RNA modifications have been known for more than half a century. To date, approximately 170 kinds of RNA modifications have been identified in different types of RNA molecules, such as messenger RNA (mRNA), transfer RNA (tRNA), and ribosomal RNA (rRNA) [1]. Among these modifications, N6-methyladenosine (m6A) is the most abundant and well characterized covalent modification [2,3]. In order to deciphering the biological functions of m6A, a series of high-throughput sequencing techniques [2–6] have been proposed. These techniques not only deepened our understanding of m6A, but also stimulated and promoted the researches on the other kinds of RNA modifications (i.e., adenosine to inosine, N1-methyladenosine, N6-methyladenosine, 5-methylcytidine, N7-methylguanosine, mRNA cap modification, and pseudouridine). The 3D structures, domains, subcellular locations, and biological functions of these enzymes were also integrated in innovative resource for RNA modification enzymes. The current version of RNAME deposits more than 21,000 manually curated RNA modification enzymes, which are from 456 species and covers the 7 common kinds of RNA modifications (i.e., adenosine to inosine, N1-methyladenosine, N6-methyladenosine, 5-methylcytidine, N7-methylguanosine, mRNA cap modification, and pseudouridine). The 3D structures, domains, subcellular locations, and biological functions of these enzymes were also integrated in RNAME. It is anticipated that RNAME will facilitate the researches on RNA modifications.

© 2022 The Author(s). Published by Elsevier B.V. on behalf of Research Network of Computational and Structural Biotechnology. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).
Unfortunately, only a very small number of experimentally validated RNA modification enzymes have been reported. Hence, our knowledge about RNA modification enzymes is only a tip of the iceberg. In order to enrich our knowledge about RNA modification enzymes and also speed up researches on RNA modifications, it is fundamental to identify RNA modification enzymes and construct a database depositing them. Although some representative databases regarding RNA modifications have been built, such as Modomics [1], m5C-Atlas [32], m6A-Atlas [33], RMDisease [34], RMVar [35], m6AVar [36], most of them focus on the associations between RNA modification and diseases, and few of them concerned on RNA modification enzymes. Although Modomics collected the RNA modification enzymes, the provided information is far from satisfactory. For example, the information about protein–protein interaction network, domains, subcellular localization, and structure of most RNA modification enzymes were not provided in Modomics.

Considering the fact that an integrative database of RNA modification enzymes is still lacking, we developed the RNAME (RNA modification enzyme) database aiming to provide a resource for efficient manipulation, browse and analysis of RNA modification enzymes, which is available at https://chenweiab.cn/rname/. The current version of RNAME deposits 21,931 RNA modification enzymes, which are from 456 species and covers the 7 main kinds of RNA modifications, namely Cap, I, m1A, m6A, m5C, mRNA cap modification (Cap), I, Ψ, and m7G. Among the 21,931 RNA modification enzymes, 439 were manually annotated and reviewed data from UniProt [37], 17,041 were annotated data from UniProt, 4,451 were predicted entries based on the data from Ensembl [38]. The framework for constructing RNAME is described in Fig. 1.

2. Material and methods

2.1. Data collection

To establish RNAME, we collected RNA modification effectors of the 7 main kinds of RNA modifications, namely Cap, I, m1A, m6A, m5C, m7G and Ψ, from RNAWRE [39] and the review articles [8, 25, 31, 40–42], and obtained 81 kinds of RNA modification effectors participating the deposition, recognition, and removal of RNA modifications. Their domains were obtained from UniProt database, Pfam database [43] and CDD database [44] and listed in Supplementary Table 1. By using the names of these effectors as the key words, we searched the UniProt database [37] and retrieved 451 proteins after manually curating their information. Among these proteins, 381 are from animals, 61 from fungi, and 9 from plants. Their UniProt IDs and sequences were provided in Supplementary Table 2. In addition, the sequences of RNA modification effectors in plants reported in the review article [42] were also collected (Supplementary Table 3). These sequences were used as the seeds to identify RNA modification effectors from other species.

2.2. Identification the effectors

In order to identify the effectors from multiple species, we firstly downloaded the complete proteomes of 456 eukaryotes including 295 animals, 94 plants and 67 fungi from Ensembl (release version 106, https://www.ensembl.org/) [38], EnsemblPlants (release version 52, https://plants.ensembl.org/) and EnsemblFungi (release version 53, https://fungi.ensembl.org/), respectively. Then, we matched the seed sequences to the protein sequences by using the BLASTP program (blast-2.13.0 + [45] with an E-value cut-off of 1e-3. To determine whether the sequences from BLAST hits are potential RNA modification effectors, we further performed domain analysis by using Batch CD-Search [44] and validated the domains on the basis of Supplementary Table 1. Accordingly, the hits with the best Blast score and CDD score were retained for animals and fungi. Since polyplody is common in plant, if a query sequence from plants yielded multiple blast hits, we sorted the Blast scores and CDD scores and retained the top ten hits. Finally, we obtained 21,931 RNA modification enzymes, among them 439 were the manually annotated and reviewed entries from UniProt, 17,041 were annotated entries from UniProt, 4,451 were predicted entries based on the data from Ensembl.

2.3. Data annotation

The basic information, structure, protein–protein interaction network (PPI network), domains, subcellular location, substrate RNA, function, genome annotation database and sequence information of these enzymes were also collected and reorganized. ‘Basic Information’ contains the protein name, UniProtID, Organism, Gene Symbol, substrate RNA, etc. ‘Structure’ displayed the 3D-structures of each effector by using the Molstar tool [46]. The 3D-structures were collected from RCSB-PDB database [47] or AlphaFold Protein Structure Database [48]. For proteins without structures in these two databases, we predicted their structures by using the SWISS-MODEL software [49]. ‘PPI network’ displayed the interaction partners by using the script echarts (https://echarts.apache.org/). Because a protein may have many interaction partners, we collected its top ten partners according to the combined score from String database [50]. ‘Domains’ displayed the protein domain, which was obtained by using the Pfam domain graphics (https://www.ebi.ac.uk/pfam/pfam_domigraphy). ‘Subcellular location’ indicated the subcellular location of the effec-

Table 1
Statistics of RNA modification enzymes in RNAME.

| Species  | Cap | I  | m1A | m6A | m5C | Ψ  | m7G |
|----------|-----|----|-----|-----|-----|----|-----|
| Animals  | 1118| 1001| 2188| 7519| 2701| 2664| 818 |
| Plant    | 140 | 0  | 1   | 2250| 263 | 2   | 0   |
| Fungi    | 132 | 3  | 191 | 515 | 66  | 291 | 68  |
tors, which was collected from UniProt Database or predicted by using SherLoc2[51]. ‘Function’ displayed the GO annotations of the effectors, which was collected from UniProt Database. ‘Genome annotation database’ displayed the annotation ids collected from Ensembl. ‘Sequence Information’ displayed the sequence of the effectors collected from Ensembl.

3. Results

3.1. Web interface implementation

The RNAME database was developed based on the python-based Django framework. The user-friendly web page was written in html, CSS and JavaScript, with all data stored and organized in MySQL[52]. An interactive web interface was constructed to allows users to access, visualize, analyze, and obtain required information of RNA modification effectors (Fig. 2).

3.2. Statistics of RNAME

The RNAME database contains 21,931 manually curated writers, readers and erasers for Cap, I, m^1A, m^6A, m^5C, W, and m^7G modifications. They are from 456 species covering animals, plants and fungi. The detail information was enumerated in Table 1 and Sup-
Complementary Table 4. Among these effectors, 22,343 have 3D structures, 21,931 have domains, 7,829 have PPI information. The detail information based on species was shown in Fig. 3.

Determining the subcellular localizations of RNA modification effectors will shed light on understanding its biological functions [53]. Thus, we analyzed the subcellular localizations of RNA modification effectors in animals (Table 2). It was found that the effectors for Cap, I, m1A, m6A, m5C, and \( \Psi \) modifications were mainly located in cytoplasm. The subcellular localizations of the effectors regulating m7G exhibit a similar proportion in both cytoplasm and nucleus. However, some effectors mostly located in nuclear but rarely in cytoplasm. For example, 262 of the 267 YTHDC1 in animals were located in nuclear with a proportion of 98.13%. This phenomenon may be associated with the function of YTHDC1 that mediates the nuclear export of m\( \Psi \)A[54].

3.3. Usage

In the “Search” module (Fig. 4A), users can search RNA modification enzymes through Exact search, Fuzzy search or Batch search. Users can search RNA modification enzymes through four paths, including “By organism” (search any key information about organism, supporting exact, fuzzy and batch search), “By Effector” (select “writers”, “readers”, or “erasers”, both exact and fuzzy search supported), “By Modification” (select m\( \Psi \)A, m\( \Psi \)A, m\( \Psi \)C, Cap, I, \( \Psi \) or m7G, supporting exact, fuzzy and batch search), and “By Gene Symbol” (select gene names of RNA modification enzymes, supporting exact, fuzzy and batch search). For all these search approaches, after inputting the search item (such as Homo sapiens) related to the concerned keyword and clicking the search button, the brief information of searching results will be presented in the result page (Fig. 4B).

In the “Browse” module (Fig. 4C), users can browse RNA modification effectors by “Effector”, “Gene Symbol”, “Modification” or “Organism”. Through the option of “By organism”, users can browse the database by “Animal”, “Fungi” or “Plant”. After clicking the browse keyword, taking Homo sapiens as an example, the brief information of searching results will be displayed in Fig. 4B.

As shown in Fig. 4B, users can further filter the results by using “Organism”, “Effector”, “Modification” and “Gene Symbol”. Detail information for each effector could be further displayed by clicking the “More” button. The detail page presents not only the basic information of an RNA modification enzyme (Fig. 5A), but also the information about its structure (Fig. 5B), protein–protein inter-
actions (Fig. 5C), domains and subcellular location (Fig. 5D), function (Fig. 5E), sequence (Fig. 5F), etc.

Users can analyze query sequences through the “Blast” page. First, enter the query sequences in FASTA format. Second, select the type of enzymes. Besides default parameters, users could also set additional parameters in the “Advanced parameters” box, such as the e-value, score, and output format. Finally, clicking “BLAST” to obtain the results. The results between query sequences and object sequences will be presented in a variety of visualization methods and available for download in vectorial (SVG) and PNG formats.

It is inevitable that our collections may not cover all RNA modification enzymes. In the “Submit” page, researchers can submit new entries that are not included in RNAME. In “Help” page, the operations manual was provided to make RNAME convenient to be used. In “Download” page, users can download the sequences of RNA modification enzymes. In “Links” page, useful links related to RNA modification were provided.

4. Conclusion

The dynamic RNA modifications were orchestrated by a series of enzymes. However, only a very small number of RNA modification enzymes have been reported and scattered across various resources, which hindered functional researches on RNA modifications. Although a bunch of databases regarding RNA modifications have been published, few of them concerned on RNA modification enzymes. Thus, there is an urgent need to develop a database for depositing RNA modification enzymes.

In the present work, by collecting and reorganizing the information from multiple resources, we developed the RNAME database unique for depositing RNA modification enzymes. The current version of RNAME is available at https://chenweilab.cn/rname/, and includes 21,931 RNA modification enzymes that the seven main kinds of RNA modifications, namely m1A, m6A, m5C, Cap, I, Ψ, and m7G. Besides the sequences of the enzymes, their 3D structures, domains, subcellular locations, and biological functions were also provided in RNAME. It is expected that the RNAME database will benefit the researches on RNA modifications.

With the advancements of experimental techniques, novel RNA modification enzymes will be reported. Therefore, we will continue to collect new data and store them in RNAME in time. Moreover, besides collecting new effectors reported in literatures, we will also collect the effectors from other domains of life and integrate them into RNAME. In addition, RNAME always welcome researchers to contribute and share their novel findings. Taken together, we hope that RNAME will provide valuable insights into researches on RNA modifications.

Funding

Natural Science Foundation of Sichuan (No.2022NSFSC1770) and National Natural Science Foundation of China No.31771471.

W. Chen and P.M. Feng conceived and designed the study. F.L. Nie, Q. Tang, Y. Liu, H.Z. Qin, S.L. Liu, and M. Wu collected the data and constructed the web-server. All authors performed the analysis, wrote the paper, and approved the final manuscript.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.
Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cbsj.2022.11.022.

References

[1] Boccaletto P, Stefanakis F, Ray A, et al. MODOMICS: a database of RNA modification pathways. 2021 update. Nucleic Acids Res 2022;50:D221–5.
[2] Dominissini D, Moshtchik-Moshkovitz S, Schwartz S, et al. Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. Nature 2012;485:201–6.
[3] Meyer KD, Salatore Y, Zumbo P, et al. Comprehensive analysis of mRNA methylation reveals enrichment in 3’ UTRs and near stop codons. Cell 2012;149:1635–46.
[4] Linder B, Grozhik AV, Olarerin-George AO, et al. Single-nucleotide-resolution mapping of m6A and m4Am throughout the transcriptome. Nat Methods 2015;12:767–72.
[5] Ke S, Alemu EA, Mertens C, et al. A majority of m6A residues are in the last exons, allowing the potential for 3’ UTR regulation. Genes Dev 2015;29:2013–53.
[6] Molnie B, Wang J, Lim KS, et al. m6A-LAC-seq reveals the census and complexity of the m6A epitranscriptome. Nat Methods 2016;13:692–8.
[7] Wiener D, Schwartz S. The epitranscriptome beyond m6A. Nat Rev Genet 2021;22:118–31.
[8] Nishikura K. A-to-t1 editing of coding and non-coding RNAs by ADARs. Nat Rev Mol Cell Biol 2016;17:83–96.
[9] Bakin A, Ongend J. Four newly located pseudouridylate residues in Escherichia coli 23S ribosomal RNA are all at the peptidyltransferase center. Analysis by the application of a new sequencing technique. Biochemistry 1993;32:9754–62.
[10] Li X, Zhu P, Ma S, et al. Chemical pulldown reveals dynamic pseudouridylation of the mammalian transcriptome. Nat Chem Biol 2015;11:592–7.
[11] Dominissini D, Nachtergaele S, Moshtich-Moshkovitz S, et al. The dynamic N(1)-methyladenosine methylome in eukaryotic messenger RNA. Nature 2016;530:441–6.
[12] Grozhik AV, Olarerin-George AO, Sindelar M, et al. Antibody cross-reactivity accounts for widespread appearance of m1A in 5’UTRs. Nat Commun 2019;10:5126.
[13] Squires JE, Patel HR, Nousch M, et al. Widespread occurrence of 5’-methylcytosine in human coding and non-coding RNA. Nucleic Acids Res 2012;40:5023–33.
[14] Khoddami V, Cairns BR. Identification of direct targets and modified bases of RNA cytosine methyltransferases. Nat Biotechnol 2013;31:458–64.
[15] Euroth C, Poulsen LD, Iversen S, et al. Detection of internal N7-methylguanosine (m7G) RNA modifications by mutational profiling sequencing. Nucleic Acids Res 2019;47:e126.
[16] Malbec L, Zhang T, Chen YS, et al. Dynamic methylation of internal mRNA N(7)-methylguanosine and its regulatory role in translation. Cell Res 2019;29:927–41.
[17] Sakurai M, Ueda H, Yano T, et al. A biochemical landscape of A-to-I RNA editing in the human brain transcriptome. Genome Res 2014;24:522–34.
[18] Wang X, Lu Z, Gomez A, et al. N6-methyladenosine-dependent regulation of messenger RNA stability. Nature 2014;505:117–20.
[19] Xiao W, Adhikari S, Dahal U, et al. Nuclear m6A Reader YTHDC1 Regulates mRNA Splicing. Mol Cell 2016;61:507–19.
[20] Chen T, Hao YJ, Zhang Y, et al. m6A RNA methylation is regulated by microRNAs and promotes reprogramming to pluripotency. Stem Cell 2021;16:289–301.
[21] Zhong X, Yu J, Frazier K, et al. Circadian Clock Regulation of Hepatic Lipid Metabolism by Modulation of m6A RNA Methylation. Cell Rep 2018;25 (1816–28):e4.
[22] Pastin JM, Dori M, Yamaguchi Y, et al. RNA-methylation-dependent RNA processing controls the speed of the circadian clock. Cell 2013;155:793–806.
[23] Yin L, Zhu X, Novak P, et al. The epitranscriptome of long noncoding RNAs in metabolic diseases. Clin Chim Acta 2021;515:80–9.
[24] Lin X, Chai G, Wu Y, et al. RNA m6A methylation regulates the epithelial mesenchymal transition of cancer cells and translation of Snail. Nat Commun 2019;10:2065.
[25] Barberi I, Kouzarides T. Role of RNA modifications in cancer. Nat Rev Cancer 2020;20:303–22.
[26] Wu Y, Zhan S, Xu Y, et al. RNA modifications in cardiovascular diseases, the potential therapeutic targets. Life Sci 2021;278:119565.
[27] Frayling TM, Timpson NJ, Weedon MN, et al. A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. Science 2007;316:889–94.
[28] Hernandez-Caballerio ME, Sierra-Ramirez JA. Single nucleotide polymorphisms of the FTO gene and cancer risk: an overview. Mol Biol Rep 2015;42:699–704.
[29] Campbell TM, Castro MAA, de Oliveira KG, et al. m6A binding by transcription factors NFkB and YBX1 enables FGFR2 signaling to modulate estrogen responsiveness in breast cancer. Cancer Res 2018;78:410–21.
[30] Lu L, Gaffney SC, Cannatara VL, et al. Transfer RNA methyltransferase gene NSUN2 mRNA expression modifies the effect of T cell activation score on patient survival in head and neck squamous carcinoma. Oral Oncol 2020;101:104554.
[31] Nombela P, Miguel-López B, Blanco S. The role of m6A, m5C and m1A in the epitranscriptome of long noncoding RNAs. Elife 2017;6:2018.