Method

Allele-specific RNA N^6-methyladenosine modifications reveal functional genetic variants in human tissues

Shuo Cao,[1,3] Haoran Zhu,[1,3] Jinru Cui,[1,3] Sun Liu,[1,3] Yuhe Li,[1] Junfang Shi,[1] Junyuan Mo,[1] Zihan Wang,[1] Hailan Wang,[1] Jiaxin Hu,[1] Lizhi Chen,[1] Yuan Li,[1] Laixin Xia,[1] and Shan Xiao[1,2]

[1] *Department of Developmental Biology, School of Basic Medical Sciences, Southern Medical University, Guangzhou 510515, China;* [2] *Guangdong Provincial Key Laboratory of Cardiac Function and Microcirculation, Guangzhou 510515, China*

An intricate network of *cis*- and *trans*-elements acts on RNA N^6-methyladenosine (m^6A), which in turn may affect gene expression and, ultimately, human health. A complete understanding of this network requires new approaches to accurately measure the subtle m^6A differences arising from genetic variants, many of which have been associated with common diseases. To address this gap, we developed a method to accurately and sensitively detect transcriptome-wide allele-specific m^6A (ASm^6A) from MeRIP-seq data and applied it to uncover 12,056 high-confidence ASm^6A modifications from 25 human tissues. We also identified 1184 putative functional variants for ASm^6A regulation, a subset of which we experimentally validated. Importantly, we found that many of these ASm^6A-associated genetic variants were enriched for common disease–associated and complex trait–associated risk loci, and verified that two disease risk variants can change m^6A modification status. Together, this work provides a tool to detangle the dynamic network of RNA modifications at the allelic level and highlights the interplay of m^6A and genetics in human health and disease.

[Supplemental material is available for this article.]

The surge of human genome sequencing data has revealed billions of genetic variants in the population. Genetic variants contribute to the phenotypic diversity in the population, and more than 10,000 genetic variants have been revealed to be associated with diseases by genome-wide association study (GWAS) (Maurano et al. 2012; Wang et al. 2021). At the molecular level, genetic variants can affect a number of processes, including DNA methylation, histone modification, gene expression, and post-transcriptional splicing (McDaniell et al. 2010; Shoemaker et al. 2010; The ENCODE Project Consortium 2012; McVicker et al. 2013; Leung et al. 2015; Roadmap Epigenomics Consortium et al. 2015; Schultz et al. 2015; GTEx Consortium 2017; Gate et al. 2018; Yang et al. 2019; Liu et al. 2020c; Amoah et al. 2021). However, the effect of many variants is often subtle, making it challenging to systematically identify and interpret functional variants, leaving a great deal of genetic variants with unknown functions.

N^6-Methyladenosine (m^6A), the most abundant modification on mammalian mRNA and long noncoding RNA (lncRNA), is catalyzed by the METTL3/METTL14/WTAP methyltransferase complex preferentially at RRACH motifs (Liu et al. 2014; Ping et al. 2014), is removed by the demethylases FTO and ALKBH5 (Jia et al. 2011; Zheng et al. 2013), and has been implicated in the regulation of a series of cotranscriptional and post-transcriptional processes (Dominissini et al. 2012; Batista et al. 2014; Wang et al. 2014a,b, 2015c; Alarcón et al. 2015; Geula et al. 2015; Liu et al. 2015; Meyer et al. 2015; Zhou et al. 2015; Vu et al. 2017; Ke et al. 2017; Xiang et al. 2017; Xu et al. 2017; Yoon et al. 2017; Zhang et al. 2017; Li et al. 2017b; Huang et al. 2018; Liu et al. 2020a). m^6A has been widely explored in human tissues via high-throughput sequencing (MeRIP/m^6A-seq), revealing an enrichment for expression quantitative trait loci (eQTLs) and GWAS single-nucleotide polymorphisms (SNPs) (Xiao et al. 2019; Liu et al. 2020b; Zhang et al. 2020a). Many human genetic variants may affect m^6A levels, leading to changes in gene expression (Zhang et al. 2020b; Xiong et al. 2021). However, accurately measuring the m^6A changes in response to genetic variants and determining the interplay between m^6A and genetic variation remains challenging.

Allele-specific analyses measure the effects of the imbalance between allelic variants in the same genome, mitigating context-specific effects (McDaniell et al. 2010; Onuchic et al. 2018; Zhou et al. 2018; Yang et al. 2019; Sun and Zhang 2020; Amoah et al. 2021). We hypothesized that an allele-specific m^6A modification (ASm^6A) analysis may greatly improve the accuracy and sensitivity of conventional MeRIP/m^6A-seq assays (McIntyre et al. 2020), providing improved resolution of m^6A quantification, especially in human tissue samples.

In this study, we developed a transcriptome-wide ASm^6A identification approach to accurately and sensitively detect m^6A imbalances between two alleles in human tissues, which could shed light on the exploration of functional genetic variants associated with human health and disease and the *trans*-factors involved in m^6A regulation.

[3] **These authors contributed equally to this work.**
Corresponding authors: asdfg@smu.edu.cn, xialx@smu.edu.cn, gavin@smu.edu.cn
Article published online before print. Article, supplemental material, and publication date are at https://www.genome.org/cgi/doi/10.1101/gr.277704.123.

Results

Generation and evaluation of an ASm^{6}A analysis pipeline

To measure the m^{6}A changes in response to genetic variation, we developed a pipeline to identify genome-wide ASm^{6}A from m^{6}A sequencing (MeRIP-seq) data (Fig. 1A). First, high-confidence heterozygous variants were identified by variant calling followed by stringent filtering to minimize the bias introduced through transcription and mapping (Castel et al. 2015). Then the allelic read counts from m^{6}A IP and input sequencing data were calculated, and a Fisher's exact test was performed to assess the modification difference between the two alleles at each heterozygous site. Last, the significant allelic m^{6}A imbalance events within m^{6}A peaks were retained and called as ASm^{6}As (Fig. 1A).

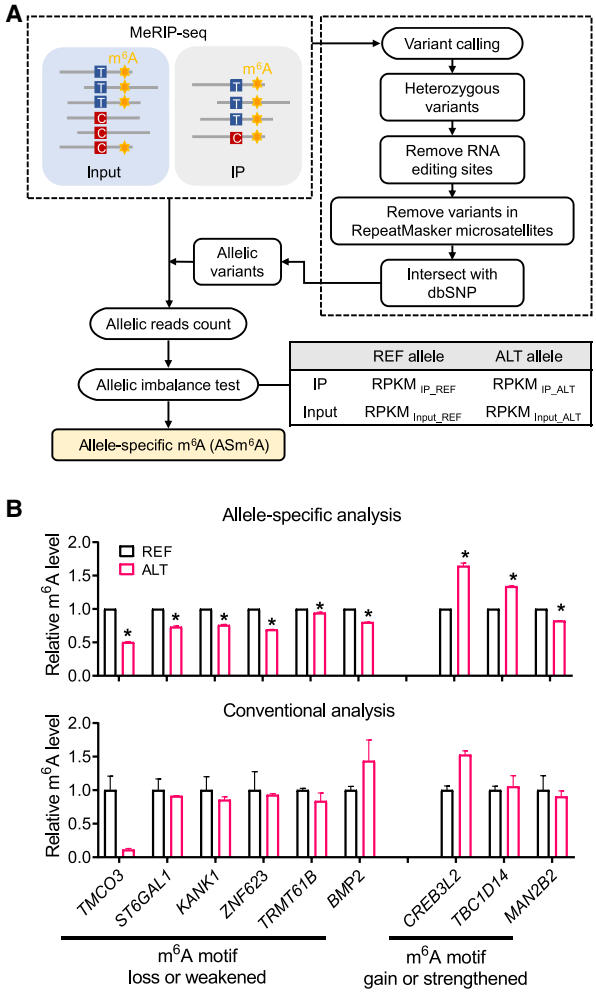

Figure 1. ASm^{6}As analysis pipeline generation and evaluation. (*A*) Pipeline for ASm^{6}As identification. Allelic variants were identified by variant calling and stringent filtering. Allelic reads were then counted from sequencing data of input and IP, and ASm^{6}As were predicted using a Fisher's exact test. (*B*) Comparison between allele-specific analysis and conventional analysis. In the m^{6}A motif loss or weakened group, the potential m^{6}A motifs (reference to alternative allele) were found in *TMCO3* (AAACT to AAATT), *ST6GAL1* (GGACT to GGGCT), *KANK1* (GAACT to GAGCT), *ZNF623* (GGACC to AGACC), *TRMT61B* (GAACT to GAGCT), and *BMP2* (AGACT to AGTCT). In the m^{6}A motif gain or strengthen group, they were found in *CREB3L2* (GCACT to GAACT), *TBC1D14* (CGACT to GGACT), and *MAN2B2* (GAGCA to GAACA). The relative m^{6}A level was calculated by dividing the reads of IP by input, then normalized to the reference allele. Data are mean ± SD from two independent experiments. In the allele-specific analysis, (*) $P < 0.05$ in both replicates as assessed by Fisher's exact test.

To compare the performance of ASm^{6}A analysis with conventional MeRIP analysis in identifying genetic variant–associated m^{6}A change, we developed a reporter assay using alleles predicted to affect m^{6}A through changes of the m^{6}A motif. In the "m^{6}A motif loss or weakened" group, the alternative allele disrupted the m^{6}A motif RRACH, such as rs11164135 on *TMCO3*, which contains a C-to-T SNP in the fourth position of the motif (i.e., AAATT instead of the reference AAACT sequence). In the "m^{6}A motif gain or strengthened" group, the alternative allele contained a stronger m^{6}A motif. We cloned the alternative and reference m^{6}A peak regions into reporter constructs and cotransfected them into HEK293T cells. The RNA was then extracted, fragmented, and subjected to MeRIP and ASm^{6}A analysis. In the conventional analysis group, the REF and ALT reporters were transfected into cells and subjected to the MeRIP procedure separately, and then, the m^{6}A levels on the REF and ALT reporters were compared. We found that the ASm^{6}A analysis showed expected m^{6}A-level changes in eight of nine reporters (Fisher's exact test), even with minor differences between the two alleles, across two replicates (Fig. 1B). In contrast, the conventional MeRIP analysis showed more unexpected m^{6}A changes and required more replicates to analyze the statistical difference in m^{6}A changes between the two alleles. Thus, our ASm^{6}A analysis pipeline can sensitively and accurately detect m^{6}A changes associated with genetic variants.

ASm^{6}A profiling in human tissues

We next used this pipeline to explore the profile of ASm^{6}As across diverse human tissues. We collected 90 reported MeRIP-seq data sets covering 25 tissues from 18 individuals (Xiao et al. 2019; Liu et al. 2020b; Zhang et al. 2020a). After passing these data through our pipeline, we obtained a total of 12,056 ASm^{6}As (Fig. 2A). A representative ASm^{6}A, shown in Figure 2B, had a higher level of m^{6}A on allele T than on allele C across three tissues (brain cerebellum, heart, and liver) from the same individual. To validate these ASm^{6}As, we randomly chose several ASm^{6}As from a kidney sample and performed m^{6}A immunoprecipitation followed by Sanger sequencing. An imbalance in the m^{6}A levels on two alleles was observed (Fig. 2C; Supplemental Fig. S1A), which was consistent with the results from the pipeline analysis.

We have also applied our method to the m^{6}A methylome generated by the base-resolution and quantitative m^{6}A sequencing method GLORI-seq and m^{6}A-SAC-seq in HEK293T and HeLa cells (Supplemental Fig. S1B; Ge et al. 2023; Liu et al. 2023). A total of 91 ASm^{6}As were identified. Sixty-six of them have at least one nearby m^{6}A site with >10% m^{6}A imbalance between the two alleles. And 27 ASm^{6}As identified by MeRIP-seq in the human tissues were supported by this analysis. Two representative ASm^{6}A sites are shown in Supplemental Figure S1C.

Analysis of the full set of ASm^{6}As in human tissues showed that they were enriched for a GGAC motif (Supplemental Fig. S2A) and preferentially found at the CDS and 3′ UTR, peaking at stop codon regions (Supplemental Fig. S2B,C), properties that are similar to those of typical m^{6}A modifications (Dominissini et al. 2012). As allelic imbalance analysis can capture the effect of rare variants (Sun et al. 2016; Onuchic et al. 2018), we annotated these ASm^{6}A variants with allele frequency from the Genome Aggregation Database (gnomAD) (Karczewski et al. 2020) and found 2123 were rare variants (minor allele frequency [MAF] < 5%). The

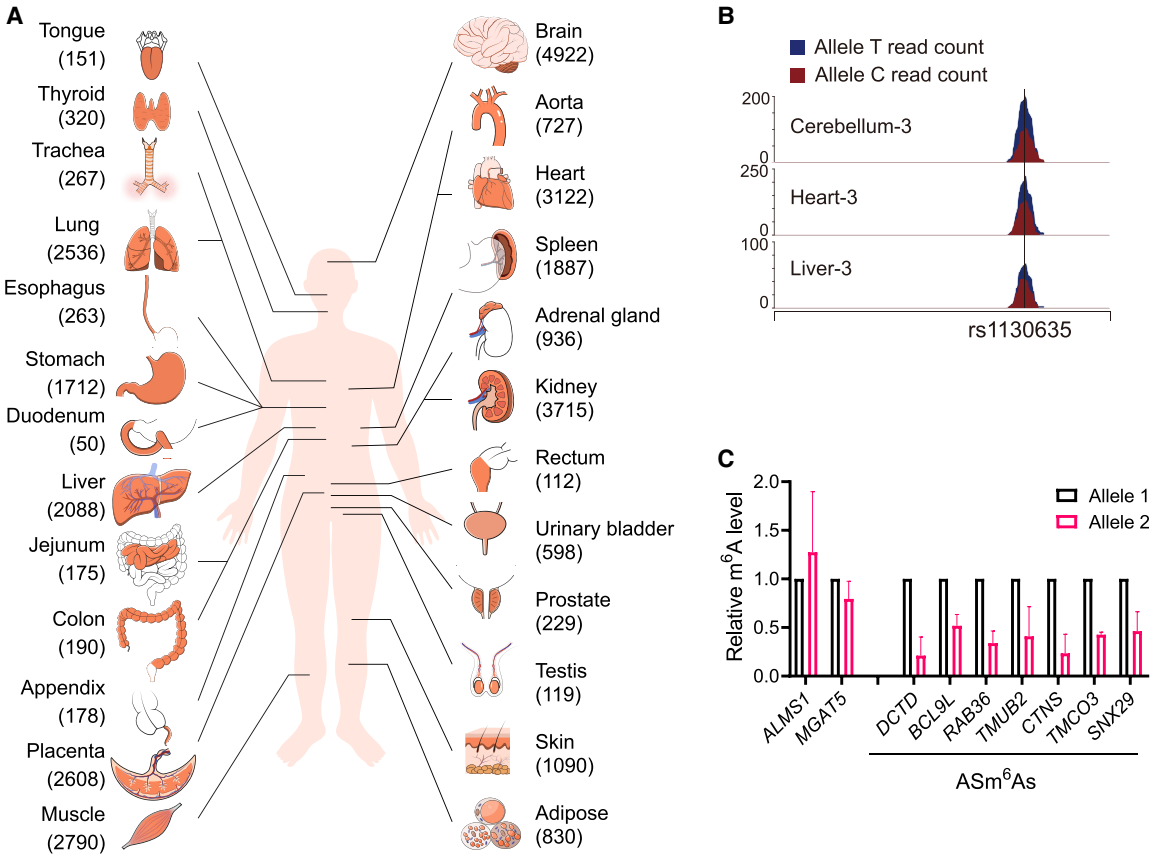

Figure 2. ASm^{6}A profiling in human tissues. (*A*) Numbers of ASm^{6}As identified in 25 human tissues. (*B*) ASm^{6}A modification at the rs1130635 locus in three different tissue samples from the same individual. The horizontal coordinate represents the genomic position, and the vertical coordinate represents the number of sequence reads supporting two alleles (T and C) in the IP sample. The T/C read counts in input samples were 40/36, 26/31, and 40/45 in Cerebellum-3, Heart-3, and Liver-3, respectively. (*C*) MeRIP followed by Sanger sequencing showing the relative m^{6}A level of allele 2 versus allele 1 at genes in the Kidney-2 sample.

fraction of rare variants in ASm^{6}A variants ranged from zero to 24.3% in different samples (Supplemental Fig. S2D). We then analyzed the host genes of ASm^{6}As in different tissues and found that these genes were enriched in tissue-specific processes, suggesting a potential tissue-specific role for ASm^{6}A. For example, the ASm^{6}A host genes in the brain were enriched in neuron differentiation, and the ASm^{6}A host genes in the heart were enriched in heart morphogenesis (Supplemental Fig. S2E).

Identification of putative functional variants for ASm^{6}A regulation

We next sought to determine the genetic variants that can regulate the m^{6}A level at the heterozygous sites. We reasoned that if a variant in a heterozygous site was functional in m^{6}A regulation, we would consistently observe an allelic imbalance in the m^{6}A modification across samples carrying it. To identify such sites, we adapted reported methods (Fig. 3A; Amoah et al. 2021; Li et al. 2021). For each ASm^{6}A, the concordance of m^{6}A methylation bias in different samples was analyzed. Genetic variants with high concordance (>90%) were considered as putative functional variants. A total of 1184 variants were identified across samples (Supplemental Table S1). For example, rs1739047003 showed an obvious imbalance in m^{6}A between the reference allele C and the alternative allele T (Fig. 3B). Among these, 217 were rare variants.

In addition, the majority of these putative functional variants (n = 770, 65.0%) overlapped with genetic variants associated with other molecular phenotypes, including all eQTLs, splicing quantitative trait loci (sQTLs) in previous work (GTEx Consortium 2017), allele-specific expression (ASE) in AlleleDB, allele-specific RBP binding (ASB) (Bahrami-Samani and Xing 2019; Yang et al. 2019), or GWAS causal variants in the CAUSALdb database (Fig. 3C; Wang et al. 2020). Moreover, these variants were significantly enriched in the functional genetic variants (Fig. 3D). To test the influence of tissue specificity in the analysis, we have also conducted enrichment analysis using the ASm^{6}As and QTLs derived from the same tissues. The results showed enrichment of ASm^{6}As for eQTLs and sQTLs in most tissues (Supplemental Fig. S3A,B). The ASE data we used did not provide tissue information, so we did not perform tissue-specific enrichment analysis.

Together, these results suggest that the putative functional variants may be involved in the regulation of other molecular phenotypes.

Validation of putative functional variants for ASm^{6}A regulation by reporter assay

To validate our analysis method and the predicted functional variants for ASm^{6}A, we generated 26 pairs of reporters containing ASm^{6}A variants that have different concordance ratios across the

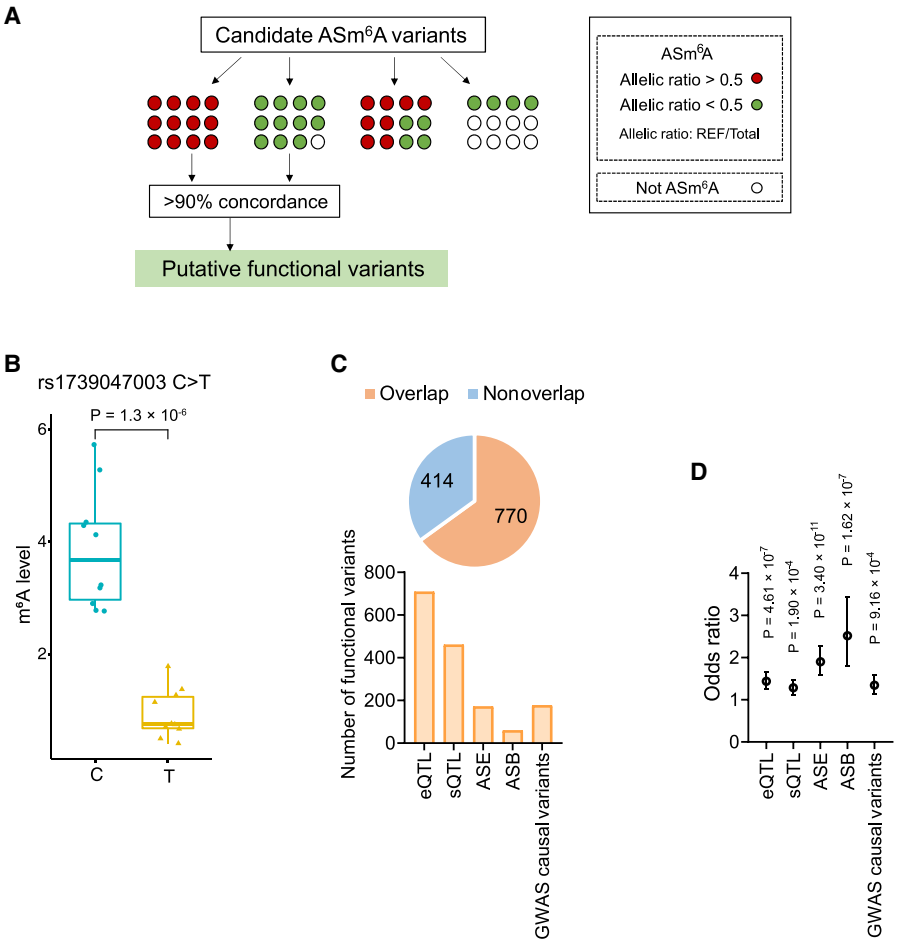

Figure 3. Prediction of functional variants for ASm^{6}A regulation. (*A*) Diagram showing the identification strategy of putative functional variants from all ASm^{6}A variants. (*B*) m^{6}A levels at the reference allele C or alternative allele T of rs1739047003 site in 20 samples. (*C*) The total proportion (*top*) and numbers (*bottom*) of putative functional variants that overlap with annotated functional genetic variants, including eQTL, sQTL, ASE, allele-specific RBP binding (ASB), and GWAS causal variants. The odds ratio, confidence interval, and *P*-value are shown. (*D*) Enrichment of putative functional variants in five genetic annotations. The error bars indicate the upper and lower bounds of the 95% confidence interval.

samples (ranging from >50% to >90%) and eight pairs of reporter-harboring heterozygous sites not identified as ASm^{6}As in any human tissues (not-ASm^{6}A variants), and cotransfected them into HEK293T cells, followed by MeRIP. We found that with a rising concordance ratio, the validated fraction also augmented. Also, all putative functional variants (concordance ratio > 90%) gave rise to a similar allelic m^{6}A bias, as was observed in the tissue samples (Fig. 4A). In contrast, only one of the eight (12.5%) not-ASm^{6}A variants generated allele-specific modifications in HEK293T cells (Fig. 4A). The allelic ratio of the validated variants, including functional variants, in tissue samples and HEK293T cells showed a positive correlation ($R^2 = 0.74$) (Fig. 4B). For example, we found the alternative allele G of functional variant rs117097571 located in the gene *TBC1D14* was more likely to be modified by m^{6}A in the tissues. In the validation reporter assay, the m^{6}A level of the G allele was also higher than that of the reference allele C (Fig. 4C). Similarly, functional variant rs370724501, which is a rare C-to-T variant in the protein-coding gene *ZNF217* (MAF = 9.848×10^{-5}), had a higher level of m^{6}A at the reference allele C than at the alternative allele T in both the tissue samples we analyzed and our reporter assay (Fig. 4D).

Prediction of *trans*-factors involved in ASm^{6}A regulation

To further investigate the regulation of ASm^{6}A, we annotated the putative functional variants for ASm^{6}A and the surrounding sequence for RNA-related features (Fig. 5A); 10.85% of the variants were enriched in the m^{6}A consensus motif RRACH (Fig. 5A,B), which may directly mediate the binding of methyltransferases or demethylases to regulate ASm^{6}A.

Moreover, we found that 18.98% of the variants were enriched in RBP binding sites (Fig. 5A,B). To find what RBPs are associated with ASm^{6}A, RBP binding sites contained in the ENCODE Project Consortium (The ENCODE Project Consortium 2012) were used. In the ENCODE Project, almost all the RBP binding sites data were from K562 and/or HepG2 cell lines. So, we performed RBP enrichment analysis using functional variants for ASm^{6}A regulation, which is a subset of variants that has high concordance in various tissue samples. We found that the binding sites of 15 RBPs were significantly enriched for putative functional variants (Fig. 5C). We have also analyzed the enrichment of RBP binding sites in HepG2 cells for all the ASm^{6}As in liver samples (Supplemental Fig. S4A) and found that 10/15 of the RBPs enriched for functional

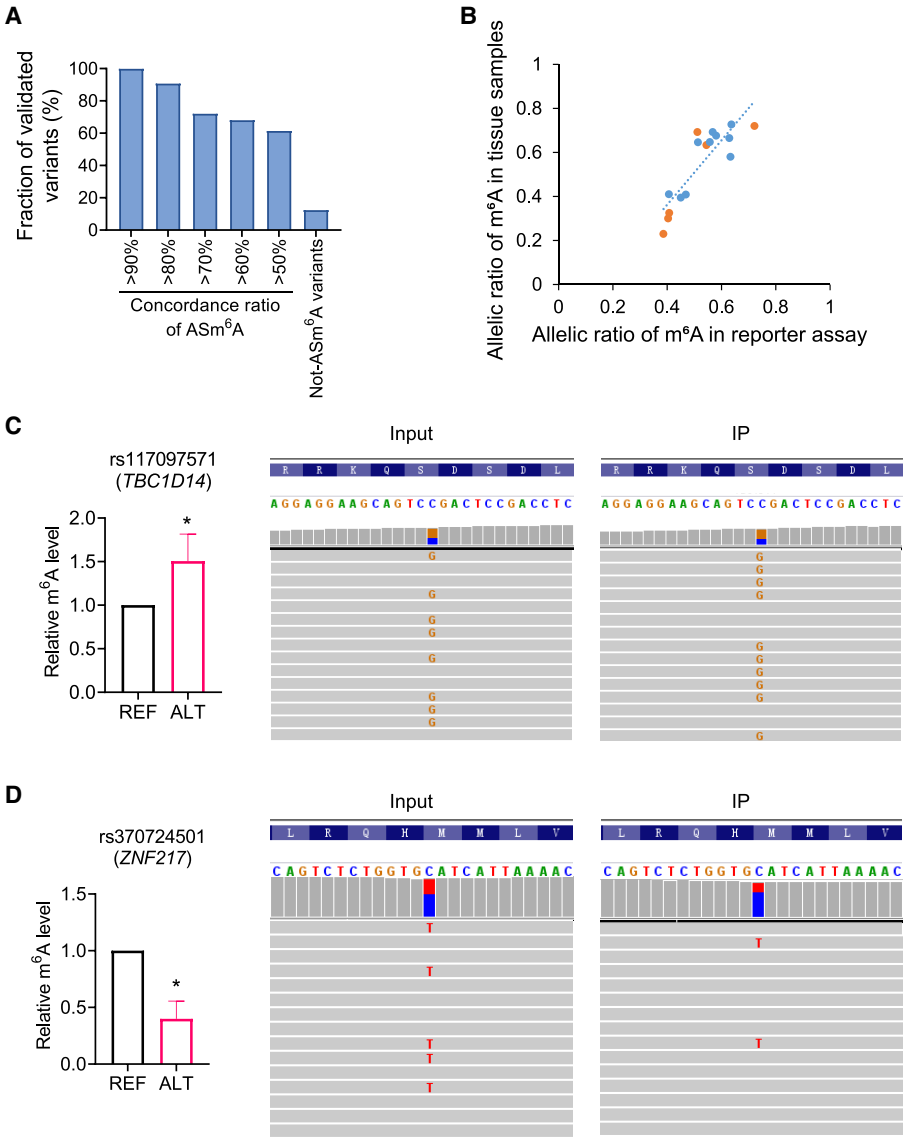

Figure 4. Validation of putative functional variants for ASm^{6}As by reporter assay. (*A*) Fraction of validated ASm^{6}As in the reporter assay. ASm^{6}A reporters with different concordance ratios across samples are shown. (*B*) Correlation of the allelic m^{6}A ratio (REF/total) of validated ASm^{6}A variants in tissue samples and HEK293T reporter assay. The orange dots indicate the functional variants. (*C*,*D*) Examples of validated variants. (*Left*) The relative m^{6}A level of ref versus alt. (*Right*) Captures showing the overall allelic reads and 15 representative reads.

variants also enriched in this analysis. In addition, 80 RBPs were found to have allelic binding sites overlapping with functional variants (Supplemental Table S2). These RBPs included known m^{6}A readers such as IGF2BP1 and IGF2BP2 (Fig. 5C), which are involved in the regulation of mRNA stability and translation (Huang et al. 2018), suggesting that ASm^{6}A may facilitate allelic stability and translation of RNA through altering the binding of m^{6}A readers.

It has been reported that the interactions of RBPs with m^{6}A methyltransferases, demethylases, or readers can alter m^{6}A levels (Zhou et al. 2015; Zheng et al. 2017; Wang et al. 2019; An et al. 2020) or facilitate m^{6}A-mediated translation (Alarcón et al. 2015; Meyer et al. 2015; Choe et al. 2018), so we analyzed the interaction network of the enriched RBPs. We found that 59 RBPs were predicted to interact with the methyltransferase complex, demethylases, or m^{6}A reader proteins (Fig. 5D; Supplemental Table S2). And in these RBPs, 35 have a significant correlation between the expression level of RBPs and the m^{6}A levels of the binding regions in the tissue samples (Fig. 5E; Supplemental Fig. S4B; Supplemental Table S2). We speculate that these RBPs may be involved in ASm^{6}A regulation or function. Supporting these findings, among these RBPs, RBM15 is a known m^{6}A regulator on *Xist* RNA through interaction with METTL3 (Patil et al. 2016), and EIF3H can interact with METTL3 to enhance translation and promote oncogenesis (Choe et al. 2018).

ASm^{6}A is associated with risk variants in common diseases

It has been previously reported that genetic regions corresponding to m^{6}A peaks in human tissues are enriched for disease-associated variants (Xiao et al. 2019; Zhang et al. 2020a). We therefore sought to explore if there is a link between ASm^{6}A and human complex traits and common diseases. To assess this, we considered GWAS risk SNPs collected from the GWAS catalog (MacArthur et al. 2017) and SNPs in linkage disequilibrium (LD) with them. Of all the ASm^{6}As in human tissues, 2560 were also GWAS risk loci (Fig. 6A). Notably, the GWAS risk loci were significantly enriched for all ASm^{6}A variants (odds ratio = 1.58, 95% CI [1.51–1.66], $P = 5.04 \times 10^{-75}$, Fisher's exact test). Furthermore, ASm^{6}A variants in 22 out of the 25 tissues were enriched in GWAS risk loci (Supplemental Fig. S5A). Also, 228 putative functional variants for ASm^{6}A regulation were GWAS risk loci.

To interpret the roles of ASm^{6}A variants in human common diseases, the GWAS risk loci were divided into 17 categories (Xiao et al. 2019). We found that ASm^{6}A variants were enriched in cardiovascular measurement, hematological measurement, and neurological disorder categories in a tissue-specific manner (Fig. 6B). Importantly, the host gene of ASm^{6}A-associated risk variants in the heart were enriched in atrial fibrillation and coronary artery disease (Fig. 6C), whereas those in the brain were enriched in schizophrenia and epilepsy (Supplemental Fig. S5B).

We then performed reporter assays to validate if the ASm^{6}A-associated risk loci could regulate m^{6}A variance in cells. rs2738464 in the 3′ UTR of the *LDLR* gene, a probable GWAS causal variant associated with low-density lipoprotein cholesterol levels and coronary artery disease (Wang et al. 2020), showed higher m^{6}A levels on the alternative allele C (Fig. 6D). We also found that rs1131351 in the 3′ UTR of *SDC1* gene, which is associated with seasonality measurement trait, had a higher m^{6}A level on the alternative allele G (Fig. 6D).

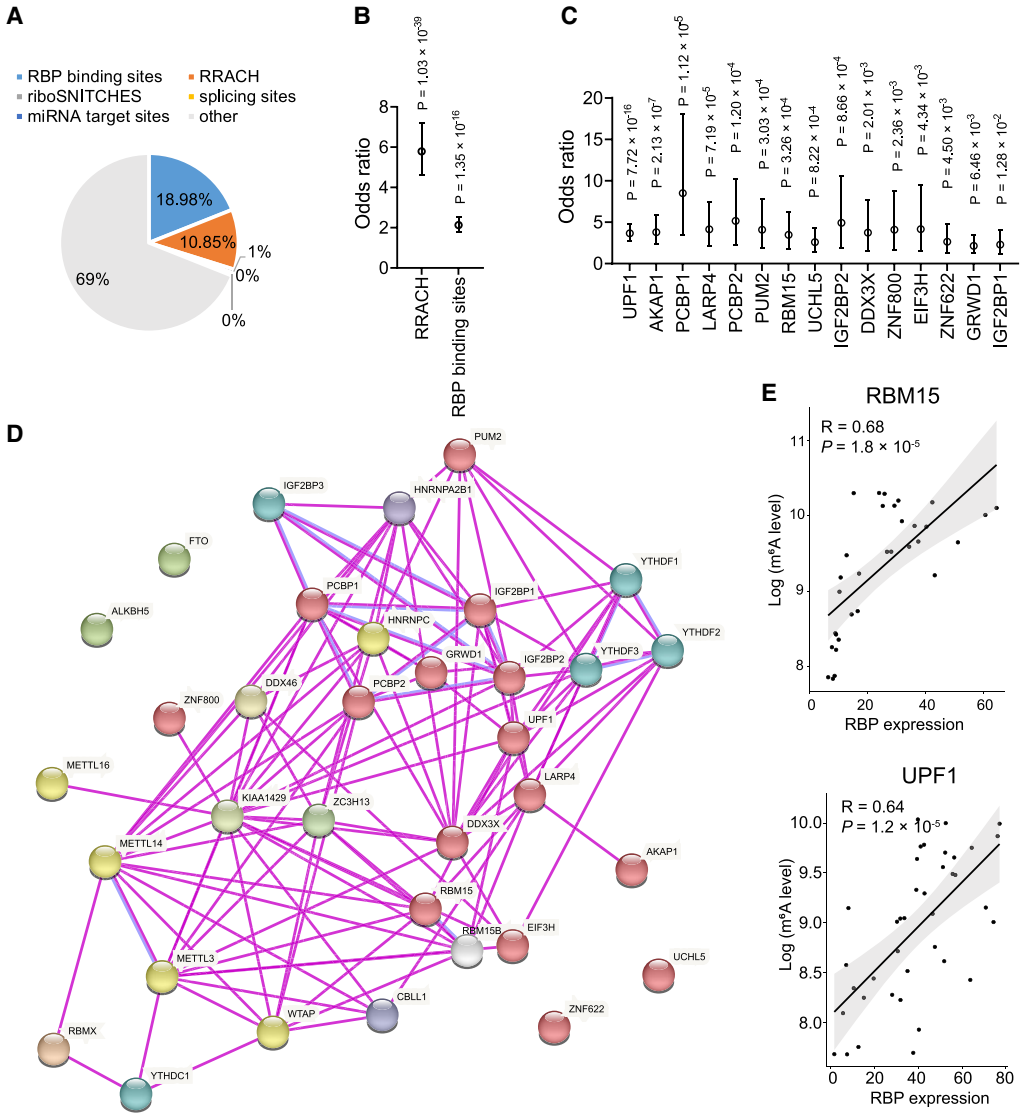

Figure 5. Prediction of *trans*-factors involved in ASm^{6}A regulation. (*A*) Proportion of the putative functional variants for ASm^{6}As that overlap with RNA-related features, including RBP binding sites, RRACH sequence, miRNA target sites, riboSNITCHES variants, and splicing sites. (*B*) Enrichment of the putative functional variants for ASm^{6}As versus random control variants in RRACH sequence and RBP binding sites by Fisher's exact test. (*C*) Enrichment of the putative functional variants for ASm^{6}As in binding sites of different RBPs by Fisher's exact test. (*D*) Protein–protein interactions between the RBPs in *C* and known m^{6}A writers, readers, and erasers. The red sphere indicates the RBPs in *C*. (*E*) Correlation between RBP expression and aggregative m^{6}A level of its targets for RBM15 (*top*) and UPF1 (*bottom*).

We further found that the biological processes of the host genes of ASm^{6}A-associated risk loci in the heart were enriched in cardiocyte differentiation, artery development, and heart morphogenesis (Supplemental Fig. S5C) and that, in the brain, they were enriched in axonogenesis (Supplemental Fig. S5C), suggesting that ASm^{6}As may affect diseases or traits by impacting the genes in which they are located. It is known that m^{6}A can affect the RNA level by affecting transcription and stability (Wang et al. 2014a; Huang et al. 2018; Li et al. 2020; Liu et al. 2020a), and we therefore asked whether ASm^{6}A-associated risk loci can affect host gene expression. We generated the ASE sites from input sequencing data and found that 90 ASm^{6}A-associated risk loci were also ASE sites. This suggested a potential mechanism by which m^{6}As could contribute to disease through regulation of host gene expression.

Discussion

Recent studies have shown that m^{6}A loci are enriched in human genetic variants. Here, we accurately and sensitively detected m^{6}A imbalances between two alleles in human tissues, identified a large number of high-confidence ASm^{6}As, and validated them using a reporter assay. We further investigated the potential for *trans*-proteins acting on ASm^{6}As. Last, we found that ASm^{6}As are enriched in human common traits and diseases.

Technical and biological variations in sample preparation, immunoprecipitation efficiency, and other experimental conditions render conventional MeRIP/m^{6}A-seq analysis methods highly challenging for the accurate detection of m^{6}A changes (McIntyre et al. 2020). The ASm^{6}A analysis we developed overcomes these limitations by directly measuring the imbalance

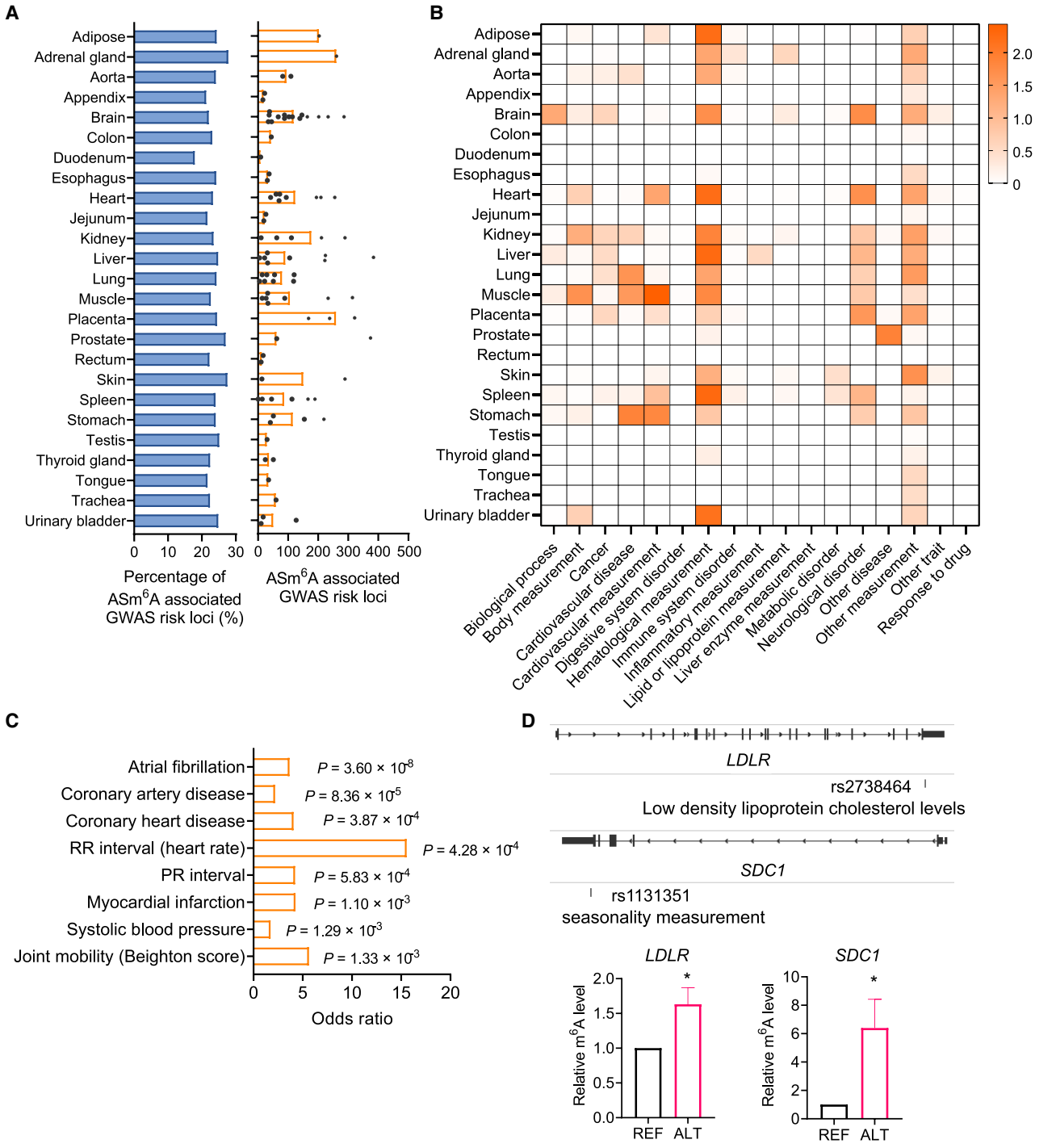

Figure 6. ASm^{6}A analysis reveals m^{6}A-regulating risk variants in common diseases. (*A*) The fraction of ASm^{6}A variant–associated GWAS risk loci in each tissue (*left*), and the number of ASm^{6}A variant–associated GWAS risk loci in each sample (*right*). (*B*) Enrichment of ASm^{6}As from various tissues in 17 GWAS categories. Fisher's exact test, $P < 0.05$. (*C*) Odds ratio showing the enrichment of host gene of ASm^{6}A-associated risk loci from the heart in GWAS traits and diseases. The most significant traits/diseases are shown. (*D*) Validation of the allelic m^{6}A modifications on GWAS risk variants rs2738464 and rs1131351 by reporter assay.

effect between two alleles in exactly the same cellular environment. Thus, ASm^{6}A enables accurate and sensitive detection of m^{6}A changes, facilitating the systematic identification of functional variants and *trans*-factors involved in m^{6}A regulation.

Allelic imbalance analysis is performed in a single sample, which is also beneficial for the detection of effects of rare and de novo variants in diseases (Sun et al. 2016; Onuchic et al. 2018). We found 2123 rare variants are associated with ASm^{6}As; 217 rare variants are putative functional variants; and four were

validated using a reporter assay. Future applications of our approach may be investigating the rare or de novo genetic mutations causing m^{6}A modification changes in disease contexts, especially developmental diseases or tumorigenesis, in which m^{6}A has been reportedly implicated (Batista et al. 2014; Wang et al. 2014b; Geula et al. 2015; Li et al. 2017b; Vu et al. 2017; Xu et al. 2017; Yoon et al. 2017; Zhang et al. 2017).

A large number of putative variants have been implicated in m^{6}A regulation, which presents a challenge for validating their impact on m^{6}A at scale. To address this challenge, we generated a reporter system to validate the putative functional variants for ASm^{6}A in cultured cells. Most of the tested variants observed in human tissues were recapitulated in the reporter system, and the inconsistency of the remaining variants may be owing to the lack of cell type–specific factors in HEK293T cells or the limited insert size cloned into the reporters, as discussed in other genetic variant reporter assays (Amoah et al. 2021).

Up to now, the number of m^{6}A sequencing data in human tissues is relatively small, and base-resolution and quantitative m^{6}A sequencing data in human tissues have yet to emerge. Although ASm^{6}A analysis is not constrained by the sample size, the discovery of functional variants for ASm^{6}A regulation by concordance analysis performs better when having a bigger sample size. We have collected the available m^{6}A sequencing data in human tissues in this study to identify the functional variants, and validated a small part of them by a reporter assay. However, there still may be false-positive or false-negative variants in this analysis, and the tissue specificity of functional variants was not analyzed owing to the limited sample size from different tissues. With the growth of sequencing data, the accuracy of functional variants discovery would increase, and the tissue specificity of the functional variants would offer more fine regulating details for m^{6}A in human tissues.

In summary, our work provides a new method to detect the dynamic network of RNA modifications at the allelic level. Accompanied with the growing atlas of human omics data, ASm^{6}As analysis will contribute to increased resolution in our understanding of the role of m^{6}A and associated genetic variants in human health and diseases.

Methods

MeRIP-seq data collection and alignment

MeRIP-seq raw sequencing reads were downloaded from the NCBI Gene Expression Omnibus (GEO; https://www.ncbi.nlm.nih.gov/geo/; accession numbers GSE114150, GSE122744) and Genome Sequence Archive in BIG Data Center (https://ngdc.cncb.ac.cn/gsa/) accession number CRA001315. Trim_galore (version 0.5.0; http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/) was used to trim adaptors and control read quality. Then, the clean reads were mapped to the human genome (hg19) using STAR (version 2.7.3a) (Dobin et al. 2013) with parameters set as --twopassMode Basic. The hg19 reference genome was used primarily owing to the comprehensive nature of relevant data and tools available at the time. Furthermore, the regions of interest in this study correspond to genic loci that have been thoroughly annotated in hg19, with minimal changes observed in updated reference genomes such as GRCh38. Therefore, the use of alternative reference genomes is unlikely to significantly alter the conclusions made herein. The Picard suite (http://broadinstitute.github.io/picard/) was used to perform "AddOrReplaceReadGroups" and "MarkDuplicates REMOVE_DUPLICATES = true" procedures to add group information and remove duplicate reads. Last, the unique reads were picked out through SAMtools (Li et al. 2009) with "-q 20."

Variant calling from input sequencing data

The SplitNCigarReads procedure in the GATK suite (Van der Auwera et al. 2013) was used to split reads into exon segments. The recalibration of bases was performed by BaseRecalibrator based on known SNPs and indels in The 1000 Genomes Project Consortium (2015). Then, ApplyBQSR was performed based on the covariate file generated by the previous step to apply base quality score recalibration. Variants were called using the HaplotypeCaller procedure. Only variants passing the filter "-stand-call-conf 20 --minimum-mapping-quality 20" were kept for downstream analysis. The variants were retained if they matched the criteria "neither found in UCSC RepeatMasker microsatellites (Tarailo-Graovac and Chen 2009) nor in RNA editing sites (RADAR database)" but were contained in the dbSNP database.

Mapping bias correction using STAR-WASP pipeline

SNVs were called based on the input sequencing data alignments using the GATK suite. Samples from the same individual were combined and used for SNV calling. Then, the SNVs were used as an input list to WASP (van de Geijn et al. 2015), which was embedded in the STAR suite through the "--varVCFfile" parameter.

Identification of ASm^{6}As

First, all biallelic heterozygous sites were picked out using VCFtools (Danecek et al. 2011). The ASEReadCounter procedure contained in GATK was executed to count the reads on the two alleles for each SNV. Only those variants that satisfied the minimum mapping reads on both alleles were considered as effective candidate heterozygous sites (each allele ≥ 2, the sum of two alleles ≥ 10). A Fisher's exact test was used to identify ASm^{6}As by comparing the normalized reads count (RPKM) on reference and alternative allele for IP and input samples. All the *P*-values produced above were supplied to FDR correction using the Python module "fdrcorrection." Significant ASm^{6}As were retained if their adjusted *P*-value was less than 0.05.

We then filtered candidate allele-specific events for location within the m^{6}A peaks. Two peak callers, MACS2 (Zhang et al. 2008) and MeTPeak (Cui et al. 2016a), were used to increase the confidence of m^{6}A peak identification. For MACS2, the *P*-value cutoff was set to 10^{-5}, and -f BAMPE, --nomodel, --call-summits was used. For MeTPeak peak calling, the default parameter settings were used. After "MaxMinNormalization" normalization, consensus m^{6}A peaks were constructed from the above two methods by MSPC. To avoid the influence of N^6,2′-*O*-dimethyladenosine (m^{6}Am) modifications, peak regions containing an adenosine at the transcription start site and having more than five reads in both the IP and input spanning −100 to +100 nt of the transcription start site were identified as m^{6}Am peaks and were removed (Schwartz et al. 2014).

To analyze the m^{6}A imbalance events from the base-resolution quantitative m^{6}A sequencing method, the GLORI-seq (GSM6432590, GSM6432591, GSM6432592, GSM6432595 GSM6432596, and GSM6432597) (Liu et al. 2023) and m^{6}A-SAC-seq data (GSM4950342, GSM4950344, GSM4950345, GSM4950347, GSM4950349, and GSM4950350) (Hu et al. 2022) were download. The allelic variants were identified from the untreated data as mentioned above. The reads from GLORI- or SAC-treated data were assigned to the reference or alternative allele file based on the allelic variants generated. The coverage for A + G in GLORI seq data should be 15

or more in both the REF and ALT files, and reads containing four or more unconverted A's were filtered. For m^{6}A-SAC-seq, the read coverage should be greater than five in both the REF and ALT files. Then the conversion rates calculation and m^{6}A site calling were performed as reported (Hu et al. 2022; Liu et al. 2023). The candidate site should be identified as m^{6}A in at least one of either the REF or ALT files. Then a two-tailed Fisher's exact test was performed to assess the reads of m^{6}A in REF and ALT files, and the significant allelic sites were considered as ASm^{6}A sites.

Determination of ASE

Determination of ASE was performed as reported (Ge et al. 2009; GTEx Consortium 2013). For each heterozygous site that satisfied the minimum read number criteria in input, we calculated a binomial *P*-value by comparing with the expected null ratio (0.5). After FDR correction, significant (FDR < 0.05) ASEs were retained.

Inference of putative functional variants for ASm^{6}A regulation

All of the ASm^{6}A variants were considered as candidate variants. The following criteria were used to identify putative regulating variants for ASm^{6}As: The variants (1) occurred in at least two samples, (2) met the minimal read counts requirement, (3) were identified as ASm^{6}A variants and having concordant allelic direction in at least 90% of the samples overlapping the m^{6}A peaks, and (4) showed no significant difference (*P*-value > 0.1, Fisher's exact test) of allelic m^{6}A level in the concordant samples (Li et al. 2021).

Motif analysis

For each ASm^{6}A, flanking regions from 4-bp upstream to 4-bp downstream (e.g., all 5-mer positions that include the variant) were extracted from hg19 reference genome by BEDTools "getfasta" (Quinlan and Hall 2010) and then used to search for motifs. The strand of each flanking region sequence was assigned based on the overlapping transcript strand information. For the sets of sequences, the RRACH motifs were searched using HOMER findMotifi.pl.

The metagene profile of ASm^{6}A variant sites was generated using the R package GUITAR (Cui et al. 2016b) with default parameters.

ASm^{6}A validation by Sanger sequencing

The tissue samples (Xiao et al. 2019) were used following the ethical regulations of the medicine ethics committee of Nanfang Hospital, Southern Medical University. The RNA extraction and MeRIP were performed as previously described (Xiao et al. 2019). The input and immunoprecipitated RNA were reverse-transcribed into cDNA using a GoScript reverse transcription system (Promega). The MeRIP-qPCR was performed as previously described (Xiao et al. 2019). The sequences around ASm^{6}A variants were amplified by 30 rounds of PCR, cloned into T-vectors, and transfected into DH5α-competent cells. Single clones were subjected to Sanger sequencing using an ABI 3730XL DNA analyzer, and the reads of each ASm^{6}A variant were counted and calculated. The results are shown in Figure 2C.

Reporter assay

Genomic DNA of HEK293T cells were extracted, and then the genomic regions containing the alleles and the m^{6}A peak were cloned into CS2 vectors via homologous recombination after BamHI and XbaI digestion to linearize the vectors. Alleles different from the extracted genome were introduced by site-directed mutagenesis. The reporter plasmids were pooled and transfected into HEK293 cells. Then, RNA was isolated using TRNzol reagent (Tiangen). After DNase I digestion, the RNA was fragmented and subjected to MeRIP as previously described (Xiao et al. 2019). The immunoprecipitated and input RNAs were subjected to library generation. Sequencing was performed on an Illumina Nova platform. The predicted alleles in Figure 1B were from kidney samples and were located in the m^{6}A peaks in HEK293T cells. Two batches of reporter assay were performed in Figure 1B, and five were performed in Figure 4. The primers used are supplied in Supplemental Table S3.

Trim_galore was used for adapter trimming as well as quality control. The "umi_tools extract" was used to extract UMIs and append to the read name. The RT primer sequence was trimmed by fastp (Chen et al. 2018). Then, these clean FASTQ files were aligned to the human reference genome using STAR. "umi_tools dedup" was used to deduplicate reads based on the mapping coordinates and the UMI attached to the reads. The count of reads from IP and input sample BAM files were obtained using BEDTools "multicov," and the m^{6}A level of each fragment was quantified as: $RPKM_{IP}/RPKM_{Input}$. Only the fragments whose m^{6}A level exceeded the average m^{6}A level of *POU5F1* and *GAPDH* were considered to have m^{6}A modifications. Heterozygous variants located within these fragments were regarded as valid variants and used for downstream analysis. For each valid variant, the number of reads on each allele was counted. Loci with reads on any allele below two were discarded. Fisher's exact tests for each heterozygous site were performed to check the significance of m^{6}A modification difference between two alleles. Variants having a significant m^{6}A imbalance and similar modification bias with the majority of the tissue samples in 80% of the replicates were considered validated variants.

For the reporter assay analysis using conventional MeRIP analysis, the REF and ALT allele reporters were transfected to the HEK293T cells separately. Then the two groups of RNA were extracted and subjected to a MeRIP-seq procedure as previously described (Xiao et al. 2019). The m^{6}A level was calculated by $RPKM_{IP}/RPKM_{Input}$ (Wan et al. 2015; Xiao et al. 2019), and then the m^{6}A level of REF and ALT alleles was compared.

Gene Ontology analysis of ASm^{6}A host genes

ASm^{6}As were assigned to genes referring to the human reference annotation (gff3 file). The clusterProfiler (Yu et al. 2012) R package was used to determine the overrepresentation of biological process categories of ASm^{6}A host genes.

Enrichment analysis of ASm^{6}As

The eQTL and sQTL data sets were obtained from the GTEx Consortium (2017). The ASE data sets were obtained from AlleleDB (http://alleledb.gersteinlab.org/). The allele-specific RBP binding data sets were from reported papers (Bahrami-Samani and Xing 2019; Yang et al. 2019). The causal variants were downloaded from CAUSALdb (http://mulinlab.org/causaldb). RBP binding sites were downloaded from The ENCODE Project Consortium (2012). To generate the control variant set, the SNPsnap website (Pers et al. 2015) was used to sample 100 sets of variants, and the MAF, number of variants in linkage disequilibrium (LD buddies), distance to nearest gene, and gene density were considered. To match the location of ASm^{6}A variants, the EAS data set of SNPsnap was divided into promoter-TSS, 5′ UTR, exon, intron, 3′ UTR, TTS, intergenic, and noncoding regions. Then, a two-tailed Fisher's exact test was used to compare the number of ASm^{6}A variants and control variants that overlap with RBP binding sites or other indicated data sets as reported (Zhang et al. 2020b). The

overlaps between ASm^{6}As and rare variants and GWAS-related ASm^{6}A and ASE were not significant.

Enrichment of ASm^{6}A variants in GWAS SNP data sets

For GWAS enrichment analysis, the total ASm^{6}A data set was used. Control variants were randomly sampled by SNPsnap. The GWAS SNPs were collected from the NHGRI GWAS Catalog (MacArthur et al. 2017). Only autosomal SNPs were retained. SNiPA (Arnold et al. 2015) was used to obtain the LD mutations within a 500-kb window with a parameter of $r^2 > 0.8$ in the East Asian population for each GWAS SNP. A Fisher's exact test was used to test enrichment of ASm^{6}A variants in GWAS SNP data sets (Zhang et al. 2020b). Enrichr was used in the enrichment analysis of the host gene of ASm^{6}A-associated risk loci in the GWAS traits/diseases (Kuleshov et al. 2016). R v4.2.0 was used (R Core Team 2022).

Enrichment of ASm^{6}A variants in 17 GWAS categories

GWAS SNPs were grouped into 17 categories according to the NHGRI GWAS catalog (MacArthur et al. 2017). One hundred control SNP sets were generated as in the enrichment analysis above. The number of SNPs in each GWAS category and its control SNP set were determined. Then 100 Fisher's exact tests were performed. If the *P*-value was greater than 0.05 in one test, the corresponding odds ratio was set to zero. The mean odds ratio of 100 Fisher's exact tests was shown (Xiao et al. 2019).

Construction of protein–protein interaction network

Interactions with ASm^{6}A-enriched RBPs and known m^{6}A writers, readers, and erasers (Jia et al. 2011; Zheng et al. 2013; Liu et al. 2014; Ping et al. 2014; Wang et al. 2014a, 2015; Xiao et al. 2016; Li et al. 2017a; Yin et al. 2022) were visualized through the STRINGdb R package (Szklarczyk et al. 2019). Among these active interaction sources, only "experiment" records were shown in the network.

Correlation between RBP expression and m^{6}A level of its targets

The TPM of each RBP in different samples was calculated as was the aggregated m^{6}A score of m^{6}A peaks that overlapped with the RBP-binding sites. Based on the expression of a given RBP and its corresponding aggregated m^{6}A score, a regression analysis was performed, and the *P*-value was calculated.

Data access

All the raw and processed sequencing data generated in this study have been submitted to the NCBI BioProject database (https://www.ncbi.nlm.nih.gov/bioproject/) under accession number PRJNA794937. The custom codes are provided as Supplemental Code.

Competing interest statement

The authors declare no competing interests.

Acknowledgments

This work was supported by the National Key R&D Program of China (2019YFA0111200, 2021YFA0805400), the National Natural Science Foundation of China (82230053, 32222016, 32170845, 32200454), the Guangdong Basic and Applied Basic Research Foundation (2022B1515020107), and the Science and Technology Program of Guangzhou, China (202201010922).

References

The 1000 Genomes Project Consortium. 2015. A global reference for human genetic variation. *Nature* **526:** 68–74. doi:10.1038/nature15393

Alarcón CR, Lee H, Goodarzi H, Halberg N, Tavazoie SF. 2015. N^6-methyladenosine marks primary microRNAs for processing. *Nature* **519:** 482–485. doi:10.1038/nature14281

Amoah K, Hsiao YE, Bahn JH, Sun Y, Burghard C, Tan BX, Yang EW, Xiao X. 2021. Allele-specific alternative splicing and its functional genetic variants in human tissues. *Genome Res* **31:** 359–371. doi:10.1101/gr.265637.120

An S, Huang W, Huang X, Cun Y, Cheng W, Sun X, Ren Z, Chen Y, Chen W, Wang J. 2020. Integrative network analysis identifies cell-specific *trans* regulators of m^{6}A. *Nucleic Acids Res* **48:** 1715–1729. doi:10.1093/nar/gkz1206

Arnold M, Raffler J, Pfeufer A, Suhre K, Kastenmüller G. 2015. *SNiPA*: an interactive, genetic variant-centered annotation browser. *Bioinformatics* **31:** 1334–1336. doi:10.1093/bioinformatics/btu779

Bahrami-Samani E, Xing Y. 2019. Discovery of allele-specific protein-RNA interactions in human transcriptomes. *Am J Hum Genet* **104:** 492–502. doi:10.1016/j.ajhg.2019.01.018

Batista PJ, Molinie B, Wang J, Qu K, Zhang J, Li L, Bouley DM, Lujan E, Haddad B, Daneshvar K, et al. 2014. m^{6}A RNA modification controls cell fate transition in mammalian embryonic stem cells. *Cell Stem Cell* **15:** 707–719. doi:10.1016/j.stem.2014.09.019

Castel SE, Levy-Moonshine A, Mohammadi P, Banks E, Lappalainen T. 2015. Tools and best practices for data processing in allelic expression analysis. *Genome Biol* **16:** 195. doi:10.1186/s13059-015-0762-6

Chen S, Zhou Y, Chen Y, Gu J. 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34:** i884–i890. doi:10.1093/bioinformatics/bty560

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. doi:10.1038/s41586-018-0538-8

Cui X, Meng J, Zhang S, Chen Y, Huang Y. 2016a. A novel algorithm for calling mRNA m^{6}A peaks by modeling biological variances in MeRIP-seq data. *Bioinformatics* **32:** i378–i385. doi:10.1093/bioinformatics/btw281

Cui X, Wei Z, Zhang L, Liu H, Sun L, Zhang SW, Huang Y, Meng J. 2016b. Guitar: an R/Bioconductor package for gene annotation guided transcriptomic analysis of RNA-related genomic features. *Biomed Res Int* **2016:** 8367534. doi:10.1155/2016/8367534

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, et al. 2011. The variant call format and VCFtools. *Bioinformatics* **27:** 2156–2158. doi:10.1093/bioinformatics/btr330

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29:** 15–21. doi:10.1093/bioinformatics/bts635

Dominissini D, Moshitch-Moshkovitz S, Schwartz S, Salmon-Divon M, Ungar L, Osenberg S, Cesarkas K, Jacob-Hirsch J, Amariglio N, Kupiec M, et al. 2012. Topology of the human and mouse m^{6}A RNA methylomes revealed by m^{6}A-seq. *Nature* **485:** 201–206. doi:10.1038/nature11112

The ENCODE Project Consortium. 2012. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489:** 57–74. doi:10.1038/nature11247

Gate RE, Cheng CS, Aiden AP, Siba A, Tabaka M, Lituiev D, Machol I, Gordon MG, Subramaniam M, Shamim M, et al. 2018. Genetic determinants of co-accessible chromatin regions in activated T cells across humans. *Nat Genet* **50:** 1140–1150. doi:10.1038/s41588-018-0156-2

Ge B, Pokholok DK, Kwan T, Grundberg E, Morcos L, Verlaan DJ, Le J, Koka V, Lam KC, Gagné V, et al. 2009. Global patterns of *cis* variation in human cells revealed by high-density allelic expression analysis. *Nat Genet* **41:** 1216–1222. doi:10.1038/ng.473

Ge R, Ye C, Peng Y, Dai Q, Zhao Y, Liu S, Wang P, Hu L, He C. 2023. m^{6}A-SAC-seq for quantitative whole transcriptome m^{6}A profiling. *Nat Protoc* **18:** 626–657. doi:10.1038/s41596-022-00765-9

Geula S, Moshitch-Moshkovitz S, Dominissini D, Mansour AA, Kol N, Salmon-Divon M, Hershkovitz V, Peer E, Mor N, Manor YS, et al. 2015. Stem cells. m^{6}A mRNA methylation facilitates resolution of naïve pluripotency toward differentiation. *Science* **347:** 1002–1006. doi:10.1126/science.1261417

GTEx Consortium. 2013. The Genotype-Tissue Expression (GTEx) project. *Nat Genet* **45:** 580–585. doi:10.1038/ng.2653

GTEx Consortium. 2017. Genetic effects on gene expression across human tissues. *Nature* **550:** 204–213. doi:10.1038/nature24277

Hu L, Liu S, Peng Y, Ge R, Su R, Senevirathne C, Harada BT, Dai Q, Wei J, Zhang L, et al. 2022. m^{6}A RNA modifications are measured at single-base resolution across the mammalian transcriptome. *Nat Biotechnol* **40:** 1210–1219. doi:10.1038/s41587-022-01243-z

Huang H, Weng H, Sun W, Qin X, Shi H, Wu H, Zhao BS, Mesquita A, Liu C, Yuan CL, et al. 2018. Recognition of RNA N^6-methyladenosine by IGF2BP proteins enhances mRNA stability and translation. *Nat Cell Biol* **20:** 285–295. doi:10.1038/s41556-018-0045-z

Jia G, Fu Y, Zhao X, Dai Q, Zheng G, Yang Y, Yi C, Lindahl T, Pan T, Yang YG, et al. 2011. N^6-Methyladenosine in nuclear RNA is a major substrate of the obesity-associated FTO. *Nat Chem Biol* **7:** 885–887. doi:10.1038/nchembio.687

Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q, Collins RL, Laricchia KM, Ganna A, Birnbaum DP, et al. 2020. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **581:** 434–443. doi:10.1038/s41586-020-2308-7

Ke S, Pandya-Jones A, Saito Y, Fak JJ, Vågbø CB, Geula S, Hanna JH, Black DL, Darnell JE Jr, Darnell RB. 2017. m^{6}A mRNA modifications are deposited in nascent pre-mRNA and are not required for splicing but do specify cytoplasmic turnover. *Genes Dev* **31:** 990–1006. doi:10.1101/gad.301036.117

Kuleshov MV, Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z, Koplev S, Jenkins SL, Jagodnik KM, Lachmann A, et al. 2016. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res* **44:** W90–W97. doi:10.1093/nar/gkw377

Leung D, Jung I, Rajagopal N, Schmitt A, Selvaraj S, Lee AY, Yen CA, Lin S, Lin Y, Qiu Y, et al. 2015. Integrative analysis of haplotype-resolved epigenomes across human tissues. *Nature* **518:** 350–354. doi:10.1038/nature14217

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Genome Project Data Processing S. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25:** 2078–2079. doi:10.1093/bioinformatics/btp352

Li A, Chen YS, Ping XL, Yang X, Xiao W, Yang Y, Sun HY, Zhu Q, Baidya P, Wang X, et al. 2017a. Cytoplasmic m^{6}A reader YTHDF3 promotes mRNA translation. *Cell Res* **27:** 444–447. doi:10.1038/cr.2017.10

Li HB, Tong J, Zhu S, Batista PJ, Duffy EE, Zhao J, Bailis W, Cao G, Kroehling L, Chen Y, et al. 2017b. m^{6}A mRNA methylation controls T cell homeostasis by targeting the IL-7/STAT5/SOCS pathways. *Nature* **548:** 338–342. doi:10.1038/nature23450

Li Z, Weng H, Su R, Weng X, Zuo Z, Li C, Huang H, Nachtergaele S, Dong L, Hu C, et al. 2017c. FTO plays an oncogenic role in acute myeloid leukemia as a N^6-methyladenosine RNA demethylase. *Cancer Cell* **31:** 127–141. doi:10.1016/j.ccell.2016.11.017

Li Y, Xia L, Tan K, Ye X, Zuo Z, Li M, Xiao R, Wang Z, Liu X, Deng M, et al. 2020. N^6-methyladenosine co-transcriptionally directs the demethylation of histone H3K9me2. *Nat Genet* **52:** 870–877. doi:10.1038/s41588-020-0677-3

Li Q, Wang Z, Zong L, Ye L, Ye J, Ou H, Jiang T, Guo B, Yang Q, Liang W, et al. 2021. Allele-specific DNA methylation maps in monozygotic twins discordant for psychiatric disorders reveal that disease-associated switching at the *EIPR1* regulatory loci modulates neural function. *Mol Psychiatry* **26:** 6630–6642. doi:10.1038/s41380-021-01126-w

Liu J, Yue Y, Han D, Wang X, Fu Y, Zhang L, Jia G, Yu M, Lu Z, Deng X, et al. 2014. A METTL3–METTL14 complex mediates mammalian nuclear RNA N^6-adenosine methylation. *Nat Chem Biol* **10:** 93–95. doi:10.1038/nchembio.1432

Liu N, Dai Q, Zheng G, He C, Parisien M, Pan T. 2015. N^6-methyladenosine-dependent RNA structural switches regulate RNA–protein interactions. *Nature* **518:** 560–564. doi:10.1038/nature14234

Liu J, Dou X, Chen C, Chen C, Liu C, Xu MM, Zhao S, Shen B, Gao Y, Han D, et al. 2020a. N^6-methyladenosine of chromosome-associated regulatory RNA regulates chromatin state and transcription. *Science* **367:** 580–586. doi:10.1126/science.aay6018

Liu J, Li K, Cai J, Zhang M, Zhang X, Xiong X, Meng H, Xu X, Huang Z, Peng J, et al. 2020b. Landscape and regulation of m^{6}A and m^{6}Am methylome across human and mouse tissues. *Mol Cell* **77:** 426–440.e6. doi:10.1016/j.molcel.2019.09.032

Liu Y, Li C, Shen S, Chen X, Szlachta K, Edmonson MN, Shao Y, Ma X, Hyle J, Wright S, et al. 2020c. Discovery of regulatory noncoding variants in individual cancer genomes by using *cis*-X. *Nat Genet* **52:** 811–818. doi:10.1038/s41588-020-0659-5

Liu C, Sun H, Yi Y, Shen W, Li K, Xiao Y, Li F, Li Y, Hou Y, Lu B, et al. 2023. Absolute quantification of single-base m^{6}A methylation in the mammalian transcriptome using GLORI. *Nat Biotechnol* **41:** 355–366. doi:10.1038/s41587-022-01487-9

MacArthur J, Bowler E, Cerezo M, Gil L, Hall P, Hastings E, Junkins H, McMahon A, Milano A, Morales J, et al. 2017. The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). *Nucleic Acids Res* **45:** D896–D901. doi:10.1093/nar/gkw1133

Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E, Wang H, Reynolds AP, Sandstrom R, Qu H, Brody J, et al. 2012. Systematic localization of common disease-associated variation in regulatory DNA. *Science* **337:** 1190–1195. doi:10.1126/science.1222794

McDaniell R, Lee BK, Song L, Liu Z, Boyle AP, Erdos MR, Scott LJ, Morken MA, Kucera KS, Battenhouse A, et al. 2010. Heritable individual-specific and allele-specific chromatin signatures in humans. *Science* **328:** 235–239. doi:10.1126/science.1184655

McIntyre ABR, Gokhale NS, Cerchietti L, Jaffrey SR, Horner SM, Mason CE. 2020. Limits in the detection of m^{6}A changes using MeRIP/m^{6}A-seq. *Sci Rep* **10:** 6590. doi:10.1038/s41598-020-63355-3

McVicker G, van de Geijn B, Degner JF, Cain CE, Banovich NE, Raj A, Lewellen N, Myrthil M, Gilad Y, Pritchard JK. 2013. Identification of genetic variants that affect histone modifications in human cells. *Science* **342:** 747–749. doi:10.1126/science.1242429

Meyer KD, Patil DP, Zhou J, Zinoviev A, Skabkin MA, Elemento O, Pestova TV, Qian SB, Jaffrey SR. 2015. 5′ UTR m^{6}A promotes cap-independent translation. *Cell* **163:** 999–1010. doi:10.1016/j.cell.2015.10.012

Onuchic V, Lurie E, Carrero I, Pawliczek P, Patel RY, Rozowsky J, Galeev T, Huang Z, Altshuler RC, Zhang Z, et al. 2018. Allele-specific epigenome maps reveal sequence-dependent stochastic switching at regulatory loci. *Science* **361:** eaar3146. doi:10.1126/science.aar3146

Patil DP, Chen CK, Pickering BF, Chow A, Jackson C, Guttman M, Jaffrey SR. 2016. m^{6}A RNA methylation promotes *XIST*-mediated transcriptional repression. *Nature* **537:** 369–373. doi:10.1038/nature19342

Pers TH, Timshel P, Hirschhorn JN. 2015. SNPsnap: a web-based tool for identification and annotation of matched SNPs. *Bioinformatics* **31:** 418–420. doi:10.1093/bioinformatics/btu655

Ping XL, Sun BF, Wang L, Xiao W, Yang X, Wang WJ, Adhikari S, Shi Y, Lv Y, Chen YS, et al. 2014. Mammalian WTAP is a regulatory subunit of the RNA N^6-methyladenosine methyltransferase. *Cell Res* **24:** 177–189. doi:10.1038/cr.2014.3

Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26:** 841–842. doi:10.1093/bioinformatics/btq033

R Core Team. 2022. *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna. https://www.R-project.org/.

Roadmap Epigenomics Consortium, Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, Heravi-Moussavi A, Kheradpour P, Zhang Z, Wang J, et al. 2015. Integrative analysis of 111 reference human epigenomes. *Nature* **518:** 317–330. doi:10.1038/nature14248

Schultz MD, He Y, Whitaker JW, Hariharan M, Mukamel EA, Leung D, Rajagopal N, Nery JR, Urich MA, Chen H, et al. 2015. Human body epigenome maps reveal noncanonical DNA methylation variation. *Nature* **523:** 212–216. doi:10.1038/nature14465

Schwartz S, Mumbach MR, Jovanovic M, Wang T, Maciag K, Bushkin GG, Mertins P, Ter-Ovanesyan D, Habib N, Cacchiarelli D, et al. 2014. Perturbation of m^{6}A writers reveals two distinct classes of mRNA methylation at internal and 5′ sites. *Cell Rep* **8:** 284–296. doi:10.1016/j.celrep.2014.05.048

Shoemaker R, Deng J, Wang W, Zhang K. 2010. Allele-specific methylation is prevalent and is contributed by CpG-SNPs in the human genome. *Genome Res* **20:** 883–889. doi:10.1101/gr.104695.109

Sun M, Zhang J. 2020. Allele-specific single-cell RNA sequencing reveals different architectures of intrinsic and extrinsic gene expression noises. *Nucleic Acids Res* **48:** 533–547. doi:10.1093/nar/gkz1134

Sun W, Poschmann J, Cruz-Herrera Del Rosario R, Parikshak NN, Hajan HS, Kumar V, Ramasamy R, Belgard TG, Elanggovan B, Wong CCY, et al. 2016. Histone acetylome-wide association study of autism spectrum disorder. *Cell* **167:** 1385–1397.e11. doi:10.1016/j.cell.2016.10.031

Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, Simonovic M, Doncheva NT, Morris JH, Bork P, et al. 2019. STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res* **47:** D607–D613. doi:10.1093/nar/gky1131

Tarailo-Graovac M, Chen N. 2009. Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr Protoc Bioinformatics* **Chapter 4:** Unit 4.10. doi:10.1002/0471250953.bi0410s05

van de Geijn B, McVicker G, Gilad Y, Pritchard JK. 2015. WASP: allele-specific software for robust molecular quantitative trait locus discovery. *Nat Methods* **12:** 1061–1063. doi:10.1038/nmeth.3582

Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, Del Angel G, Levy-Moonshine A, Jordan T, Shakir K, Roazen D, Thibault J, et al. 2013. From FastQ data to high confidence variant calls: the genome analysis toolkit best practices pipeline. *Curr Protoc Bioinformatics* **43:** 11.10.11–11.10.33. doi:10.1002/0471250953.bi1110s43

Vu LP, Pickering BF, Cheng Y, Zaccara S, Nguyen D, Minuesa G, Chou T, Chow A, Saletore Y, MacKay M, et al. 2017. The N^6-methyladenosine (m^{6}A)-forming enzyme METTL3 controls myeloid differentiation of normal hematopoietic and leukemia cells. *Nat Med* **23:** 1369–1376. doi:10.1038/nm.4416

Wan Y, Tang K, Zhang D, Xie S, Zhu X, Wang Z, Lang Z. 2015. Transcriptome-wide high-throughput deep m^{6}A-seq reveals unique

differential m^{6}A methylation patterns between three organs in *Arabidopsis thaliana*. *Genome Biol* **16:** 272. doi:10.1186/s13059-015-0839-2

Wang X, Lu Z, Gomez A, Hon GC, Yue Y, Han D, Fu Y, Parisien M, Dai Q, Jia G, et al. 2014a. N^6-methyladenosine-dependent regulation of messenger RNA stability. *Nature* **505:** 117–120. doi:10.1038/nature12730

Wang Y, Li Y, Toth JI, Petroski MD, Zhang Z, Zhao JC. 2014b. N^6-methyladenosine modification destabilizes developmental regulators in embryonic stem cells. *Nat Cell Biol* **16:** 191–198. doi:10.1038/ncb2902

Wang X, Zhao BS, Roundtree IA, Lu Z, Han D, Ma H, Weng X, Chen K, Shi H, He C. 2015. N^6-methyladenosine modulates messenger RNA translation efficiency. *Cell* **161:** 1388–1399. doi:10.1016/j.cell.2015.05.014

Wang L, Wen M, Cao X. 2019. Nuclear hnRNPA2B1 initiates and amplifies the innate immune response to DNA viruses. *Science* **365:** eaav0758. doi:10.1126/science.aav0758

Wang J, Huang D, Zhou Y, Yao H, Liu H, Zhai S, Wu C, Zheng Z, Zhao K, Wang Z, et al. 2020. CAUSALdb: a database for disease/trait causal variants identified using summary statistics of genome-wide association studies. *Nucleic Acids Res* **48:** D807–D816. doi:10.1093/nar/gkz1026

Wang Q, Dhindsa RS, Carss K, Harper AR, Nag A, Tachmazidou I, Vitsios D, Deevi SVV, Mackay A, Muthas D, et al. 2021. Rare variant contribution to human disease in 281,104 UK Biobank exomes. *Nature* **597:** 527–532. doi:10.1038/s41586-021-03855-y

Xiang Y, Laurent B, Hsu CH, Nachtergaele S, Lu Z, Sheng W, Xu C, Chen H, Ouyang J, Wang S, et al. 2017. RNA m^{6}A methylation regulates the ultraviolet-induced DNA damage response. *Nature* **543:** 573–576. doi:10.1038/nature21671

Xiao W, Adhikari S, Dahal U, Chen YS, Hao YJ, Sun BF, Sun HY, Li A, Ping XL, Lai WY, et al. 2016. Nuclear m^{6}A reader YTHDC1 regulates mRNA splicing. *Mol Cell* **61:** 507–519. doi:10.1016/j.molcel.2016.01.012

Xiao S, Cao S, Huang Q, Xia L, Deng M, Yang M, Jia G, Liu X, Shi J, Wang W, et al. 2019. The RNA N^6-methyladenosine modification landscape of human fetal tissues. *Nat Cell Biol* **21:** 651–661. doi:10.1038/s41556-019-0315-4

Xiong X, Hou L, Park YP, Molinie B, GTEx Consortium, Gregory RI, Kellis M. 2021. Genetic drivers of m^{6}A methylation in human brain, lung, heart and muscle. *Nat Genet* **53:** 1156–1165. doi:10.1038/s41588-021-00890-3

Xu K, Yang Y, Feng GH, Sun BF, Chen JQ, Li YF, Chen YS, Zhang XX, Wang CX, Jiang LY, et al. 2017. Mettl3-mediated m^{6}A regulates spermatogonial differentiation and meiosis initiation. *Cell Res* **27:** 1100–1114. doi:10.1038/cr.2017.100

Yang EW, Bahn JH, Hsiao EY, Tan BX, Sun Y, Fu T, Zhou B, Van Nostrand EL, Pratt GA, Freese P, et al. 2019. Allele-specific binding of RNA-binding proteins reveals functional genetic variants in the RNA. *Nat Commun* **10:** 1338. doi:10.1038/s41467-019-09292-w

Yin R, Chang J, Li Y, Gao Z, Qiu Q, Wang Q, Han G, Chai J, Feng M, Wang P, et al. 2022. Differential m^{6}A RNA landscapes across hematopoiesis reveal a role for IGF2BP2 in preserving hematopoietic stem cell function. *Cell Stem Cell* **29:** 149–159.e7. doi:10.1016/j.stem.2021.09.014

Yoon KJ, Ringeling FR, Vissers C, Jacob F, Pokrass M, Jimenez-Cyrus D, Su Y, Kim NS, Zhu Y, Zheng L, et al. 2017. Temporal control of mammalian cortical neurogenesis by m^{6}A methylation. *Cell* **171:** 877–889.e17. doi:10.1016/j.cell.2017.09.003

Yu G, Wang LG, Han Y, He QY. 2012. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* **16:** 284–287. doi:10.1089/omi.2011.0118

Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, et al. 2008. Model-based Analysis of ChIP-Seq (MACS). *Genome Biol* **9:** R137. doi:10.1186/gb-2008-9-9-r137

Zhang C, Chen Y, Sun B, Wang L, Yang Y, Ma D, Lv J, Heng J, Ding Y, Xue Y, et al. 2017. m^{6}A modulates haematopoietic stem and progenitor cell specification. *Nature* **549:** 273–276. doi:10.1038/nature23883

Zhang H, Shi X, Huang T, Zhao X, Chen W, Gu N, Zhang R. 2020a. Dynamic landscape and evolution of m^{6}A methylation in human. *Nucleic Acids Res* **48:** 6251–6264. doi:10.1093/nar/gkaa347

Zhang Z, Luo K, Zou Z, Qiu M, Tian J, Sieh L, Shi H, Zou Y, Wang G, Morrison J, et al. 2020b. Genetic analyses support the contribution of mRNA N^6-methyladenosine (m^{6}A) modification to human disease heritability. *Nat Genet* **52:** 939–949. doi:10.1038/s41588-020-0644-z

Zheng G, Dahl JA, Niu Y, Fedorcsak P, Huang CM, Li CJ, Vågbå CB, Shi Y, Wang WL, Song SH, et al. 2013. ALKBH5 is a mammalian RNA demethylase that impacts RNA metabolism and mouse fertility. *Mol Cell* **49:** 18–29. doi:10.1016/j.molcel.2012.10.015

Zheng Q, Hou J, Zhou Y, Li Z, Cao X. 2017. The RNA helicase DDX46 inhibits innate immunity by entrapping m^{6}A-demethylated antiviral transcripts in the nucleus. *Nat Immunol* **18:** 1094–1103. doi:10.1038/ni.3830

Zhou J, Wan J, Gao X, Zhang X, Jaffrey SR, Qian SB. 2015. Dynamic m^{6}A mRNA methylation directs translational control of heat shock response. *Nature* **526:** 591–594. doi:10.1038/nature15377

Zhou ZY, Hu Y, Li A, Li YJ, Zhao H, Wang SQ, Otecko NO, Zhang D, Wang JH, Liu Y et al. 2018. Genome wide analyses uncover allele-specific RNA editing in human and mouse. *Nucleic Acids Res* **46:** 8888–8897. doi:10.1093/nar/gky613

Received January 15, 2023; accepted in revised form June 13, 2023.

Allele-specific RNA N^6-methyladenosine modifications reveal functional genetic variants in human tissues

Shuo Cao, Haoran Zhu, Jinru Cui, et al.

Genome Res. published online September 15, 2023
Access the most recent version at doi:10.1101/gr.277704.123