

FUNDING

REFERENCES

1. Li,S. and Mason,C.E. (2014) The pivotal regulatory landscape of RNA modifications. *Annu. Rev. Genom. Hum. G*, **15**, 127–150.
2. Meyer,K.D. and Jaffrey,S.R. (2014) The dynamic epitranscriptome: N6-methyladenosine and gene expression control. *Nat. Rev. Mol. Cell Biol.*, **15**, 313–326.
3. Kirchner,S. and Ignatova,Z. (2015) Emerging roles of tRNA in adaptive translation, signalling dynamics and disease. *Nat. Rev. Genet.*, **16**, 98–112.
4. Machnicka,M.A., Milanowska,K., Osman Oglou,O., Purta,E., Kurkowska,M., Olchowik,A., Januszewski,W., Kalinowski,S., Dunin-Horkawicz,S., Rother,K.M. *et al.* (2013) MODOMICS: a database of RNA modification pathways–2013 update. *Nucleic Acids Res.*, **41**, D262–D267.
5. Roundtree,I.A., Evans,M.E., Pan,T. and He,C. (2017) Dynamic RNA modifications in gene expression regulation. *Cell*, **169**, 1187–1200.
6. Gilbert,W.V., Bell,T.A. and Schaening,C. (2016) Messenger RNA modifications: form, distribution, and function. *Science*, **352**, 1408–1412.
7. Jaffrey,S.R. (2014) An expanding universe of mRNA modifications. *Nat. Struct. Mol. Biol.*, **21**, 945–946.
8. Song,C.X., Yi,C. and He,C. (2012) Mapping recently identified nucleotide variants in the genome and transcriptome. *Nat. Biotechnol.*, **30**, 1107–1116.
9. 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 m^{6}A RNA methylomes revealed by m^{6}A-seq. *Nature*, **485**, 201–206.
10. Geula,S., Moshitch-Moshkovitz,S., Dominissini,D., Mansour,A.A., Kol,N., Salmon-Divon,M., Hershkovitz,V., Peer,E., Mor,N., Manor,Y.S. *et al.* (2015) Stem cells. m^{6}A mRNA methylation facilitates resolution of naive pluripotency toward differentiation. *Science*, **347**, 1002–1006.
11. Chen,T., Hao,Y.J., Zhang,Y., Li,M.M., Wang,M., Han,W., Wu,Y., Lv,Y., Hao,J., Wang,L. *et al.* (2015) m(6)A RNA methylation is regulated by microRNAs and promotes reprogramming to pluripotency. *Cell Stem Cell*, **16**, 289–301.
12. Meyer,K.D., Saletore,Y., Zumbo,P., Elemento,O., Mason,C.E. and Jaffrey,S.R. (2012) Comprehensive analysis of mRNA methylation reveals enrichment in 3' UTRs and near stop codons. *Cell*, **149**, 1635–1646.
13. Schwartz,S., Bernstein,D.A., Mumbach,M.R., Jovanovic,M., Herbst,R.H., Leon-Ricardo,B.X., Engreitz,J.M., Guttman,M., Satija,R., Lander,E.S. *et al.* (2014) Transcriptome-wide mapping reveals widespread dynamic-regulated pseudouridylation of ncRNA and mRNA. *Cell*, **159**, 148–162.
14. Li,X., Xiong,X., Wang,K., Wang,L., Shu,X., Ma,S. and Yi,C. (2016) Transcriptome-wide mapping reveals reversible and dynamic N(1)-methyladenosine methylome. *Nat. Chem. Biol.*, **12**, 311–316.
15. Li,X., Zhu,P., Ma,S., Song,J., Bai,J., Sun,F. and Yi,C. (2015) Chemical pulldown reveals dynamic pseudouridylation of the mammalian transcriptome. *Nat. Chem. Biol.*, **11**, 592–597.
16. Dai,Q., Moshitch-Moshkovitz,S., Han,D., Kol,N., Amariglio,N., Rechavi,G., Dominissini,D. and He,C. (2017) Nm-seq maps 2'-O-methylation sites in human mRNA with base precision. *Nat. Methods*, **14**, 695–698.
17. 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.
18. Linder,B., Grozhik,A.V., Olarerin-George,A.O., Meydan,C., Mason,C.E. and Jaffrey,S.R. (2015) Single-nucleotide-resolution mapping of m^{6}A and m^{6}Am throughout the transcriptome. *Nat. Methods*, **12**, 767–772.
19. Dominissini,D., Nachtergaele,S., Moshitch-Moshkovitz,S., Peer,E., Kol,N., Ben-Haim,M.S., Dai,Q., Di Segni,A., Salmon-Divon,M., Clark,W.C. *et al.* (2016) The dynamic N(1)-methyladenosine methylome in eukaryotic messenger RNA. *Nature*, **530**, 441–446.
20. Konig,J., Zarnack,K., Luscombe,N.M. and Ule,J. (2012) Protein-RNA interactions: new genomic technologies and perspectives. *Nat. Rev. Genet.*, **13**, 77–83.
21. Barrett,T., Wilhite,S.E., Ledoux,P., Evangelista,C., Kim,I.F., Tomashevsky,M., Marshall,K.A., Phillippy,K.H., Sherman,P.M., Holko,M. *et al.* (2013) NCBI GEO: archive for functional genomics data sets–update. *Nucleic Acids Res.*, **41**, D991–D995.
22. Martin,M. (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.*, **17**, doi:10.14806/ej.17.1.200.
23. Kim,D., Landmead,B. and Salzberg,S.L. (2015) HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods*, **12**, 357–360.
24. Li,H. (2011) A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics*, **27**, 2987–2993.
25. Sun,W.J., Li,J.H., Liu,S., Wu,J., Zhou,H., Qu,L.H. and Yang,J.H. (2016) RMBase: a resource for decoding the landscape of RNA modifications from high-throughput sequencing data. *Nucleic Acids Res.*, **44**, D259–D265.
26. Tyner,C., Barber,G.P., Casper,J., Clawson,H., Diekhans,M., Eisenhart,C., Fischer,C.M., Gibson,D., Gonzalez,J.N., Guruvadoo,L. *et al.* (2017) The UCSC genome browser database: 2017 update. *Nucleic Acids Res.*, **45**, D626–D634.
27. Harrow,J., Frankish,A., Gonzalez,J.M., Tapanari,E., Diekhans,M., Kokocinski,F., Aken,B.L., Barrell,D., Zadissa,A., Searle,S. *et al.* (2012) GENCODE: The reference human genome annotation for The ENCODE Project. *Genome Res.*, **22**, 1760–1774.
28. Yates,A., Akanni,W., Amode,M.R., Barrell,D., Billis,K., Carvalho-Silva,D., Cummins,C., Clapham,P., Fitzgerald,S., Gil,L. *et al.* (2016) Ensembl 2016. *Nucleic Acids Res.*, **44**, D710–D716.
29. Meng,J., Lu,Z.L., Liu,H., Zhang,L., Zhang,S.W., Chen,Y.D., Rao,M.K. and Huang,Y.F. (2014) A protocol for RNA methylation differential analysis with MeRIP-Seq data and exomePeak R/Bioconductor package. *Methods*, **69**, 274–281.
30. Schwartz,S., Mumbach,M.R., Jovanovic,M., Wang,T., Maciag,K., Bushkin,G.G., Mertins,P., Ter-Ovanesyan,D., Habib,N., Cacchiarelli,D. *et al.* (2014) Perturbation of m^{6}A writers reveals two distinct classes of mRNA methylation at internal and 5' sites. *Cell Rep.*, **8**, 284–296.
31. 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.
32. 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.
33. Yang,J.H., Li,J.H., Shao,P., Zhou,H., Chen,Y.Q. and Qu,L.H. (2011) starBase: a database for exploring microRNA-mRNA interaction maps from Argonaute CLIP-Seq and Degradome-Seq data. *Nucleic Acids Res.*, **39**, D202–D209.
34. Forbes,S.A., Beare,D., Boutselakis,H., Bamford,S., Bindal,N., Tate,J., Cole,C.G., Ward,S., Dawson,E., Ponting,L. *et al.* (2017) COSMIC: somatic cancer genetics at high-resolution. *Nucleic Acids Res.*, **45**, D777–D783.
35. Lawrence,M.S., Stojanov,P., Mermel,C.H., Robinson,J.T., Garraway,L.A., Golub,T.R., Meyerson,M., Gabriel,S.B., Lander,E.S. and Getz,G. (2014) Discovery and saturation analysis of cancer genes across 21 tumour types. *Nature*, **505**, 495–501.
36. Alexandrov,L.B., Nik-Zainal,S., Wedge,D.C., Aparicio,S.A., Behjati,S., Biankin,A.V., Bignell,G.R., Bolli,N., Borg,A., Borresen-Dale,A.L. *et al.* (2013) Signatures of mutational processes in human cancer. *Nature*, **500**, 415–421.
37. Li,J.H., Liu,S., Zheng,L.L., Wu,J., Sun,W.J., Wang,Z.L., Zhou,H., Qu,L.H. and Yang,J.H. (2014) Discovery of protein-lncRNA interactions by integrating large-scale CLIP-Seq and RNA-Seq datasets. *Front. Bioeng. Biotechnol.*, **2**, 88.
38. Welter,D., MacArthur,J., Morales,J., Burdett,T., Hall,P., Junkins,H., Klemm,A., Flicek,P., Manolio,T., Hindorff,L. *et al.* (2014) The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res.*, **42**, D1001–D1006.
39. Johnson,A.D. and O'Donnell,C.J. (2009) An open access database of genome-wide association results. *BMC Med. Genet.*, **10**, 6.
40. Tryka,K.A., Hao,L., Sturcke,A., Jin,Y., Wang,Z.Y., Ziyabari,L., Lee,M., Popova,N., Sharopova,N., Kimura,M. *et al.* (2014) NCBI's database of genotypes and phenotypes: dbGaP. *Nucleic Acids Res.*, **42**, D975–D979.

41. Becker,K.G., Barnes,K.C., Bright,T.J. and Wang,S.A. (2004) The genetic association database. *Nat. Genet.*, **36**, 431–432.

42. International HapMap,C., Altshuler,D.M., Gibbs,R.A., Peltonen,L., Altshuler,D.M., Gibbs,R.A., Peltonen,L., Dermitzakis,E., Schaffner,S.F., Yu,F. *et al.* (2010) Integrating common and rare genetic variation in diverse human populations. *Nature*, **467**, 52–58.

43. 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.

44. Klungland,A. and Dahl,J.A. (2014) Dynamic RNA modifications in disease. *Curr. Opin. Genet. Dev.*, **26**, 47–52.

45. Blanco,S. and Frye,M. (2014) Role of RNA methyltransferases in tissue renewal and pathology. *Curr. Opin. Cell Biol.*, **31**, 1–7.

46. Patil,D.P., Chen,C.K., Pickering,B.F., Chow,A., Jackson,C., Guttman,M. and Jaffrey,S.R. (2016) m(6)A RNA methylation promotes XIST-mediated transcriptional repression. *Nature*, **537**, 369–373.

47. Edupuganti,R.R., Geiger,S., Lindeboom,R.G.H., Shi,H., Hsu,P.J., Lu,Z., Wang,S.Y., Baltissen,M.P.A., Jansen,P., Rossa,M. *et al.* (2017) N6-methyladenosine (m^{6}A) recruits and repels proteins to regulate mRNA homeostasis. *Nat. Struct. Mol. Biol.*, **24**, 870–878.

48. Ray,D., Kazan,H., Cook,K.B., Weirauch,M.T., Najafabadi,H.S., Li,X., Gueroussov,S., Albu,M., Zheng,H., Yang,A. *et al.* (2013) A compendium of RNA-binding motifs for decoding gene regulation. *Nature*, **499**, 172–177.