

# RNAmod: an integrated system for the annotation of mRNA modifications

Qi Liu<sup>1,2,\*</sup> and Richard I. Gregory<sup>1,2,3,4,5,\*</sup>

<sup>1</sup>Stem Cell Program, Division of Hematology/Oncology, Boston Children's Hospital, Boston, MA 02115, USA, <sup>2</sup>Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, MA 02115, USA, <sup>3</sup>Department of Pediatrics, Harvard Medical School, Boston, MA 02115, USA, <sup>4</sup>Harvard Initiative for RNA Medicine, Boston, MA 02115, USA and <sup>5</sup>Harvard Stem Cell Institute, Cambridge, MA 02138, USA

Received March 12, 2019; Revised May 02, 2019; Editorial Decision May 14, 2019; Accepted May 17, 2019

## ABSTRACT

Dynamic and reversible RNA modifications such as *N*<sup>6</sup>-methyladenosine (*m*<sup>6</sup>A) can play important roles in regulating messenger RNA (mRNA) splicing, export, stability and translation. Defective mRNA modification through altered expression of the methyltransferase and/or demethylases results in developmental defects and cancer progression. Identifying modified mRNAs, annotating the distribution of modification sites across the mRNA, as well as characterizing and comparing other modification features are essential for studying the function and elucidating the mechanism of mRNA modifications. Several methods including methylated RNA immunoprecipitation and sequencing (MeRIP-seq) are available for the detection of mRNA modifications. However, a convenient and comprehensive tool to annotate diverse kinds of mRNA modifications in different species is lacking. Here, we developed RNAmod (<https://bioinformatics.sc.cn/RNAmod>), an interactive, one-stop, web-based platform for the automated analysis, annotation, and visualization of mRNA modifications in 21 species. RNAmod provides intuitive interfaces to show outputs including the distribution of RNA modifications, modification coverage for different gene features, functional annotation of modified mRNAs, and comparisons between different groups or specific gene sets. Furthermore, sites of known RNA modification, as well as binding site data for hundreds of RNA-binding proteins (RBPs) are integrated in RNAmod to help users compare their modification data with known modifications and to explore the relationship with the bind-

ing sites of known RBPs. RNAmod is freely available and meets the emerging need for a convenient and comprehensive analysis tool for the fast-developing RNA modification field.

## INTRODUCTION

There are over 160 known RNA modifications and together this repertoire of different modifications has been collectively termed the 'epitranscriptome' (1). Several of these modifications have been identified in mRNA, indicating possible roles in regulating gene expression (2). For example *N*<sup>6</sup>-methyladenosine (*m*<sup>6</sup>A), the most abundant modification on mRNA, has been linked to the control of mRNA splicing, export, stability, and translation (3). *m*<sup>6</sup>A modification can be dynamically and reversibly regulated by the action of the RNA methyltransferase METTL3 and demethylases ALKBH5 and FTO (4). Accordingly, misregulation of mRNA modifications results in multi-tissue developmental defects and cancer progression (5–9). Modifications on mRNAs are not randomly distributed, rather they occur on specific sequence motifs at certain locations on mRNA. For example, *m*<sup>6</sup>A is mainly located near the translation start codon and translation stop codon (10); while *N*<sup>1</sup>-methyladenosine (*m*<sup>1</sup>A) is highly enriched at the 5' untranslated region (UTR) of mRNA (11,12). Additional mRNA modifications have recently been identified, including cytidine *N*<sup>4</sup>-acetylation (*ac*<sup>4</sup>C) (13) and internal *N*<sup>7</sup>-methylguanosine (*m*<sup>7</sup>G) (14).

The localization of mRNA modification correlates with its molecular functions, therefore, determining the distribution of modification across the mRNA and annotating the modified gene targets are essential for studying the function of modifications. Next-generation sequencing (NGS) based technologies has offered an unprecedented opportunity to discover and profile RNA modifications under various conditions, such as MeRIP-seq (methylated RNA immunoprecipitation and sequencing (15),

\*To whom correspondence should be addressed. Tel: +1 617 919 2273; Email: rgregory@enders.tch.harvard.edu  
Correspondence may also be addressed to Qi Liu. Tel: +1 617 355 0204; Email: qi.liu2@childrens.harvard.edu

miCLIP-seq (m<sup>6</sup>A individual-nucleotide-resolution cross-linking and immunoprecipitation and sequencing) (16) and detection of m<sup>7</sup>G by TRAC-seq (tRNA reduction and cleavage-sequencing) (17) and m<sup>7</sup>G-seq (14). Meanwhile, several prediction tools based on machine learning were developed to systematically predict RNA modifications, including WHISTLE (18), SRAMP (19), M6AMRFS (20), RAM-ESVM (21), DeepM6ASeq (22). All these methods have greatly facilitated the identification of mRNA modifications. However, the detailed annotation of these dynamic modifications is still a bottleneck for better understanding of their function. There are some tools available for the annotation of RNA modifications, such as RNAModR (23) and Guitar (24) that generate meta-gene plots of modification sites/peaks, and RCAS (25) for meta-gene plots and RNA based annotations, and some databases, such as RM-Base (26) also provide the function to perform meta-gene plots. However, to the best of our knowledge, all available tools pay more attention to conducting just meta-gene analysis and there is no convenient web server available for the automated analysis and annotation of mRNA modifications. A comprehensive web-based tool to annotate diverse kinds of mRNA modifications and their potential functions for different species is still lacking.

Here, we developed RNAMod, a very convenient web-based platform, for the analysis and functional annotation of mRNA modifications in up to 21 species, including human, mouse, rat, zebrafish, fly, yeast and Arabidopsis, etc. RNAMod uses the commonly used BED format of chromosomal location of RNA modifications as input, which can be generated by almost all site/peak calling tools, such as MACS (27), exomePeak (28) and BayesPeak (29). RNAMod first maps individual modification sites onto RNAs and then performs various analysis and annotations, including peak/site distribution among different RNA features, distribution among different gene biotypes, peak/site coverage for different gene features, coverage around transcription start/end sites and translation start/end sites, coverage around splicing sites, mRNA metagene analysis, and gene characteristics and functional analyses of modified mRNAs, thereby providing a comprehensive annotation and visualization of the distribution of mRNA modifications. Besides, RNAMod allows the user to perform comparative analysis between groups or between specific gene sets and other background genes. Moreover, JBrowse is integrated in RNAMod and contains known modification sites and binding sites of RBP, which will help users to compare their modification sites with known modifications and RBP binding sites. It is noted that RNAMod can also be used for the annotation and visualization of CLIP-seq data to study the binding features of RBPs. In the download page, RNAMod provides the annotation for known modification sites and binding sites of nearly 500 RBPs from different species. The web server pages of RNAMod utilize several intuitive JavaScript libraries to show the results in very interactive ways. Meanwhile, users can download all the outputs using a one-click link. Overall, RNAMod is a one-stop online interactive platform for the annotation, analysis, and visualization of mRNA modifications. The website is freely available and open access with no login re-

quired (<https://bioinformatics.sc.cn/RNAMod> or <http://61.147.117.195/RNAMod> or <https://rnabioinform.tch.harvard.edu/RNAMod>). We believe that it will meet the emerging need for a convenient and comprehensive analysis tool for the fast-developing RNA biology research field.

## ANALYSIS WORKFLOW

### Data sources

The gene annotation files in GTF format and mRNA transcript sequences in FASTA format were downloaded from GENCODE (<https://www.encodegenes.org>) for human (hg38 and hg19) and mouse (mm10), while for the other 19 species, the annotation files were retrieved from FTP site of Ensembl genome database (30) in which, compared with other databases, the annotations contain not only mRNA but also other types of RNAs. To accelerate data processing and gene features extraction in the RNAMod web server, all GTF gene annotations were stored in TxDB objects in the SQLite database by using GenomicFeatures (31) and RSQLite packages in R (<https://www.r-project.org>). The reference genomes used for the sequences extraction were stored as BSgenome packages. For the species without an available BSgenome package in Bioconductor (<https://www.bioconductor.org>), we constructed them using in-house developed programs based on genome sequence download from Ensembl. We split the mRNA sequences into three different parts (5' UTR, CDS and 3' UTR) and calculated their gene characteristics separately. The characteristics include sequence length, GC content, and minimum free energy, which are used in the comparative characteristics analysis of modified genes. RNAfold (32) was used to calculate the minimum free energy. For the functional annotation of modified genes, the Gene Ontology (GO) terms and KEGG pathways are provided by clusterProfiler package (33), Reactome pathways are packaged with ReactomePA packages (34), while Disease Ontology, Network of Cancer Gene and DisGeNET disease genes are available in DOSE packages (35). Other gene sets used in functional enrichments include chemical and genetic perturbation genes, microRNA target motif genes, cancer gene neighborhoods, cancer module genes, oncogenic signature genes, and immunologic signature genes, which are all downloaded from The Molecular Signatures Database (MSigDB) (36). The known RNA modification sites data were obtained from RMBase 2.0 (26) and MeT-DB V2.0 (37) while RBP binding sites data were obtained from POSTAR (38) database. The known RNA modification sites in RMBase contain different kinds of modification sites identified from a large number of high-throughput sequencing studies. We separated them into individual files by study and samples. MeT-DB V2.0 is an important database for m<sup>6</sup>A modification and contains 185 samples from 26 independent studies for seven species. RBP binding site information in POSTAR database are already organized by samples. The modification sites information and RBP binding site data for each sample were transformed into BED format, which were then stored in JBrowse using flatfile-to-json.pl script and annotated using RNAMod.

## RNA modification data processing in RMBase 2.0 and Met-DB 2.0

In RMBase 2.0 (26), raw sequencing data of m<sup>6</sup>A-seq or MeRIP-seq is obtained from the Gene Expression Omnibus (GEO) (39) and the Sequence Read Archive (SRA) (40) were first cleaned using cutadapt (<https://cutadapt.readthedocs.io/en/stable>) and FASTX-toolkit ([http://hannonlab.cshl.edu/fastx\\_toolkit](http://hannonlab.cshl.edu/fastx_toolkit)). The clean data were then mapped to the reference genomes using hisat2 (41) with default parameters and the converted BAM mapping results were used to call modification peaks using exomePeak (28) with the parameters ‘fold change (FC) >2, P-value <0.01 and false discovery rate (FDR) <0.05’. Based on the identified peaks, the exact m<sup>6</sup>A positions were predicted by searching consensus RRACH motifs (R denotes A or G while H represents A, C or U) (16). By using a similar strategy, the m<sup>1</sup>A modification sites were detected based on the peaks from m<sup>1</sup>A-seq using GAAGAAG motif (42). The peaks/sites of other less abundant modifications, such as pseudouridine and m<sup>5</sup>C, were manually collected from the respective studies. In Met-DB 2.0, the raw sequencing data of MeRIP-seq IP/Input collected from SRA (40) were cleaned using Trim Galore (v0.4.2) ([https://www.bioinformatics.babraham.ac.uk/projects/trim\\_galore](https://www.bioinformatics.babraham.ac.uk/projects/trim_galore)) and mapped to the reference genomes using Tophat2 (v2.1.0) (43) with default parameters. Based on the BAM mapping files of IP and input samples, m<sup>6</sup>A peaks were then called using exomePeak (28) with default parameters.

## Overview of RNAmoD

The overall workflow of RNAmoD is shown in Figure 1. RNAmoD first extracts gene features from the reference genome annotation and then maps the submitted modification sites onto different RNA features, it then performs various coverage calculations, metagene analyses, and annotations focusing on mRNAs. The annotations include: (i) site distribution among different gene features and gene biotypes, (ii) coverage analysis among RNA features, (iii) site distribution around transcription start/end site, (iv) site distribution around translation start/end sites, (v) site distribution around splicing junction sites, (vi) comparison of gene characteristics between modified genes and other genes, (vii) modified site heatmap around translation start/end sites and transcription start/end sites, (viii) mRNA metagene analysis, (ix) motif enrichment analysis and (x) functional enrichments for modified genes. To facilitate users who have different analysis requirements, the web server of RNAmoD provides three functional modules, including single case, group case and gene case. The single case module allows users to annotate RNA modifications for a single sample. The group case modules allow users to annotate and compare the distribution of modifications between two samples or even more groups. Considering the situation in which users need to know the modification site distribution and modified gene features for a specific gene or set of genes, such as translation down-regulated genes, cell-cycle genes, or neural development genes, the gene analysis module can be used to achieve this goal by allowing users to input a specific gene group. Meanwhile, we integrated JBrowse tool (44) to help users to check and compare the modification

in the context of known RNA modification sites and RNA-binding sites of RBPs.

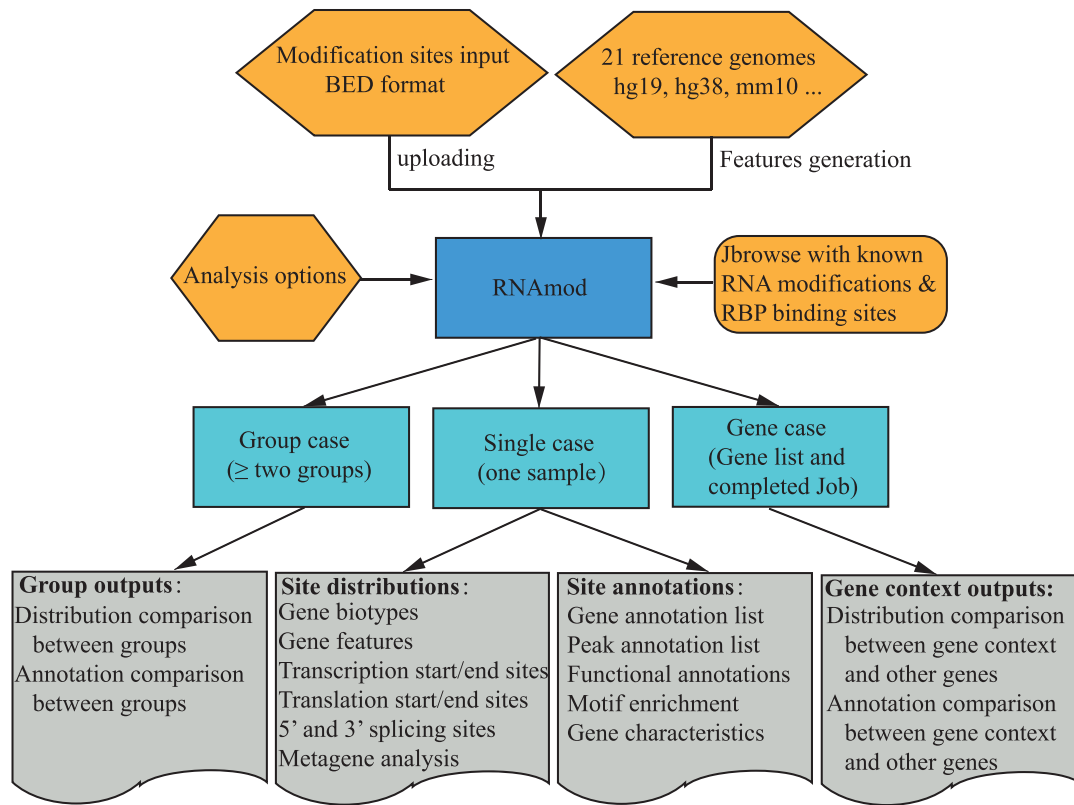
## Data input of RNAmoD

In the single case, RNAmoD utilizes the commonly used BED format of chromosomal location of RNA modifications as input, which can be generated by almost all modification sites/peak calling tools based on MeRIP-seq, such as MACS (27), exomePeak (28), BayesPeak (29) and MetPeak (45). BED format can also be easily converted from other text outputs that are generated from the data analyses of other protocols, such as miCLIP-m6A (16) and PseudoSeq (46). In the links web page, RNAmoD also provides an R script for users to call the modification peaks/sites from MeRIP-seq data. To speed up the uploading, users are recommended to submit a compressed BED file in .gz or .zip format. After submission of data, the data analysis queue system provides users with a job ID, a string consisting of 16 random characters, which can be used to retrieve the results once the job is finished. In the group case module, two types of inputs are supported: BED format files and single case job IDs. When a single case analysis job ID is given, the web server will retrieve the completed annotation automatically. In the gene case, users are required to input the single case or group case job IDs, and the gene list (Figure 2).

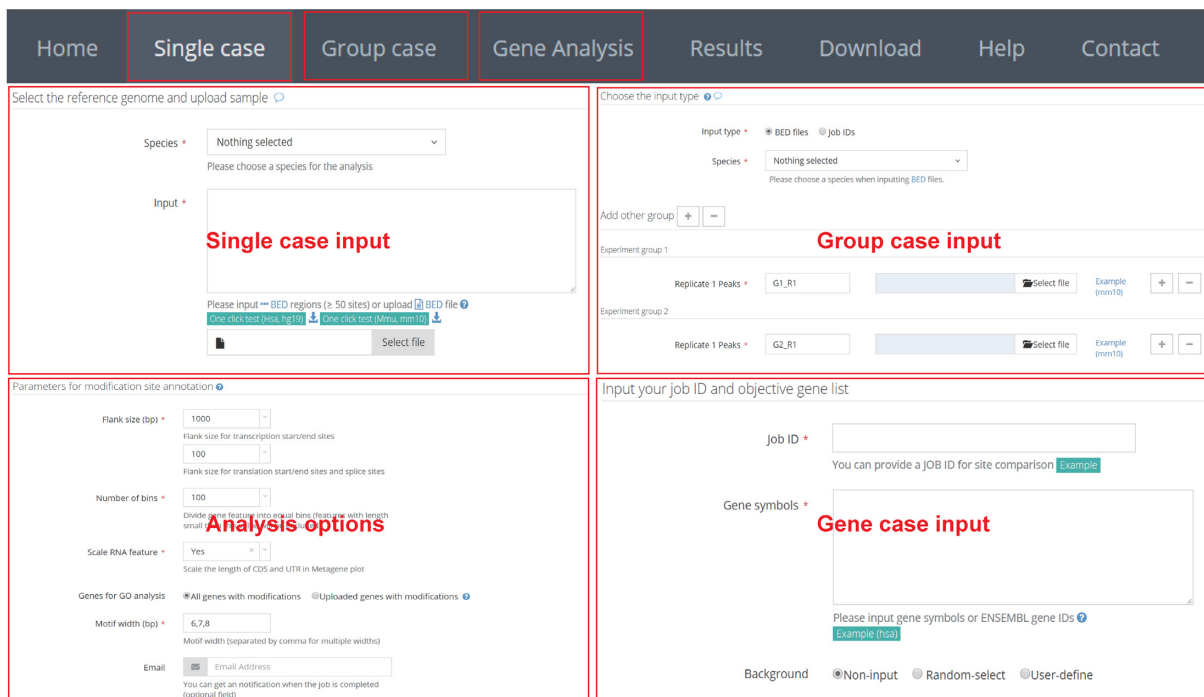
RNAmoD provides users several parameters to flexibly re-define the mRNA features. The first is flank length interval (upstream and downstream), which can be used to define the transcription site/end site genomic regions, translation start/end site, and splicing site transcript regions. The second is the number of bins in which to divide the gene features for coverage plots and mRNA metagene analyses. The third is the parameter controlling the shape of metagene plot curve. The fourth is the length of the motif in the enriched motif identification. Meanwhile, RNAmoD allows users to submit a group of genes to define the subset of modified genes used to conduct functional annotations. In the group case, an additional parameter is provided to control whether the intersection or union set in each group is used in the annotation. Notably, all input webpages are organized with examples to help users with correct inputs.

## Annotation and analysis workflow

RNAmoD first extracts different gene features from the reference genome annotations, including (i) genomic regions of 5'UTR, CDS, 3'UTR, promoter regions, downstream regions, transcription start sites; (ii) transcript regions of translation start/end site and splicing sites; (iii) gene characteristics of GC content, length and minimum free energy. Next, based on extracted gene features, RNAmoD performs various calculations based on the genomic locations of uploaded modification sites, including the intersection with different gene biotypes, intersection with gene features, modification distribution around specific site boundaries, and motif enrichment for the peaks/locations. For metagene analyses for each mRNA feature or the whole mRNA, after excluding the features with length shorter than the bins number, each gene feature is divided into equal size



**Figure 1.** Overall workflow of RNAmod. RNAmod provides three functional modules, including single case, group case, and gene case. The single case module allows users to annotate RNA modifications for a single sample. The group case module allows users to annotate and compare the modification distribution between two samples or even more groups. The gene case analysis module can be used to analyze the modification distribution in the context of specific genes.



**Figure 2.** Inputs of RNAmod. In the single case input, RNAmod utilizes the commonly used BED format of the chromosomal location of RNA modifications as input. In the group case, two types of inputs are supported: BED format files and single case job IDs. In the gene case input, users are required to input the single case or group case job IDs and the object gene list.

bins (100 bins by default) and the average number of sites is counted for each bin. For the coverage distribution near specific site boundaries, including transcription start/end sites, translation start/end site, and splicing site, RNAmoD calculates the average coverage of peaks/sites for each nucleotide both upstream and downstream of the location. For motif analysis, sequences around modification sites and random genomic regions are extracted from constructed BSgenome package using getSeq function in Biostrings package. Then the extracted sequences are used as input for Homer software (47) (<http://homer.ucsd.edu/homer>) to identify enriched motifs (with motifs lengths of 6, 7 and 8 bp set as default). Finally, based on the genes with modifications or the gene subset intersected from a user submitted list, RNAmoD performs various functional annotations, including GO and KEGG pathway for all supported species, and Reactome pathway, Disease Ontology, Network of Cancer Gene, DisGeNET disease genes and MSigDB functional gene sets for human, mouse, rat, zebrafish, fly and C.elegans.

### The output of RNAmoD

The outputs of RNAmoD are presented in intuitive web interfaces, which typically contain the following types of information: (i) The overall statistics of modifications across different RNA features, with the default priority of CDS, 3' UTR, 5' UTR, intron, promoter, and downstream if overlap annotations are found. (ii) The overall statistics of modifications across different RNA biotypes, such as mRNA, miRNA, tRNA, mt-tRNA, and lincRNA. (iii) Plot of modification site coverage at mRNA transcription start/end sites (with 1000 bp flank size as default). (iv) Plot of modification site coverage around translation start/end sites (with 100 bp flank size as default). (v) Plot of modification site coverage at the mRNA 5' and 3' exon/intron splicing junction sites (with 100 bp flank size as default). (vi) The coverage of modification sites among different gene features, including promoter, 5' UTR, CDS, 3'UTR and downstream regions. (vii) The metagene plot of modification sites across whole mRNA. (viii) The top five enriched motifs for modification sites. (ix) The density heatmaps of modifications across transcription start/end site regions. (x) The density heatmaps of modifications across translation start/end site regions. (xi) The GO functional annotation barplots and detailed list of enriched GO terms. (xii) The pathway functional annotation barplots and detailed list of enriched pathways. (xiii) Detailed lists of modified genes and modification sites. (xiv) JBrowse integration for the visualization of modification sites in comparison to known RNA modifications and RBP binding sites. In the coverage plots, 95% confidence interval (mean  $\pm$  standard error of the mean  $\times$  1.96) is also shown. In the group-case study, besides the basic statistics and site annotation list for each group, the outputs also contain statistics for modification distribution comparison and the overlapping modification sites between groups. In the gene-case study, instead of the comparison between groups, the comparison between input genes and other background genes will be performed, with the output results shown in a similar format to that in the group case. All of the results are shown in interactive tables and

figures on the web page, and these are all available for downloading as different formats (Figure 3). It is also noted that RNAmoD provides users different ways to show the distribution charts and provides a link to download all the results.

### Implementation

The web server is hosted within PHP/Apache environment under a Linux system and is equipped with two Octa-cores AMD processors (2.6 GHz each) and 64GB of RAM. The back-end pipeline is implemented in the Python/Perl language, and the plots are drawn by R (<http://www.r-project.org>). Several JavaScript libraries, such as JQuery, Bootstrap.js, DataTable.js, and Highchart.js libraries, were used for creating dynamic and interactive data visualization web browser interfaces, which provide users with a highly intuitive interface for manipulating the tool and viewing the analysis results.

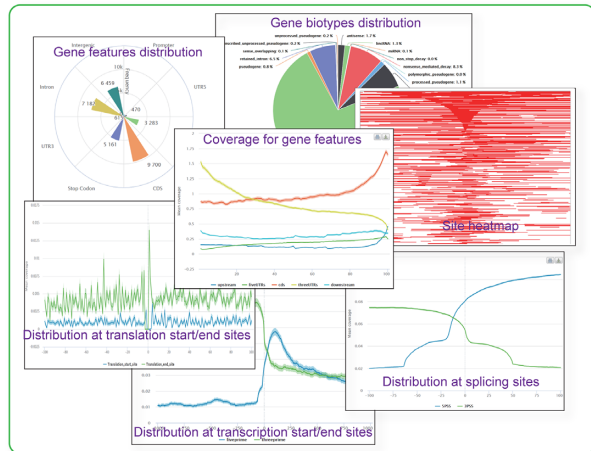
### Comparison with other web servers

There are several web servers that are based on machine-learning technologies for identifying diverse mRNA modifications, especially m<sup>6</sup>A (Supplementary Table S1). For example, SRAMP (19) uses a machine-learning framework to predict mammalian m<sup>6</sup>A sites by extracting and integrating the sequence and structural features around m<sup>6</sup>A sites. HAMR (48) is a method to detect and differentiate different classes of modifications in RNA-seq data based on the principle that RNA modification site can affect RT-activity that leads to sequence mismatches or truncations. RNAm5Cfinder (49) is a web-server that uses Random Forest and RNA sequence features to predict RNA m<sup>5</sup>C modification sites in mouse and human. However, there is still a lack of a convenient and comprehensive web server for the downstream annotation of mRNA modifications. Although some tools or databases provide modification annotation functions (Supplementary Table S1), their functionality is still limited. For instance, RMBase (26) is a comprehensive database for RNA modifications and their associations with post-transcriptional regulation and it provides a convenient web tool—modTools (modMetagene and modAnnotation) (26) to annotate modification sites. modMetagene can be used to draw the metagene plot while modAnnotation displays the table of modification site with gene annotations. Both available annotation tools pay more attention to conducting meta-gene analysis and basic gene annotations.

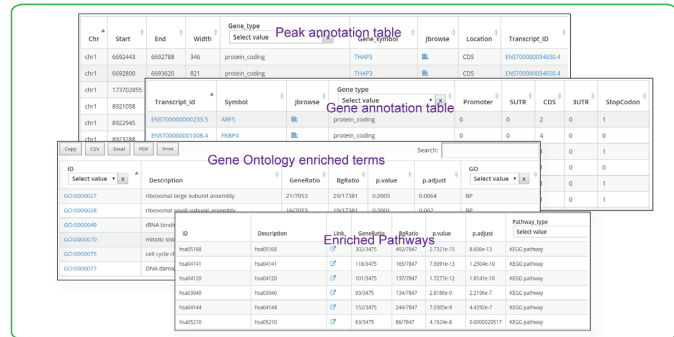
### CASE STUDIES

RNAmoD was used to annotate five mRNA modifications, including N<sup>6</sup>-methyladenosine (m<sup>6</sup>A), N<sup>1</sup>-methyladenosine (m<sup>1</sup>A), 5-methylcytosine (m<sup>5</sup>C), pseudouridine ( $\varphi$ ) and N<sup>4</sup>-acetylcytidine (ac<sup>4</sup>C) modifications (Supplementary Figure S1). The results showed that the m<sup>6</sup>A sites are mainly located within the CDS and UTRs, with a density peak near the translation end sites (stop codons). Other modifications show similar distribution of sites with enrichment near the translation start sites (start codons), especially ac<sup>4</sup>C modification, which is in accordance with their reported roles in regulating translation. In contrast, a large proportion of

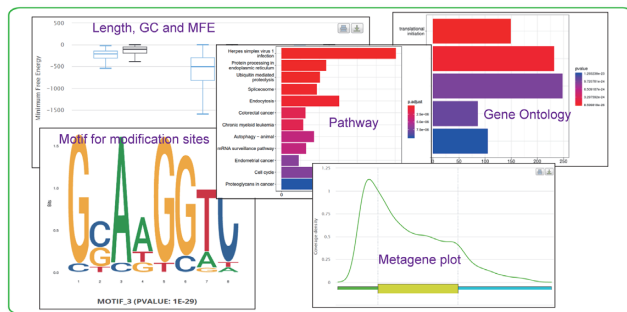
## Modification site distribution



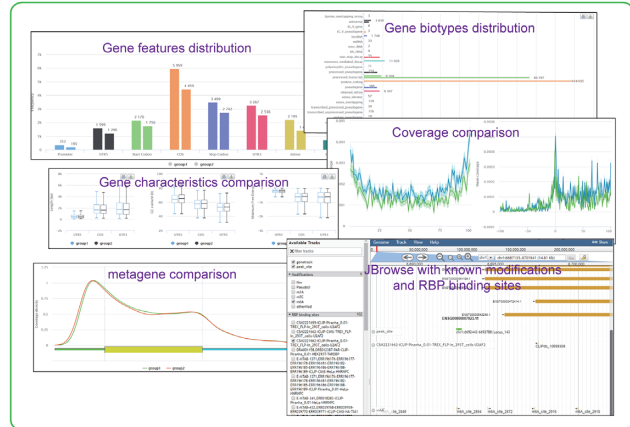
## Gene and site annotation details



## Gene characteristic statistics



## Modification sites comparison



**Figure 3.** Screenshots of the RNAmod outputs. The typical outputs contain four types of information: the distribution of modification sites, statistics of gene characteristics, gene and site annotation details, and comparison of modification sites.

pseudouridine modifications are located in 3'UTR regions, which may indicate specific functions in mRNA regulation.

RNAmod can also be used to conveniently annotate RBP binding sites and analyze their binding patterns. Insulin-like growth factor 2 mRNA-binding proteins (IGF2BPs) have been identified as  $m^6A$  readers that can guard thousands of mRNA transcripts from decay (50). RNAmod annotation of IGF2BP3 (a member of IGF2BP gene family) binding sites showed that it mainly binds to 3'UTR regions of protein coding genes (Supplementary Figure S2A, S2B and S2G) with a GGATG motif (Supplementary Figure S2H). The coverage plot, metagene plot, and heatmaps all show that the binding sites showed obvious peak density near the stop codon (Supplementary Figure S2C, S2E and S2I). Differential binding was observed at exons and intron at splice sites (Supplementary Figure S2D). The gene characteristics comparison between IGF2BP3 binding genes and background genes indicates that the 3'UTR of binding genes have longer length and less minimum free energy. The visualization of a specific mRNA example using Jbrowse further showed that the IGF2BP3 binding sites are in close proximity to the known  $m^6A$  modification site (Supplementary Figure S2J).

### Perspective

High-throughput sequencing greatly facilitates RNA epitranscriptomics studies and offers an effective method

to comprehensively investigate mRNA modification in genome scale. Meanwhile, many software or web servers are developed to predict mRNA modification using machine-learning technologies. However, the convenient and integrated annotation of these mRNA modifications from large amounts of data still remains a challenge. Therefore, we developed an automated and easy-to-use web service, RNAmod, for research communities to analyze and annotate mRNA modifications. Currently, RNAmod supports 21 model reference genomes across vertebrates, insects, nematodes and plants. More species will be supported in the further. Besides, the data in RNAmod will be updated regularly to keep up with the source databases.

Meanwhile, RNAmod can also be used for the annotation and visualization of CLIP-Seq data to study the binding features of RNA binding proteins (RBPs). We used RNAmod to annotate the known RNA modification sites and nearly 500 RBP binding sites for human, mouse, fly, yeast, zebrafish, Arabidopsis, etc., whose results are availed in the download page of RNAmod. In addition, RNAmod allows users to perform comparative analysis between groups or among specific gene contexts. Overall, RNAmod is a one stop online interactive platform for the annotation, analysis, and visualization of mRNA modifications. It will meet the emerging need for a convenient, and comprehensive analysis tool for the research field.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

## ACKNOWLEDGEMENTS

*Author contributions:* Q.L and R.I.G. conceived the web server. Q.L. designed the bioinformatic pipeline and constructed the web server. Q.L. and R.I.G. wrote the manuscript.

## FUNDING

US National Institute of General Medical Sciences [R01GM086386]; National Cancer Institute [R01CA211328]. Funding for open access charge: National Cancer Institute [R01CA211328].

*Conflict of interest statement.* None declared.

## REFERENCES

- Saletore, Y., Meyer, K., Korlach, J., Vilfan, I.D., Jaffrey, S. and Mason, C.E. (2012) The birth of the Epitranscriptome: deciphering the function of RNA modifications. *Genome Biol.*, **13**, 175.
- Davalos, V., Blanco, S. and Esteller, M. (2018) SnapShot: messenger RNA modifications. *Cell*, **174**, 498–498.
- Shi, H., Wei, J. and He, C. (2018) Where, When, and How: Context-Dependent Functions of RNA Methylation Writers, Readers, and Erasers. *Mol. cell*, **74**, 640–650.
- Meyer, K.D. and Jaffrey, S.R. (2017) Rethinking m(6A) Readers, Writers, and Erasers. *Annu. Rev. Cell Dev. Biol.*, **33**, 319–342.
- Choe, J., Lin, S., Zhang, W., Liu, Q., Wang, L., Ramirez-Moya, J., Du, P., Kim, W., Tang, S., Sliz, P. et al. (2018) mRNA circularization by METTL3-eIF3h enhances translation and promotes oncogenesis. *Nature*, **561**, 556–560.
- Vu, L.P., Pickering, B.F., Cheng, Y., Zaccara, S., Nguyen, D., Minuesa, G., Chou, T., Chow, A., Saletore, Y., MacKay, M. et al. (2017) The N(6)-methyladenosine (m(6A))-forming enzyme METTL3 controls myeloid differentiation of normal hematopoietic and leukemia cells. *Nat. Med.*, **23**, 1369–1376.
- Frye, M. and Blanco, S. (2016) Post-transcriptional modifications in development and stem cells. *Development*, **143**, 3871–3881.
- Hausmann, I.U., Bodi, Z., Sanchez-Moran, E., Mongan, N.P., Archer, N., Fray, R.G. and Soller, M. (2016) m(6A) potentiates Sxl alternative pre-mRNA splicing for robust *Drosophila* sex determination. *Nature*, **540**, 301–304.
- Zhao, B.S., Wang, X., Beadell, A.V., Lu, Z., Shi, H., Kuuspalu, A., Ho, R.K. and He, C. (2017) m(6A)-dependent maternal mRNA clearance facilitates zebrafish maternal-to-zygotic transition. *Nature*, **542**, 475–478.
- Lin, S., Choe, J., Du, P., Triboulet, R. and Gregory, R.I. (2016) The m(6A) Methyltransferase METTL3 promotes translation in human cancer cells. *Mol. Cell*, **62**, 335–345.
- Safra, M., Sas-Chen, A., Nir, R., Winkler, R., Nachshon, A., Bar-Yaacov, D., Erlacher, M., Rossmann, W., Stern-Ginossar, N. and Schwartz, S. (2017) The m1A landscape on cytosolic and mitochondrial mRNA at single-base resolution. *Nature*, **551**, 251–255.
- 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.
- Arango, D., Sturgill, D., Alhusaini, N., Dillman, A.A., Sweet, T.J., Hanson, G., Hosogane, M., Sinclair, W.R., Nanan, K.K., Mandler, M.D. et al. (2018) Acetylation of Cytidine in mRNA promotes translation efficiency. *Cell*, **175**, 1872–1886.
- Zhang, L.-S., Liu, C., Ma, H., Dai, Q., Sun, H.-L., Luo, G., Zhang, Z., Zhang, L., Hu, L., Dong, X. et al. (2019) Transcriptome-wide mapping of internal N<sup>7</sup>-methylguanosine methylome in mammalian mRNA. *Mol. Cell*, doi:10.1016/j.molcel.2019.03.036.
- 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.
- 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.
- Lin, S., Liu, Q., Lelyveld, V.S., Choe, J., Szostak, J.W. and Gregory, R.I. (2018) Mett1/Wdr4-Mediated m(7)G tRNA methylome is required for normal mRNA translation and embryonic stem cell self-renewal and differentiation. *Mol. Cell*, **71**, 244–255.
- Zhang, Q., Chen, K., Wu, X., Wei, Z., Rong, R., Lu, Z., Meng, J., de Magalhães, J.P., Su, J. and Rigden, D.J. (2019) WHISTLE: A high-accuracy map of the human N6-methyladenosine (m6A) epitranscriptome predicted using a machine learning approach. *Nucleic Acids Res.*, **47**, e41.
- Zhou, Y., Zeng, P., Li, Y.H., Zhang, Z. and Cui, Q. (2016) SRAMP: prediction of mammalian N6-methyladenosine (m6A) sites based on sequence-derived features. *Nucleic Acids Res.*, **44**, e91.
- Qiang, X., Chen, H., Ye, X., Su, R. and Wei, L. (2018) M6AMRFS: robust prediction of N6-methyladenosine sites with sequence-based features in multiple species. *Front. Genet.*, **9**, 495.
- Chen, W., Xing, P. and Zou, Q. (2017) Detecting N(6)-methyladenosine sites from RNA transcriptomes using ensemble Support Vector Machines. *Sci. Rep.*, **7**, 40242.
- Zhang, Y. and Hamada, M. (2018) DeepM6ASeq: prediction and characterization of m6A-containing sequences using deep learning. *BMC Bioinformatics*, **19**, 524.
- Evers, M., Shafik, A., Schumann, U. and Preiss, T. (2016) RNAModR: functional analysis of mRNA modifications in R. bioRxiv doi: <https://doi.org/10.1101/080051>, 10 October 2016, preprint: not peer reviewed.
- Cui, X., Wei, Z., Zhang, L., Liu, H., Sun, L., Zhang, S.W., Huang, Y. and Meng, J. (2016) Guitar: An R/Bioconductor package for gene annotation guided transcriptomic analysis of RNA-related genomic features. *BioMed Res. Int.*, **2016**, 8367534.
- Uyar, B., Yusuf, D., Wurmus, R., Rajewsky, N., Ohler, U. and Akalin, A. (2017) RCAS: an RNA centric annotation system for transcriptome-wide regions of interest. *Nucleic Acids Res.*, **45**, e91.
- Xuan, J.J., Sun, W.J., Lin, P.H., Zhou, K.R., Liu, S., Zheng, 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.
- Zhang, Y., Liu, T., Meyer, C.A., Eeckhoutte, J., Johnson, D.S., Bernstein, B.E., Nussbaum, C., Myers, R.M., Brown, M., Li, W. et al. (2008) Model-based analysis of ChIP-Seq (MACS). *Genome Biol.*, **9**, R137.
- Meng, J., Lu, Z., Liu, H., Zhang, L., Zhang, S., Chen, Y., Rao, M.K. and Huang, Y. (2014) A protocol for RNA methylation differential analysis with MeRIP-Seq data and exomePeak R/Bioconductor package. *Methods*, **69**, 274–281.
- Spyrou, C., Stark, R., Lynch, A.G. and Tavare, S. (2009) BayesPeak: Bayesian analysis of ChIP-seq data. *BMC Bioinformatics*, **10**, 299.
- Zerbino, D.R., Achuthan, P., Akanni, W., Amode, M.R., Barrell, D., Bhai, J., Billis, K., Cummins, C., Gall, A., Giron, C.G. et al. (2018) Ensembl 2018. *Nucleic Acids Res.*, **46**, D754–D761.
- Lawrence, M., Huber, W., Pages, H., Aboyoun, P., Carlson, M., Gentleman, R., Morgan, M.T. and Carey, V.J. (2013) Software for computing and annotating genomic ranges. *PLoS Comput. Biol.*, **9**, e1003118.
- Lorenz, R., Bernhart, S.H., Honer, Z., Siederdisen, C., Tafer, H., Flamm, C., Stadler, P.F. and Hofacker, I.L. (2011) ViennaRNA Package 2.0. *Algorithms Mol. Biol.*: *AMB*, **6**, 26.
- Yu, G., Wang, L.G., Han, Y. and He, Q.Y. (2012) clusterProfiler: an R package for comparing biological themes among gene clusters. *OmicS*, **16**, 284–287.
- Yu, G. and He, Q.Y. (2016) ReactomePA: an R/Bioconductor package for reactome pathway analysis and visualization. *Mol. Biosyst.*, **12**, 477–479.
- Yu, G., Wang, L.G., Yan, G.R. and He, Q.Y. (2015) DOSE: an R/Bioconductor package for disease ontology semantic and enrichment analysis. *Bioinformatics*, **31**, 608–609.

36. Liberzon, A., Birger, C., Thorvaldsdottir, H., Ghandi, M., Mesirov, J.P. and Tamayo, P. (2015) The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst.*, **1**, 417–425.
37. 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.
38. Zhu, Y., Xu, G., Yang, Y.T., Xu, Z., Chen, X., Shi, B., Xie, D., Lu, Z.J. and Wang, P. (2019) POSTAR2: deciphering the post-transcriptional regulatory logics. *Nucleic Acids Res.*, **47**, D203–D211.
39. 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.
40. Leinonen, R., Sugawara, H. and Shumway, M. (2011) The sequence read archive. *Nucleic Acids Res.*, **39**, D19–D21.
41. Kim, D., Langmead, B. and Salzberg, S.L. (2015) HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods*, **12**, 357–360.
42. 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.
43. Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R. and Salzberg, S.L. (2013) TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.*, **14**, R36.
44. 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.
45. Cui, X., Meng, J., Zhang, S., Chen, Y. and Huang, Y. (2016) A novel algorithm for calling mRNA m6A peaks by modeling biological variances in MeRIP-seq data. *Bioinformatics*, **32**, i378–i385.
46. 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.
47. 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.
48. Ryvkin, P., Leung, Y.Y., Silverman, I.M., Childress, M., Valladares, O., Dragomir, I., Gregory, B.D. and Wang, L.S. (2013) HAMR: High-throughput annotation of modified ribonucleotides. *RNA*, **19**, 1684–1692.
49. Li, J., Huang, Y., Yang, X. and Zhou, Y. (2018) RNAm5Cfinder: a web-server for Predicting RNA 5-methylcytosine (m5C) sites based on random forest. *Sci. Rep.*, **8**, 17299.
50. Huang, H., Weng, H., Sun, W., Qin, X., Shi, H., Wu, H., Zhao, B.S., Mesquita, A., Liu, C., Yuan, C.L. *et al.* (2018) Recognition of RNA N(6)-methyladenosine by IGF2BP proteins enhances mRNA stability and translation. *Nat. Cell Biol.*, **20**, 285–295.