

# Dynamic landscape and evolution of m<sup>6</sup>A methylation in human

Hui Zhang<sup>1,†</sup>, Xinrui Shi<sup>1,†</sup>, Tao Huang<sup>1,\*</sup>, Xueni Zhao<sup>1</sup>, Wanying Chen<sup>1</sup>, Nannan Gu<sup>1</sup> and Rui Zhang<sup>1,2,\*</sup>

<sup>1</sup>Key Laboratory of Gene Engineering of the Ministry of Education, State Key Laboratory of Biocontrol, School of Life Sciences, Sun Yat-Sen University, Guangzhou 510275, PR China and <sup>2</sup>RNA Biomedical Institute, Sun Yat-Sen Memorial Hospital, Sun Yat-Sen University, Guangzhou 510120, PR China

Received February 29, 2020; Revised April 23, 2020; Editorial Decision April 24, 2020; Accepted April 24, 2020

## ABSTRACT

m<sup>6</sup>A is a prevalent internal modification in mRNAs and has been linked to the diverse effects on mRNA fate. To explore the landscape and evolution of human m<sup>6</sup>A, we generated 27 m<sup>6</sup>A methylomes across major adult tissues. These data reveal dynamic m<sup>6</sup>A methylation across tissue types, uncover both broadly or tissue-specifically methylated sites, and identify an unexpected enrichment of m<sup>6</sup>A methylation at non-canonical cleavage sites. A comparison of fetal and adult m<sup>6</sup>A methylomes reveals that m<sup>6</sup>A preferentially occupies CDS regions in fetal tissues. Moreover, the m<sup>6</sup>A sub-motifs vary between fetal and adult tissues or across tissue types. From the evolutionary perspective, we uncover that the selection pressure on m<sup>6</sup>A sites varies and depends on their genic locations. Unexpectedly, we found that ~40% of the 3'UTR m<sup>6</sup>A sites are under negative selection, which is higher than the evolutionary constraint on miRNA binding sites, and much higher than that on A-to-I RNA modification. Moreover, the recently gained m<sup>6</sup>A sites in human populations are clearly under positive selection and associated with traits or diseases. Our work provides a resource of human m<sup>6</sup>A profile for future studies of m<sup>6</sup>A functions, and suggests a role of m<sup>6</sup>A modification in human evolutionary adaptation and disease susceptibility.

## INTRODUCTION

Chemical modifications on RNA have been recently appreciated as an important regulatory feature (1). Recent technological breakthroughs, driven mainly by the sequencing-

based approaches, have enabled the genome-wide profiling of such RNA modifications, particularly the RNA deamination and methylation (2–6). However, except for A-to-I (adenosine to inosine) RNA editing, which is the predominant type of RNA deamination in animals (7–12), less is known about the dynamics and evolution of most RNA modifications.

m<sup>6</sup>A is one of the most prevalent internal modifications in mRNAs (2,13–16). It is present among eukaryotic species that range from yeast, plants, flies to mammals. m<sup>6</sup>A RNA methylation is catalyzed by a multicomponent methyltransferase complex, including METTL3, METTL14 and WTAP (17,18). It has a consensus motif RRACH (in which R represents A or G, and H represents A, C or U). m<sup>6</sup>A methylation regulates the splicing, expression, decay and translation of mRNAs (19–21), and plays crucial roles in various cellular pathways and processes such as cell differentiation, development and metabolism (15). To date, m<sup>6</sup>A has been identified in several thousand human protein-coding genes. Although m<sup>6</sup>A profile of many cultured human cell lines and fetal human tissues have been reported (22,23), we still have limited information about the global landscape and dynamics of m<sup>6</sup>A in adult human tissues.

It has been hypothesized that gene regulation, ranging from transcriptional processing to post-transcriptional regulation, has a central role in phenotypic evolution (24–26). Therefore, a fundamental question in biology is to understand how natural selection has shaped the evolution of gene regulation (27,28), including RNA modifications. Some studies have shown that m<sup>6</sup>A peaks, which typically span one to several hundred nt, are conserved between human and mouse (2) and the m<sup>6</sup>A peak regions have much higher sequence conservation scores than those of randomly selected regions (13). While others suggest that only 37% of the m<sup>6</sup>A peaks are conserved between human and

\*To whom correspondence should be addressed. Tel: +86 20 84111711; Email: zhangrui3@mail.sysu.edu.cn  
Correspondence may also be addressed to Tao Huang. Tel: +86 20 84111710; Email: huangt93@mail.sysu.edu.cn  
†The authors wish it to be known that, in their opinion, the first three authors should be regarded as joint First Authors.

rhesus macaque (29), and the sequence of m<sup>6</sup>A RAC central motif is only slightly conserved than the control RAC sites (30). In addition, a recent study suggested that most m<sup>6</sup>A sites in CDS regions are evolutionarily unconserved (31). However, those studies were limited in scope and scale, thus, a systematic investigation of the selection pressure on individual m<sup>6</sup>A sites is needed.

## MATERIALS AND METHODS

### Sample procurement

Samples of nine human adult tissues were obtained from Chinese Brain Bank Center (Wuhan, China). These tissues were collected post-mortem from individuals with no known medical history. The consent of human tissue samples using autopsy was obtained from the patients' families. Samples were lysed and homogenized in TRIzol Reagent (Invitrogen) using Precellys evolution tissue homogenizer (Bertin). Total RNA was extracted using chloroform and isopropanol following the manufacturer's protocol. The quality of the total RNAs was determined by agarose gel electrophoresis and three biological replicates of RNA samples that with thick 28S and 18S ribosomal RNA (rRNA) gel bands at an approximate mass ratio of 2:1 were selected. These tissues are from five donors (N1–N5), including frontal cortex (N1–N3), cerebellum (N1–N3), heart (N–N3), liver (N1–N3), lung (N1, N3, N5), kidney (N1, N2, N5), spleen (N1, N2, N5), muscle (N2–N4) and testis (N1–N3). N1, male, age 39; N2, male, age 44; N3, male, age 47; N4, male, age 57; N5, male, age 44.

### m<sup>6</sup>A-seq library preparation

m<sup>6</sup>A immunoprecipitation and library construction were performed as described previously with some modification (2). In brief, samples were lysed and homogenized in TRIzol Reagent (Invitrogen) using Precellys evolution tissue homogenizer (Bertin). Total RNA was extracted using chloroform and isopropanol following the manufacturer's protocol. Polyadenylated mRNA was enriched from total RNA using GenElute mRNA miniprep kit (Sigma-Aldrich). RNA samples were fragmented in 1X Next Magnesium RNA Fragmentation Buffer (NEB) at 94°C for 5min and fragmented RNA was then cleaned up using ethanol precipitation. 10ng fragmented RNA was used to construct input control library with VAHTS Stranded mRNA-seq library prep kit (Vazyme). 15–40 ug fragmented RNA was further incubated with 5ug rabbit anti-m<sup>6</sup>A polyclonal antibody (Synaptic Systems, catalog number 202 003) in IPP buffer (150 mM NaCl, 0.1% Igepal CA-630, 10 mM Tris-HCl, Ph7.4) overnight at 4°C. The m<sup>6</sup>A-Ab mixture was then immunoprecipitated by incubation with protein-G magnetic beads (Thermo Fisher, pre-blocked with 0.5 mg ml<sup>-1</sup> BSA at 4°C for 2h) at 4°C for another 2h. After washing with IPP buffer, bound RNA was competitively eluted from the beads with 0.5 mg ml<sup>-1</sup> N<sup>6</sup>-methyladenosine (Sigma-Aldrich), followed by ethanol precipitation. RNA was resuspended in 8 µl water and used for library construction. Libraries were sequenced on HiSeq X (Illumina) to produce paired-end 150 bp reads.

### MeRIP-seq of HEK293T cells

MeRIP-seq was performed as described previously (23). In brief, total RNA of HEK293T cells was extracted and fragmented in Next Magnesium RNA Fragmentation Buffer (NEB) at 94°C for 5 min. 10 ng fragmented RNA was used to construct the input library. 300 ug of fragmented RNA was incubated with 5 ug rabbit anti-m<sup>6</sup>A polyclonal antibody (Abcam, catalog number ab151230) overnight at 4°C. After stringent washing, bound RNA was eluted by competition with N<sup>6</sup>-methyladenosine (Sigma-Aldrich), followed by rRNA removal with QIAseq FastSelect RNA Removal Kits (Qiagen). Both the input and IP libraries were constructed using NEBNext Ultra II Directional RNA Library Prep Kit for Illumina (NEB). Libraries were sequenced on HiSeq X (Illumina) to produce paired-end 150 bp reads.

### METTL3 knockout cell generation

METTL3 knockout cell was generated via CRISPR/Cas9-induced mutagenesis. In brief, a sgRNA sequence (GCAGAAGCGGCGTGCAGAAC) was designed using CRISPR-ERA (<http://CRISPR-ERA.stanford.edu>). The sgRNA template oligonucleotide was synthesized and cloned into lentiCRISPR v2 plasmid (Addgene#52961). The plasmid was transfected into HEK293T cells. Transfected cells were selected using puromycin (1 µg /ml). Mutant clones were selected by Sanger sequencing. The loss of METTL3 protein expression was verified with the METTL3 antibody (Proteintech, catalog number 15073-1-AP) by western blot. The m<sup>6</sup>A levels of HEK293T wild type and METTL3 knockout cells were measured using EpiQuik m<sup>6</sup>A RNA Methylation Quantification Kit (Epigentek). In brief, polyadenylated RNA was separated from total RNA using Oligo dT Magnetic Beads (Vazyme). 300ng polyadenylated mRNA was used for m<sup>6</sup>A quantification following the manufacturer's protocol.

### m<sup>6</sup>A-seq data analysis

m<sup>6</sup>A peak identification was performed as previously described (32). In brief, we trimmed the adaptor and low quality reads using Cutadapt (33) (-e 0.3 -minimum-length 25; -trim-n -q 20,20) and fastx toolkit (fastx\_trimmer -f 6; -t 5 -m 20). rRNA reads were then removed using SortMeRNA (34). Next, clean reads were mapped to the human genome (hg19) using TopHat2 (version 2.1.0) (35). Enriched peaks were identified by scanning each gene using 100-nt sliding windows, and calculating an enrichment score for each sliding window (winscore).

$$\text{winscore} = \log_2 \left( \frac{\text{MeanWinIP}/\text{MedianGeneIP}}{\text{MeanWinControl}/\text{MedianGeneControl}} \right)$$

MeanWinIP and MeanWinControl are the mean coverage for each window for immunoprecipitation and input control, respectively. MedianGeneIP and MedianGeneControl are gene median coverages for immunoprecipitation and input control, respectively. Windows with RPKM  $\geq 10$  in the IP sample, enrichment score  $\geq 2$  in genes with RPKM in the input sample  $\geq 1$  were defined as enriched windows. Last, only the peaks with winscore  $\geq 2$  in

at least two samples of a tissue type were considered as real m<sup>6</sup>A peaks.

To generate the metagene profile of m<sup>6</sup>A site distribution across transcripts, we first determined the number of bins that need to be divided for a given gene based on relative lengths between 5'UTR, CDS and 3'UTR of the human transcriptome (GENCODE v26). The relative lengths between 5'UTR, CDS and 3'UTR are 10:50:40, thus for each gene, 10, 50 and 40 bins of equal length were made for 5'UTR, CDS and 3'UTR, respectively. Next, for each m<sup>6</sup>A site, we assigned it to the longest isoform of the corresponding gene and determined which bin it is located in. Last, the number of sites for each bin was summarized and the curve was fitted with polyfit.

The mapping statistics of all datasets were summarized in Supplementary Table S1.

### m<sup>6</sup>A peak call via exomePeak and MeTPeak

IP and input reads were mapped as described above. For each tissue type, the consistent peaks in all replicates were called using exomePeak (36) or MeTPeak (37) with default parameters.

### Tissue specificity analysis

For each site, we first calculated its average winscore in each tissue type. Next, we calculated its tissue specificity index tau (38) using the average winscore of each site:

$$\text{tau} = \frac{\sum_{i=1}^n 1 - \hat{x}_i}{n - 1}; \quad \hat{x}_i = \frac{x_i}{\max_{1 \leq i \leq n} x_i}$$

$x_i$  is the average winscore of a site in tissue  $i$ ;  $n$  is the number of tissues.

Shared m<sup>6</sup>A sites were defined as sites with tau < 0.15. Tissue-specific m<sup>6</sup>A sites were defined as sites with tau > 0.6.

### m<sup>6</sup>A-RIP qPCR

RIP was performed as described above. Both the IP and input RNAs were reverse transcribed and the m<sup>6</sup>A marked mRNAs and NC (GAPDH) mRNA were detected by qPCR. The enrichment fold of IP versus input of each gene was calculated and normalized to NC. Primers were listed as follows:

```
ATP5C1-F GGGAGCTTCGGCGCAT
ATP5C1-R CGCGCGAGAGAACATGGTAG
DCTN1-F GCACGGTTCCTGACAAGTCTA
DCTN1-R GACACAGAATCCTGCTTGCC
PSMB4-F ATGGAAGCGTTTTTGGGGTC
PSMB4-R GAGTGGACGGAATGCGGTAA
SDHAF2-F GCCTTGCTTCCGGCTTCTTA
SDHAF2-R TGTCATCCTTGAGGCAGG
GAPDH-F TGCCAAATATGATGACATCAAGAA
GAPDH-R GGAGTGGGTGTCGCTGTTG
```

### APA data analysis

Human APA cleavage sites were downloaded from PolyA\_DB version 3.2 (<http://exon.umdj.edu/polya-db/v3/>). PolyA\_DB version 3.2 catalogs polyA sites using deep sequencing data.

### 3' processing efficiency measurement assay

We used a bicistronic luciferase reporter construct, pPAS-PORT, to measure 3' processing efficiency (39). 3'UTR of each selected gene was inserted into pPASPORT, between Renilla and Firefly luciferase genes. Plasmids were transfected into wild-type and METTL3 knockout HEK293T cells using Lipofectamine 3000 (Thermo Fisher Scientific), respectively. Renilla and Firefly luminescences were measured 24 h later using Dual-Glo Luciferase Assay System (Promega) on GloMax -96 Microplate Luminometer (Promega). All primers used to construct the reporter genes are listed in Supplementary Table S2.

### Sub-motif analysis

To obtain the expected numbers of windows with both GGACH and AAACH sub-motifs, the m<sup>6</sup>A sub-motif sequences were shuffled within all m<sup>6</sup>A peaks of a given sample. Next, the number of windows with both sub-motifs was calculated. We repeated the shuffling analysis for 10 000 times and obtained 10 000 expected numbers. To plot and compare results from different samples, we performed normalization by mean-centering the values. In brief, for a given sample, we first calculated the mean value of expected window numbers ( $M_{\text{expected}}$ ). Next, both observed number and the 10 000 expected numbers were divided by  $M_{\text{expected}}$ .

### Rejected substitution score acquisition

The rejected substitution score data were from Sidow lab (<http://mendel.stanford.edu/SidowLab/downloads/gerp/>).

### Cross-species analysis

To conduct the CDS m<sup>6</sup>A conservation analysis, we used the method from a previous study (31) with some modifications. First, we required that the control A sites and the m<sup>6</sup>A sites were in genes with similar dN/dS values (human versus mouse). Second, we required that the control A sites and the m<sup>6</sup>A sites were with similar distances to the stop codon. The human-mouse pairwise alignment file was downloaded from UCSC genome browser, and human-mouse dN and dS value table was obtained via Ensembl biomart.

The proportion of m<sup>6</sup>A sites in the third codon that are under evolutionary constraints between human and mouse was calculated as: (the evolutionary rate at m<sup>6</sup>A sites (1 - 0.644) - the mean evolutionary rate at control sites (1 - 0.621))/(1 - 0.621).

To determine the age of individual m<sup>6</sup>A sites, pairwise alignment files were downloaded from UCSC genome browser.

### RNA editing site selection

Human RNA editing sites were downloaded from RADAR database (40). RADAR version 2, which includes 16 464 human RNA editing sites located at 3'UTR region, was used for analysis.

### miRNA target site selection

miRNA binding sites were downloaded from TargetScan database (Release 7.1 <http://www.targetscan.org/>). The Tar-

getScan algorithm predicts biological targets of miRNAs by searching for the presence of 8mer and 7mer sites that match the seed region of each miRNA (41). A total of 669 927 and 601 858 target sites of conserved and broadly conserved miRNA families were analyzed, separately. Conserved miRNA families are conserved across most mammals, but usually not beyond placental mammals. Broadly conserved miRNA families are conserved across most vertebrates, usually to zebrafish.

### SNP data

We downloaded SNP and genotype data from the 1000 Genome Project (<http://www.internationalgenome.org/>). We discarded all insertion and deletion polymorphisms, SNPs with more than two alleles, SNPs monomorphic (that is, having only one allele) in all populations and SNPs that did not map uniquely to the human genome (hg19). Finally, a total of 77 664 537 SNPs were used for analysis.

### Derived allele frequency (DAF) analysis

For each SNP, we extracted the ancestral allele information from the downloaded VCF files of the 1000 Genome Project. For an RRAC motif containing an SNP, we defined it as a gain of an m<sup>6</sup>A site if the derived allele created an RRAC motif.

### Fst calculation

VCFtools was used to calculate Fst between populations (42).

### Haplotype homozygosity-based tests

The VCF files were obtained from 1000 Genome Project. For each m<sup>6</sup>A SNP, we extracted SNPs within  $\pm 2$  Mb to generate individual VCF files. The Perl script `vcf2impute.legend_haps.pl` from `impute2` (<https://mathgen.stats.ox.ac.uk/impute/impute.v2.html#scripts>) was used to convert a VCF file into reference panel format: one legend file and one haplotype file. The VCF files of some SNPs were failed to convert so these SNPs were excluded from the further analysis. The hap format file and map format file required by R package 'rehh' and `selscan` were formatted using the legend and haplotype files by in-house Perl scripts. `iHS` was calculated using the 'ihh2ihs' function of `rehh` package in R (43). `XP-EHH` was calculated using the 'ies2xpehh' function of `rehh` package.

### Overlap with GTEx eQTL SNPs

GTEx v6p eQTL file was downloaded from GTEx website (<https://gtexportal.org/home/>). Only significantly associated SNPs were used.

### Overlap with GWAS data

The NHGRI-EBI Catalog of Published GWAS (44), which contains 68 741 trait-associated SNPs, was used to query for overlaps with the 102 m<sup>6</sup>A SNPs.

To find proxy SNPs (nearby SNPs in linkage disequilibrium) of m<sup>6</sup>A SNPs that are GWAS hits, we identified all SNPs in strong linkage disequilibrium with high `Fst` (`Fst` > 0.15) SNPs. This step was implemented via the VCFtools by extracting all SNPs with pairwise  $r^2$  > 0.8 based on CEU, YRI, CHB, PUR and GIH populations from the 1000 Genome Project.

### Gene ontology (GO) term analysis

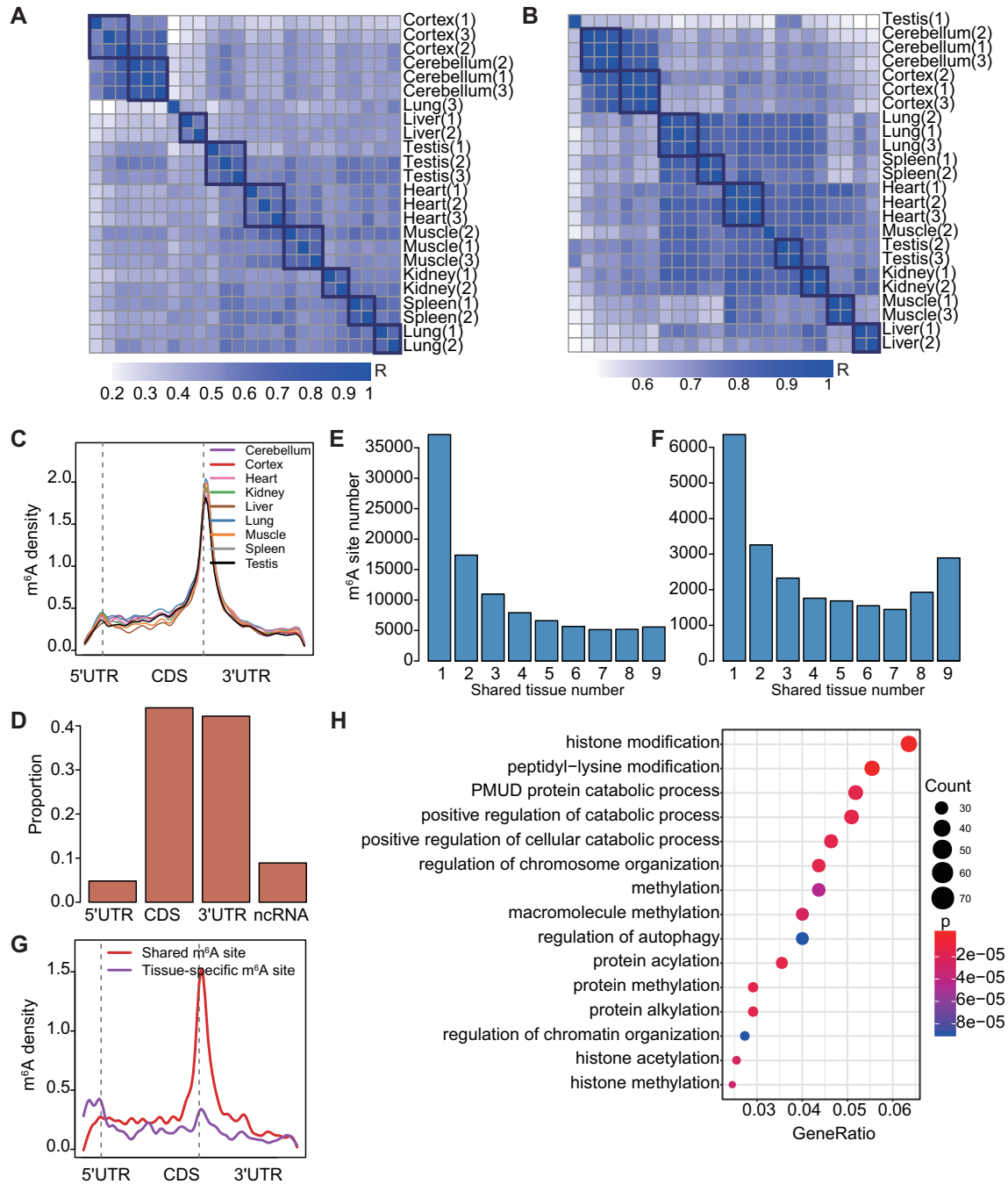
GO term enrichment analysis was performed using R packages `clusterProfiler` (45).

## RESULTS

### m<sup>6</sup>A profiles across human adult tissues

To explore the landscape of m<sup>6</sup>A in adult human tissues, we constructed m<sup>6</sup>A-seq libraries from nine human adult tissues, including cerebellum, frontal cortex, heart, kidney, liver, lung, muscle, spleen and testis. For each tissue, three individuals were profiled. m<sup>6</sup>A peaks were called in each sample using a winscore approach as previously described (32). m<sup>6</sup>A peaks found in at least 2 samples of a tissue type were considered as real m<sup>6</sup>A peaks. On average, we found 19 100 m<sup>6</sup>A peaks per tissue (Supplementary Table S3). The samples of the same tissue type clustered well according to either their m<sup>6</sup>A levels or gene expression levels (Figure 1A, B). Sequence logo analysis of all datasets confirmed that the called peaks are enriched in the m<sup>6</sup>A consensus motif RRACH (Supplementary Figure S1), consistent with the previous observation (2). We next inferred the m<sup>6</sup>A sites by searching for the RRACH motifs within the peaks (Materials and methods). As expected, m<sup>6</sup>A sites preferentially appeared around stop codons (Figure 1C). To evaluate the performance of the winscore peak call method we used, we applied two additional peak calling algorithms (`exomePeak` (36) and `MeTPeak` (37)) for m<sup>6</sup>A site call and compared m<sup>6</sup>A sites called from different methods. Most sites identified in our method overlapped with those identified in `exomePeak` or `MeTPeak`, and both `exomePeak` and `MeTPeak` called a lot more uniquely identified peaks (Supplementary Figure S2A-B). Compared with sites called by `exomePeak` or `MeTPeak`, as expected, sites called from our method had higher winscores (Supplementary Figure S2C). Thus, sites called using the winscore approach seemed to represent a more stringent set of m<sup>6</sup>A sites and were used for all following analyses. Combining all data together, we obtained a union of 101 340 m<sup>6</sup>A sites that span various classes of genic regions (Figure 1D, Supplementary Supplementary Table S4).

To reveal the methylation landscape across adult tissues, we examined to what extent the m<sup>6</sup>A sites are shared between tissues. We found that more than 36.7% of the sites were found in one specific tissue type and only 5.5% of the sites were shared in all tissues we studied (Figure 1E), thus m<sup>6</sup>A methylation seems to have high tissue-specificity. The profiles of four selected m<sup>6</sup>A peaks across tissues, as well as the m<sup>6</sup>A-RIP qPCR validation results, were shown in Supplementary Figure S3. To further investigate the effect of differential gene expression on the tissue-specificity of m<sup>6</sup>A



**Figure 1.** m<sup>6</sup>A profile in human adult tissues. (A, B) Heatmap of Pearson correlation on m<sup>6</sup>A peak winscores (A) or gene expression levels (B) of protein-coding genes. Gene expression levels were quantified as the number of RNA-seq reads per kilobase of transcript per million mapped reads (RPKM). (C) The distribution of m<sup>6</sup>A sites across the length of mRNA transcripts for nine adult tissues. 5'UTRs, CDSs and 3'UTRs of protein-coding genes were individually grouped into 10, 50 and 40 bins of their total length, and the percentage of m<sup>6</sup>A sites that fall within each bin was determined. (D) Genic locations of m<sup>6</sup>A sites of adult tissues. Sites in nine adult tissues were merged for analysis. Sites were annotated using Ensembl gene annotations and ANNOVAR software. ncRNA, Noncoding RNA. (E) The distribution of the numbers of m<sup>6</sup>A sites that are methylated in one or more tissues. (F) The distribution of the numbers of m<sup>6</sup>A sites that are methylated in one or more tissues. Only m<sup>6</sup>A sites within the ubiquitously expressed genes (FPKM > 3 in all tissues) were analyzed. (G) The distribution of tissue-specific or shared m<sup>6</sup>A sites across the length of mRNA transcripts for 9 adult tissues. Tissue-specific m<sup>6</sup>A sites, sites that are within the ubiquitously expressed genes and have a tau > 0.6; shared m<sup>6</sup>A sites, sites that are within the ubiquitously expressed genes and have a tau < 0.15. (H) GO terms enriched in the ubiquitously expressed genes with shared m<sup>6</sup>A sites (tau < 0.15). GO term analysis was performed using clusterProfiler. All ubiquitously expressed genes were used as the background. P values were corrected by Bonferroni adjustments and the top 15 enriched go terms were shown.

sites, we examined m<sup>6</sup>A sites that were located in the ubiquitously expressed genes across human tissues. Of the 4063 ubiquitously expressed genes, 23 167 sites were identified. Of these sites, 27.4% was methylated in only one tissue type and 12.4% was shared by all tissue types (Figure 1F). These data together suggest that both tissue-specific gene expression and tissue-specific methylation contribute to the observed high tissue-specificity of m<sup>6</sup>A sites.

To better investigate the tissue-specificity of individual m<sup>6</sup>A sites, we applied the widely used tissue specificity index tau to measure the tissue-specific methylation (Materials and Methods). Tau varies from 0 to 1, where 0 means broadly expressed, and 1 is specific. Interestingly, when examining the genic locations of tissue-specific (tau > 0.6) and shared m<sup>6</sup>A sites (tau < 0.15), we found that shared sites tended to be located around the stop codon, while tissue-specific sites tended to be away from the stop codon (Figure 1G). This result suggests the m<sup>6</sup>A sites away from stop codon may perform tissue-specific functions, while m<sup>6</sup>A sites around stop codon are more likely to be required for the maintenance of basic cellular function, thus methylated in all cells of an organism. Consistently, the genes with shared sites were enriched in essential functions such as chromatin organization, cellular catabolic process and histone modification (Figure 1H). This tissue-specific methylation pattern was confirmed using RNA-endoribonuclease-facilitated sequencing data that identify m<sup>6</sup>A sites in three human tissue types (Supplementary Figure S4). Moreover, to control the potential difference in RIP efficiencies between samples, we normalized the winscores based on the top 50 peaks in each sample (Supplementary Figure S5A) and repeated the analysis. We found that the tissue-specific methylation pattern still holds (Supplementary Figure S5B). Notably, an enrichment of tissue-specific sites was found in 5'UTR (Figure 1G). To ask whether the signal in 5'UTR is m<sup>6</sup>Am, we performed the sequence logo analysis for tissue-specific peaks in 5'UTR regions. We found that the called peaks are enriched in the m<sup>6</sup>A consensus motif RRACH (Supplementary Figure S5C). Moreover, we examined the distance of the tissue-specific m<sup>6</sup>A sites in the 5'UTR regions to the TSS. We found that <20% of the sites are very close to TSS (Supplementary Figure S5D). These results together suggest that most tissue-specific m<sup>6</sup>A peaks in 5'UTRs are m<sup>6</sup>Am, although we cannot exclude the possibility that some of the peaks that are very close to TSS are m<sup>6</sup>Am.

We also applied a lower or higher stringency to define m<sup>6</sup>A sites (i.e. we required that the sites were found in at least one sample or all samples of a tissue type) and repeated the tissue-specificity analysis. In the low stringency condition, 45.6% of the sites were methylated in only one tissue type and 4.3% of the sites were shared in all tissues; in the high stringency condition, 50.9% of the sites were methylated in only one tissue type and 0.9% of the sites were shared in all tissues. In both conditions, we consistently observed that shared sites tended to be located around the stop codon, while tissue-specific sites tended to be away from the stop codon (Supplementary Figure S6A, B). Moreover, the genes with shared sites were enriched in essential functions such as chromatin organization, cellular

catabolic process and histone modification (Supplementary Figure S6C, D).

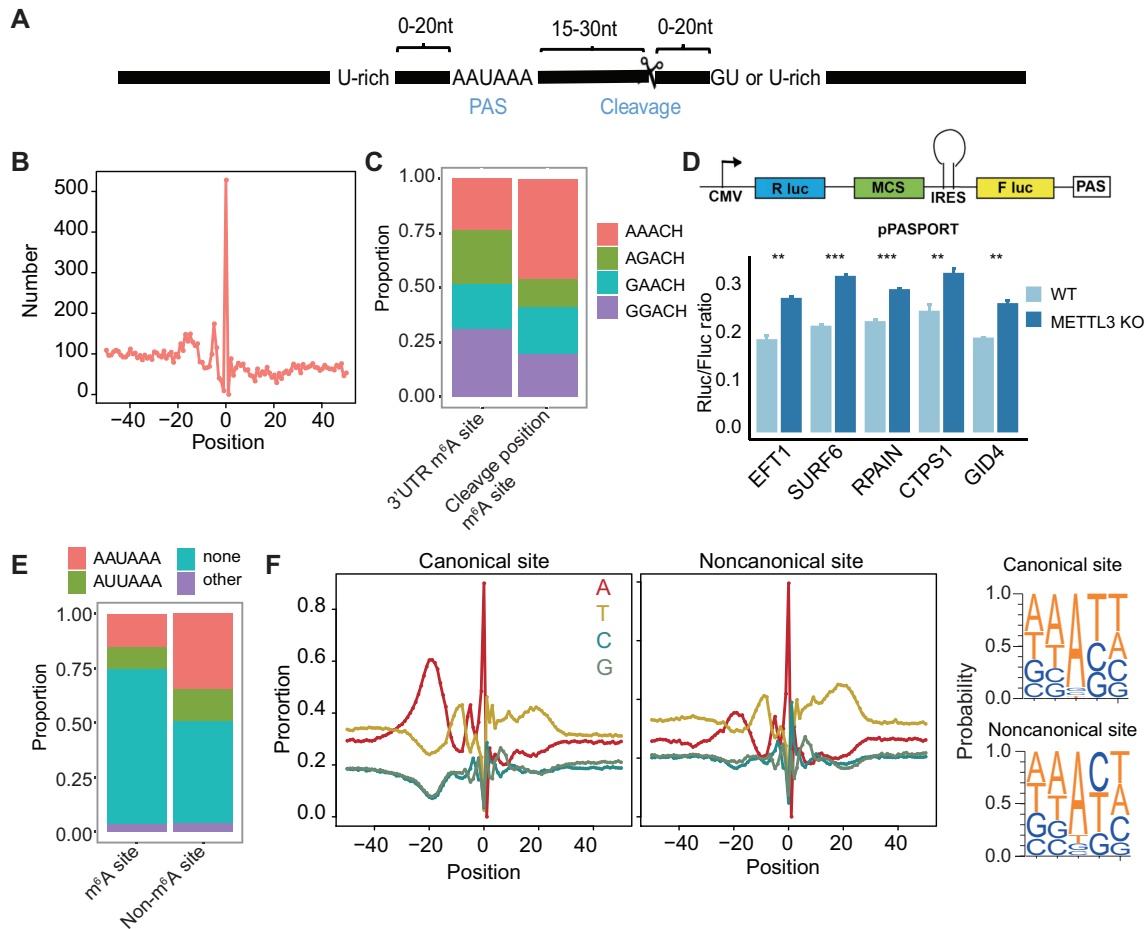
Taken together, these data reveal dynamic m<sup>6</sup>A methylation across tissue types, uncover both broadly or tissue-specifically methylated sites, and highlight the potentially distinct regulatory effects for m<sup>6</sup>A sites around and away from the stop codon.

### m<sup>6</sup>A methylation is enriched at non-canonical cleavage sites in 3'UTR

Polyadenylation processing of pre-mRNAs is an essential step in the generation of mature mRNAs. It includes an endonucleolytic cleavage followed by polyadenylation (46). A cleavage site is typically located in the downstream 15–30nt of the poly(A) signal (PAS) (Figure 2A). Most eukaryotic genes harbor multiple PASs, leading to expression of alternative polyadenylation (APA) isoforms. m<sup>6</sup>A is known to be associated with APA selection (30,47,48), however, whether m<sup>6</sup>A methylation directly regulates polyadenylation is unknown. The generation of the comprehensive list of m<sup>6</sup>A sites, along with the map of poly(A) cleavage sites in human (49), enables us to systematically examine the relationship between m<sup>6</sup>A methylation and cleavage. To do so, we first examined the distribution of the distance between a cleavage site and the nearest m<sup>6</sup>A site. Unexpectedly, we found that m<sup>6</sup>A sites are highly enriched in the cleavage site position (Figure 2B). This observation was confirmed using miCLIP and RNA-endoribonuclease-facilitated sequencing data that identify m<sup>6</sup>A sites with single-nucleotide-resolution (Supplementary Figure S7A). In addition, compared with 3'UTR m<sup>6</sup>A sites that were not located at the cleavage position, m<sup>6</sup>A sites at the cleavage position tended to be enriched in AAACH sub-motif (Figure 2C).

To examine the effect of m<sup>6</sup>A on PAS regulation, we utilized a PAS reporter assay to measure the impact of m<sup>6</sup>A on poly(A) site processing efficiency (Figure 2D). We first generated METTL3 knockout HEK293T cells (Supplementary Figure S7B, C) and confirmed the reduced m<sup>6</sup>A level (Supplementary Figure S7D). Next, we selected five m<sup>6</sup>A sites that were in the cleavage site position and methylated in HEK293T cells (Supplementary Figure S7E). We sub-cloned ~300 bp region around each m<sup>6</sup>A sites into the reporter plasmid and transfected each of the reporters into both wild-type and METTL3 knockout cells. We found that these m<sup>6</sup>A located PAS regions had a higher processing efficiency in the METTL3 knockout cells (Figure 2D). This result suggests that m<sup>6</sup>A may repress polyadenylation processing.

APA events can lead to the production of noncanonical mRNA isoforms, affecting the fate of the transcript and the nature of the products of translation (50). To ask if m<sup>6</sup>A may be involved in such regulation, we divided APA sites into canonical and noncanonical groups and examined their association with m<sup>6</sup>A. Interestingly, we found that m<sup>6</sup>A sites tended to be located at the noncanonical APA sites (Figure 2E). An examination of the nucleotide compositions around the cleavage sites revealed that, compared with canonical APA sites, noncanonical APA sites had an enrichment of Cs at the immediate downstream of the cleav-



**Figure 2.** m<sup>6</sup>A methylation is enriched at non-canonical cleavage sites in 3'UTR. (A) Schematic representation of a poly(A) site and polyadenylation configuration. PAS, poly(A) signals. (B) The distribution of m<sup>6</sup>A sites around the cleavage sites. Position 0 means the cleavage position. (C) The proportion of m<sup>6</sup>A sub-motifs at the cleavage position or 3'UTR. (D) Top: Schematic diagram showing the bicistronic luciferase reporter system to measure 3' processing efficiency. Fragment of interest was inserted into the MCS position (i.e. between Renilla (R luc) and Firefly luciferase (F luc) genes). Efficient mRNA 3' processing at the tested PAS leads to high expression of Renilla luciferase gene and low expression of Firefly luciferase gene, while poor mRNA 3' processing results in the opposite mode of gene expression. Therefore, the Renilla/Firefly ratio provides a quantitative measurement of the processing efficiency at the tested PAS. Bottom: Luciferase reporter assays to determine the relative PAS activity of 5 selected sites in wild-type and METTL3 knockout cells. Four biological replicates were performed and statistical significance was calculated using Student's t-test. The relative PAS activities are represented as mean  $\pm$  sd. \*\* $P < 0.01$ ; \*\*\* $P < 0.001$ . (E) The proportion of canonical and noncanonical PAS for m<sup>6</sup>A and non-m<sup>6</sup>A cleavage sites. Canonical groups: AAUAAA, AUUAAA and other; noncanonical group, none. (F) Left: nucleotide sequence composition around all 3P-seq-identified canonical and noncanonical poly(A) sites. Position 0 means the cleavage position. Cleavage sites located in 3'UTR or 3'UTR downstream 1kb regions were analyzed. Right: The sequence context 2nt upstream and downstream of the cleavage position.

age position (Figure 2F), thus are more likely to form the RRACH motif required for m<sup>6</sup>A methylation.

Taken together, we unexpectedly observed an enrichment of m<sup>6</sup>A sites in the cleavage position, particularly for the noncanonical APA events. This observation raises the possibility that the methylation status of the cleavage position may affect the cleavage efficiency directly, therefore regulating APA selection. Consequently, the dynamic methylation of m<sup>6</sup>A at cleavage position across tissue types may contribute to the dynamic regulation of APA across tissue types.

### Developmental dynamics of m<sup>6</sup>A methylation

To understand the developmental dynamics of m<sup>6</sup>A methylation, we compared the m<sup>6</sup>A profile between fetal and adult tissues. m<sup>6</sup>A profiles of seven human fetal tissues, includ-

ing brain, heart, kidney, liver, lung, muscle and stomach, were used for analysis (23). Among these tissues, 5 tissue types were in common between fetal and adult samples. m<sup>6</sup>A peaks and sites of fetal tissues were called as we did in our adult tissue data. We confirmed that the samples of the same tissue type clustered together according to either their m<sup>6</sup>A levels or gene expression levels (Supplementary Figure S8A, B). In addition, the called peaks were enriched in the m<sup>6</sup>A consensus motif RRACH (Supplementary Figure S8C). In total, we obtained a union of 60 440 fetal m<sup>6</sup>A sites. These fetal m<sup>6</sup>A sites also preferentially appeared around stop codons (Figure 3A) and spanned various classes of genic regions (Figure 3B). Interestingly, fetal and adult tissues showed distinct distributions of m<sup>6</sup>A sites along the transcripts. Although m<sup>6</sup>A sites preferentially appeared around stop codons in both fetal and adult tissues,

the CDS regions of fetal tissues showed clearly higher m<sup>6</sup>A proportions than that of adult tissues (Figure 3C). Because the m<sup>6</sup>A profile data of fetal tissues were generated using a different RIP procedure and antibody from our method, to exclude the possibility that the observed difference is due to technical issue, we performed two analyses. First, we constructed m<sup>6</sup>A-seq libraries using the same RNA sample with two different methods. We found that sites called from both methods had the same distribution across the transcripts (Supplementary Figure S9A), suggesting that the use of different library construction methods had no significant impact on m<sup>6</sup>A distribution analysis. Second, we analyzed an independent fetal tissue m<sup>6</sup>A-seq data, which include 3 post-conception week 11 (PCW11) fetal human brain samples, 3 mouse developing brain (E13.5) samples and 2 human 47 day forebrain organoid samples (51). These m<sup>6</sup>A-seq data were generated with the same RIP procedure as our data. The enrichment of CDS sites in fetal tissues was confirmed in this data set (Supplementary Figure S9B). VIRMA is known to mediate preferential m<sup>6</sup>A mRNA methylation in 3'UTR and near stop codon (48). An examination of VIRMA expression revealed that, compared with adult tissues, fetal tissues had higher expression levels (Supplementary Figure S9C, D), thus the observed difference between fetal and adult tissues may be due to other unknown regulators.

### The regulation of m<sup>6</sup>A motifs across tissue types or developmental stages

We noted that although the m<sup>6</sup>A consensus motif RRACH is enriched in all tissue types examined, the detailed motifs vary between tissues (Supplementary Figure S1 and Supplementary Figure S8C). To examine the motif dynamics and regulation between tissues, we divided the RRACH motif into four sub-motifs (GGACH, AGACH, GAACH and AAACH) for analysis. We found that different tissues had different sub-motif preferences (Figure 3D). For example, AAACH was overrepresented in both fetal brain and adult frontal cortex. Interestingly, we found that the proportions of AAACH and GGACH sub-motifs were most variable between tissues, while the proportion of AGACH and GAACH sub-motifs were consistent across tissues (Figure 3E and Supplementary Figure S10A). Moreover, during development, the sub-motifs of some tissues were changing but others were not (Supplementary Figure S10B).

m<sup>6</sup>A is installed by a multicomponent methyltransferase complex. Besides the core methyltransferase subunits, it also contains other proteins that interact with core subunits to methylate specific positions (52). The combination of core methyltransferase subunits with different interacting proteins may lead to different motif preference. The observation above suggests that GGACH and AAACH sub-motifs may be installed by core methyltransferase subunits with distinct interacting proteins. If so, we expect that these two sub-motifs may tend to occur in different peak regions. To ask if this is true, we shuffled the m<sup>6</sup>A sub-motif sequence within the peaks and calculated the numbers of peaks with both AAACH and GGACH sub-motifs (Materials and methods). We found that the observed number of peaks with both AAACH and GGACH sub-motifs were

significantly less than the expected numbers in all tissues we examined (Figure 3F), consistent with our expectation.

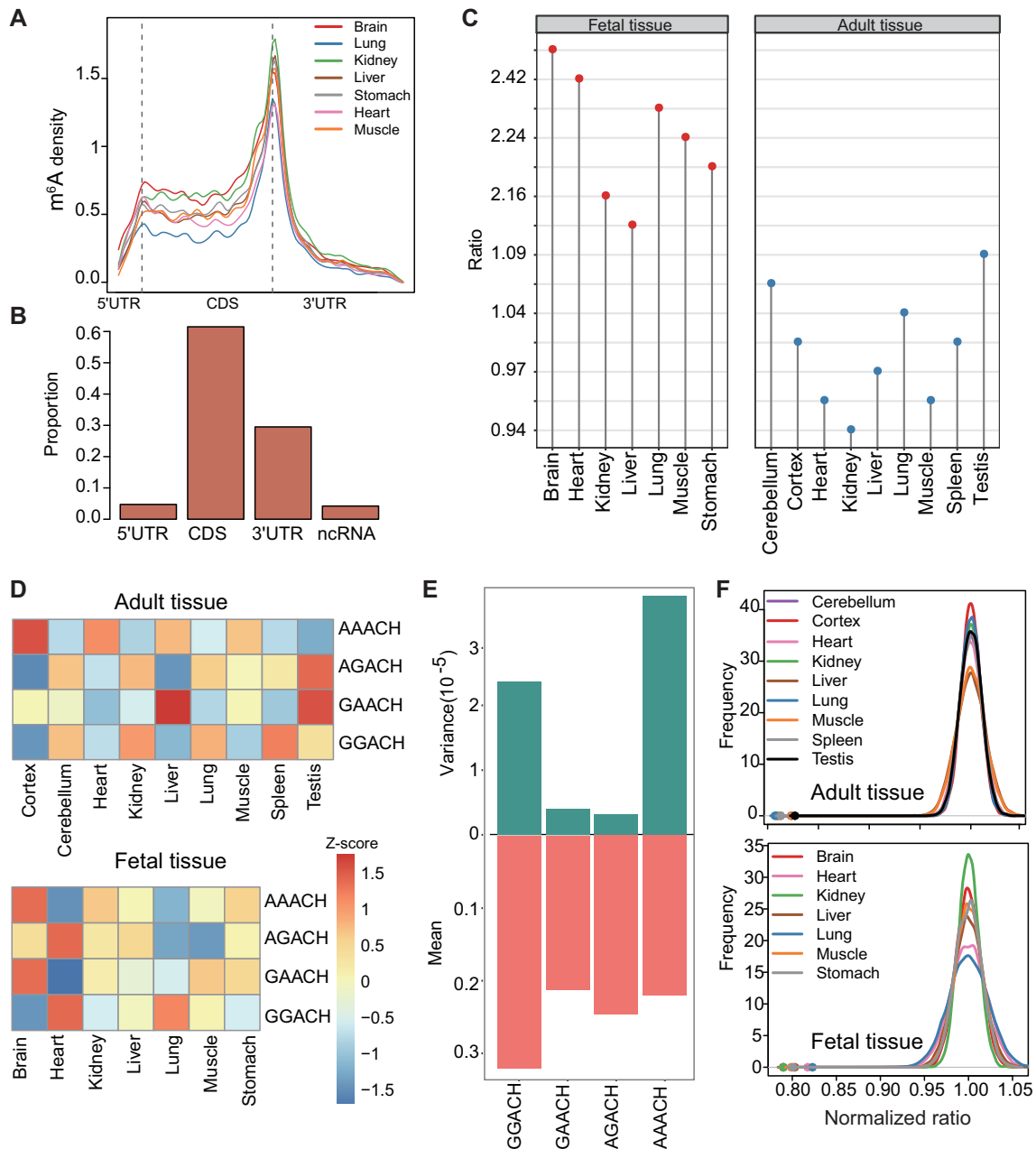
### The evolutionary landscape of human m<sup>6</sup>A methylation

Having revealed the dynamics of m<sup>6</sup>A methylation across human tissues, we next investigated its evolution. First, we examined the cross-species conservation of m<sup>6</sup>A sites to assess the strength of selective pressure on individual m<sup>6</sup>A sites.

For CDS sites, we examined the selection pressure of m<sup>6</sup>A sites in different codon positions, as they may be subject to different evolutionary constraints. To estimate the strength of selection pressure, we chose to compare the fraction of conserved m<sup>6</sup>A sites between human and mouse with that of the control A sites, as previously described (31). It is known that different genes are subject to different selection pressure. To control such effect, for control A sites in non-m<sup>6</sup>A RRACH motifs, we selected sites in genes with similar selection pressure, i.e. similar dN/dS ratio, as the m<sup>6</sup>A sites. In addition, m<sup>6</sup>A sites tended to be located in the 3' end of the CDS region. An examination of evolutionary constraint across the CDS region revealed that As in different CDS regions are subject to different levels of evolutionary constraint (Supplementary Figure S11). To control such effect, for each gene, we grouped the CDS regions into 20 bins and required that the control A sites were located in the same bin as the m<sup>6</sup>A sites. We found that m<sup>6</sup>A sites in different codon positions had different conservation patterns, consistent with a previous study (31). The m<sup>6</sup>A sites in the first codon position were less conserved than control A sites, while m<sup>6</sup>A sites in the third codon were much conserved than control A sites (Figure 4A). We estimated that 6% of the m<sup>6</sup>A sites in the third codon are under evolutionary constraints (Materials and methods), therefore likely functional. m<sup>6</sup>A methylation is known as a barrier to tRNA accommodation and translation elongation, and m<sup>6</sup>A in the first codon position has the strongest effect on delaying tRNA accommodation (53). Therefore, the effect of m<sup>6</sup>A modification in the first codon position may be generally detrimental, and more likely to be less conserved.

For 3'UTR sites, we used INSIGHT (Inference of Natural Selection from Interspersed Genomically coHerent elements) (54), a method for measuring the influence of natural selection for short, widely scattered noncoding elements, to estimate the proportion of m<sup>6</sup>A sites that are under negative selection (in other words, are functional). INSIGHT obtains information about natural selection by contrasting patterns of polymorphism and divergence in m<sup>6</sup>A motifs (RRACH) with those in flanking neutral regions. We obtained estimate of  $\rho$  that range from 0.33 to 0.56 for different tissues (Figure 4B). As a comparison, we examined two additional classes of regulatory elements in 3'UTR region: A-to-I RNA editing sites and miRNA binding sites. For RNA editing sites, we examined the editing site triplet motif (11,55) and estimated  $\rho = 0.03$  (Figure 4B). For miRNA binding sites identified using Targetscan algorithm (41), we obtained an average estimate of  $\rho = 0.21$  and 0.24 for conserved and broadly conserved miRNA families, respectively (Figure 4B). Thus, unexpectedly, we detected a much stronger signature of natural selection in m<sup>6</sup>A motifs com-



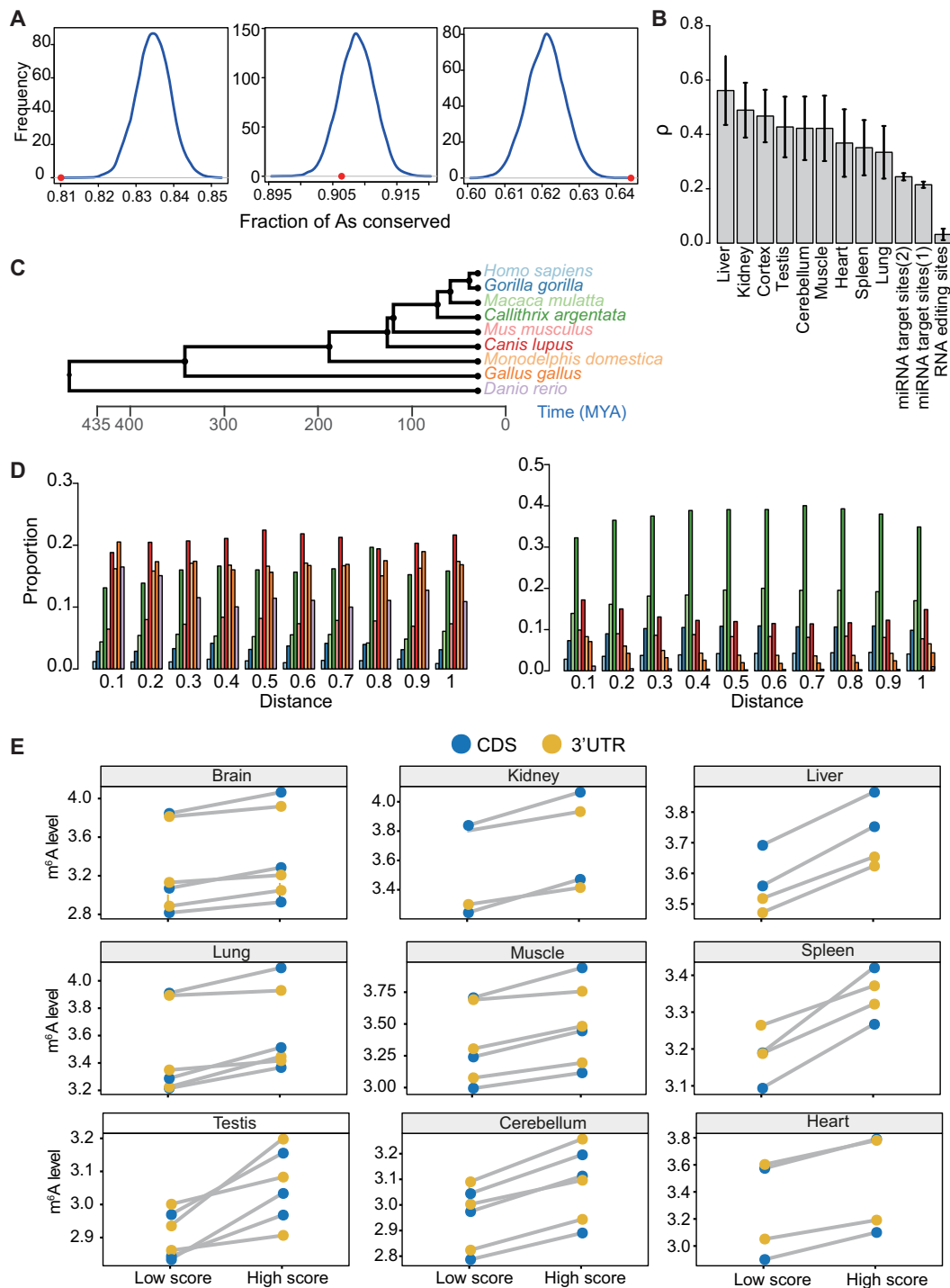


**Figure 3.** Developmental dynamics of m<sup>6</sup>A profile. (A) The distribution of m<sup>6</sup>A sites across the length of mRNA transcripts for 7 fetal tissues. (B) Genic locations of m<sup>6</sup>A sites of fetal tissues. Sites in seven fetal tissues were merged for analysis. (C) The ratio between the CDS m<sup>6</sup>A site number and 3'UTR m<sup>6</sup>A site number in fetal and adult tissues. (D) Heatmap showing the normalized proportion (the values were centered and scaled in the row direction) of 4 m<sup>6</sup>A sub-motifs in adult and fetal tissues. (E) Variance and mean value of the proportion of 4 sub-motifs across human adult tissues. (F) GGACH and AAACH sub-motifs tend to locate in different windows for all tissue types we studied. Color dots indicate the observed numbers of m<sup>6</sup>A peaks with both GGACH and AAACH sub-motifs. The distribution of the expected numbers was generated by the shuffled data. P is the fraction of the distribution on the left side of the dots. It is found that the observed numbers of m<sup>6</sup>A peaks with both GGACH and AAACH sub-motifs are significantly smaller than that of the shuffled data ( $P < 0.0001$ ). The x-axis is the ratio of the number of windows with both sub-motifs over the mean number of windows with both sub-motifs calculated using the shuffled data.

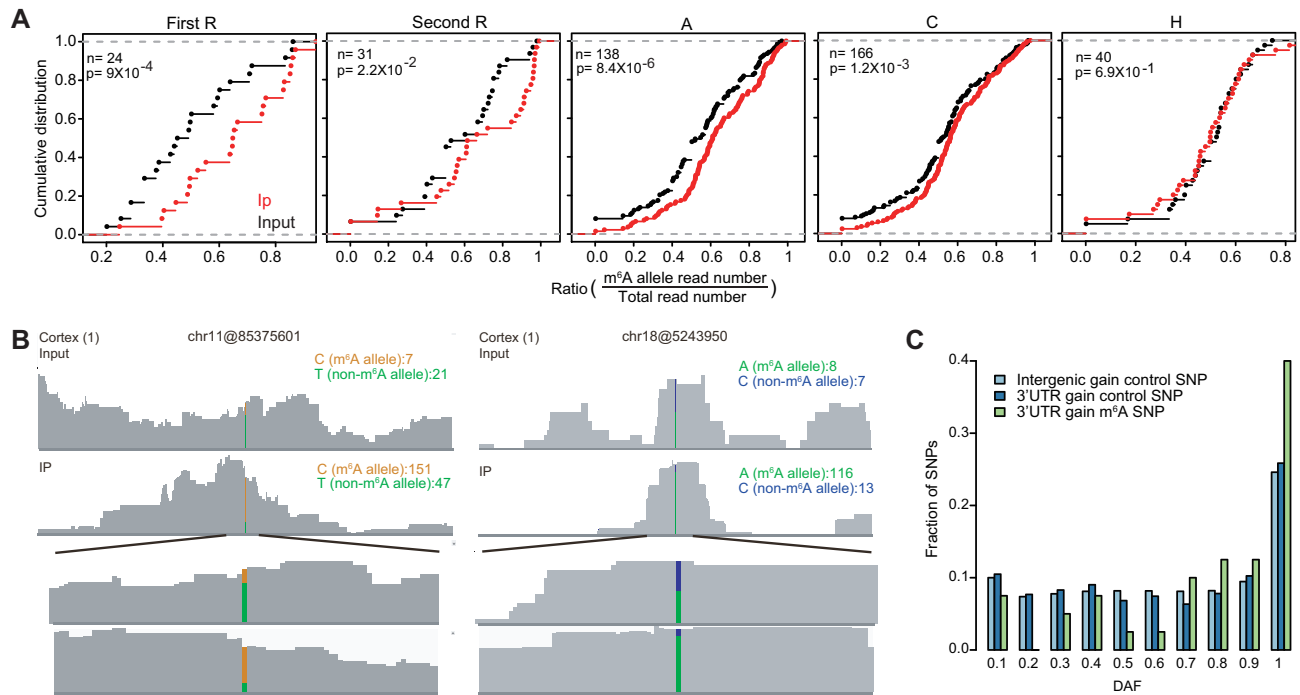
pared with other post-transcriptional regulatory elements, with about half of the nucleotides estimated to be under negative selection. Next, we ask if the selection pressure on 3'UTR m<sup>6</sup>A sites is associated with their locations. We grouped m<sup>6</sup>A sites into 10 bins based on their locations and used phylostratigraphy data (Figure 4C) to examine the age of human m<sup>6</sup>A sites in each bin. We found that m<sup>6</sup>A sites

had an older age than the control sites in all bins, suggesting that 3'UTR m<sup>6</sup>A sites are generally subject to negative selection (Figure 4D).

Second, we examined the relationship between m<sup>6</sup>A level and m<sup>6</sup>A site conservation. We obtained the rejected substitution scores, a score to measure the nucleotide-level constraint (56), of all m<sup>6</sup>A sites, and compared the methylation



**Figure 4.** Natural selection on  $m^6A$  inferred by cross-species analysis. (A) Comparison of evolutionary conservation between  $m^6A$  and control A sites (non- $m^6A$  RRACH). Frequency distributions of the fraction of conserved control sites in 10,000 random sets with the sample size equal to the number of  $m^6A$  sites at three codon positions were plotted separately. Red dots indicate the fraction of conserved  $m^6A$  sites.  $P$  is the fraction of the distribution on the right side of the dots. First codon,  $P = 1$ ; second codon position,  $P = 0.89$ ; third codon position,  $P < 0.0001$ . (B) Estimates of  $\rho$  for  $m^6A$  RRACH motifs, A-to-I RNA editing triplet motif and miRNA binding sites.  $\rho$  represents the fraction of sites under selection within functional elements, which is calculated by INSIGHT.  $m^6A$  sites,  $m^6A$  RRACH motifs in 3'UTR region; RNA editing sites, nonrepetitive A-to-I RNA editing sites in 3'UTR region; miRNA target sites, 3'UTR miRNA binding sites for conserved (1) or broadly conserved (2) miRNA families. The regions that match the seed region of the miRNAs were used for analysis. (C) A tree representing the schematic phylogeny of the species studied. (D) The age distribution of  $m^6A$  sites (left) and control non- $m^6A$  RRACH sites (right) across the 3'UTR region.  $m^6A$  sites and control sites were grouped into 10 bins based on their distance to stop codon. The age of a site was based on the most distantly related species in which the site was conserved. The color codes for the age are the same as in (C). (E) Comparison of methylation levels between  $m^6A$  sites under strong and weak constraints in CDS and 3'UTR.  $m^6A$  sites with top 25% (High score, strong constraint) and bottom 25% (Low score, weak constraint) rejection scores were compared.  $m^6A$  peak winscore of a site was used to represent  $m^6A$  level. Each dot represents the median methylation level of the CDS or 3'UTR sites; biological replicates of each tissue type were plotted, separately.



**Figure 5.** Positive selection inferred from SNP genotype data. (A) Comparison of the m<sup>6</sup>A allele ratio between IP and input samples of heterozygote m<sup>6</sup>A SNPs. Only heterozygote SNPs with one genotype matching the RRACH motif and the other genotype not matching the RRACH motif were considered. SNPs located in each position of RRACH motifs were analyzed, separately. *P* values were calculated with the Kolmogorov-Smirnov test. (B) Examples of m<sup>6</sup>A peaks with SNPs. For chr11@85375601 m<sup>6</sup>A site, the C allele (GAACT, m<sup>6</sup>A site is underlined and the SNP is marked in italics) matches the m<sup>6</sup>A motif and the T allele (GAATT) disrupts the motif. For chr18@5243950 m<sup>6</sup>A site, the A allele (AAACA) matches the m<sup>6</sup>A motif and the C allele (ACACA) disrupts the motif. For both sites, compared with the non-m<sup>6</sup>A allele, the m<sup>6</sup>A allele is highly enriched in the IP sample. (C) DAF distributions of 1,000 Genome Project SNPs of m<sup>6</sup>A RRAC motifs for 3'UTR sites. The derived alleles that create m<sup>6</sup>A motifs were analyzed. Two groups of control sites were selected: (1) 'RRAC' from all non-m<sup>6</sup>A RRACH motifs located at 3'UTR region; (2) 'RRAC' from all RRACH motifs in the intergenic region. *P* values were calculated with a one-sided Mann-Whitney U test comparing the DAF distribution of SNPs in m<sup>6</sup>A RRAC motif with the distribution of SNPs in control sites. 3'UTR control group versus m<sup>6</sup>A group, *P* = 0.00037; intergenic group versus m<sup>6</sup>A group, *P* = 0.000013.

levels between m<sup>6</sup>A sites under stronger constraints and weaker constraints. Interestingly, we found that m<sup>6</sup>A sites under stronger constraints had higher methylation levels (Figure 4E). This observation suggests that the conserved m<sup>6</sup>A sites are optimized for m<sup>6</sup>A writer binding and methylation, thus likely functional.

### Positive selection of m<sup>6</sup>A sites inferred from population genomic analysis

To ask if the m<sup>6</sup>A sites that were recently gained during human evolution were under positive selection, we analyzed SNP genotype data from the 1,000 Genome Project (Materials and methods). To prevent the ascertainment bias between functional classes, 5'UTR, 3'UTR of protein-coding genes and ncRNAs were analyzed. CDS region was excluded from this analysis, because it is difficult to distinguish if the selection signals are from m<sup>6</sup>A methylation or other factors unrelated to m<sup>6</sup>A methylation, such as amino acid changes. First, we asked if SNPs located within the RRACH motifs do affect methylation. To do so, we first identified all heterogeneous m<sup>6</sup>A SNPs in 27 adult tissue samples using the input RNA-seq data. Next, we calculated the m<sup>6</sup>A allele ratio (m<sup>6</sup>A allele read number/ total read

number) using reads covered the selected SNPs for both IP and input samples. Last, we compared the m<sup>6</sup>A allele ratios between IP and input samples and each position of the RRACH motif was examined separately. We found that m<sup>6</sup>A allele had higher ratios in IP samples for SNPs located in the RRAC positions (Figure 5A). Thus SNPs located at the RRAC positions but not the H position affect methylation status. The read coverages of two representative heterogeneous m<sup>6</sup>A SNPs in IP and input samples were shown in Figure 5B. Next, we examined the DAF spectrums for a derived allele that are located in the RRAC positions and create an m<sup>6</sup>A motif. Different DAF distributions were directly compared using a Mann-Whitney test, as previously described (27). Interestingly, the DAF distribution for the SNPs of which the derived allele creates m<sup>6</sup>A motifs was significantly skewed toward high-frequency alleles relative to matched control or neutral sites, suggesting that some of these SNPs are subject to positive selection (Figure 5C).

The excess of high-frequency derived alleles that create m<sup>6</sup>A motifs promotes us to further characterize the m<sup>6</sup>A site SNPs that are likely under positive selection. To have a comprehensive scan of positively selected SNPs related to m<sup>6</sup>A modification, we examined not only the SNPs of which the derived allele is m<sup>6</sup>A allele but also those of which the

ancestral allele is m<sup>6</sup>A allele. Because it is plausible that selection may sometimes switch to favor an ancestral allele that has been segregating in the population. We first identified m<sup>6</sup>A SNPs that have been highly differentiated among populations (measured by the *Fst* parameter) and then determined where these differentiation events might be driven by positive selection during human evolution. In total, we identified 102 highly differentiated SNPs (*Fst* > 0.15) (Supplementary Table S5). Of these SNPs, 37 showed evidence of selection in iHS (the integrated Haplotype Scores) (57) and/or XP-EHH (Cross Population Extended Haplotype Homozygosity) (58) tests (Supplementary Tables S6–S7).

Among these differentiated SNPs, 81 SNPs are located in protein-coding genes. The remaining 21 SNPs are located in ncRNA genes, suggesting that m<sup>6</sup>A sites in ncRNA genes may be another class of targets of recent positive selection. The protein-coding gene list contains a number of genes involved in biological pathways thought to be recently targeted by positive selection, such as metabolism of carbohydrates, lipids and brain development (57). Particularly, it includes 25 genes that have previously been characterized as positively selected genes in the human lineage or across human populations (Supplementary Table S5).

As m<sup>6</sup>A is known to affect mRNA stability, to further understand the potential regulatory effects of these SNPs, we examined the overlaps between these variants and eQTL SNPs from the Genotype-Tissue Expression (GTEx) project. Of the 102 SNPs, 60 are associated with gene expression abundance (Supplementary Table S8), suggesting that these m<sup>6</sup>A SNPs may contribute to the expression changes. Last, we characterized the potential phenotypic effect of these variants. Using Genome-wide association study (GWAS) data, we found that a substantial proportion of the 102 m<sup>6</sup>A SNPs is disease- or trait- associated SNPs (Supplementary Table S8), such as those associated with HDL cholesterol levels, body mass index, and atherosclerosis.

## DISCUSSION

The importance of m<sup>6</sup>A as a post-transcriptional modification has been appreciated, but the evolution, function and regulation of individual m<sup>6</sup>A sites remain largely unknown, in part because of insufficient data about its prevalence and dynamics. Here, we compiled the m<sup>6</sup>A methylomes of major adult human tissues, providing a valuable resource for future studies of the regulation and functions of this modification. We reveal that the distribution and motifs of m<sup>6</sup>A vary across tissue types or during development, suggesting that m<sup>6</sup>A is widely regulated by trans factors and involved in human development. Notably, it is known that the variation of postmortem conditions in different samples may affect RNA integrity and gene expression quantification. Since it is unknown that whether m<sup>6</sup>A in post-mortem tissues represents the in-situ state and whether m<sup>6</sup>A is more unstable than RNA itself, caution needs to be taken when using the m<sup>6</sup>A maps generated with postmortem tissues.

We used comparative genomics and population genetics approaches to show that a significant negative selection has acted on m<sup>6</sup>A sites, particularly the ones in the third codon position and 3'UTR. Current opinion on m<sup>6</sup>A modification

believes that despite the functional importance of this modification, the single m<sup>6</sup>A site seems dispensable, as long as a transcript is methylated and can be recognized by a member of the major m<sup>6</sup>A reader YTH family (20,59,60). Our data, however, do not support this view and suggest that many of the single sites should be functionally important, most likely because their functions are position-dependent. Furthermore, with 1,000 Genome Project data, we identified a number of m<sup>6</sup>A SNPs whose patterns of allelic variation are not consistent with neutrality. There is a functional difference between alleles and finally the functional difference would result in a phenotypic effect that would be influenced by selection. These SNPs are enriched in the genes related to the immune system, dietary fatty acid processing and neuronal functions, consistent with the recently identified functions of m<sup>6</sup>A modification (16,61–64).

In summary, our work provides a resource of m<sup>6</sup>A profile in humans for future studies of m<sup>6</sup>A regulation and functions. Furthermore, our data provide independent evidence for the functional importance of m<sup>6</sup>A modification from the evolutionary perspective, and also suggests an unexpected role of m<sup>6</sup>A modification in recent human evolutionary adaptation and disease susceptibility.

## DATA AVAILABILITY

The sequence data have been deposited in the NCBI GEO database under the accession code GSE122744.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

## ACKNOWLEDGEMENTS

We thank Dr Jinkai Wang for sharing the m<sup>6</sup>A peak identification codes, Dr Chengguo Yao and Dr Yongsheng Shi for sharing the pPASPORT vector, and SYSU EEB Sequencing Core Facility for sequencing service.

*Author contribution:* T.H., X.N.Z., W.Y.C., H.Z. and N.N.G. carried out the experiments. H.Z. and X.R.S. performed computational analyses. R.Z., H.Z., X.R.S., T.H. and N.N.G. wrote the paper.

## FUNDING

National Natural Science Foundation of China [91631108, 31571341 to R.Z.]; Guangdong Innovative and Entrepreneurial Research Team Program [2016ZT06S638 to R.Z.]; Thousand Talents Plan-The Recruitment Program for Young Professionals (to R.Z.). Funding for open access charge: National Natural Science Foundation of China [91631108 and 31571341 to R.Z.].

*Conflict of interest statement.* None declared.

## REFERENCES

- Li, S. and Mason, C.E. (2014) The pivotal regulatory landscape of RNA modifications. *Annu. Rev. Genomics Hum. Genet.*, **15**, 127–150.
- Dominianni, 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<sup>6</sup>A RNA methylomes revealed by m<sup>6</sup>A-seq. *Nature*, **485**, 201–206.

3. Ramaswami,G., Zhang,R., Piskol,R., Keegan,L.P., Deng,P., O'Connell,M.A. and Li,J.B. (2013) Identifying RNA editing sites using RNA sequencing data alone. *Nat. Methods*, **10**, 128–132.
4. Edelheit,S., Schwartz,S., Mumbach,M.R., Wurtzel,O. and Sorek,R. (2013) Transcriptome-wide mapping of 5-methylcytidine RNA modifications in bacteria, archaea, and yeast reveals m5C within archaeal mRNAs. *PLoS Genet.*, **9**, e1003602.
5. Li,X., Xiong,X. and Yi,C. (2016) Epitranscriptome sequencing technologies: decoding RNA modifications. *Nat. Methods*, **14**, 23–31.
6. Huang,T., Chen,W.Y., Liu,J.H., Gu,N.N. and Zhang,R. (2019) Genome-wide identification of mRNA 5-methylcytosine in mammals. *Nat. Struct. Mol. Biol.*, **26**, 380–388.
7. Nishikura,K. (2010) Functions and regulation of RNA editing by ADAR deaminases. *Annu. Rev. Biochem.*, **79**, 321–349.
8. Xu,G. and Zhang,J. (2014) Human coding RNA editing is generally nonadaptive. *Proc. Natl. Acad. Sci. U.S.A.*, **111**, 3769–3774.
9. Yu,Y., Zhou,H., Kong,Y., Pan,B., Chen,L., Wang,H., Hao,P. and Li,X. (2016) The landscape of A-to-I RNA editome is shaped by both positive and purifying selection. *PLoS Genet.*, **12**, e1006191.
10. Duan,Y., Dou,S., Luo,S., Zhang,H. and Lu,J. (2017) Adaptation of A-to-I RNA editing in *Drosophila*. *PLoS Genet.*, **13**, e1006648.
11. Zhang,R., Deng,P., Jacobson,D. and Li,J.B. (2017) Evolutionary analysis reveals regulatory and functional landscape of coding and non-coding RNA editing. *PLoS Genet.*, **13**, e1006563.
12. Liscovitch-Brauer,N., Alon,S., Porath,H.T., Elstein,B., Unger,R., Ziv,T., Admon,A., Levanon,E.Y., Rosenthal,J.J.C. and Eisenberg,E. (2017) Trade-off between transcriptome plasticity and genome evolution in cephalopods. *Cell*, **169**, 191–202.
13. 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.
14. Meyer,K.D. and Jaffrey,S.R. (2014) The dynamic epitranscriptome: N6-methyladenosine and gene expression control. *Nat. Rev. Mol. Cell Biol.*, **15**, 313–326.
15. Zhao,B.S., Roundtree,I.A. and He,C. (2017) Post-transcriptional gene regulation by mRNA modifications. *Nat. Rev. Mol. Cell Biol.*, **18**, 31–42.
16. Fu,Y., Dominissini,D., Rechavi,G. and He,C. (2014) Gene expression regulation mediated through reversible m6A RNA methylation. *Nat. Rev. Genet.*, **15**, 293–306.
17. Bokar,J.A., Shambaugh,M.E., Polayes,D., Matera,A.G. and Rottman,F.M. (1997) Purification and cDNA cloning of the AdoMet-binding subunit of the human mRNA (N6-adenosine)-methyltransferase. *RNA*, **3**, 1233–1247.
18. 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 N6-adenosine methylation. *Nat. Chem. Biol.*, **10**, 93–95.
19. Wang,X., Lu,Z., Gomez,A., Hon,G.C., Yue,Y., Han,D., Fu,Y., Parisien,M., Dai,Q., Jia,G. *et al.* (2014) N6-methyladenosine-dependent regulation of messenger RNA stability. *Nature*, **505**, 117–120.
20. Wang,X., Zhao,B.S., Roundtree,I.A., Lu,Z., Han,D., Ma,H., Weng,X., Chen,K., Shi,H. and He,C. (2015) N(6)-methyladenosine modulates messenger RNA translation efficiency. *Cell*, **161**, 1388–1399.
21. 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 m6A Reader YTHDC1 Regulates mRNA Splicing. *Mol. Cell*, **61**, 507–519.
22. Sun,W.-J., Li,J.-H., Liu,S., Wu,J., Zhou,H., Qu,L.-H. and Yang,J.-H. (2015) RMBase: a resource for decoding the landscape of RNA modifications from high-throughput sequencing data. *Nucleic Acids Res.*, **44**, D259–D265.
23. 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.
24. King,M. and Wilson,A. (1975) Evolution at two levels in humans and chimpanzees. *Science*, **188**, 107–116.
25. Barbosa-Morais,N.L., Irimia,M., Pan,Q., Xiong,H.Y., Gueroussov,S., Lee,L.J., Slobodenic,V., Kutter,C., Watt,S., Colak,R. *et al.* (2012) The evolutionary landscape of alternative splicing in vertebrate species. *Science*, **338**, 1587–1593.
26. Xu,J., Zhang,R., Shen,Y., Liu,G., Lu,X. and Wu,C.I. (2013) The evolution of evolvability in microRNA target sites in vertebrates. *Genome Res.*, **23**, 1810–1816.
27. Chen,K. and Rajewsky,N. (2006) Natural selection on human microRNA binding sites inferred from SNP data. *Nat. Genet.*, **38**, 1452–1456.
28. Arbiza,L., Gronau,I., Aksoy,B.A., Hubisz,M.J., Gulko,B., Keinan,A. and Siepel,A. (2013) Genome-wide inference of natural selection on human transcription factor binding sites. *Nat. Genet.*, **45**, 723–729.
29. Ma,L., Zhao,B., Chen,K., Thomas,A., Tuteja,J.H., He,X., He,C. and White,K.P. (2017) Evolution of transcript modification by N6-methyladenosine in primates. *Genome Res.*, **27**, 385–392.
30. Ke,S., Alemu,E.A., Mertens,C., Gaotman,E.C., Fak,J.J., Mele,A., Haripal,B., Zucker-Scharff,I., Moore,M.J., Park,C.Y. *et al.* (2015) A majority of m6A residues are in the last exons, allowing the potential for 3' UTR regulation. *Genes Dev.*, **29**, 2037–2053.
31. Liu,Z. and Zhang,J. (2018) Most m6A RNA modifications in protein-coding regions are evolutionarily unconserved and likely nonfunctional. *Mol. Biol. Evol.*, **35**, 666–675.
32. Batista,Pedro J., Molinie,B., Wang,J., Qu,K., Zhang,J., Li,L., Bouley,Donna M., Lujan,E., Haddad,B., Daneshvar,K. *et al.* (2014) m6A RNA modification controls cell fate transition in mammalian embryonic stem cells. *Cell Stem Cell*, **15**, 707–719.
33. Martin,M. (2011) Cutadapt removes adapter sequences from High-Throughput sequencing reads. *EMBnet.journal*, **17**, 10–12.
34. Kopylova,E., Noe,L. and Touzet,H. (2012) SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. *Bioinformatics*, **28**, 3211–3217.
35. Kim,D., Perteau,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.
36. Meng,J., Cui,X., Rao,M.K., Chen,Y. and Huang,Y. (2013) Exome-based analysis for RNA epigenome sequencing data. *Bioinformatics*, **29**, 1565–1567.
37. 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.
38. Kryuchkova-Mostacci,N. and Robinson-Rechavi,M. (2016) A benchmark of gene expression tissue-specificity metrics. *Brief. Bioinform.*, **18**, 205–214.
39. Yao,C., Biesinger,J., Wan,J., Weng,L., Xing,Y., Xie,X. and Shi,Y. (2012) Transcriptome-wide analyses of CstF64-RNA interactions in global regulation of mRNA alternative polyadenylation. *PNAS*, **109**, 18773–18778.
40. Ramaswami,G. and Li,J.B. (2014) RADAR: a rigorously annotated database of A-to-I RNA editing. *Nucleic Acids Res.*, **42**, D109–D113.
41. Agarwal,V., Bell,G.W., Nam,J.W. and Bartel,D.P. (2015) Predicting effective microRNA target sites in mammalian mRNAs. *eLife*, **4**, e05005.
42. Danecek,P., Auton,A., Abecasis,G., Albers,C.A., Banks,E., DePristo,M.A., Handsaker,R.E., Lunter,G., Marth,G.T., Sherry,S.T. *et al.* (2011) The variant call format and VCFtools. *Bioinformatics*, **27**, 2156–2158.
43. Gautier,M. and Vitalis,R. (2012) rehh: an R package to detect footprints of selection in genome-wide SNP data from haplotype structure. *Bioinformatics*, **28**, 1176–1177.
44. Welter,D., MacArthur,J., Morales,J., Burdett,T., Hall,P., Junkins,H., Klemm,A., Flicek,P., Manolio,T., Hindorf,L. *et al.* (2014) The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res.*, **42**, D1001–D1006.
45. Yu,G., Wang,L.-G., Han,Y. and He,Q.-Y. (2012) clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS*, **16**, 284–287.
46. Colgan,D.F. and Manley,J.L. (1997) Mechanism and regulation of mRNA polyadenylation. *Genes Dev.*, **11**, 2755–2766.
47. Molinie,B., Wang,J., Lim,K.S., Hillebrand,R., Lu,Z.-x., Van Wittenberghe,N., Howard,B.D., Daneshvar,K., Mullen,A.C., Dedon,P. *et al.* (2016) m6A-LAIC-seq reveals the census and complexity of the m6A epitranscriptome. *Nat Meth.*, **13**, 692–698.
48. Yue,Y., Liu,J., Cui,X., Cao,J., Luo,G., Zhang,Z., Cheng,T., Gao,M., Shu,X., Ma,H. *et al.* (2018) VIRMA mediates preferential m(6)A mRNA methylation in 3'UTR and near stop codon and associates with alternative polyadenylation. *Cell discovery*, **4**, 10.

49. Wang,R., Nambiar,R., Zheng,D. and Tian,B. (2018) PolyA-DB 3 catalogs cleavage and polyadenylation sites identified by deep sequencing in multiple genomes. *Nucleic Acids Res.*, **46**, D315–D319.
50. Tian,B. and Manley,J.L. (2016) Alternative polyadenylation of mRNA precursors. *Nat. Rev. Mol. Cell Biol.*, **18**, 18.
51. Yoon,K.-J., Ringeling,F.R., Vissers,C., Jacob,F., Pokrass,M., Jimenez-Cyrus,D., Su,Y., Kim,N.-S., Zhu,Y., Zheng,L. *et al.* (2017) Temporal control of mammalian cortical neurogenesis by m6A Methylation. *Cell*, **171**, 877–889.
52. Yang,Y., Hsu,P.J., Chen,Y.S. and Yang,Y.G. (2018) Dynamic transcriptomic m(6)A decoration: writers, erasers, readers and functions in RNA metabolism. *Cell Res.*, **28**, 616–624.
53. Choi,J., Jeong,K.W., Demirci,H., Chen,J., Petrov,A., Prabhakar,A., O’Leary,S.E., Dominissini,D., Rechavi,G., Soltis,S.M. *et al.* (2016) N(6)-methyladenosine in mRNA disrupts tRNA selection and translation-elongation dynamics. *Nat. Struct. Mol. Biol.*, **23**, 110–115.
54. Gronau,I., Arbiza,L., Mohammed,J. and Siepel,A. (2013) Inference of natural selection from interspersed genomic elements based on polymorphism and divergence. *Mol. Biol. Evol.*, **30**, 1159–1171.
55. Eggington,J.M., Greene,T. and Bass,B.L. (2011) Predicting sites of ADAR editing in double-stranded RNA. *Nat. Commun.*, **2**, 319.
56. Davydov,E.V., Goode,D.L., Sirota,M., Cooper,G.M., Sidow,A. and Batzoglou,S. (2010) Identifying a high fraction of the human genome to be under selective constraint using GERP++. *PLoS Comput. Biol.*, **6**, e1001025.
57. Voight,B.F., Kudaravalli,S., Wen,X. and Pritchard,J.K. (2006) A map of recent positive selection in the human genome. *PLoS Biol.*, **4**, e72.
58. Sabeti,P.C., Varilly,P., Fry,B., Lohmueller,J., Hostetter,E., Cotsapas,C., Xie,X., Byrne,E.H., McCarroll,S.A., Gaudet,R. *et al.* (2007) Genome-wide detection and characterization of positive selection in human populations. *Nature*, **449**, 913–918.
59. Geula,S., Moshitch-Moshkovitz,S., Dominissini,D., Mansour,A.A., Kol,N., Salmon-Divon,M., Hershkovitz,V., Peer,E., Mor,N., Manor,Y.S. *et al.* (2015) Stem cells. m6A mRNA methylation facilitates resolution of naive pluripotency toward differentiation. *Science*, **347**, 1002–1006.
60. Zhou,J., Wan,J., Gao,X., Zhang,X., Jaffrey,S.R. and Qian,S.-B. (2015) Dynamic m6A mRNA methylation directs translational control of heat shock response. *Nature*, **526**, 591–594.
61. Frayling,T.M., Timpson,N.J., Weedon,M.N., Zeggini,E., Freathy,R.M., Lindgren,C.M., Perry,J.R.B., Elliott,K.S., Lango,H., Rayner,N.W. *et al.* (2007) A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. *Science*, **316**, 889–894.
62. 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.
63. Zheng,Q., Hou,J., Zhou,Y., Li,Z. and Cao,X. (2017) The RNA helicase DDX46 inhibits innate immunity by entrapping m6A-demethylated antiviral transcripts in the nucleus. *Nat. Immunol.*, **18**, 1094–1103.
64. Li,H.B., Tong,J., Zhu,S., Batista,P.J., Duffy,E.E., Zhao,J., Bailis,W., Cao,G., Kroehling,L., Chen,Y. *et al.* (2017) m6A mRNA methylation controls T cell homeostasis by targeting the IL-7/STAT5/SOCS pathways. *Nature*, **548**, 338–342.