Published online 3 October 2017 *Nucleic Acids Research, 2018, Vol. 46, Database issue* ***D139–D145***
doi: 10.1093/nar/gkx895

m6AVar: a database of functional variants involved in m^{6}A modification

Yueyuan Zheng[1,2,†], Peng Nie[2,†], Di Peng[2,†], Zhihao He[2], Mengni Liu[2], Yubin Xie[2], Yanyan Miao[2], Zhixiang Zuo[1,*] and Jian Ren[1,2,3,*]

[1]Sun Yat-sen University Cancer Center, State Key Laboratory of Oncology in South China, Collaborative Innovation Center for Cancer Medicine, Sun Yat-sen University, Guangzhou 510060, China, [2]State Key Laboratory of Biocontrol, School of Life Sciences, Sun Yat-sen University, Guangzhou, Guangdong 510275, China and [3]Collaborative Innovation Center of High Performance Computing, National University of Defense Technology, Changsha 410073, China

Received August 15, 2017; Revised September 17, 2017; Editorial Decision September 21, 2017; Accepted September 23, 2017

ABSTRACT

Identifying disease-causing variants among a large number of single nucleotide variants (SNVs) is still a major challenge. Recently, *N*^{6}-methyladenosine (m^{6}A) has become a research hotspot because of its critical roles in many fundamental biological processes and a variety of diseases. Therefore, it is important to evaluate the effect of variants on m^{6}A modification, in order to gain a better understanding of them. Here, we report m6AVar (http://m6avar.renlab.org), a comprehensive database of m^{6}A-associated variants that potentially influence m^{6}A modification, which will help to interpret variants by m^{6}A function. The m^{6}A-associated variants were derived from three different m^{6}A sources including miCLIP/PA-m^{6}A-seq experiments (high confidence), MeRIP-Seq experiments (medium confidence) and transcriptome-wide predictions (low confidence). Currently, m6AVar contains 16 132 high, 71 321 medium and 326 915 low confidence level m^{6}A-associated variants. We also integrated the RBP-binding regions, miRNA-targets and splicing sites associated with variants to help users investigate the effect of m^{6}A-associated variants on post-transcriptional regulation. Because it integrates the data from genome-wide association studies (GWAS) and ClinVar, m6AVar is also a useful resource for investigating the relationship between the m^{6}A-associated variants and disease. Overall, m6AVar will serve as a useful resource for annotating variants and identifying disease-causing variants.

INTRODUCTION

Rapid improvement in high-throughput sequencing technology has resulted in the identification of millions of single nucleotide variants (SNVs) across multiple genomes. A major challenge in delineating these variants is to distinguish the functional variants from the rest. In recent years, numerous studies have been undertaken to explore disease-associated nonsynonymous SNVs that alter amino acid at the protein level (1). Nevertheless, there is growing evidence showing many synonymous SNVs, which do not alter the amino acid sequences of proteins and are considered 'silent' mutations, also affect the function of genes and cause various diseases, suggesting a role in transcriptional or post-transcriptional regulation (2). Many studies have shown that variants have the capacity to alter the secondary structure of RNA, influence RNA–protein interactions (3), and change the splicing sites of exonic splicing enhancers and silencers (4) as well as genetic information by means of RNA editing (5). We speculated that variants might also influence RNA modification (e.g. m^{6}A) by changing the RNA sequences of the target sites or key flanking nucleotides.

N^6-Methyladenosine (m^{6}A) is a pervasive RNA modification in eukaryotes, that is involved in various biological processes such as embryonic development (6), cell apoptosis (7), spermatogenesis (8) and circadian rhythms (9). Recent development of the high-throughput sequencing techniques for m^{6}A (known as Methylated RNA Immunoprecipitation Sequencing (MeRIP-Seq), Photo-Crosslinking-Assisted m^{6}A Sequencing Strategy (PA-m^{6}A-seq) and m^{6}A individual-nucleotide-resolution cross-linking and immunoprecipitation sequencing (miCLIP)) has provided thousands of m^{6}A sites and deep insights into the m^{6}A machinery (10–13), revealing the essential regulatory roles of

*To whom correspondence should be addressed. Tel: +86 20 87342325; Fax: +86 20 87342325; Email: renjian.sysu@gmail.com
Correspondence may also be addressed to Zhixiang Zuo. Email: zuozhx@sysucc.org.cn
†These authors contributed equally to this work as first authors.

m^{6}A in RNA splicing, miRNA function and RNA stability (7,8,14,15).

An increasing number of studies have revealed that dysregulation of m^{6}A modification may impact various diseases. It has been found that the knockout of *METTL3* in human cancer cells decreased the invasion of tumor cells (16). The activation of *ALKBH5* in hypoxic breast cancer cells would promote cancer stem cell enrichment (17). In addition, previous studies have suggested that the m^{6}A eraser FTO is related to metabolism dysfunction (18) and acts as an oncogenic role in Acute Myeloid Leukemia (19). Furthermore, a previous study in mice indicated that m^{6}A might be important in neurodevelopmental processes (10). To further investigate the potential pathogenesis of m^{6}A modification, it is necessary to evaluate the effect of variants on m^{6}A modification. This will be helpful for both an understanding of the variants' pathogenic molecular mechanisms and the identification of additional disease-causing variants.

As result of the intensifying researches and accumulating data on the m^{6}A machinery, databases on m^{6}A modification have emerged in recent years. In 2015, Liu *et al.* collected 74 samples from 22 different m^{6}A-seq experiments and constructed MeT-DB, the first comprehensive m^{6}A database of the mammalian transcriptome (20). Later, Sun *et al.* developed the RNA modification database called 'RMBase' that includes 226 000 m^{6}A sites and 10 005 m^{5}C sites (21). Although the above databases have greatly aided research of m^{6}A functions, there is still no specific resource that would help study of influence of variants on m^{6}A modification.

In this study, we present m6AVar (http://m6avar.renlab.org), a comprehensive database that allows the annotation, visualization and exploration of m^{6}A-associated variants in humans and mice (Figure 1). A great number of the m^{6}A-associated variants were derived from millions of germlines and somatic variants as well as three different m^{6}A sources that included miCLIP experiments, PA-m^{6}A-seq experiments, MeRIP-Seq experiments and transcriptome-wide predictions. We further annotated the m^{6}A-associated variants by checking whether they localized in regions with RBP binding sites, as well as miRNA targets and splicing sites. Moreover, disease-associated data from GWAS and ClinVar database were also integrated into m6AVar, which allows users to explore the underlying relationship between the m^{6}A machinery and diseases.

MATERIALS AND METHODS

Data resource

Germline and somatic variants were obtained from dbSNP and TCGA, respectively (Supplementary Table S1). We preserved those variants within the exonic regions for subsequent analysis. All of the m^{6}A sites were derived from seven miCLIP experiments, two PA-m^{6}A-seq experiments, 244 MeRIP-Seq experiments (Supplementary Table S2) and a transcriptome-wide prediction based on Random Forest algorithm. To identify the potential roles of m^{6}A-associated variants in post-transcriptome regulation, the RBP binding sites from starBase2 (22) and CLIPdb (23) (Supplementary Table S3), the miRNA–RNA interactions from starBase2 and the canonical splice sites (GT-AG) from Ensembl annotations were collected. In addition, we obtained a large number of disease-associated SNPs from different data sets (GWAS catalog (24), Johnson and O'Donnel (25), dbGAP (26), GAD (27) and ClinVar (28)) to perform disease-association analysis. The detailed description and statistics for these data resources can be found in Supplementary Table S4.

Data preprocessing

As the raw data collected from the diverse databases utilized different data formats, it is essential to unify them under standard procedures. To do this, the genomic coordinates of all of the data resources were converted to GRCh37 for the human and GRCm38 for the mouse using the LiftOver (29). The location of each m^{6}A site was then annotated by the transcript structure, including the CDS, 3′ UTR, 5′ UTR, start codon and stop codon etc. All the genomic information on the non-coding genes are from DASHR (30), miRBase (v21) (31), GtRNAdb (32) and piRNABank (33). Furthermore, all of the SNPs were annotated by ANNOVAR (updated to 1 February 2016) in two steps (34). First, we studied the conservation of evolutionary sequence in the m^{6}A-associated variants using phastCons 100-way and 60-way gene conservation scores for the human and mouse respectively (35). Second, we measured the deleterious level of each variant by integrating the results from five predictors of variant function (SIFT (36), PolyPhen2 HVAR (37), PolyPhen2 HDIV (37), LRT (38) and FATHMM (39)). Each variant was scored from 0 to 5 scale by counting deleterious levels of the variants obtained by the above five methods according to their thresholds curated by the dbNSFP database (40).

Derivation of the m^{6}A sites

The m^{6}A sites in m6AVar were derived using three different strategies with confidence levels ranging from high to low as illustrated below:

1. The m^{6}A sites having a high confidence level were extracted from the published single-nucleotide resolution m^{6}A sites in the miCLIP experiments. Besides, we also obtained m^{6}A sites that conformed to the DRACH (where D = A, G or U; R = G or A; H = A, C or U) motif from PA-m^{6}A-seq experiments (10,11,41).
2. The m^{6}A sites having a medium confidence level were predicted from the previously published MeRIP-seq data. We first downloaded all the MeRIP-Seq samples from the GEO database as raw data. Quality control was performed with FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc) and the sequencing adaptors were removed using Trimmomatic (v0.33) (42). A minimum of 25 nucleotides was required for unambiguous alignment. All qualified reads were mapped to reference genomes (GRCh37 for human and GRCm38 for mouse) by Tophat (v2.1.1) using default parameters (43). We applied three peak callers (MACS2 (44), MeTPeak (45) and Meyer's method (10)) to identify the m^{6}A peaks separately. MSPC (46) was then ap-

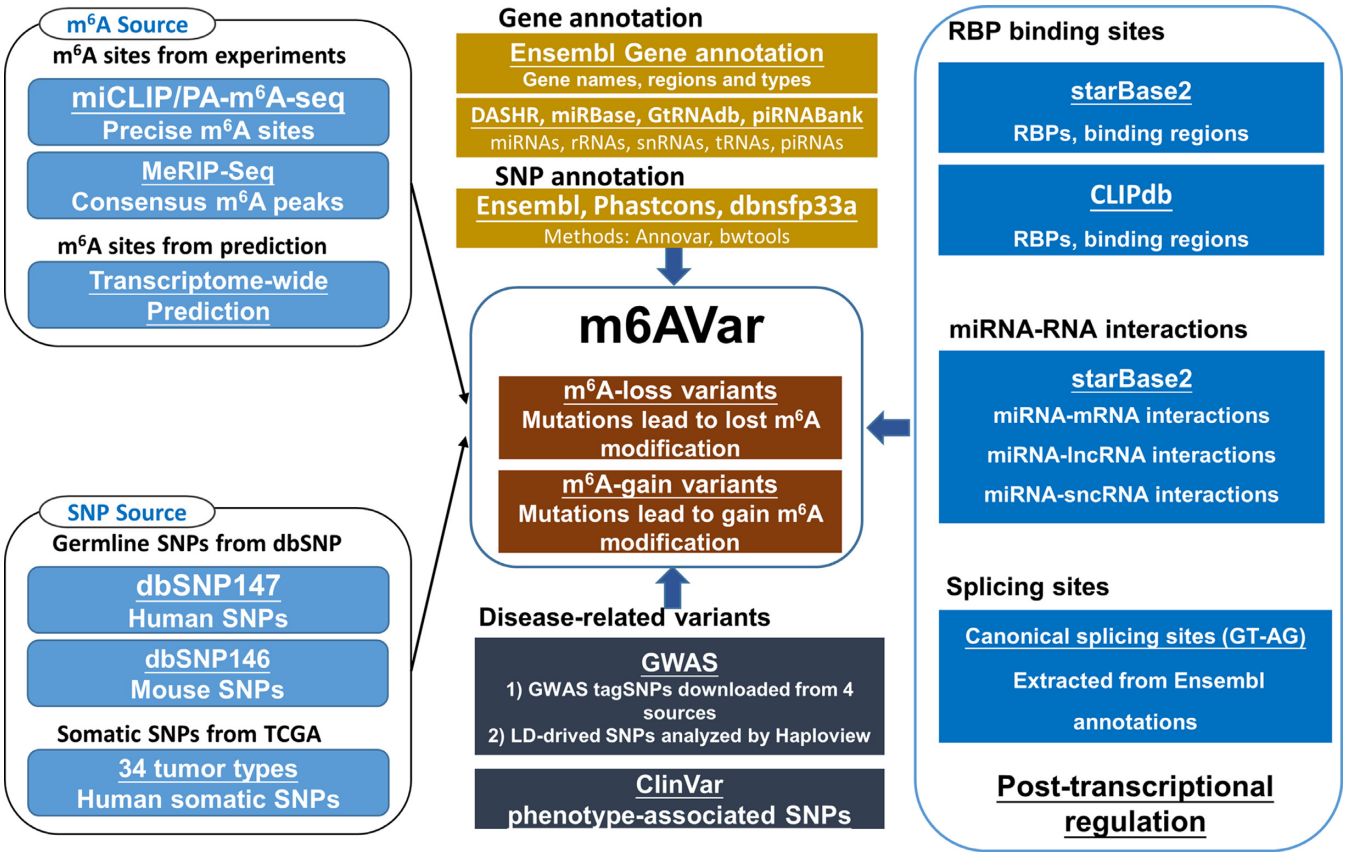

Figure 1. Overall design and construction of m6AVar.

plied to construct consensus peaks from the three methods (Supplementary method, Supplementary Table S5). We then predicted single-nucleotide resolution m^{6}A sites using m^{6}AFinder based on Random Forest algorithm from these consensus peaks (Supplementary method, Supplementary Figure S1).

3. In addition, to cover all potential m^{6}A sites, we also obtain m^{6}A sites with low confidence level by transcriptome-wide prediction using 'm6AFinder' (Supplementary method).

Derivation of the m^{6}A-associated variants

We defined a variant as an m^{6}A-associated variant by evaluating whether it has the potential to alter the DRACH motif or other sequence features essential for m^{6}A modification. According to various levels of confidence, we extracted the corresponding m^{6}A-associated variants as follows.

1. For m^{6}A sites having a high confidence level, we retained the variants that located nearby the m^{6}A sites and then looked for the variants that disrupt DRACH motif around the m^{6}A sites, such as changing from D(A/G/U) to C, R(G/A) to C/T, A to C/G/U, C to G/A/U, H (A/C/U) to G.
2. For m^{6}A sites with a medium confidence level, the m^{6}A-associated variants were derived from the intersection between the variants and the m^{6}A sites generated from MeRIP-Seq experiments. The Random Forest prediction model was subsequently applied to find the variants in the m^{6}A site region that change the DRACH motif or other sequence features.
3. For m^{6}A sites with a low confidence level, we separately predicted the m^{6}A status for the sequence around the variants in both the reference sequence and mutant sequence by the Random Forest prediction model based on DRACH motif and other sequence features. The variants result in loss of m^{6}A sites in mutant sequence compared to reference sequence were defined as m^{6}A-loss variants. In the opposite case, they were defined as m^{6}A-gain variants.

Post-transcriptional regulation association analysis

First, m^{6}A-associated variants were intersected with RNA-binding proteins (RBPs) regions for the same sample. In terms of miRNA targets, we matched all of the m^{6}A-associated variants with miRNA targets to obtain the m^{6}A-associated variants which potentially impacted miRNA-target interactions. Additionally, we extracted 100 bp upstream from the 5′ splicing sites and 100 bp downstream from the 3′ splicing sites. Subsequently we matched all of the m^{6}A-associated variants with these regions to obtain the splicing sites affected by the m^{6}A-associated variants.

Table 1. m^{6}A-associated variants in m6AVar

	Human dbSNP147			Mouse dbSNP146			TCGA			Total		
	Loss[a] variants	Gain[b] variants	All	Loss variants	Gain variants	All	Loss variants	Gain variants	All	Loss variants	Gain variants	All
miCLIP/PA-m^{6}A-Seq (High confidence)	13 703	—	13 703	935	—	935	1 494	—	1 494	16 132	—	16 132
MeRIP-Seq (Medium confidence)	54 222	—	54 222	9404	—	9404	7695	—	7695	71 321	—	71 321
Prediction (Low confidence)	144 534	100 542	243 880	17 739	12 382	29 982	32 069	21 298	53 053	194 342	134 222	326 915
Total	212 360	100 542	311 706	28 065	12 382	40 308	41 243	21 298	62 227	281 668	134 222	414 241

[a]Loss variants were those variants resulting in loss of m^{6}A sites in mutant sequence compared to reference sequence.
[b]Gain variants were conversely formed.

Table 2. Statistics of associated data in m6AVar

	RBP-binding regions		miRNA Targets		Splicing sites		Disease-related variants	
	Variants	RBPs	Variants	miRNAs	Variants	Genes	ClinVar	GWAS
Human dbSNP147	183 960 (59.02%)	68	6371 (2.04%)	268	158 469 (50.84%)	17 921	1919	372
Mouse dbSNP146	3370 (8.36%)	29	196 (0.49%)	173	13 382 (33.20%)	7781	0	0
TCGA	31 899 (51.26%)	66	338 (0.54%)	219	39 605 (63.65%)	12 273	178	168

Disease association analysis

LD analysis was performed for each GWAS disease-associated SNP. We used Haploview to obtain its LD mutations with a parameter of $r^2 > 0.8$ in at least one of the four populations from CHB, CEU, JPT and TSI (19). Then we selected all the m^{6}A-associated variants by mapping them with GWAS disease-associated SNPs and their LD mutations. Moreover, we also collected ClinVar data in order to annotate the m^{6}A-associated variants with specific functions.

Database and web interface implementation

All the metadata in m6AVar were stored and managed in MySQL tables. The web interfaces were implemented in Hyper Text Markup Language (HTML), Cascading Style Sheets (CSS) and Hypertext Preprocessor (PHP). In order to provide visualization of all the analysis results, multiple statistical diagrams were shown by EChars and genome browser was implemented using Jbrowser (47).

RESULTS

Database content

m6AVar contains three different confidence levels of m^{6}A-associated variants for human and mouse (Table 1). The m^{6}A-associated variants with high confidence level were derived from miCLIP or PA-m^{6}A-seq experiments. For human, there are 13 703 and 1494 high confidence level m^{6}A-associated germline and somatic variants from dbSNP and TCGA, respectively. For mouse, there are 935 high confidence level m^{6}A-associated germline variants from dbSNP. The m^{6}A-associated variants with medium confidence level were derived from MeRIP-Seq experiments. For human, there are 54 222 and 7695 medium confidence level m^{6}A-associated germline and somatic variants from dbSNP and TCGA, respectively. For mouse, there are 9,404 medium confidence level m^{6}A-associated germline variants from dbSNP. In addition, a genome-wide prediction based on Random Forest algorithm was performed for the sequences around all the collected variants from dbSNP and TCGA to find the variants that cause potential gain or loss of m^{6}A sites. As a result, we obtained 296 933 and 29 982 low confidence level m^{6}A-associated variants in human and mouse, respectively.

Moreover, m6AVar contains many associated data, such as RBPs, miRNA and splicing sites, as well as disease information (Table 2). For human, 183 960 and 31 899 m^{6}A-associated variants from dbSNP and TCGA are related to 68 and 66 RBPs. 6371 and 338 m^{6}A-associated variants from dbSNP and TCGA are related to 268 and 219 miRNAs. 158 469 and 39 605 m^{6}A-associated variants from dbSNP and TCGA are related to the splicing sites of 17 921 and 12 273 genes. Moreover, there are 2097 and 540 disease-related m^{6}A-associated variants recorded in ClinVar and GWAS, respectively. For mouse, there are 3370 m^{6}A-associated variants related to 29 RBPs, 196 m^{6}A-associated variants related to 173 miRNAs, and 13 382 m^{6}A-associated variants related to the splicing sites of 7781 genes.

Web interface and usage

m6AVar provides user-friendly web interfaces that enable users to browse, search and download all of the m^{6}A-associated variants in the database.

Search. m6AVar provides four modes to query the database, i.e. by RsID, Gene, Chromosome region and Disease. Here, we illustrate an example to show how to utilize m6AVar by search function (Figure 2). We sought to undertake an investigation of m^{6}A modification in breast cancer using m6AVar. Through the 'RsID' search mode, users can check whether a variant of interest functionally affects the m^{6}A status. In addition, m^{6}A-associated variants in known breast cancer-related genes may be obtained by using the 'Gene' mode (Figure 2A). Taking the human tumor suppressor gene *BRCA1* as an example, 102 m^{6}A-associated variants in *BRCA1* are presented as a table in the search results page (Figure 2B). Among them, 11, 26 and 65 m^{6}A-associated variants were derived from miCLIP (high confidence), MeRIP-Seq (medium confidence) and prediction (low confidence), respectively (Figure 2C). A statistical plot shows the number of germline and somatic m^{6}A-associated variants (Figure 2D). Users may obtain the related RBPs, miRNA targets, splicing sites and diseases from

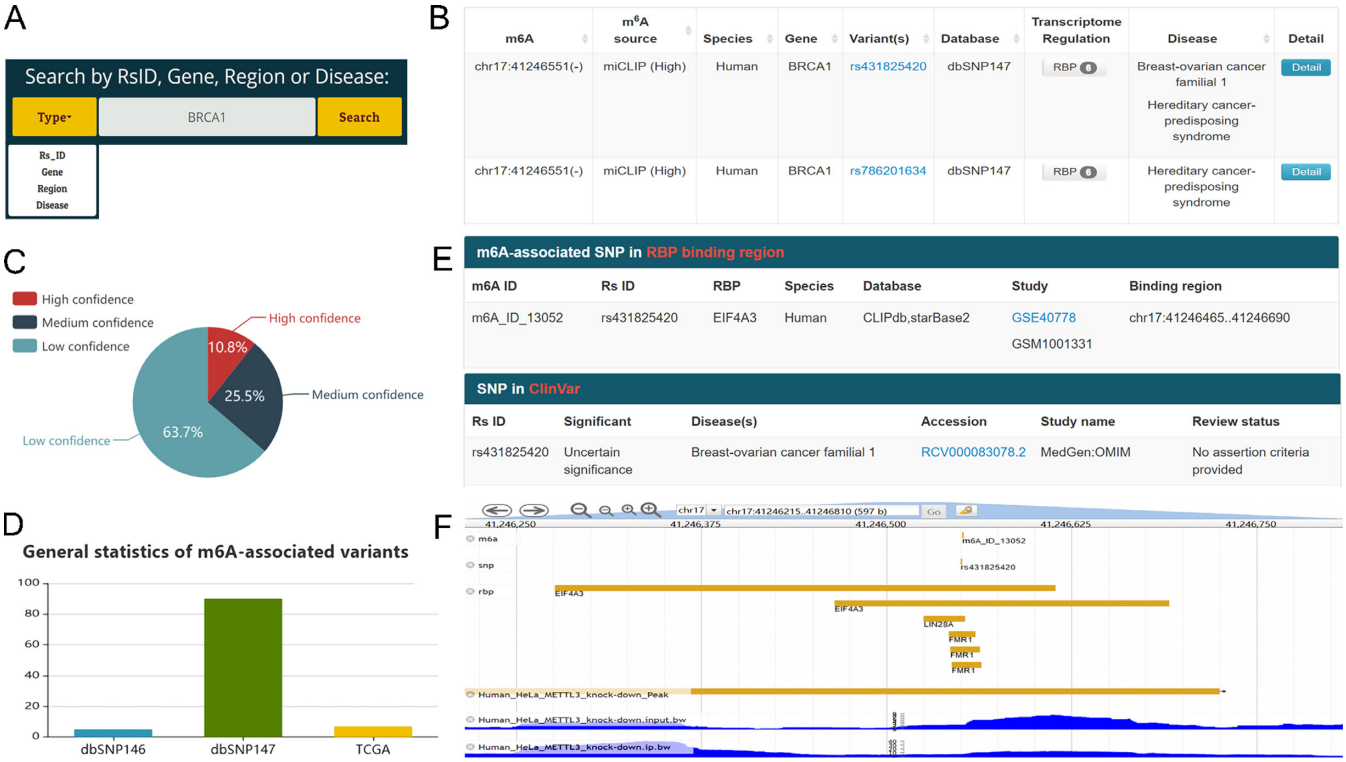

Figure 2. A schematic workflow of the search interface in m6AVar. (**A**) m6AVar provides the four search modes of Rs ID, Gene, Region and Disease. (**B**) Snapshot of search results for '***BRCA1***' using the 'Gene' search mode. Basic information on all of the m^{6}A-associated variants located in the '***BRCA1***' output presented as a table. (**C** and **D**) Distribution of m^{6}A-associated variants in the different sources and databases. (**E**) Detailed information on m^{6}A-associated variants related to post-transcriptional regulation and disease. (**F**) Visualization of specific m^{6}A-associated variants with JBrowse.

the detailed information on each variant (Figure 2E). Furthermore, m6AVar also allows users to find more disease-related variants directly through 'Disease' mode. In order to facilitate follow-up experimental studies, it allows users to customize results with the advanced search and to sort the table by clicking on the column names. Furthermore, we applied the JBrowse Genome Browser to visualize every m^{6}A-associated variant. Users can select the tracks of interest to be shown, such as gene information, SNP site, m^{6}A site, RBP binding regions, miRNA targets and the MeRIP-Seq peak level from the different samples (Figure 2F).

Browse. The 'Browse' page displays: (i) Summary of m^{6}A associated variants from three m^{6}A sources (with a high, medium and low confidence level) (Supplementary Figure S2). (ii) Statistical graphs showing the overall functional gain and loss variants' frequency distribution in a circular layout (Supplementary Figure S3), and m^{6}A-associated variants' distribution in gene regions and gene types as well as other databases (Supplementary Figure S4). (iii) Browse m^{6}A-associated variants by gene types. To retrieve data more efficiently, various filters, such as gene types, associations and confidence levels are provided (Supplementary Figure S5).

Download. all data in the database can be downloaded from the 'Download' page, and a detailed introduction of m6AVar database as well as tutorial are available on the 'Help' page.

DISCUSSION

m6AVar is a comprehensive database of the m^{6}A-associated variants that localize in the vicinity of m^{6}A sites and potentially influence m^{6}A modification in human and mouse. Currently, m6AVar holds ∼352 000 m^{6}A-associated germline variants and ∼62 000 m^{6}A-associated somatic variants, most of them were enriched in protein-coding genes (dbSNP147, 95.77%; dbSNP146, 92.12% and TCGA, 98.89%). The m^{6}A-associated variants that can potentially affect RBP-binding regions, miRNA-targets and splicing sites were discovered by systematic association analyses. Furthermore, diseased-related variants from GWAS and ClinVar have been intersected with the m^{6}A-associated variants to identify the pathogenic variations contributing to dysregulation of m^{6}A modification.

m6AVar has the following advantages in comparison with MeT-DB and RMBase. (i) m6AVar is a specific database dedicated to the investigation of the functional association between variants and m^{6}A modification. (ii) m6AVar integrates somatic variants of 34 cancers from TCGA, which will help to reveal the potential mechanisms of m^{6}A in cancer. (iii) m6AVar provides detail annotations and genomic coordinates for each variant and related m^{6}A site. This will help biologists determine its relevant biological features. (iv) m6AVar integrates the results from association analyses with RBP-binding regions, miRNA-targets and splicing sites, revealing the potential relationship among variants, m^{6}A modification and other post-transcriptional reg-

ulation. (v) More than 2000 disease-related variants have been identified by linking the m^{6}A-associated variants with GWAS and ClinVar data, which may assist the community in identifying the functional disease-causing variants. (vi) m6AVar is a user-friendly database with multiple statistical diagrams and genome browser through which users can browse all of the m^{6}A-associated variants and search interested data by various criteria.

In conclusion, m6AVar provides useful information on m^{6}A-associated variants to help experimental biologists interpret the disease-related variants by m^{6}A function and explore the molecular mechanism of m^{6}A modification. m6AVar will be continually updated whenever new high-throughput m^{6}A sites data and variants data are made available in public databases.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

FUNDING

National Natural Science Foundation of China [31771462, 81772614, 31471252, U1611261]; National Key Research and Development Program [2017YFA0106700]; Guangdong Natural Science Foundation [2014TQ01R387, 2014A030313181]; Science and Technology Program of Guangzhou, China [201604020003]. Funding for open access charge: National Natural Science Foundation of China [31471252].

Conflict of interest statement. None declared.

REFERENCES

1. Haraksingh,R.R. and Snyder,M.P. (2013) Impacts of variation in the human genome on gene regulation. *J. Mol. Biol.*, **425**, 3970–3977.
2. Sauna,Z.E. and Kimchi-Sarfaty,C. (2011) Understanding the contribution of synonymous mutations to human disease. *Nat. Rev. Genet.*, **12**, 683–691.
3. Mao,F., Xiao,L., Li,X., Liang,J., Teng,H., Cai,W. and Sun,Z.S. (2016) RBP-Var: a database of functional variants involved in regulation mediated by RNA-binding proteins. *Nucleic Acids Res.*, **44**, D154–D163.
4. 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.
5. Ramaswami,G., Deng,P., Zhang,R., Anna Carbone,M., Mackay,T.F. and Li,J.B. (2015) Genetic mapping uncovers cis-regulatory landscape of RNA editing. *Nat. Commun.*, **6**, 8194.
6. Zhong,S., Li,H., Bodi,Z., Button,J., Vespa,L., Herzog,M. and Fray,R.G. (2008) MTA is an Arabidopsis messenger RNA adenosine methylase and interacts with a homolog of a sex-specific splicing factor. *Plant Cell*, **20**, 1278–1288.
7. Ping,X.L., Sun,B.F., Wang,L., Xiao,W., Yang,X., Wang,W.J., Adhikari,S., Shi,Y., Lv,Y., Chen,Y.S. *et al.* (2014) Mammalian WTAP is a regulatory subunit of the RNA N6-methyladenosine methyltransferase. *Cell Res.*, **24**, 177–189.
8. Zheng,G., Dahl,J.A., Niu,Y., Fedorcsak,P., Huang,C.M., Li,C.J., Vagbo,C.B., Shi,Y., Wang,W.L., Song,S.H. *et al.* (2013) ALKBH5 is a mammalian RNA demethylase that impacts RNA metabolism and mouse fertility. *Mol. Cell*, **49**, 18–29.
9. 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.
10. 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.
11. 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.
12. 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.
13. 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.
14. Jia,G., Fu,Y., Zhao,X., Dai,Q., Zheng,G., Yang,Y., Yi,C., Lindahl,T., Pan,T., Yang,Y.G. *et al.* (2011) N6-methyladenosine in nuclear RNA is a major substrate of the obesity-associated FTO. *Nat. Chem. Biol.*, **7**, 885–887.
15. Xiao,W., Adhikari,S., Dahal,U., Chen,Y.S., Hao,Y.J., Sun,B.F., Sun,H.Y., Li,A., Ping,X.L., Lai,W.Y. *et al.* (2016) Nuclear m(6)A reader YTHDC1 regulates mRNA splicing. *Mol. Cell*, **61**, 507–519.
16. Lin,S., Choe,J., Du,P., Triboulet,R. and Gregory,R.I. (2016) The m(6)A methyltransferase METTL3 promotes translation in human cancer cells. *Mol. Cell*, **62**, 335–345.
17. Zhang,C., Samanta,D., Lu,H., Bullen,J.W., Zhang,H., Chen,I., He,X. and Semenza,G.L. (2016) Hypoxia induces the breast cancer stem cell phenotype by HIF-dependent and ALKBH5-mediated m(6)A-demethylation of NANOG mRNA. *Proc. Natl. Acad. Sci. U.S.A.*, **113**, E2047–E2056.
18. Merkestein,M., Laber,S., McMurray,F., Andrew,D., Sachse,G., Sanderson,J., Li,M., Usher,S., Sellayah,D., Ashcroft,F.M. *et al.* (2015) FTO influences adipogenesis by regulating mitotic clonal expansion. *Nat. Commun.*, **6**, 6792.
19. Li,Z., Weng,H., Su,R., Weng,X., Zuo,Z., Li,C., Huang,H., Nachtergaele,S., Dong,L., Hu,C. *et al.* (2017) FTO plays an oncogenic role in acute myeloid leukemia as a N6-methyladenosine RNA demethylase. *Cancer Cell*, **31**, 127–141.
20. Liu,H., Flores,M.A., Meng,J., Zhang,L., Zhao,X., Rao,M.K., Chen,Y. and Huang,Y. (2015) MeT-DB: a database of transcriptome methylation in mammalian cells. *Nucleic Acids Res.*, **43**, D197–D203.
21. Sun,W.J., Li,J.H., Liu,S., Wu,J., Zhou,H., Qu,L.H. and Yang,J.H. (2016) RMBase: a resource for decoding the landscape of RNA modifications from high-throughput sequencing data. *Nucleic Acids Res.*, **44**, D259–D265.
22. Li,J.H., Liu,S., Zhou,H., Qu,L.H. and Yang,J.H. (2014) starBase v2.0: decoding miRNA-ceRNA, miRNA-ncRNA and protein-RNA interaction networks from large-scale CLIP-Seq data. *Nucleic Acids Res.*, **42**, D92–D97.
23. Yang,Y.C., Di,C., Hu,B., Zhou,M., Liu,Y., Song,N., Li,Y., Umetsu,J. and Lu,Z.J. (2015) CLIPdb: a CLIP-seq database for protein-RNA interactions. *BMC Genomics*, **16**, 51.
24. Welter,D., MacArthur,J., Morales,J., Burdett,T., Hall,P., Junkins,H., Klemm,A., Flicek,P., Manolio,T., Hindorff,L. *et al.* (2014) The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res.*, **42**, D1001–D1006.
25. Johnson,A.D. and O'Donnell,C.J. (2009) An open access database of genome-wide association results. *BMC Med. Genet.*, **10**, 6.
26. Mailman,M.D., Feolo,M., Jin,Y., Kimura,M., Tryka,K., Bagoutdinov,R., Hao,L., Kiang,A., Paschall,J., Phan,L. *et al.* (2007) The NCBI dbGaP database of genotypes and phenotypes. *Nat. Genet.*, **39**, 1181–1186.
27. Becker,K.G., Barnes,K.C., Bright,T.J. and Wang,S.A. (2004) The genetic association database. *Nat. Genet.*, **36**, 431–432.
28. Landrum,M.J., Lee,J.M., Benson,M., Brown,G., Chao,C., Chitipiralla,S., Gu,B., Hart,J., Hoffman,D., Hoover,J. *et al.* (2016) ClinVar: public archive of interpretations of clinically relevant variants. *Nucleic Acids Res.*, **44**, D862–D868.
29. Tyner,C., Barber,G.P., Casper,J., Clawson,H., Diekhans,M., Eisenhart,C., Fischer,C.M., Gibson,D., Gonzalez,J.N., Guruvadoo,L. *et al.* (2017) The UCSC Genome Browser database: 2017 update. *Nucleic Acids Res.*, **45**, D626–D634.

30. Leung,Y.Y., Kuksa,P.P., Amlie-Wolf,A., Valladares,O., Ungar,L.H., Kannan,S., Gregory,B.D. and Wang,L.S. (2016) DASHR: database of small human noncoding RNAs. *Nucleic Acids Res.*, **44**, D216–D222.
31. Kozomara,A. and Griffiths-Jones,S. (2014) miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res.*, **42**, D68–D73.
32. Chan,P.P. and Lowe,T.M. (2016) GtRNAdb 2.0: an expanded database of transfer RNA genes identified in complete and draft genomes. *Nucleic Acids Res.*, **44**, D184–D189.
33. Sai Lakshmi,S. and Agrawal,S. (2008) piRNABank: a web resource on classified and clustered Piwi-interacting RNAs. *Nucleic Acids Res.*, **36**, D173–D177.
34. Wang,K., Li,M. and Hakonarson,H. (2010) ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.*, **38**, e164.
35. Siepel,A., Bejerano,G., Pedersen,J.S., Hinrichs,A.S., Hou,M., Rosenbloom,K., Clawson,H., Spieth,J., Hillier,L.W., Richards,S. *et al.* (2005) Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res.*, **15**, 1034–1050.
36. 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.
37. 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.
38. Chun,S. and Fay,J.C. (2009) Identification of deleterious mutations within three human genomes. *Genome Res.*, **19**, 1553–1561.
39. 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.
40. Liu,X., Wu,C., Li,C. and Boerwinkle,E. (2016) dbNSFP v3.0: a one-stop database of functional predictions and annotations for human nonsynonymous and splice-site SNVs. *Hum. Mutat.*, **37**, 235–241.
41. Schwartz,S., Mumbach,M.R., Jovanovic,M., Wang,T., Maciag,K., Bushkin,G.G., Mertins,P., Ter-Ovanesyan,D., Habib,N., Cacchiarelli,D. *et al.* (2014) Perturbation of m6A writers reveals two distinct classes of mRNA methylation at internal and 5' sites. *Cell Rep.*, **8**, 284–296.
42. Bolger,A.M., Lohse,M. and Usadel,B. (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, **30**, 2114–2120.
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. Zhang,Y., Liu,T., Meyer,C.A., Eeckhoute,J., Johnson,D.S., Bernstein,B.E., Nusbaum,C., Myers,R.M., Brown,M., Li,W. *et al.* (2008) Model-based analysis of ChIP-Seq (MACS). *Genome Biol.*, **9**, R137.
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. Jalili,V., Matteucci,M., Masseroli,M. and Morelli,M.J. (2015) Using combined evidence from replicates to evaluate ChIP-seq peaks. *Bioinformatics*, **31**, 2761–2769.
47. 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.