D1396–D1404 *Nucleic Acids Research, 2021, Vol. 49, Database issue* *Published online 3 October 2020*
doi: 10.1093/nar/gkaa790

RMDisease: a database of genetic variants that affect RNA modifications, with implications for epitranscriptome pathogenesis

Kunqi Chen [1,2,†], Bowen Song [3,4,†], Yujiao Tang [1,3,†], Zhen Wei [1,3,*], Qingru Xu [1], Jionglong Su [4], João Pedro de Magalhães [2], Daniel J. Rigden [3] and Jia Meng [1,3,5,*]

[1]Department of Biological Sciences, Xi'an Jiaotong-Liverpool University, Suzhou, Jiangsu 215123, China, [2]Institute of Ageing & Chronic Disease, University of Liverpool, L7 8TX Liverpool, UK, [3]Institute of Systems, Molecular and Integrative Biology, University of Liverpool, L7 8TX Liverpool, UK, [4]Department of Mathematical Sciences, Xi'an Jiaotong-Liverpool University, Suzhou, Jiangsu 215123, China and [5]AI University Research Centre, Xi'an Jiaotong-Liverpool University, Suzhou, Jiangsu 215123, China

Received July 20, 2020; Revised September 08, 2020; Editorial Decision September 09, 2020; Accepted September 11, 2020

ABSTRACT

Deciphering the biological impacts of millions of single nucleotide variants remains a major challenge. Recent studies suggest that RNA modifications play versatile roles in essential biological mechanisms, and are closely related to the progression of various diseases including multiple cancers. To comprehensively unveil the association between disease-associated variants and their epitranscriptome disturbance, we built RMDisease, a database of genetic variants that can affect RNA modifications. By integrating the prediction results of 18 different RNA modification prediction tools and also 303,426 experimentally-validated RNA modification sites, RMDisease identified a total of 202,307 human SNPs that may affect (add or remove) sites of eight types of RNA modifications (m^6A, m^5C, m^1A, m^5U, Ψ, m^6Am, m^7G and Nm). These include 4,289 disease-associated variants that may imply disease pathogenesis functioning at the epitranscriptome layer. These SNPs were further annotated with essential information such as post-transcriptional regulations (sites for miRNA binding, interaction with RNA-binding proteins and alternative splicing) revealing putative regulatory circuits. A convenient graphical user interface was constructed to support the query, exploration and download of the relevant information. RMDisease should make a useful resource for studying the epitranscriptome impact of genetic variants via multiple RNA modifications with emphasis on their potential disease relevance. RMDisease is freely accessible at: www.xjtlu.edu.cn/biologicalsciences/rmd.

INTRODUCTION

With the advances in the high-throughput sequencing technique, millions of single nucleotide polymorphisms (SNPs) have been identified from multiple species and in multiple human cancers, suggesting their critical roles in diverse biological functions and human health. However, deciphering if and how SNPs lead to functional changes is still a major challenge. Even synonymous SNPs, which do not change the amino acid sequence and so are sometimes considered 'silent' mutations, can still play critical roles during transcriptional and post-transcriptional regulation (1), such as changing splicing sites (2), influencing RNA-protein interactions (3) and alteration of RNA secondary structures.

Substantial efforts have been made to relate the genetic variants to their immediate biological consequences, including with regard to transcriptional regulation (4,5), post-transcriptional protein modification (6–12), RNA–protein interaction (13), calpain cleavage (14), ceRNA networks (15), polyadenylation (16) and RNA modifications (17,18). Most of these works were based on a widely adopted computational framework, i.e. a machine learning model is firstly trained to capture the characteristics of a specific type of epigenetic mark (or interaction) from gold standard experimental datasets. With the model, it is then possible to assess the probability of the mark being associated with a given (DNA, RNA or protein) sequence, and further, predicting whether a genetic mutation can affect the status of epigenetic mark by comparing the probabilities of the mark being associated with the original and the mutated se-

*To whom correspondence should be addressed. Tel: +86 512 81880492; Fax: +86 512 88161899; Email: jia.meng@xjtlu.edu.cn
Correspondence may also be addressed to Zhen Wei. Email: zhen.wei01@xjtlu.edu.cn
†The authors wish it to be known that, in their opinion, the first three authors should be regarded as Joint First Authors.

quences. Importantly, this analysis not only explains how a genetic variant regulates an epigenetic mark, but also helps explain the phenotypes associated with the SNP, as those identified from GWAS analysis. For example, if a SNP is known to increase the risk of a disease, and can also destroy a transcription factor binding site, it is often natural to speculate a transcription-related disease mechanism, even though the two may not be directly causal, and additional experimental validation is necessary.

The epitranscriptome has emerged as an important layer for gene expression regulation (19,20). Recent studies suggest that various RNA modifications occur widely in the transcriptome, play versatile roles in essential biological mechanisms, and are closely related to the progression of various diseases (21–24). For example, N^6-methyladenosine (m^6A) controls the speed of the circadian clock (25), affects RNA stability (26) and regulates the progression of multiple cancers (27–30). N^4-acetylcytidine (ac4C) promotes translation efficiency (31), and 2′-*O*-methylation (Nm) of HIV transcripts help the virus to avoid innate immune sensing (32).

A number of high-throughput approaches have been developed for profiling the transcriptome-wide distribution of different types of RNA modifications, including m^6A-seq (or MeRIP-seq) (33,34), PA-m^6A-seq (35), miCLIP (36) and m^6A-CLIP (37). These approaches have generated a large number of high-quality epitranscriptome datasets, on the basis of which the properties of the modification-carrying RNA sequences can be characterized. This enables the prediction of RNA modification sites from the primary sequences (38,39), and hence allows prediction of the impact of genetic variants on RNA modification by comparing the potential for modification of the original and the mutated sequences. iRNA-methyl (40) and SRAMP (41) are two of the earliest and most widely adopted approaches for predicting RNA methylation sites from the primary RNA sequences. We previously developed a high-accuracy predictor WHISTLE (42) for prediction of RNA methylation sites by taking advantage of conventional sequence features as well as 35 additional genomic features (43).

There exist several databases of RNA modifications with different focuses. The MODOMICS database concerns mainly RNA modification pathways. MeT-DB (44), CVm6A (45) and REPIC (46) collect and annotate the transcriptome m^6A sites under different experimental conditions; while RMBase is currently the most comprehensive database containing well annotated sites of multiple types of RNA modifications identified in multiple species. m^6AVar (17) and m7GDiseaseDB (18) contain disease-associated SNPs that can affect m^6A and internal m^7G RNA modification, respectively. Both were based on their respective customized RNA modification site prediction tools (m6AFinder and m7GFinder). These efforts together greatly facilitated research into RNA modifications. However, to the best of our knowledge, the impact of genetic mutation on most transcriptome modifications such as m^5C, Ψ and Nm, has not been studied, and a centralized platform is not yet available for systematically deciphering the association of genetic variants with multiple RNA modification types as predicted by multiple independent tools.

To address this, we present here RMDisease, a database of genetic variants that affect RNA modifications with a focus on their potential disease association. By integrating 303,426 RNA modification sites, 40,915,548 somatic and germline SNPs and 18 prediction tools, RMDisease represents the most comprehensive available mapping from genome variants to their epitranscriptome disturbance. RMdisease contains a total of 202,307 human SNPs that can effect (add or remove) eight types of RNA modifications (m^6A, m^5C, m^1A, m^5U, Ψ, m^6Am, m^7G and Nm), including 4,289 disease-associated variants that may imply disease pathogenesis functioning at the epitranscriptome layer. Additionally, the RNA modification-affecting SNPs were further annotated with putative post-transcriptional machinery including RNA-binding protein (RBP) binding sites, miRNA targets and splicing sites. A graphical user interface was constructed to support the query, exploration and download of the database. The overall design of RMDisease is summarized in Figure 1.

MATERIALS AND METHODS

Data resource

We considered in this study only the RNA modifications that widely occur in the transcriptome. Since there is not yet available a relatively complete high-confidence collection of such data, we manually collected from 32 studies the sites of eight types of RNA modifications, including m^6A (178 049 sites), m^5C (95 391 sites), m^1A (16 346 sites), m^5U (3696 sites), Ψ (3137 sites), m^6Am (2447 sites), m^7G (2525 sites) and Nm (1835 sites), respectively. These sites were reported from 68 high-throughput sequencing experiments generated by 18 base-resolution technologies, including m6A-REF-seq (47), MAZTER-seq (48), miCLIP (36), m6A-CLIP-seq (37), PA-m6A-seq (35), Ψ-seq (49), Pseudo-seq (50), CeU-Seq (51), RBS-Seq (52), m1A-MAP (53), m1A-seq (54), Aza-IP (55), RNA-BisSeq (56), FICC-Seq (57), Nm-seq (58), m7G-seq (59), m7G-miCLIP-Seq (60). Detailed information regarding these sequencing samples is provided in Supplementary Table S1.

We obtained 3 820 716 somatic variants and 37 094 832 germline variants from dbSNP (v151) and TCGA (v15.0), respectively (see Supplementary Table S3). Only the variants located within the mature transcripts were kept for further analysis.

Derivation of RNA modification-associated variants

The RNA modification-associated SNPs (RM-SNPs) are defined as SNPs that may lead to the gain or loss of an RNA modification site, as reported by the prediction tools via comparing the modification status of the original and the mutated sequences. RM-SNPs were further classified into 3 sub-groups based on their reliability, including: (i) high: a SNP directly alters the experimentally validated RNA modification site, leading to its loss; (ii) medium: a SNP alters a nucleotide within the 41 bp flanking window of an experimentally validated RNA modification site (but not directly the modifiable nucleotide itself), causing its loss as predicted by a machine learning model; (iii) Low: a SNP alters a nucleotide within the 41 bp flanking window of an

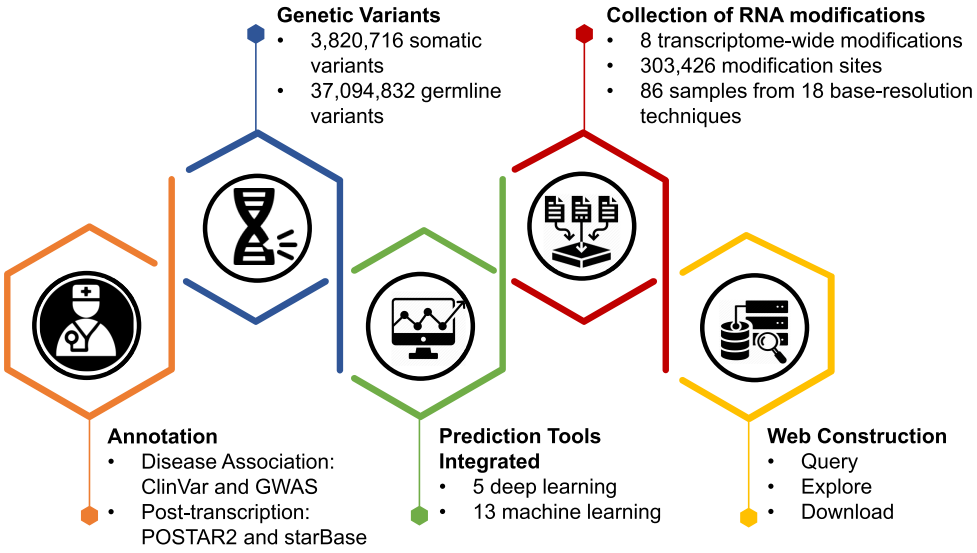

Figure 1. The overall design of RMDisease. RMDisease integrates 303,426 high-confidence RNA modification sites experimentally detected by 18 base-resolution technologies and 18 independent *in silico* machine learning tools to evaluate systematically the potentials of somatic and germline variants to affect eight types of transcriptome modifications. Disease association and post-transcriptional regulations were further integrated to unveil potential epitranscriptome pathogenesis and putative regulatory machinery. A graphical user interface was constructed to support the query, exploration and download of the database.

RNA modification site (may or may not directly the modifiable nucleotide itself), causing significant increase or decrease in the probability of RNA modification. It may be worth noting that the identification of RM-SNPs of high or medium confidence level both require experimentally validated RNA modification sites in the very beginning.

We integrated 18 prediction tools developed for eight RNA modifications to perform comprehensive and independent evaluations, including five deep learning-based methods: DeepPromise (38), DeepM6Aseq (61), DeepM-RMD (62), iPseU-CNN (63), Deep-2′-O-Me (64) and 13 machine learning-based methods: WHISTLE(65), SRAMP(66), iRNA-3type (67), iMRM (68), iRNA-2OM (69), iRNA-m5C (70), RAMPred (71), iRNA-PseColl (72), iRNA-m7G (73), m7GFinder (74), iRNA-PseU (75), PIANO (76), ISGm1A (77).

The association level (AL) between SNP and an RNA modification is calculated as follows:

$$\mathrm{AL} = \begin{cases} 2P_{SNP} - 2\max(0.5,\ P_{WT}) \text{ for gain} \\ 2P_{WT} - 2\max(0.5,\ P_{SNP}) \text{ for loss} \end{cases} \quad (1)$$

where, P_{WT} and P_{SNP} represent the probability of modification for the wild type and mutated sequences, respectively, as obtained from individual prediction tool (or experiment data if available). The association level (AL) ranges from 0 to 1, with 1 indicating the greatest impact on RNA modification. The statistical significance was assessed by comparing to the ALs of all mutations, with which the upper bound of the *P*-value can be calculated. Different prediction tools were employed for the same modification to obtain the association level for their corresponding target modification. As shown in previous studies, the RNA modification prediction tools that integrate both the sequence and genome-derived features outperform those based on sequence features only (38,65,74,76–77), and were thus used as the primary method. We retained the RM-SNPs with association level >0.4 and *P*-value <0.05 as predicted by approaches that took advantage of both sequence and genome-derived features. The results obtained from other methods (based on sequence only) were provided in RMDisease as well for reference purpose (see Supplementary Table S2).

Additional annotation

To annotate the basic genomic information of RNA modification-associated variants, the transcript structure from UCSC genome browser including CDS, 3′UTR, 5′UTR, start codon and stop codon, etc. were used, and the genomic conservation were annotated by the phastCons 60-way. In addition, the deleterious level of each RNA modification-associated variant was analyzed by SIFT (78), PolyPhen2 HVAR (79), PolyPhen2HDIV (79), LRT (80) and FATHMM (81) using the ANNOVAR package (82). Furthermore, to provide annotation of putative post-transcriptional regulatory machinery, information on RBP binding sites from POSTAR2 (83), miRNA-RNA interaction from miRanda (84) and startBase2 (85), and splicing sites from UCSC annotation with GT-AG role within 100 bp upstream and downstream of RNA modification-associated variants was integrated into the database. To unveil potentially epitranscriptome-related pathogenesis, the association between disease and RNA modification-associated variants was extracted from GWAS catalog (86), ClinVar (87) and Johnson and O'Donnel's data (88). Additionally, the RNA molecule–drug sensitivity associations from RNAactDrug (89) were obtained to provide a drug suggestion for each RNA modification-associated SNP.

Database and web interface implementation

MySQL tables were used for the storage and management of the metadata in RMDisease. Hyper Text Markup Language (HTML), Cascading Style Sheets (CSS) and Hyper-

text Preprocessor (PHP) were applied in the construction of web interfaces. The multiple statistical diagrams were created by EChars and the genome browser was implemented using Jbrowse (90) for the exploration of all the analysis results.

RESULTS

Database content

We firstly evaluated the potentials of genetic variants to add or remove an RNA modification site directly or indirectly. In the end, a total of 57 622, 23 463, 61 563, 23 875, 24 822, 640, 5047 and 5275 genetic variants were found to be associated with m^6A, m^1A, m^5C, Ψ, m^7G, Nm, m^5U and m^6Am, respectively, providing so far the most comprehensive map of genetic factors of epitranscriptome disturbance (Table 1).

We then obtained disease annotations of SNPs from GWAS catalog, Johnson and O'Donnel's data, and ClinVar, and mapped them to RM-SNPs. These SNPs may link disease pathogenesis and clinical relevance to epitranscriptome regulations (Table 2). We summarized in Table 3 the diseases that are associated with the most RM-SNPs of a specific RNA modification type.

We then asked whether the RM-associated SNPs are more functional relevant to important biological events compared to non-associated SNPs, and used evolutionary conservation as an indicator. For this purpose, the phastCons 100-way conservation scores from UCSC was considered to evaluate the conservation of individual site, which was calculated for human genome derived from genome-wide multiple alignments with 99 other vertebrate species. Interestingly, we found that the RM-associated variants were more conserved than non-associated variants (Figure 2), suggesting that the RM-associated SNPs underwent stronger selection pressure than the other variants, and may be related to important biological events that can be regulated at the epitranscriptome layer.

Website interface and usage

The user-friendly web interfaces provided in RMDisease enable the search, browse and download RNA modification associated-SNPs by modification type, gene, disease, chromosome region, RsID and post-transcriptional regulations. A genome browser was integrated for interactive exploration of genome regions of interest. All data provided in the RMDisease database can be freely downloaded. For the convenience of users, detailed instructions on how to use RMDisease were placed in the 'help' page. RMDisease is freely accessible at: www.xjtlu.edu.cn/biologicalsciences/rmd.

Case study: MATR3

MATR3 provides structural support for the nucleus and aids in several important nuclear functions. A mutation on MATR3 with Rs ID: rs185734839 at chr5:138 665 490 is known to be associated with 'Distal myopathy' from GWAS study according to the ClinVar database. As an anonymous SNP located on 3′ untranslated region, this mutation doesn't affect the encoded protein sequence; however, it can directly destroy a known m^6A RNA methylation site located at the same position, which was previously detected by two MAZTER-seq experiments in human ESC cell line (48). Post-transcriptional annotations suggest that, the m^6A site eliminated by rs185734839 falls within the target regions of RNA binding protein CSTF2 and four microRNAs (miR-24, miR-1, miR-206 and miR-613), which provided potential functional circuits of the RNA methylation. It should be of interests to explore whether the methylation status of MATR3 can significantly affect its biological functionality, especially with respect to the disease, RBP and miRNAs mentioned previously. Additional case studies were provided in the Supplementary Materials and Supplementary Table S4.

CONCLUSIONS

An increasing number of biological mechanisms and disease mechanisms have been associated with the epitranscriptome, which consists of more than 100 different types of RNA modifications (91) at tens of thousands of locus in the human transcriptome. To systematically unveil the linkage between genetic factors and their respective epitranscriptome disturbance, we developed RMDisease, a database of genetic variants with potentials to alter eight types of widely spread transcriptome modifications, with emphasis on epitranscriptome disease pathogenesis. RMDisease revealed for the first time the impacts of genetic variants on six types of RNA modifications (m^5C, m^1A, m^5U, Ψ, m^6Am and Nm), and offered substantial improvements over existing works for m^6A and m^7G (17,18).

RMDisease and m7GDiseaseDB (92) used the same inference method. Both databases were based on m7GFinder (74), which integrates the sequence as well as the genomic features. The main different between them is that, a different scoring system was implemented in RMDisease (see Materials and Methods section), which penalizes direct mutation of a putative m^6A site and makes the association level (AL) fall within the range of 0 to 1. Additionally, RMDisease integrated the results obtained from another m^7G predictor iRNA-m7G (93), and use it as an independent reference. There exists major difference between m6AVar (17) and RMDisease for m^6A-associated SNPs. Besides the aforementioned differences between m7GDiseaseDB and RMDisease, i.e. a different scoring framework and extra independent tools integrated, RMDisease also provides the statistical significance of the predicted associations, and was based on more accurate m^6A predictor WHISTLE (42) and with more reliable epitranscriptome datasets integrated. A total of 12 325 m^6A sites that can be affected by SNPs were found to be shared between RMDisease and m^6AVar (see Supplementary Figure S1 for more details).

Previous studies have shown that there exist snoRNPs that can guide the formation of Nm and Psi with the base pairing mechanism (94–98). To the best of our knowledge, none of the existing prediction approaches for RNA modification explicitly considered this mechanism, which may undermine their prediction capability. Indeed, sequence-based predictors for pseudouridine sites without considering the base-paring mechanism between target RNAs and snoRNPs yielded very limited accuracy (lower than 80%)

Table 1. RM-SNPs collected in RMDisease

Modification type	Confidence level	Germline mutation			Somatic mutation			Total		
		Loss	Gain	All	Loss	Gain	All	Loss	Gain	All
m^6A	High	1405	-	1405	4276	-	4276	5681	-	5681
	Medium	13118	-	13118	33666	-	33666	46784	-	46784
	Low	38	654	692	161	4304	4465	199	4958	5157
m^1A	High	104	-	104	61	-	61	165	-	165
	Medium	856	-	856	3134	-	3134	3990	-	3990
	Low	1066	2654	3720	5730	9858	15588	6796	12512	19308
m^5C	High	596	-	596	1518	-	1518	2114	-	2114
	Medium	11350	-	11350	28682	-	28682	40032	-	40032
	Low	1338	3862	5200	6448	7769	14217	7786	11631	19417
Ψ	High	3	-	3	19	-	19	22	-	22
	Medium	412	-	412	1271	-	1271	1683	-	1683
	Low	1133	2726	3859	4324	13987	18311	5457	16713	22170
m^7G	High	10	-	10	82	-	82	92	-	92
	Medium	253	-	253	922	-	922	1175	-	1175
	Low	1511	745	2256	9685	11614	21299	11196	12359	23555
Nm	High	4	-	4	21	-	21	25	-	25
	Medium	85	-	85	530	-	530	615	-	615
	Low	0	0	0	0	0	0	0	0	0
m^5U	High	12	-	12	0	-	0	12	-	12
	Medium	14	-	14	38	-	38	52	-	52
	Low	350	703	1053	433	3497	3930	783	4200	4983
m^6Am	High	12	-	12	2	-	2	14	-	14
	Medium	24	-	24	132	-	132	156	-	156
	Low	220	1598	1818	82	3205	3287	302	4803	5105

Note: RM-SNPs are further classified into two categories: (i) Direct: a SNP directly alters the modifiable nucleotide, leading to the loss of a known or predicted RNA modification site, or alters an non-modifiable nucleotide into one that can be modified. (ii) Indirect (within 41 bp): a SNP alters a nucleotide within the 41 bp flanking window of an RNA modification site (but not directly the modifiable nucleotide itself), causing significant increase or decrease in the probability of RNA modification. We considered only SNPs within the 41bp window for possible indirect effects. This is because most existing RNA modification prediction methods chose to be based on 41bp sequence or less (67–71,73,108). Increasing the length considered here may not help improve the completeness of the results but add substantially the computation load. Additionally, the Nm-SNPs of low confidence were not predicted due to the tremendous search space (Nm can happen to all nucleotide) and its relatively low abundance in the human transcriptome.

Table 2. Disease-associated RM-SNPs collected in RMDisease

Modification type	SNP source	Total RM-SNP	Disease-associated RM-SNPs					
			ClinVar			GWAS		
			SNP	Disease	Gene	SNP	Disease	Gene
m^6A	dbSNP151	15 215	989	400	453	148	77	117
	TCGA	42 407	332	187	164	0	0	0
m^1A	dbSNP151	4680	326	208	247	29	27	29
	TCGA	18 783	217	175	139	0	0	0
m^5C	dbSNP151	17 146	994	397	450	128	67	94
	TCGA	44 417	318	140	130	0	0	0
Ψ	dbSNP151	4274	238	208	207	35	32	35
	TCGA	19 601	51	63	41	0	0	0
m^7G	dbSNP151	2519	183	141	166	25	22	25
	TCGA	22 303	41	63	33	0	0	0
Nm	dbSNP151	89	5	5	5	0	0	0
	TCGA	551	2	18	2	0	0	0
m^5U	dbSNP151	1079	40	40	39	8	8	8
	TCGA	3968	32	44	20	0	0	0
m^6Am	dbSNP151	1854	83	75	77	4	4	4
	TCGA	3421	31	16	27	0	0	0

(61,99), and there exists speculation that the sequence features of pseudouridine sites may not exist at all (100). Incorporating base-pairing information into the prediction models are likely to further improve the prediction performance. Nevertheless, the primary approaches implemented in RMDisease are based on both sequence and genomic features. For pseudouridylation, the PIANO method (101), which was used in RMDisease as the primary prediction approach, has achieved substantially better performance than those based on sequence features only (92,101). One possible explanation is that, since many biological features are correlated, the base-pairing information may be indirectly and inexplicitly captured by the model after including additional genomic features, such as, secondary structure of RNA and genomic conservation. Additionally, our previous studies showed that, including additional ge-

Table 3. Diseases associated with the most RM-SNPs

Disease name	ClinVar study accession	MedGen identifier	Type	#SNP
Hereditary cancer-predisposing syndrome	RCV000129430.4	C0027672	m^6A	58
Hereditary cancer-predisposing syndrome	RCV000129430.4	C0027672	m^1A	14
Hereditary pancreatitis (PCTT)	RCV000468581.1	C0238339	m^5C	88
Hereditary cancer-predisposing syndrome	RCV000129430.4	C0027672	Ψ	12
Cardiovascular phenotype	RCV000249770.1	CN230736	m^7G	7
Adenocarcinoma of lung	RCV000439229.1	C0152013	Nm	1
Adenocarcinoma of lung	RCV000439229.1	C0152013	m^5U	8
Leigh syndrome (LS)	RCV000268982.1	C0023264	m6Am	4

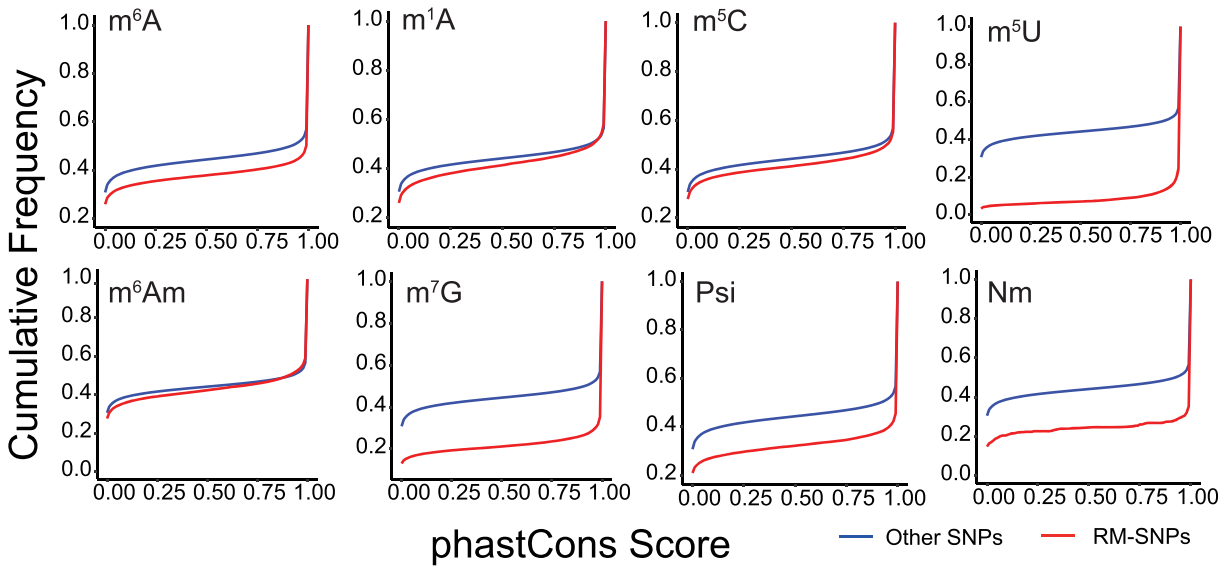

Figure 2. Comparing the PhastCons scores of RNA modification-associated and non-associated SNPs. The sites where RNA modification-associated variants localized were more conserved than non-associated variants for all the eight transcriptome modifications considered in RMDisease.

nomic features can effectively improve the accuracy of a predictor, for example, for m^6A on mRNAs (42), lncRNAs (102) and introns (103), as well as for m^1A (77), Pseudouridine (101) and m^7G (74) site prediction. It may be worth noting that, although existing methods did not explicitly model the base-pairing mechanisms between target RNAs and snoRNPs, they may still vaguely capture the relevant patterns. For example, a previous study showed that there exist snoRNAs that contain two conserved sequence motifs, namely box C (RUGAUGA) and box D (CUGA), and the 2′-*O*-methylation occurs on the target RNA precisely five nucleotides upstream of the box D. Sequence with the corresponding nucleotide contents should show higher probability for 2′-*O*-methylation. However, if there exists other more prevalent mechanisms, this relative weak pattern may not be detected, leading to false prediction related to the sites formed from this mechanism. Meanwhile, for snoRNP-guided RNA modification sites, changes due to mutation on these small RNAs cannot be captured by the analysis pipeline of RMDisease; similarly, changes due to mutation of key RNA modification enzyme genes, such as writers (e.g. METTL3 and METTL14) and erasers (e.g. FTO and ALKBH5), were not covered in this database, either. However, those mutations can potentially disturb the epitranscriptome at a much greater scale.

It is also worth noting that substantial discrepancy has been observed among different epitranscriptome profiling approaches, which can capture different bias (104–107). To minimize the impact of technology usage, special efforts have been made in this study to obtain the most comprehensive collection of the RNA modification sites including those generated from different technologies (Supplementary Table S1). Multiple machine learn models were trained with these datasets for each RNA modification to produce the most reliable results out of the data that is available.

In summary, RMDisease will serve as a useful resource for studies of genetic factors concerning the epitranscriptome regulatory circuits and their potential roles in pathogenesis.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

Author's contribution: Z.W. conceived the idea; K.C. and B.S. collected and processed the data; Y.T. built the website; K.C. and B.S. drafted the manuscript. All authors read, critically revised and approved the final manuscript.

FUNDING

National Natural Science Foundation of China [31671373]; XJTLU Key Program Special Fund [KSF-E-51]; AI University Research Centre through XJTLU Key Programme Special Fund [KSF-P-02]. Funding for open access charge: National Natural Science Foundation of China [31671373]; XJTLU Key Program Special Fund [KSF-E-51]; AI University Research Centre through XJTLU Key Programme Special Fund [KSF-P-02].

Conflict of interest statement. D.J.R. is Executive Editor of *NAR*.

REFERENCES

1. Sauna,Z.E. and Kimchi-Sarfaty,C. (2011) Understanding the contribution of synonymous mutations to human disease. *Nat. Rev. Genet.*, **12**, 683–691.
2. Wu,X. and Hurst,L.D. (2016) Determinants of the usage of splice-associated cis-motifs predict the distribution of human pathogenic SNPs. *Mol. Biol. Evol.*, **33**, 518–529.
3. Mao,F., Xiao,L., Li,X., Liang,J., Teng,H., Cai,W. and Sun,Z.S. (2015) RBP-Var: a database of functional variants involved in regulation mediated by RNA-binding proteins. *Nucleic Acids Res.*, **44**, D154–D163.
4. Andersen,M.C., Engström,P.G., Lithwick,S., Arenillas,D., Eriksson,P., Lenhard,B., Wasserman,W.W. and Odeberg,J. (2008) In silico detection of sequence variations modifying transcriptional regulation. *PLoS Comput. Biol.*, **4**, e5.
5. Riley,T.R., Lazarovici,A., Mann,R.S. and Bussemaker,H.J. (2015) Building accurate sequence-to-affinity models from high-throughput in vitro protein-DNA binding data using FeatureREDUCE. *Elife*, **4**, e06397.
6. Ryu,G.M., Song,P., Kim,K.W., Oh,K.S., Park,K.J. and Kim,J.H. (2009) Genome-wide analysis to predict protein sequence variations that change phosphorylation sites or their corresponding kinases. *Nucleic Acids Res.*, **37**, 1297–1307.
7. Ren,J., Jiang,C., Gao,X., Liu,Z., Yuan,Z., Jin,C., Wen,L., Zhang,Z., Xue,Y. and Yao,X. (2010) PhosSNP for systematic analysis of genetic polymorphisms that influence protein phosphorylation. *Mol. Cell. Proteomics*, **9**, 623–634.
8. Kim,Y., Kang,C., Min,B. and Yi,G.S. (2015) Detection and analysis of disease-associated single nucleotide polymorphism influencing post-translational modification. *BMC Med Genomics*, **8**, S7.
9. Wagih,O., Reimand,J. and Bader,G.D. (2015) MIMP: predicting the impact of mutations on kinase-substrate phosphorylation. *Nat. Methods*, **12**, 531–533.
10. Xu,H.-D., Shi,S.-P., Chen,X. and Qiu,J.-D. (2015) Systematic analysis of the genetic variability that impacts SUMO conjugation and their involvement in human diseases. *Sci. Rep.*, **5**, 10900.
11. Krassowski,M., Paczkowska,M., Cullion,K., Huang,T., Dzneladze,I., Ouellette,B.F.F., Yamada,J.T., Fradet-Turcotte,A. and Reimand,J. (2017) ActiveDriverDB: human disease mutations and genome variation in post-translational modification sites of proteins. *Nucleic Acids Res.*, **46**, D901–D910.
12. Patrick,R., Kobe,B., Lê Cao,K.-A. and Bodén,M. (2017) PhosphoPICK-SNP: quantifying the effect of amino acid variants on protein phosphorylation. *Bioinformatics*, **33**, 1773–1781.
13. Groenning,A.G.B., Doktor,T.K., Larsen,S.J., Petersen,U.S.S., Holm,L.L., Bruun,G.H., Hansen,M.B., Hartung,A.-M., Baumbach,J. and Andresen,B.S. (2020) DeepCLIP: Predicting the effect of mutations on protein-RNA binding with deep learning. *Nucleic Acids Res.*, **48**, 7099–7118.
14. Liu,Z.-X., Yu,K., Dong,J., Zhao,L., Liu,Z., Zhang,Q., Li,S., Du,Y. and Cheng,H. (2019) Precise prediction of calpain cleavage sites and their aberrance caused by mutations in cancer. *Front. Genet.*, **10**, 715.
15. Wang,P., Li,X., Gao,Y., Guo,Q., Ning,S., Zhang,Y., Shang,S., Wang,J., Wang,Y., Zhi,H. *et al.* (2019) LnCeVar: a comprehensive database of genomic variations that disturb ceRNA network regulation. *Nucleic Acids Res.*, **48**, D111–D117.
16. Yang,Y., Zhang,Q., Miao,Y.-R., Yang,J., Yang,W., Yu,F., Wang,D., Guo,A.-Y. and Gong,J. (2019) SNP2APA: a database for evaluating effects of genetic variants on alternative polyadenylation in human cancers. *Nucleic Acids Res.*, **48**, D226–D232.
17. Zheng,Y., Nie,P., Peng,D., He,Z., Liu,M., Xie,Y., Miao,Y., Zuo,Z. and Ren,J. (2017) m6AVar: a database of functional variants involved in m6A modification. *Nucleic Acids Res.*, **46**, D139–D145.
18. Song,B., Tang,Y., Chen,K., Wei,Z., Rong,R., Lu,Z., Su,J., de Magalhaes,J.P., Rigden,D.J. and Meng,J. (2020) m7GHub: deciphering the location, regulation and pathogenesis of internal mRNA N7-methylguanosine (m7G) sites in human. *Bioinformatics*, **36**, 3528–3536.
19. He,C. (2010) Grand challenge commentary: RNA epigenetics? *Nat. Chem. Biol.*, **6**, 863–865.
20. 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.
21. McCown,P.J., Ruszkowska,A., Kunkler,C.N., Breger,K., Hulewicz,J.P., Wang,M.C., Springer,N.A. and Brown,J.A. (2020) Naturally occurring modified ribonucleosides. *WIREs RNA*, **n/a**, e1595.
22. Jones,J.D., Monroe,J. and Koutmou,K.S. (2020) A molecular-level perspective on the frequency, distribution, and consequences of messenger RNA modifications. *WIREs RNA*, **n/a**, e1586.
23. Esteve-Puig,R., Bueno-Costa,A. and Esteller,M. (2020) Writers, readers and erasers of RNA modifications in cancer. *Cancer Lett.*, **474**, 127–137.
24. Zhang,Z., Luo,K., Zou,Z., Qiu,M., Tian,J., Sieh,L., Shi,H., Zou,Y., Wang,G., Morrison,J. *et al.* (2020) Genetic analyses support the contribution of mRNA N6-methyladenosine (m6A) modification to human disease heritability. *Nat. Genet.*, **52**, 939–949.
25. Fustin,J.M., Doi,M., Yamaguchi,Y., Hida,H., Nishimura,S., Yoshida,M., Isagawa,T., Morioka,M.S., Kakeya,H., Manabe,I. *et al.* (2013) RNA-methylation-dependent RNA processing controls the speed of the circadian clock. *Cell*, **155**, 793–806.
26. Wang,X., Lu,Z., Gomez,A., Hon,G.C., Yue,Y., Han,D., Fu,Y., Parisien,M., Dai,Q., Jia,G. *et al.* (2014) N6-methyladenosine-dependent regulation of messenger RNA stability. *Nature*, **505**, 117–120.
27. Yang,S., Wei,J., Cui,Y.H., Park,G., Shah,P., Deng,Y., Aplin,A.E., Lu,Z., Hwang,S., He,C. *et al.* (2019) m(6)A mRNA demethylase FTO regulates melanoma tumorigenicity and response to anti-PD-1 blockade. *Nat. Commun.*, **10**, 2782.
28. Niu,Y., Lin,Z., Wan,A., Chen,H., Liang,H., Sun,L., Wang,Y., Li,X., Xiong,X.F., Wei,B. *et al.* (2019) RNA N6-methyladenosine demethylase FTO promotes breast tumor progression through inhibiting BNIP3. *Mol. Cancer*, **18**, 46.
29. Lin,X., Chai,G., Wu,Y., Li,J., Chen,F., Liu,J., Luo,G., Tauler,J., Du,J., Lin,S. *et al.* (2019) RNA m(6)A methylation regulates the epithelial mesenchymal transition of cancer cells and translation of Snail. *Nat. Commun.*, **10**, 2065.
30. Lee,H., Bao,S., Qian,Y., Geula,S., Leslie,J., Zhang,C., Hanna,J.H. and Ding,L. (2019) Stage-specific requirement for Mettl3-dependent m(6)A mRNA methylation during haematopoietic stem cell differentiation. *Nat. Cell Biol.*, **21**, 700–709.
31. 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.
32. Ringeard,M., Marchand,V., Decroly,E., Motorin,Y. and Bennasser,Y. (2019) FTSJ3 is an RNA 2′-O-methyltransferase recruited by HIV to avoid innate immune sensing. *Nature*, **565**, 500–504.
33. 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.
34. 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.
35. Chen,K., Lu,Z., Wang,X., Fu,Y., Luo,G.Z., Liu,N., Han,D., Dominissini,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.

36. 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.
37. Gjoneska,E., Pfenning,A.R., Mathys,H., Quon,G., Kundaje,A., Tsai,L.-H. and Kellis,M. (2015) Conserved epigenomic signals in mice and humans reveal immune basis of Alzheimer/'s disease. *Nature*, **518**, 365–369.
38. Chen,Z., Zhao,P., Li,F., Wang,Y., Smith,A.I., Webb,G.I., Akutsu,T., Baggag,A., Bensmail,H. and Song,J. (2019) Comprehensive review and assessment of computational methods for predicting RNA post-transcriptional modification sites from RNA sequences. *Brief. Bioinform.*, bbz112.
39. Lv,H., Zhang,Z.-M., Li,S.-H., Tan,J.-X., Chen,W. and Lin,H. (2019) Evaluation of different computational methods on 5-methylcytosine sites identification. *Brief. Bioinform.*, **21**, 982–995.
40. Chen,W., Feng,P., Ding,H., Lin,H. and Chou,K.-C. (2015) iRNA-Methyl: identifying N 6-methyladenosine sites using pseudo nucleotide composition. *Anal. Biochem.*, **490**, 26–33.
41. 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.
42. Chen,K., Wu,X., Zhang,Q., 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.
43. Zhu,X., He,J., Zhao,S., Tao,W., Xiong,Y. and Bi,S. (2019) A comprehensive comparison and analysis of computational predictors for RNA N6-methyladenosine sites of Saccharomyces cerevisiae. *Brief. Funct. Genomics*, **19**, 367–376.
44. 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.
45. Han,Y., Feng,J., Xia,L., Dong,X., Zhang,X., Zhang,S., Miao,Y., Xu,Q., Xiao,S., Zuo,Z. *et al.* (2019) CVm6A: a visualization and exploration database for m(6)As in cell lines. *Cells*, **8**, 168.
46. Liu,S., He,C. and Chen,M. (2019) REPIC: a database for exploring N⁶-methyladenosine methylome. *Genome Biol.*, **21**, 100.
47. Huang,T., Chen,W., Liu,J., Gu,N. and Zhang,R. (2019) Genome-wide identification of mRNA 5-methylcytosine in mammals. *Nat. Struct. Mol. Biol.*, **26**, 380–388.
48. 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.
49. 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.
50. 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.
51. 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.
52. Khoddami,V., Yerra,A., Mosbruger,T.L., Fleming,A.M., Burrows,C.J. and Cairns,B.R. (2019) Transcriptome-wide profiling of multiple RNA modifications simultaneously at single-base resolution. *PNAS*, **116**, 6784–6789.
53. 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.
54. Safra,M., Sas-Chen,A., Nir,R., Winkler,R., Nachshon,A., Bar-Yaacov,D., Erlacher,M., Rossmanith,W., Stern-Ginossar,N. and Schwartz,S. (2017) The m1A landscape on cytosolic and mitochondrial mRNA at single-base resolution. *Nature*, **551**, 251–255.
55. Khoddami,V. and Cairns,B.R. (2013) Identification of direct targets and modified bases of RNA cytosine methyltransferases. *Nat. Biotechnol.*, **31**, 458–464.
56. Yang,X., Yang,Y., Sun,B.F., Chen,Y.S., Xu,J.W., Lai,W.Y., Li,A., Wang,X., Bhattarai,D.P., Xiao,W. *et al.* (2017) 5-methylcytosine promotes mRNA export - NSUN2 as the methyltransferase and ALYREF as an m(5)C reader. *Cell Res.*, **27**, 606–625.
57. Carter,J.M., Emmett,W., Mozos,I.R., Kotter,A., Helm,M., Ule,J. and Hussain,S. (2019) FICC-Seq: a method for enzyme-specified profiling of methyl-5-uridine in cellular RNA. *Nucleic Acids Res.*, **47**, e113.
58. 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.
59. 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(7)-Methylguanosine methylome in mammalian mRNA. *Mol. Cell*, **74**, 1304–1316.
60. Malbec,L., Zhang,T., Chen,Y.S., Zhang,Y., Sun,B.F., Shi,B.Y., Zhao,Y.L., Yang,Y. and Yang,Y.G. (2019) Dynamic methylome of internal mRNA N(7)-methylguanosine and its regulatory role in translation. *Cell Res.*, **29**, 927–941.
61. He,J., Fang,T., Zhang,Z., Huang,B., Zhu,X. and Xiong,Y. (2018) PseUI: Pseudouridine sites identification based on RNA sequence information. *BMC Bioinformatics*, **19**, 306.
62. Sun,P.P., Chen,Y.B., Liu,B., Gao,Y.X., Han,Y., He,F. and Ji,J.C. (2019) DeepMRMP: A new predictor for multiple types of RNA modification sites using deep learning. *Math. Biosci. Eng.: MBE*, **16**, 6231–6241.
63. Tahir,M., Tayara,H. and Chong,K.T. (2019) iPseU-CNN: Identifying RNA pseudouridine sites using convolutional neural networks. *Mol. Ther. Nucleic Acids*, **16**, 463–470.
64. Mostavi,M., Salekin,S. and Huang,Y. (2018) In: *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. pp. 2394–2397.
65. Wu,X., Wei,Z., Chen,K., Zhang,Q., Su,J., Liu,H., Zhang,L. and Meng,J. (2019) m6Acomet: large-scale functional prediction of individual m(6)A RNA methylation sites from an RNA co-methylation network. *BMC Bioinformatics*, **20**, 223.
66. 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.
67. Chen,W., Feng,P., Yang,H., Ding,H., Lin,H. and Chou,K.C. (2018) iRNA-3typeA: identifying three types of modification at RNA's adenosine sites. *Mol. Ther. Nucleic Acids*, **11**, 468–474.
68. Liu,K. and Chen,W. (2020) iMRM:a platform for simultaneously identifying multiple kinds of RNA modifications. *Bioinformatics*, **36**, 3336–3342.
69. Yang,H., Lv,H., Ding,H., Chen,W. and Lin,H. (2018) iRNA-2OM: A sequence-based predictor for identifying 2′-O-Methylation sites in homo sapiens. *J. Comput. Biol.*, **25**, 1266–1277.
70. Lv,H., Zhang,Z.M., Li,S.H., Tan,J.X., Chen,W. and Lin,H. (2019) Evaluation of different computational methods on 5-methylcytosine sites identification. *Brief. Bioinform.*, **21**, 982–995.
71. Chen,W., Feng,P., Tang,H., Ding,H. and Lin,H. (2016) RAMPred: identifying the N(1)-methyladenosine sites in eukaryotic transcriptomes. *Sci. Rep.*, **6**, 31080.
72. Feng,P., Ding,H., Yang,H., Chen,W., Lin,H. and Chou,K.C. (2017) iRNA-PseColl: identifying the occurrence sites of different RNA modifications by incorporating collective effects of nucleotides into PseKNC. *Mol. Ther. Nucleic Acids*, **7**, 155–163.
73. Chen,W., Feng,P., Song,X., Lv,H. and Lin,H. (2019) iRNA-m7G: identifying N(7)-methylguanosine sites by fusing multiple features. *Mol. Ther. Nucleic Acids*, **18**, 269–274.
74. Song,B., Tang,Y., Chen,K., Wei,Z., Rong,R., Lu,Z., Su,J., de Magalhães,J.P., Rigden,D.J. and Meng,J. (2020) m7GHub: deciphering the location, regulation and pathogenesis of internal mRNA N7-methylguanosine (m7G) sites in human. *Bioinformatics*, **36**, 3528–3536.
75. Chen,W., Tang,H., Ye,J., Lin,H. and Chou,K.C. (2016) iRNA-PseU: Identifying RNA pseudouridine sites. *Mol. Ther. Nucleic Acids*, **5**, e332.

76. Song,B., Tang,Y., Wei,Z., Liu,G., Su,J., Meng,J. and Chen,K. (2020) PIANO: a web server for pseudouridine-site (Ψ) identification and functional annotation. *Front. Genet.*, **11**, 88.
77. Lian,L., Lei,X., Meng,J. and Wei,Z. (2020) ISGm1A: integration of sequence features and genomic features to improve the prediction of human m1A RNA methylation sites. *IEEE Access*, **8**, 81971–81977.
78. Kumar,P., Henikoff,S. and Ng,P.C. (2009) Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat. Protoc.*, **4**, 1073–1081.
79. Adzhubei,I.A., Schmidt,S., Peshkin,L., Ramensky,V.E., Gerasimova,A., Bork,P., Kondrashov,A.S. and Sunyaev,S.R. (2010) A method and server for predicting damaging missense mutations. *Nat. Methods*, **7**, 248–249.
80. Chun,S. and Fay,J.C. (2009) Identification of deleterious mutations within three human genomes. *Genome Res.*, **19**, 1553–1561.
81. Shihab,H.A., Gough,J., Cooper,D.N., Stenson,P.D., Barker,G.L., Edwards,K.J., Day,I.N. and Gaunt,T.R. (2013) Predicting the functional, molecular, and phenotypic consequences of amino acid substitutions using hidden Markov models. *Hum. Mutat.*, **34**, 57–65.
82. Wang,K., Li,M. and Hakonarson,H. (2010) ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.*, **38**, e164.
83. Zhu,Y., Xu,G., Yang,Y.T., Xu,Z., Chen,X., Shi,B., Xie,D., Lu,Z.J. and Wang,P. (2018) POSTAR2: deciphering the post-transcriptional regulatory logics. *Nucleic Acids Res.*, **47**, D203–D211.
84. Agarwal,V., Bell,G.W., Nam,J.-W. and Bartel,D.P. (2015) Predicting effective microRNA target sites in mammalian mRNAs. *eLife*, **4**, e05005.
85. Li,J.-H., Liu,S., Zhou,H., Qu,L.-H. and Yang,J.-H. (2013) 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.
86. Buniello,A., MacArthur,J.A L., Cerezo,M., Harris,L.W., Hayhurst,J., Malangone,C., McMahon,A., Morales,J., Mountjoy,E., Sollis,E. *et al.* (2018) The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res.*, **47**, D1005–D1012.
87. Landrum,M.J., Lee,J.M., Benson,M., Brown,G., Chao,C., Chitipiralla,S., Gu,B., Hart,J., Hoffman,D., Hoover,J. *et al.* (2015) ClinVar: public archive of interpretations of clinically relevant variants. *Nucleic Acids Res.*, **44**, D862–D868.
88. Johnson,A.D. and O'Donnell,C.J. (2009) An open access database of genome-wide association results. *BMC Med. Genet.*, **10**, 6.
89. Dong,Q., Li,F., Xu,Y., Xiao,J., Xu,Y., Shang,D., Zhang,C., Yang,H., Tian,Z., Mi,K. *et al.* (2019) RNAactDrug: a comprehensive database of RNAs associated with drug sensitivity from multi-omics data. *Brief. Bioinform.*, bbz142.
90. Buels,R., Yao,E., Diesh,C.M., Hayes,R.D. and Holmes,I.H. (2016) JBrowse: a dynamic web platform for genome visualization and analysis. *Genome Biol.*, **17**, 66.
91. Boccaletto,P., Machnicka,M.A., Purta,E., Piątkowski,P., Bagiński,B., Wirecki,T.K., de Crécy-Lagard,V., Ross,R., Limbach,P.A., Kotter,A. *et al.* (2017) MODOMICS: a database of RNA modification pathways. 2017 update. *Nucleic Acids Res.*, **46**, D303–D307.
92. Song,B., Chen,K., Tang,Y., Ma,J., Meng,J. and Wei,Z. (2020) PSI-MOUSE: predicting mouse pseudouridine sites from sequence and genome-derived features. *Evolutionary Bioinformatics*, **16**, 1176934320925752.
93. Chen,W., Feng,P., Song,X., Lv,H. and Lin,H. (2019) iRNA-m7G: identifying N7-methylguanosine sites by fusing multiple features. *Mol/ Ther/ - Nucleic Acids*, **18**, 269–274.
94. Kiss-László,Z., Henry,Y., Bachellerie,J.-P., Caizergues-Ferrer,M. and Kiss,T. (1996) Site-specific ribose methylation of preribosomal RNA: a novel function for small nucleolar RNAs. *Cell*, **85**, 1077–1088.
95. Watkins,N.J., Gottschalk,A., Neubauer,G., Kastner,B., Fabrizio,P., Mann,M. and Lührmann,R. (1998) Cbf5p, a potential pseudouridine synthase, and Nhp2p, a putative RNA-binding protein, are present together with Gar1p in all H BOX/ACA-motif snoRNPs and constitute a common bipartite structure. *RNA*, **4**, 1549–1568.
96. Lowe,T.M. and Eddy,S.R. (1999) A computational screen for methylation guide snoRNAs in yeast. *Science*, **283**, 1168–1171.
97. Badis,G., Fromont-Racine,M. and Jacquier,A. (2003) A snoRNA that guides the two most conserved pseudouridine modifications within rRNA confers a growth advantage in yeast. *RNA*, **9**, 771–779.
98. Motorin,Y. and Helm,M. (2011) RNA nucleotide methylation. *Wiley Interdiscip/ Rev.: RNA*, **2**, 611–631.
99. Chen,W., Feng,P., Ding,H. and Lin,H. (2016) PAI: predicting adenosine to inosine editing sites by using pseudo nucleotide compositions. *Sci. Rep.*, **6**, 35123–35123.
100. Dou,L., Li,X., Ding,H., Xu,L. and Xiang,H. (2019) Is there any sequence feature in the RNA pseudouridine modification prediction problem. *Mol. Ther. - Nucleic Acids*, **19**.
101. Song,B., Tang,Y., Wei,Z., Liu,G., Su,J., Meng,J. and Chen,K. (2020) PIANO: a web server for pseudouridine-site (Ψ) identification and functional annotation. *Frontiers in Genetics*, **11**, 88.
102. Liu,L., Lei,X., Fang,Z., Tang,Y., Meng,J. and Wei,Z. (2020) LITHOPHONE: improving lncRNA methylation site prediction using an ensemble predictor. *Frontiers in Genetics*, **11**, 545.
103. Liu,L., Lei,X., Meng,J. and Wei,Z. (2020) WITMSG: large-scale prediction of human intronic m6A RNA methylation sites from sequence and genomic features. *Curr. Genomics*, **21**, 67–76.
104. Adachi,H., De Zoysa,M.D. and Yu,Y.-T. (2018) Post-transcriptional pseudouridylation in mRNA as well as in some major types of noncoding RNAs. *Biochim. Biophys. Acta (BBA)-Gene Regul. Mech.*, **1862**, 230–239.
105. Zaringhalam,M. and Papavasiliou,F.N. (2016) Pseudouridylation meets next-generation sequencing. *Methods*, **107**, 63–72.
106. Hussain,S., Aleksic,J., Blanco,S., Dietmann,S. and Frye,M. (2013) Characterizing 5-methylcytosine in the mammalian epitranscriptome. *Genome Biol.*, **14**, 215.
107. Capitanchik,C., Toolan-Kerr,P., Luscombe,N.M. and Ule,J. (2020) How do you identify m6 a methylation in transcriptomes at high resolution? a comparison of recent datasets. *Front. Genet.*, **11**, 398.
108. Chen,W., Tang,H. and Lin,H. (2017) MethyRNA: a web server for identification of N(6)-methyladenosine sites. *J. Biomol. Struct. Dyn.*, **35**, 683–687.