

Integrative network analysis identifies cell-specific *trans* regulators of m⁶A

Sanqi An^{1,2,†}, Wanxu Huang^{1,2,†}, Xiang Huang^{1,2,†}, Yixian Cun^{1,2}, Weisheng Cheng^{1,2}, Xiang Sun^{1,2}, Zhijun Ren^{1,2}, Yaxin Chen^{1,2}, Wenfang Chen^{1,2} and Jinkai Wang^{1,2,3,4,*}

¹Department of Medical Bioinformatics, Zhongshan School of Medicine, Sun Yat-sen University, Guangzhou 510080, China, ²Center for Stem Cell Biology and Tissue Engineering, Key Laboratory for Stem Cells and Tissue Engineering, Ministry of Education, Sun Yat-sen University, Guangzhou 510080, China, ³RNA Biomedical Institute, Sun Yat-sen Memorial Hospital, Sun Yat-sen University, Guangzhou 510120, China and ⁴Center for Precision Medicine, Sun Yat-sen University, Guangzhou 510080, China

Received September 22, 2019; Revised December 03, 2019; Editorial Decision December 12, 2019; Accepted December 18, 2019

ABSTRACT

N⁶-methyladenosine (m⁶A) is a reversible and dynamic RNA modification in eukaryotes. However, how cells establish cell-specific m⁶A methylomes is still poorly understood. Here, we developed a computational framework to systematically identify cell-specific *trans* regulators of m⁶A through integrating gene expressions, binding targets and binding motifs of large number of RNA binding proteins (RBPs) with a co-methylation network constructed using large-scale m⁶A methylomes across diverse cell states. We applied the framework and successfully identified 32 high-confidence m⁶A regulators that modulated the variable m⁶A sites away from stop codons in a cell-specific manner. To validate them, we knocked down three regulators respectively and found two of them (TRA2A and CAPRN1) selectively promoted the methylations of the m⁶A sites co-localized with their binding targets on RNAs through physical interactions with the m⁶A writers. Knockdown of TRA2A increased the stabilities of the RNAs with TRA2A bound near the m⁶A sites and decreased the viability of cells. The successful identification of m⁶A regulators demonstrates a powerful and widely applicable strategy to elucidate the cell-specific m⁶A regulators. Additionally, our discovery of pervasive *trans*-acting regulating of m⁶A provides novel insights into the mechanisms by which spatial and temporal dynamics of m⁶A methylomes are established.

INTRODUCTION

N⁶-methyladenosine (m⁶A) is the most prevalent internal RNA modification in mRNA and long non-coding RNAs of eukaryotes. It is a reversible RNA modification prefers to occur on DRACH motif near stop codon and in long internal exon of mRNA (1,2). A nuclear methyltransferase complex consisting of METTL3, which is the catalytic subunit, METTL14, WTAP, VIRMA, ZC3H13, RBM15 (or RBM15B) and CBL1/HAKAI catalyzes the m⁶A modifications co-transcriptionally, acting as m⁶A ‘writers’ (3–5). A specific m⁶A demethylase ALKBH5 as well as a less specific m⁶A demethylase FTO mediate the demethylation of m⁶As, acting as the m⁶A ‘erasers’ (6). A variety of proteins including YTH domain-containing proteins can specifically bind m⁶A marks as the m⁶A ‘readers’ and regulate a variety of post-transcriptional processes, such as RNA decay, alternative splicing, translation, alternative polyadenylation and nuclear export (7–9).

It is widely accepted that m⁶A RNA methylation is dynamically regulated (10). More and more studies reported the alterations of functionally important m⁶A sites caused by expression change of m⁶A writers and erasers played important roles in a variety of physiological and pathological processes (11–15). A recent study reported that 33–46% of the variability of m⁶A levels were due to *cis*-regulation, suggesting that the dynamics of m⁶A are likely through global regulation by modulating the abundances of methyltransferase components (16). However, it is still unclear whether *trans*-regulation plays important roles in site-specific dynamics of m⁶A levels. Besides global regulation, site-specific m⁶A dynamics can possibly be precisely established through the interplays of a variety of *trans*-acting m⁶A regulators with m⁶A writers and erasers at specific sites bound by the regulators. Indeed, m⁶A is deposited on nascent RNAs (17) and can be regulated co-transcriptionally through H3K36me3 histone modification

*To whom correspondence should be addressed. Tel: +86 2087335142; Fax: +86 2087331209; Email: wangjk@mail.sysu.edu.cn

†These authors contributed equally to the paper as first authors.

(18) as well as transcription factors (19,20). Transcription factor CEBPZ recruits METTL3 to the promoter of a specific set of active genes to regulate the m⁶A of the associated mRNAs involved in acute myeloid leukemia (19). Similarly, transcription factors SMAD2/3 can selectively promote the m⁶A modifications of the genes involved in early cell fate decision through co-transcriptional recruitment of the m⁶A methyltransferase complex onto the nascent RNAs (20). On the other hand, Cao *et al.* reported two RNA binding proteins (RBPs) DDX46 and HNRNPA2B1 dynamically interacted with m⁶A erasers to regulate the m⁶A of genes critical for innate immunity in response to viral infection (21,22). Nevertheless, whether specific regulation of m⁶A is prevalent remains a mystery.

Systematical analyses of large-scale m⁶A methylomes are promising to elucidate the *trans*-acting m⁶A regulators. Theoretically, there should be a correlation between the gene expression of a *trans*-acting m⁶A regulator and the m⁶A levels of the m⁶A sites regulated by the regulator. However, in practice, it is very challenging due to various technical difficulties: (i) proper quantification of m⁶A is difficult due to various technical biases of m⁶A-seq data; (ii) it is almost impossible to afford serious multiple testing correction for massive correlation tests between genes and m⁶A sites; (iii) Pearson correlation of m⁶A levels quantified using m⁶A-seq suffers seriously by the outlier issue; (iv) correlations may not reflect direct effects of regulation.

In this study, we developed a computational framework to systematically identify cell-specific *trans* regulators of m⁶A through integrating gene expressions, binding targets and binding motifs of a large number of RBPs with a co-methylation network constructed using large-scale m⁶A methylomes across diverse cell states. We applied the framework to the public available m⁶A-seq data of 25 unique cell lines and successfully identified 32 high-confidence m⁶A regulators with reasonable experimental validation rate, demonstrating a powerful and widely applicable strategy to elucidate cell-specific the m⁶A regulators. Our discovery of pervasive *trans*-acting regulating of m⁶A provided novel insights into the mechanisms by which spatial and temporal dynamics of m⁶A methylomes were established.

MATERIALS AND METHODS

Processing of the m⁶A-seq data in multiple cell lines

Raw sequence data of 104 m⁶A-seq libraries (IP and input) from 25 unique cell lines were downloaded from Sequence Read Archive (SRA, <https://www.ncbi.nlm.nih.gov/sra>) (1–2,19,23–37). The accession numbers of these data can be found in Supplementary Table S1. The reads were mapped to hg19 human genome using HISTA2 (v2.1.0) (38). We used StringTie (v1.3.4d) (39) to calculate the TPMs (Transcripts Per Million) of Ensembl annotated genes using the input libraries, followed by quantile normalization of the TPMs across all samples. m⁶A peaks were identified according to the methods as described in our previous paper (23), which was modified from the method published earlier by Dominissini *et al.* (2). Briefly, we made sliding windows of 100 bp with 50 bp overlap on the exon regions and calculated the RPKM of each window. The sliding windows with winscore (enrichment score) >2 were identified as m⁶A

peaks in each sample (2,23). To deal with the technical issue that lowly expressed windows might have unreliable winscores, we added 1 to the RPKM of each window in both IP and input before winscore calculation in order to penalize the windows with low RPKMs. We took the union of m⁶A peaks identified in these samples for further analyses. The m⁶A ratio of each peak was calculated as the RPKM (without adding 1) of IP library divided by the RPKM (without adding 1) of input library. To m⁶A ratios based on the denominators (peak RPKM of input) < 5 were treated as NAs (not available) in the downstream analyses. The m⁶A peaks with NAs in more than half of the samples were removed. The continuous m⁶A peaks in the same gene were merged, the merged peaks with more than 5 continuous sliding windows (300 bp) were then divided into multiple peaks that spanning no more than five sliding windows.

Different protocols of RNA fragmentations before immunoprecipitation in the preparations of different m⁶A-seq libraries might cause the variations of read signals at the actually same m⁶A peaks, resulting in diverse widths and centers of the actual same m⁶A peaks thus false m⁶A differences in certain regions, we therefore defined the m⁶A ratio of each merge peaks with multiple sliding windows as the maximum m⁶A ratio of all windows for each sample respectively. Global m⁶A differences among samples caused by diverse activities of m⁶A writers and erasers as well as technical variation of immunoprecipitation efficiencies would dilute and distort the signals of selective regulation of m⁶A, we therefore used quantile normalization to normalize the m⁶A ratios of the merged peaks across all samples.

Analyses of the m⁶A ratios across multiple cell lines

Hierarchical clustering of all samples was performed using 1- Pearson correlation coefficient as distance metric based on m⁶A ratios or TPMs of the merged peaks with CVs > 0.7 or 1000 genes with the largest CVs. The two hierarchical clustering dendrograms were subsequently compared using the ‘dendextend’ package (40) implemented in R. HOMER software (41) was used for motif enrichment analysis using randomly permuted sequences as the background for RNAs. To compare the overlaps of miCLIP-seq m⁶A sites (CITS + CIMS) in HEK293 cells (42) between stable m⁶A peaks and variable m⁶A peaks, we only used the m⁶A peaks identified in HEK293T cells according to the above-described pipeline. Distributions of m⁶A peaks were plotted on a mega gene with 10 bins in 5' UTR, CDS, and 3' UTR respectively using the methods as described in our previous paper (23). Radar plot was plotted using ‘fmsb’ package implemented in R.

Construction of the co-methylation network

We merged the m⁶A ratios as well as TPMs of all samples from each of the 25 unique cell lines by taking the averages. 29173 m⁶A peaks with CV of m⁶A ratio across 25 unique cell lines >0.3 were used to construct the signed weighted m⁶A co-methylation network using WGCNA package (43) implemented in R. The adjacency matrix was constructed by raising the $0.5 + 0.5 \times$ correlation matrix to the power of 7. The hierarchical clustering tree was then cut into 41

co-methylation modules using dynamic hybrid tree-cutting algorithm. The m⁶A index of each module was represented by the eigengene, which was the first component of Principal Component Analysis. The 41 modules were further clustered into 12 larger modules based on the correlation of their m⁶A indexes for the analyses required larger number of m⁶A peaks. Gene Ontology analysis was performed using DAVID with the genes in all the modules as the background (44).

Analyses of the cancer module

The gene expression, mutation, and clinical data of TCGA (<https://tcga-data.nci.nih.gov/>) were downloaded from cBioPortal (45,46). We calculated the means of log₂-transformed TPM+1 of all genes in cancer the module as the gene expression index of the cancer module. We used the Cox regression to examine the correlations between gene expression indexes of the cancer module and patient survival in each cancer type. Gene Ontology analysis was performed using DAVID with the genes in all modules as the background (44).

Identification of m⁶A regulators

We used 1442 expressed RBPs out of 1648 genes annotated under the term 'RBP' in Gene Ontology Database (47) to scan for m⁶A regulators by testing the Pearson correlation between the TPMs of RBPs and the eigengenes of the 41 m⁶A modules respectively. To exclude the spurious significances due to the outliers in Pearson correlation, for each correlation test between RBP expression and eigengene of co-methylation module, we used the maximum *P*-values of 25 Pearson correlations calculated based on 24 of the 25 cell lines (in the other word, we removed one cell line in each of the 25 correlation). The correlations with FDR < 0.2 were determined as significant correlations. The *P*-value cutoff corresponding to the FDR of 0.2 were determined based on the null distribution of *P*-values generated by 10 times permutations. In each permutation, we randomly relabeled the samples and performed the Pearson correlation between RBP expression and eigengene of co-methylation module using the above-described method. The *P*-value cutoff was further determined as the *P*-value under which the average number of significant correlations in permutations was only one fifth of the observed number of significant correlations using real data (Supplementary Figure S4A).

We downloaded the CLIP-seq peaks from starBase database (version 3) (48–50) as well as the ENCODE CLIP-seq dataset in HepG2 and K562 cells (51). Significance of the overlapping between a set of CLIP-seq peaks and an m⁶A module was calculated by testing whether the module and other modules (as background) had equal fraction of m⁶A peaks that overlap with the CLIP-seq peak (at least 1 bp) using χ^2 tests. We obtained the 110 Motifs of 89 RBPs from a published dataset based on large scale in vitro RNA-competence (52) as well as other well-known RBP motifs (53–55). Significance of the enrichment of an RBP motif in an m⁶A module was calculated by testing whether the module and other modules (as background) had equal fraction of m⁶A peaks that contain the RBP motif using χ^2 tests.

The significant overlapping was defined as the ones with Benjamini-Hochberg FDR < 0.05 based on all the tests. The final list of m⁶A regulators were those RBPs with gene expression significantly correlated with the modules and with CLIP-seq targets or motifs significantly enriched in the same modules. The proteins that interact with METTL3, METT14, WTAP, VIRMA and m⁶A based on IP mass spectrum data were directly obtained from the published papers (4,56–59).

Analyses of the low-input m⁶A-seq data

The reads of the second end were trimmed to 50 bp from 3' end for downstream analyses. We mapped the raw data to human genome and calculate m⁶A ratios for each sliding window using the above-described protocol. We used 'exomePeak' package implemented in R to identify the m⁶A peaks and determine the differentially methylated m⁶A peaks with FDR < 0.05 (60). To examine whether the m⁶A peaks within the associated module or with CLIP-seq binding showed stronger switch of m⁶A ratios upon RBP knock-down, we calculated the fold change of m⁶A ratios upon RBP knockdown for each m⁶A peaks in all the modules. If one RBP significantly correlated with multiple modules as predicted, we merged these modules together as the RBP associated module for the analyses. To filter out the ambiguous fold change values due to small denominators, only the peaks with input window RPKM > 5 in all samples and m⁶A ratio > 0.1 in both replicates of control samples were considered for the analyses. The data were visualized using the Integrative Genomics Viewer (IGV) tool (61). Differential gene expression analyses were performed based on the input data using DESeq2 (62) according to the read counts of each gene determined by HTSeq (63). The genes with FDR < 0.05 and mean CPM (Counts per Million) > 100 were determined as the differentially expressed genes.

Cell culture

Cells were maintained at 37°C with 5% CO₂ in a humidified incubator and passaged every 2–3 days. Wild type HEK293T and HepG2 cells were cultured in high-glucose Dulbecco's Modified Eagle Medium (DMEM, ThermoFisher Scientific) supplemented with 10% FBS (Ex-Cell Bio). All cells were tested for absence of mycoplasma contamination using Myco-Blue Mycoplasma Detector (Vazyme).

Plasmid constructs and transfection

For gene knocking-down, short-hairpin RNA (shRNA) oligos were synthesized, annealed and inserted into pLKO.1 vector. The pLKO.1-shRNA plasmids were then transfected into HEK293T cells with packing vectors pMD2.G and psPAX2 to produce lentiviruses. To overexpress the RBPs, we inserted the full-length coding regions amplified from HEK293T cDNA library by polymerase chain reaction (PCR) into pCDNA3.1 vector followed by adding the Flag tag. The pCDNA3.1-RBP plasmids were transfected into HEK293T cells with Lipofectamine 2000 (Invitrogen)

according to the manufacturer's instructions. All the sequences of shRNA oligos and PCR primers are listed in Supplementary Table S2.

RNA isolation and real-time quantitative PCR

Total RNA was extracted using NucleoZOL reagent (Macherey-Nagel) or MiniBEST Universal RNA Extraction Kit (Takara, Japan). First-strand cDNA was synthesized by reverse transcription of 1 μ g RNA using HiScript II 1st Strand cDNA Synthesis Kit (Vazyme, China) according to the manufacturer's protocol. Quantitative real time-PCR was performed using TB Green Premix Ex Taq (Takara, Japan) in QuantStudio 7 Flex Real-Time PCR System (Life Technologies, USA). *β -actin* and *GAPDH* were used as reference genes for input normalization. The mRNA expression was measured by quantitative PCR using the $\Delta\Delta$ CT method. Primers for quantitative PCR were listed in Supplementary Table S2.

CRISPR-Cas9 mediated *METTL3* knockout

Transiently transfected plasmid expressing two sgRNAs targeting human *METTL3* exons was adopted for internal fragment deletion in specified size according to CRISPR-Cas9-2hitKO system. Two target guide RNAs were designed using the online tool (<http://tools.genome-engineering.org>) with high scores and minor off-target effects and then subcloned into CRISPR-Cas9-2hitKO plasmid. (All the sequences of sgRNA oligos were listed in Supplementary Table S2). To establish the knockout cell lines, CRISPR-Cas9-2hitKO plasmid carrying two sgRNA-expressing cassettes were transfected into HepG2, GFP expressing cells were enriched by FACS (MoFlo Astrios EQ, Beckman Coulter) 3 days later and seeded at low density for single colony isolation. Knockout efficiency was tested by DNA sequencing and verified by western blotting.

Co-immunoprecipitation and western blot

Whole-cell extracts were extracted by directly lysing the cells with 1 \times RIPA Buffer (Cell Signaling Technology) with 1 mM PMSF (Beyotime) added immediately before use. Samples were boiled by adding 6 \times sodium dodecyl sulphate (SDS) sample buffer for 10 min at 100°C and resolved using SDS-polyacrylamide gel electrophoresis. To perform immunoprecipitation, we lysed the cells by RIPA lysis buffer. The lysates were sonicated at 4°C and cleared by centrifugation at 12 000 rpm for 15 min at 4°C. Immunoprecipitation was carried out by incubating the FLAG beads (Bimake) at 4°C with the lysate overnight. Immunoprecipitates were washed three times in cold E1A lysis buffer (250 mM NaCl, 50 mM HEPES (pH 7.5), 0.1% NP-40, 5 mM EDTA, protease inhibitor cocktail (Roche)) and boiled with 2 \times SDS sample buffer for 10 min. The proteins were probed with the following antibodies: *METTL3* Rabbit mAb (1:2000, 15073-1-AP, Proteintech), *METTL14* Rabbit mAb (1:500, 51104S, Cell Signaling Technology), Monoclonal ANTI-FLAG M2 antibody (1:1000, F1804, Sigma), *GAPDH* (1:2000, 5174, Cell

Signaling Technology), ALKBH5 (1:3000, ab195377, abcam) and FTO (1:1000, 31687, Cell Signaling Technology). Immuno-detection was performed using HRP-conjugated Affinipure Goat Anti-Mouse IgG(H+L) (1:5000, SA00001-1, Proteintech) or HRP-conjugated Affinipure Goat Anti-Rabbit IgG(H+L) (1:5000, SA00001-2, Proteintech) and ECL prime substrate (Bio-Rad) according to the manufacturer's instructions.

Low-input m⁶A-seq

Low-input m⁶A-seq was performed based on the protocols previously described by Zeng *et al.* (64) with several modifications. Briefly, a total volume of 8–10 μ g total RNA was fragmented using the 10 \times RNA Fragmentation Buffer (100 mM Tris-HCl, 100 mM ZnCl₂ in nuclease free H₂O) and purified with sodium acetate (Sigma-Aldrich), glycogen (Thermo Fisher Scientific) and 100% ethanol. A total of 30 μ l of protein-A/G magnetic beads (10002D/10004D, Thermo Fisher Scientific) were washed twice in IP buffer (150 mM NaCl, 10 mM Tris-HCl, pH 7.5, 0.1% IGEPAL CA-630 in nuclease free H₂O) and incubated with 5 μ g anti-m⁶A antibody (202003, Synaptic Systems) in 500 μ l of IP buffer at 4°C for at least 6 h. After washed twice in IP buffer, the antibody-bead mixture was resuspended by fragmented total RNAs in IP buffer and incubated at 4°C for 2 h. Then after washed twice in low-salt IP buffer, and twice in high-salt IP buffer at 4°C for 10 min each, the m⁶A enriched fragmented RNAs were eluted and purified from the beads using RNeasy Mini Kit (QIAGEN). Sequencing libraries were generated using the SMARTer Stranded Total RNA-Seq Kit v2 (634413, Takara). All libraries were sequenced on an Illumina HiSeq X Ten platform to produce 20–40 M strand-specific paired-end reads.

mRNA stability assay

Cells were seeded into 6-well plates and treated with actinomycin D (5 mg/ml) for 4, 2 and 0 h after culturing for 12 h. We used NucleoZOL reagent (Macherey-Nagel) to extract the total RNAs followed by reverse transcription. The abundances of the interest genes were detected measured in each time point by real-time quantitative PCR (qPCR) using 18S rRNA as the reference gene. The qPCR primers are listed in Supplementary Table S2.

Colony formation assay

After trypsinization and cell counting, 1200 HepG2 cells were seeded per well in 6-well plates and cultured in DMEM supplemented with 10% fetal bovine serum (FBS) for 7 days. Cells were rinsed with phosphate buffered saline once and fixed using paraformaldehyde and stained using crystal violet.

RESULTS

Development of a computational framework to systematically identify cell-specific m⁶A regulators

To overcome the above-mentioned difficulties of identifying cell-specific m⁶A regulators using large-scale m⁶A methylomes, we developed a computational framework through

integrating gene expressions, binding targets and binding motifs of a large number of RBPs with a co-methylation network constructed using large-scale m⁶A methylomes across diverse cell states (Figure 1).

First of all, a variety of technical issues of m⁶A-seq data could hinder the successful systematic analyses of the quantitative m⁶A ratios calculated based on the m⁶A-seq data. Therefore, we performed multiple processes to minimize the influences of different types of technical issues (see ‘Materials and Methods’ section for details). Besides applying stringent filters to rule out the unreliable quantifications, we also merge the peaks across multiple samples and used the with the maximum m⁶A ratio of 100 bp window to represent the m⁶A ratio of merged peaks with continuous windows in each sample, therefore, the shifting of peak centers and divergence of peak breadths due to technical biases in preparing the m⁶A-seq libraries, such as variations in RNA fragmentation lengths and sequencing lengths, will be controlled. At last, quantile normalizations of the m⁶A ratios are performed, so that not only variations of antibody efficiencies can be corrected but also we can focus on capturing the mechanisms that regulate selective subsets of m⁶A peaks other than the global regulation dictated by m⁶A writers and erasers.

We hypothesis that the m⁶A sites regulated by the same m⁶A regulators should have correlated m⁶A levels (co-methylation) across different cell states, and the m⁶A levels of the module should also be correlated with the gene expressions of their regulators. Co-methylation module-based analyses can greatly reduce the dimension of the data and noise of individual m⁶A peaks. Therefore, we use WGCNA (43) to construct a signed weighted co-methylation network. For each module, we calculate the Pearson correlations between the m⁶A indexes (the first component of principal component analysis) and the expression of 1442 RBPs annotated by GO database respectively. Since Pearson correlation is very sensitive to outliers, we perform Jackknife resampling (leave one out) and take the least significant *P* value for each test (see ‘Materials and Methods’ section for details). We then use random permutation to determine the threshold of significance (see ‘Materials and Methods’ section for details).

On the other hand, if an RBP regulate the m⁶A sites near their targets, we expect to see the co-localization of RBP binding sites with the m⁶A peaks of the module regulated by the RBP, otherwise, the correlation may reflect the indirect effects of regulation such as through regulating the abundance of another m⁶A regulator. In this framework, we integrate the CLIP-seq data of 157 RBPs obtained from starBase (version 3) (48,50) as well as ENCODE CLIP-seq dataset (65) to test whether the RBP binding sites are over-represented in the corresponding m⁶A modules. Since most RBPs do not have available CLIP-seq data, we also take advantage of the RNAcompete-derived motifs (52) as well as several well-known motifs (53–55) of 89 RBPs to test the enrichment of motifs in the associated modules. The RBPs with gene expression significantly correlate with specific modules and with binding targets or motifs significantly enriched in the same modules will be identified as the high-confidence m⁶A regulators specifically regulate the m⁶A sites in the associated modules.

Systematic analyses of m⁶A methylomes of multiple cell lines revealed credible dynamics of m⁶A

In order to systematically elucidate the cell-specific *trans*-acting regulation of m⁶A using real data, we applied the above computational framework to the public available m⁶A-seq data of 104 samples in 25 unique human cell lines (Supplementary Table S1). We obtained about 15 000 m⁶A peaks for each cell (Supplementary Figure S1A), and the m⁶A peaks for these cell lines were strongly enriched near stop codons, consistent with previous reports (1) (Figure 2A).

To evaluate the reliability of the normalized m⁶A ratios, we performed unsupervised hierarchical clustering using the m⁶A ratios of these samples (replicates were merged) (Figure 2B). We found clear variations of m⁶A ratios and that the samples were not clustered according to the technical issues including labs, RNA selection protocols and antibodies (Figure 2B). Since gene expression obtained from RNA-seq data (input of m⁶A-seq) was less affected by technical bias, we clustered these samples using gene expression to represent the real relationship among these samples (Supplementary Figure S1B). The hierarchical clustering dendrogram generated using m⁶A-seq data were in general similar as that generated using gene expression (*P* = 0.006; permutation test using ‘dendextend’ package (40); Supplementary Figure S1C). The same cell lines (MONO-MAC-6 and HEK293T) from different labs were also clustered together, whereas, the A549 cell lines from different labs were not clustered together possibly due to the diverse expression of m⁶A writers (see Supplementary Text 1 for clarification; Supplementary Figure S1D and E).

Variable m⁶A sites are away from stop codons

To systematically study the patterns and mechanisms of *trans*-regulating of m⁶A, we combined the m⁶A-seq data of the same cell lines to obtain the methylomes of 25 unique human cell lines. We found different m⁶A peaks had different levels of variations across these cell lines (Supplementary Figure S1F). The m⁶A peaks near stop codons had significantly smaller coefficient of variations (CVs) of m⁶A ratios than the other m⁶A peaks, while the peaks in the long internal exons were slightly smaller than that in UTRs and coding regions (Figure 2C). As shown in Figure 2D, the 11 949 stable m⁶A peaks (CV < 0.3) tended to be enriched near stop codons, whereas the 29 173 variable m⁶A peaks (CV > 0.3) were enriched in coding regions and completely lost the enrichment near stop codons and less enriched in long internal exons, suggesting that the m⁶A sites near stop codons are regulated mainly by *cis*-acting elements rather than *trans*-acting factors (Figure 2E; an example is shown in Figure 2F). Based on this definition, there were about 50% of stable peaks in each cell line (Supplementary Figure S1G), which was consistent with a recent report that *cis*-regulation account for 33–46% of the variability of m⁶A levels (16). In this study, we would like to focus on the variable m⁶A peaks. Compared to those stable m⁶A peaks, the variable m⁶A peaks occurred on the genes with similar gene expression but had significantly smaller maximum m⁶A ratios across all cell lines (Figure 2G and H).

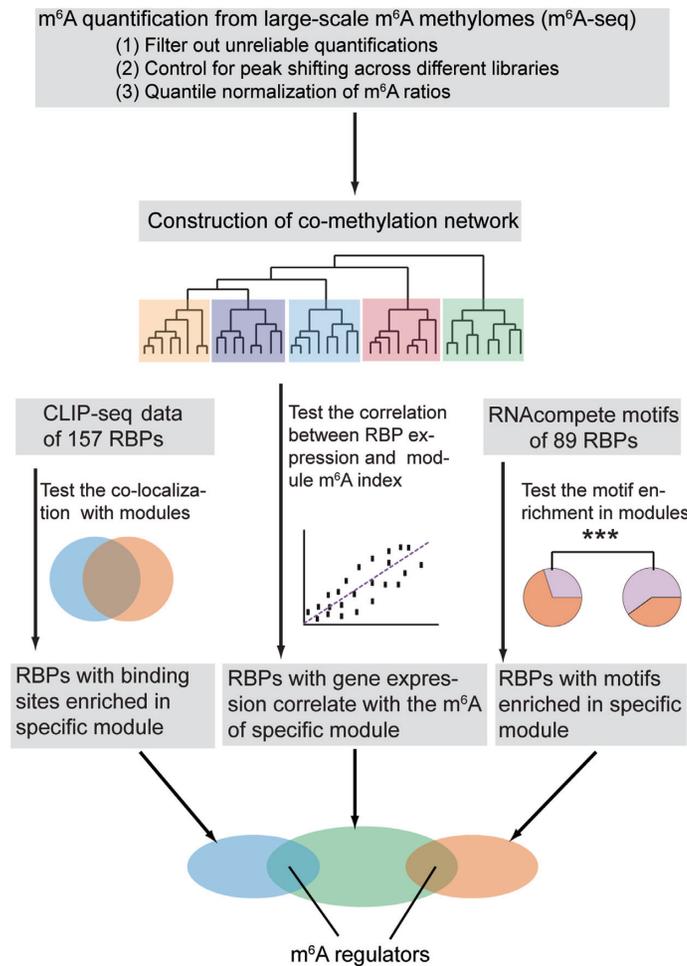


Figure 1. Schematic flow chart demonstrating the computational framework to identify *trans* regulators of m⁶A.

To test whether the variable m⁶A peaks were genuine m⁶A peaks or technical noises, we used single-nucleotide-resolution m⁶A sites in HEK293 cells obtained by miCLIP-seq technology as gold standard to evaluate the m⁶A peaks (42). We found the variable m⁶A peaks and stable m⁶A peaks identified in HEK293T cells exhibited the similar proportions that overlapped with miCLIP-seq reported m⁶A sites, indicating that the variable m⁶A peaks were as genuine as the stable peaks (Figure 2I).

Modular co-methylation of the variable m⁶A sites revealed prevalent *trans*-acting regulation of m⁶A

We constructed a signed weighted co-methylation network using WGCNA (43) based on the m⁶A ratios of 29 173 variable m⁶A peaks across the 25 unique cell lines (Figure 3A). We obtained 41 co-methylation modules, which were further merged into 12 larger modules according to the correlations of module m⁶A indexes (the first component of principal component analysis) among them (Figure 3B and Supplementary Table S3). As shown in Figure 3B, the m⁶A indexes of the m⁶A modules showed strong cell type specificities, suggesting that cell-specific m⁶A methylomes may be resulted from co-regulation of m⁶A sites by cell-specific

regulators. On the other hand, we found the co-methylation modules showed specific topology of m⁶A and that the m⁶A peaks in one module strongly enriched near translation start sites (Figure 3C), while the m⁶A peaks in another two modules strongly enriched in long internal exons (Figure 3D), suggesting that the topology of m⁶A methylomes are also dynamic and regulated by certain *trans*-acting factors. We also checked the motif enrichment of these modules and found these modules were enriched in distinct motifs (Supplementary Figure S2A). Moreover, the representative module-specific motifs tend to be lowly occurred in other modules, suggesting the different modules are regulated by diverse *trans*-acting factors (Supplementary Figure S2B). The genes in the 12 combined modules were enriched in different GO (47) categories (Supplementary Figure S2C), suggesting that the co-regulated m⁶A sites tend to play specific functional roles in specific cells.

An m⁶A module was specifically methylated in cancer cell lines

We found one of the 12 combined m⁶A co-methylation modules (blue module) was highly methylated specifically in cancer cell lines other than normal somatic cells as well as

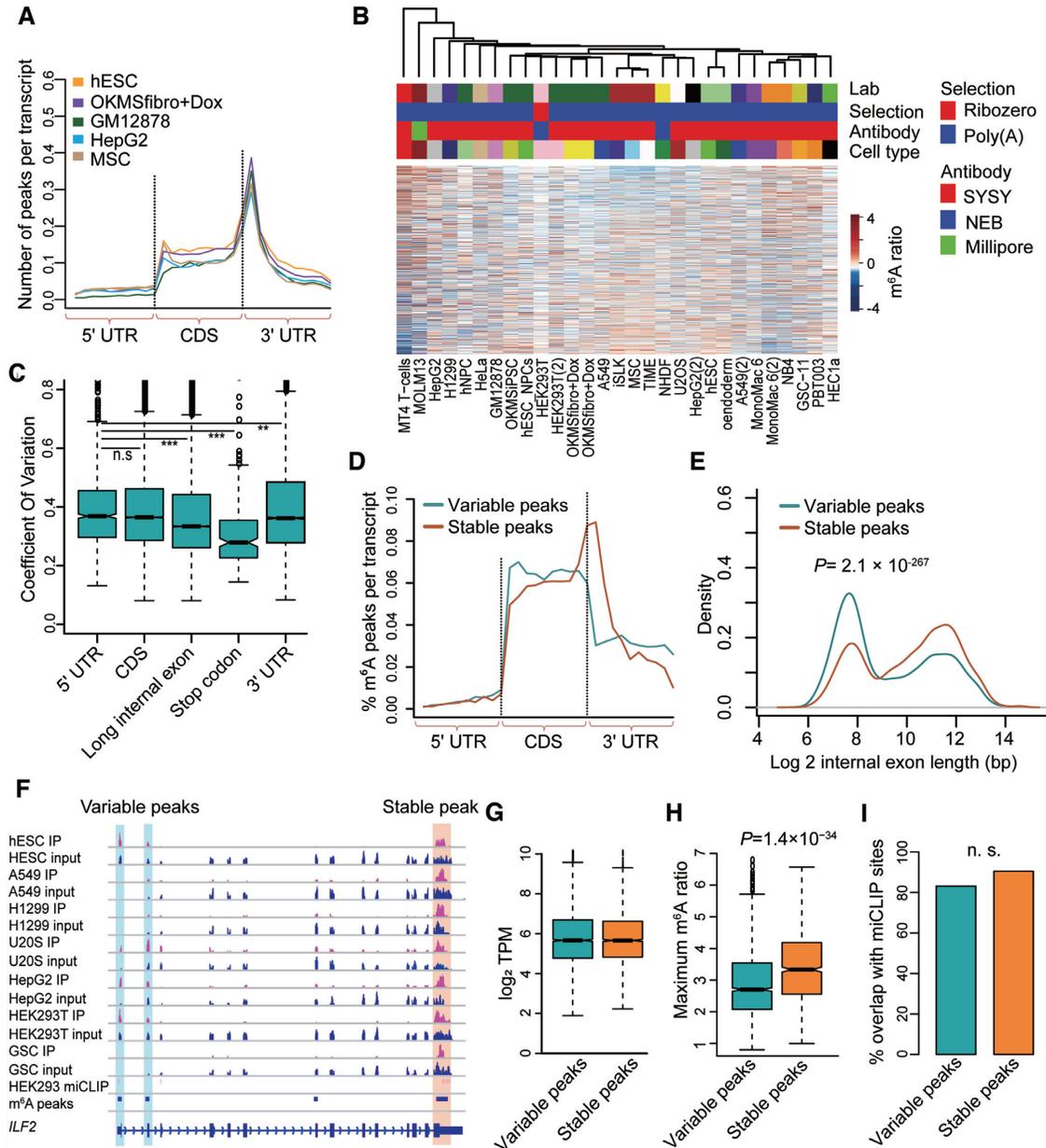


Figure 2. Analyses of m⁶A methylomes of multiple cell lines. (A) Normalized distributions of m⁶A peaks across 5' UTR, CDS and 3' UTR for representative cell lines. (B) The unsupervised hierarchical clustering and heatmap of the m⁶A ratios for the m⁶A peaks with the largest CVs across all cell lines. The technical information is indicated above the heatmap. (C) Box plot representing the CVs of m⁶A ratios for the peaks located at different regions of mRNAs. (D) Normalized distributions of variable and stable m⁶A peaks across 5' UTR, CDS and 3' UTR. (E) Densities of logarithm transformed lengths of the internal exons with variable m⁶A peaks and stable m⁶A peaks. The *P*-value of Wilcoxon test is indicated. (F) Tracks showing the read coverage of the IPs, inputs and the merged m⁶A peaks of the representative cell lines as well as the HEK293 m⁶A sites from miCLIP-seq data on *ILF2*. The tracks are shown for optimal viewing. The variable and stable m⁶A peaks are highlighted, respectively. (G) Box plots representing the logarithm transformed TPMs of variable and stable m⁶A peaks. (H), Box plot representing the maximum m⁶A ratios across all cell lines of variable and stable m⁶A peaks. (I) Bar plot showing the percentages of variable peaks and stable peaks that overlap with m⁶A sites obtained from miCLIP-seq. 'n.s.' denotes non-significant.

stem cells (Figure 4A). As shown in Figure 4A, the expression of the corresponding genes was much higher in those cancer cell lines, suggesting that the enhanced m⁶A methylation at the m⁶A sites in this module may result in elevated abundances of the mRNAs harboring these m⁶As possibly through increasing the RNA stabilities (57). Then, we took advantage of the TCGA (The Cancer Genome Atlas) clin-

ical samples to further address the oncological roles of this module. Interestingly, in 13 of the 14 cancer types included in TCGA, the gene expression indexes of this module were significantly higher in cancer tissues than in normal tissues, suggesting the up-regulation of these genes are common in cancers and may relate to the etiology of most cancers (Figure 4B). Furthermore, the high gene expression indexes

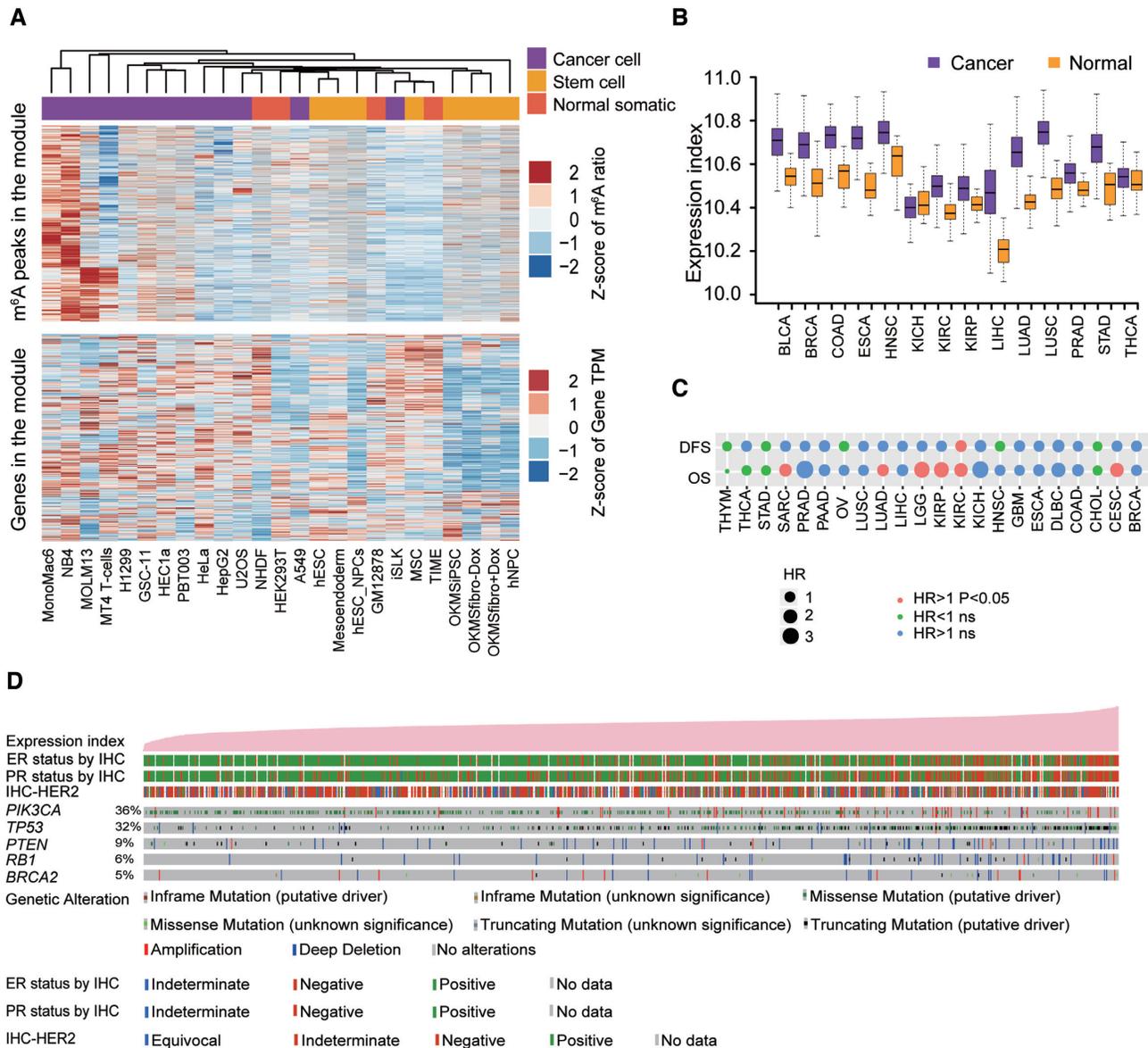


Figure 4. Discovery of a co-methylation module specifically methylated in cancer cell lines. (A) Heatmaps representing the Z-scores of m⁶A ratios (upper panel) and gene expressions (low panel) of the peaks and corresponding genes across all cell lines. The types of cell lines are indicated at the top of the upper panel. (B) Box plot representing the gene expression indexes of the genes corresponding to cancer-specific module for cancer and normal samples of 14 cancer types in TCGA. (C) Cox correlations between the gene expression indexes of the genes corresponding to cancer-specific module and the survival of cancer patients of 14 cancer types in TCGA. OS: overall survival; DFS: disease-free survival; HR: hazard ratio. (D) Tracks representing the gene expression indexes of the genes corresponding to the cancer-specific module and genetic alteration spectrum of the key markers as well as clinical phenotypes of the breast cancer patients from TCGA. The patient samples are sorted according to the gene expression indexes of the cancer-specific module.

and D), while the expression of m⁶A eraser ALKBH5 was negatively correlated with module M14 (Supplementary Figure S4E), suggesting that some components of writer complex as well as demethylase may also confer specificities of m⁶A. Besides, we also found a previously reported specific m⁶A regulator SMAD3, which can specifically promote the installation of m⁶A (20), was positively correlated with module M33 (Supplementary Figure S4F). The proteins of m⁶A regulators would possibly interplay with m⁶A writers or erasers at their binding sites on RNAs, we collected the published Mass Spectrum data of protein pull-down using the antibodies of METTL3, METTL14,

WTAP, VIRMA (4,59,66), we found 108 RBPs out of the 588 low-confidence m⁶A regulators could be pulled down by at least one of the antibodies (Supplementary Figure S4G). On the other hand, we found 44 RBPs out of the low-confidence m⁶A regulators could be pulled down by m⁶A modified oligos (58) (Supplementary Figure S4H).

Then we took advantage of the CLIP-seq data of 157 RBPs obtained from starBase (version 3) (48,50) as well as ENCODE CLIP-seq dataset (65) to test whether the RBP binding sites were over-represented in the corresponding m⁶A modules. There were 22% of the significantly correlated pairs of RBP expression and m⁶A module showed sig-

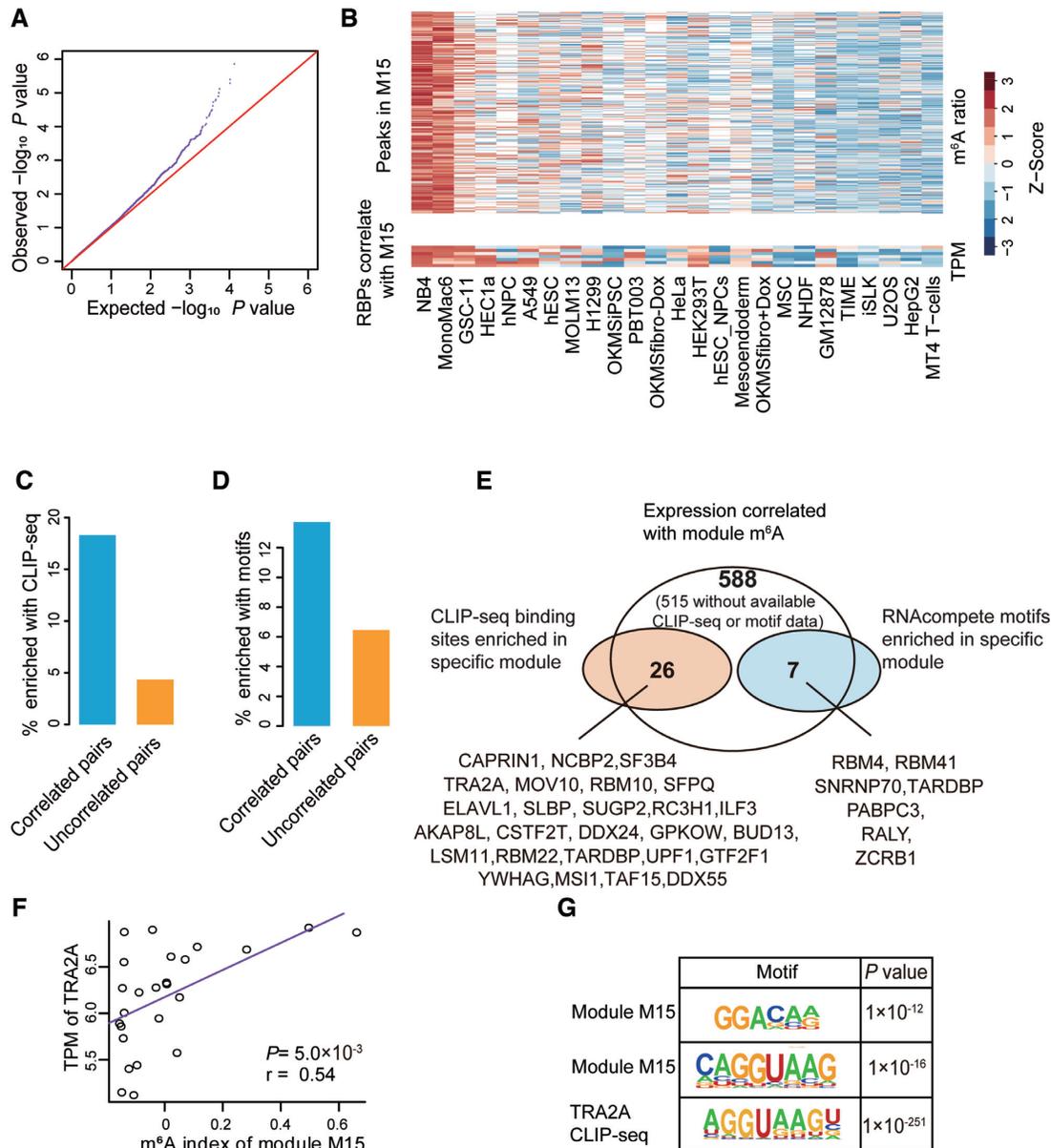


Figure 5. Systematic identification of m⁶A regulators. (A) Q-Q plot comparing the distributions of expected *P*-values and observed *P*-values of the correlations between the gene expressions of RBPs and m⁶A indexes of the co-methylation modules. (B) Heatmaps representing the m⁶A ratios of the m⁶A peaks within the module M15 (upper panel) and the gene expressions of the RBPs that significantly correlated with the m⁶A indexes of M15 (lower panel). The cell lines are sorted according to the m⁶A indexes of M15. (C and D) Barplot representing the percentages of the pairs of RBPs and modules that enriched for CLIP-seq binding sites (C) or motifs (D) of the RBPs out of the pairs that showing significant and non-significant (top 1000 least significant) correlations between gene expressions of the RBPs and the m⁶A indexes of the modules. (E) Venn diagram demonstrating the identification of 32 high-confidence m⁶A regulators. (F) Scatter plot showing the correlation between TRA2A gene expression and m⁶A indexes of module M15 across all cell lines. (G) Representative motifs enriched in module M15 and TRA2A CLIP-seq targets.

nificant enrichment of CLIP-seq peaks in the same m⁶A modules, in contrast, it was only 4% for the top 1000 least significant pairs of RBP and m⁶A modules, indicating a significantly enriched occurrences of co-localization of RBP binding sites with their significantly correlated modules ($P = 5.8 \times 10^{-8}$, two-tailed Chi-square test; Figure 5C). Since most RBPs did not have available CLIP-seq data, we utilized the RNAcompete-derived motifs (52) as well as several well-known motifs (53–55) of 89 RBPs for further evaluation. Similarly, we observed a trend that the motifs of the

RBPs were more likely to show significant enrichment in the m⁶A modules correlated with the RBPs as compared to the uncorrelated modules ($P = 0.09$, two-tailed Chi-square test; Figure 5D). The RBP motifs were mostly enriched within 50bp of m⁶A motifs, suggesting that the RBPs may tend to regulate the m⁶A sites around 50bp of its binding sites (Supplementary Figure S4I). In the end, out of the 50 and 26 low-confidence m⁶A regulators with available CLIP-seq data and known motifs respectively, 26 (52%) and 7 (27%) RBPs also showed significant co-localization with the ex-

act correlated m⁶A modules based on CLIP-seq and motif analyses respectively (Figure 5E and Supplementary Table S4). We referred these 32 RBPs to high-confidence m⁶A regulators. As exemplified, the gene expression of an RBP TRA2A significantly correlated with the m⁶A ratio of module M15 (Figure 5F), which happened to enrich for a motif resembled the TRA2A motif obtained from CLIP-seq (Figure 5G). On the other hand, since there were 516 low-confidence m⁶A regulators without available CLIP-seq datum or known motif, we would expect a dramatic increase of the high-confidence m⁶A regulators when more and more CLIP-seq data become available in the future.

Experimental validations of selected m⁶A regulators

Because eCLIP-seq of plenty of RBPs had been performed in HepG2 cells (65), we selected 3 high-confidence m⁶A regulators TRA2A, CAPRIN1 and MOV10, which were highly expressed and with corresponding module highly methylated in HepG2 cells, to experimentally validate their regulatory functions on m⁶A. First of all, we tested whether knocking down of these regulators affected the stoichiometry of some m⁶A peaks in human HepG2 cells using low-input m⁶A-seq (64). The m⁶A peaks identified in low-input m⁶A-seq were enriched in stop codons as expected (Supplementary Figure S5A). Then we performed Co-Immunoprecipitation (Co-IP) to test whether these regulators interacted with the major m⁶A writers (METTL3 and METTL14) and erasers (FTO and ALKBH5).

The m⁶A ratios of the m⁶A peaks with TRA2A binding were significantly down-regulated upon TRA2A depletion as compared to the m⁶A peaks without TRA2A binding ($P = 1.6 \times 10^{-13}$, two-tailed Wilcoxon test; Figure 6A), indicating that TRA2A promoted the installation of m⁶A through binding near the m⁶A sites other than indirect effects such as regulating another m⁶A regulator. Similarly, the m⁶A ratios of the predicted m⁶A module regulated by TRA2A were also significantly down-regulated upon TRA2A depletion as compared to other modules ($P = 3.1 \times 10^{-3}$, two-tailed Wilcoxon test; Figure 6B). The above results were very consistent with our observation that the expression of TRA2A was positively correlated with the m⁶A indexes of the corresponding modules. As exemplified in Figure 6C, the TRA2A bound m⁶A peak in the long non-coding RNA *MALAT1* was downregulated upon *TRA2A* depletion. We further found TRA2A interacted with METTL3 independent of RNAs, suggesting that TRA2A promote the installing of m⁶A near its binding sites through recruitment of METTL3 (Figure 6D and Supplementary Figure S5B).

Similar results were observed for CAPRIN1, the m⁶A ratios of the m⁶A peaks within the related module and co-localized with CAPRIN1 were significantly down-regulated upon *CAPRIN1* depletion as compared to the m⁶A peaks in other modules and without CAPRIN1 binding ($P = 5.4 \times 10^{-8}$, two-tailed Wilcoxon test; Figure 6E), though the CAPRIN1 CLIP-seq data were obtained from a different cell line. Strikingly, we found CAPRIN1 interacted with both METTL3 and METTL14, suggesting that CAPRIN1 can recruit the methyltransferase complex to promote the m⁶A installation near its binding sites (Figure 6F). Whereas,

we did not find the depletion of MOV10 changed the m⁶A ratios of the peaks within the related module (Figure 6G), nor did we find the interaction of MOV10 with any of the m⁶A writers or erasers. Therefore, MOV10 was not like a genuine m⁶A regulator.

Therefore, we finally validated TRA2A and CAPRIN1 as genuine m⁶A regulators, and MOV10 was a false positive discovery. We further identified 427 and 124 differentially methylated m⁶A peaks due to knockdown of *TRA2A* and *CAPRIN1* respectively using exomePeak software (60). Similar as the CLIP-seq binding targets of these RBPs, the differentially methylated m⁶A peaks were all enriched in protein coding regions other than near stop codons, consistent with our finding that m⁶A peaks near stop codons were stable (Supplementary Figure S5C and D).

We then tried to investigate whether these m⁶A regulators had any functional consequences by regulating m⁶A. We noticed that *TRA2A* knockdown resulted in upregulated gene expression of 470 genes and downregulated gene expression of only 79 genes, the up-regulated genes were significantly enriched in KEGG pathway related to protein processing in endoplasmic reticulum as well as metabolism (Supplementary Figure S6A). Of the 470 up-regulated genes, there were 107 genes with at least one m⁶A peak co-localized with TRA2A CLIP-seq peak, these genes were enriched in the pathway of protein processing in endoplasmic reticulum, suggesting the functional role of TRA2A on homeostasis of endoplasmic reticulum by regulation m⁶A. Since m⁶A was reported to promote the degradation of RNAs (8), we hypothesized that TRA2A induced the m⁶A modification of specific RNAs to facilitate their degradations. To test this hypothesis, we selected 11 up-regulated genes with multiple m⁶A sites co-localized with TRA2A CLIP-seq binding sites to examine the effects of TRA2A on their RNA stabilities. We found 7 of the 11 genes showed significantly increased stability in *TRA2A* knockdown HepG2 cells, including 3 genes (*HSPA8*, *RRBP1*, *UGGT1*) involved in ‘protein processing in endoplasmic reticulum’ (Supplementary Figure S6B–D). After cultured for several generations, we also noticed remarkably decreased viability of *TRA2A* knockdown HepG2 cells based on colony formation assay (Supplementary Figure S6E), which was possibly due to the induction of endoplasmic reticulum stress in the longtime culturing of cells with defects in maintaining homeostasis of endoplasmic reticulum.

DISCUSSION

In this study, we successfully developed a computational framework to systematically identify *trans* regulators of m⁶A through integrating gene expressions, binding targets, and binding motifs of a large number of RBPs with a co-methylation network constructed using large-scale m⁶A methylomes across diverse cell states. Applying the framework to the public available m⁶A-seq data of 25 unique cell lines revealed pervasive *trans*-acting regulation of m⁶A and identified 32 high-confidence m⁶A regulators with reasonable experimental validation rate.

The successful identification of m⁶A regulators using 25 distinct cell lines demonstrated a powerful and widely portable strategy to elucidate the *trans*-acting regulation of

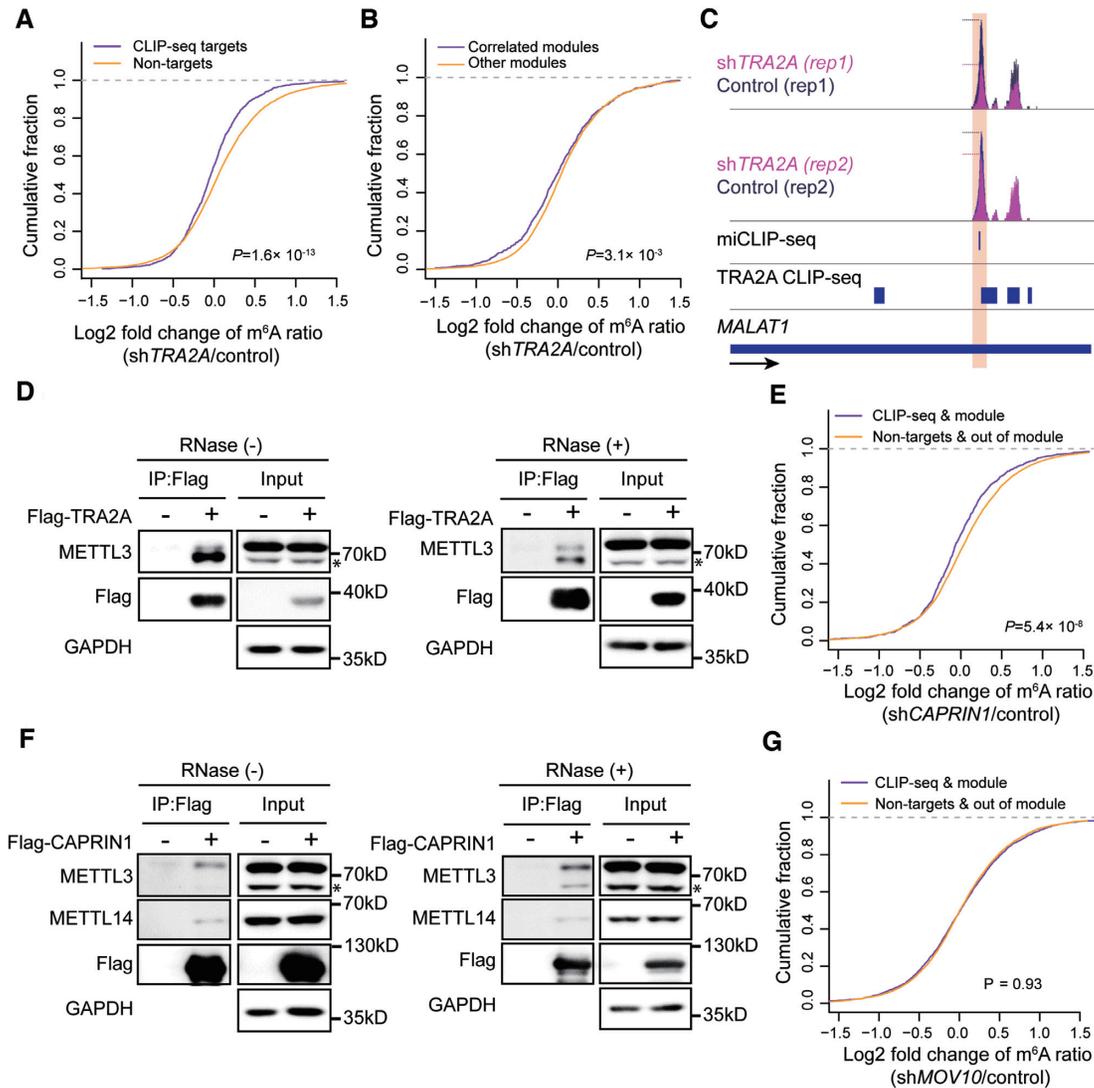


Figure 6. Experimental validation of selected m^6A regulators. (A) Plot of cumulative fraction of \log_2 fold change of m^6A ratios upon *TRA2A* knockdown for the m^6A peaks overlap or non-overlap with *TRA2A* CLIP-seq targets. *P*-value of two-tailed Wilcoxon test is indicated. (B) Plot of cumulative fraction of \log_2 fold change of m^6A ratios upon *TRA2A* knockdown for the m^6A peaks within or not within the co-methylation modules correlated with *TRA2A*. *P* value of two-tailed Wilcoxon test is indicated. (C) Tracks displaying the read coverage of IPs normalized by inputs as well as the miCLIP-seq m^6A sites and *TRA2A* CLIP-seq peaks in HepG2 cells on the long non-coding gene *MALAT1*. The m^6A peak with down-regulated m^6A ratio in sh*TRA2A* is highlighted. The dashed lines indicate the summits of the peaks. (D) Western blots showing the interaction between *TRA2A* and *METTL3* with and without RNase treatment respectively. * indicates a non-specific band (see Supplementary Figure S5B). (E) Plot of cumulative fraction of \log_2 fold change of m^6A ratios upon *CAPRIN1* knockdown comparing the m^6A peaks in the correlated modules and overlap with *CAPRIN1* CLIP-seq targets versus the peaks not within the correlated module or overlap with *CAPRIN1* CLIP-seq targets. *P*-value of two-tailed Wilcoxon test is indicated. (F) Western blots showing the interactions of *CAPRIN1* with *METTL3* and *METTL14* with and without RNase treatment respectively. * indicates a non-specific band (see Supplementary Figure S5B). (G) Plot of cumulative fraction of \log_2 fold change of m^6A ratios upon *MOV10* knockdown comparing the m^6A peaks in the correlated modules and overlap with *MOV10* CLIP-seq targets versus the peaks not within the correlated module or overlap with *MOV10* CLIP-seq targets. *P*-value of two-tailed Wilcoxon test is indicated.

m^6A based on a batch of m^6A methylomes. In this study, we definitely underestimated the prevalence of m^6A regulators due to technical limitations, such as the limited number of cell lines, lack of available CLIP-seq data for most RBPs. Moreover, we probably also missed the m^6A regulators that played important roles in extremely specific cells or physiological and pathological processes. Since m^6A -seq technology becomes more and more applicable and affordable (64), large-scale m^6A -seq data in specific biological systems will be available in near future. It is of great advantage to apply

our computational framework to these data in order to uncover the *trans*-acting mechanisms that may be important for specific systems. For example, applying the framework to a population of cancer samples may reveal novel m^6A regulators specifically regulate specific m^6A sites involved in tumorigenesis in certain types of cancers.

We noticed that there were two types of m^6A sites according to their variation among multiple cell lines. It is interesting that the m^6A s around stop codons tend to be hard wired thus very stable among different cell lines, these m^6A

sites could be considered as indispensable ‘structural m⁶A sites’. They are installed at specific positions and are important for the basic functions and biogenesis of mRNAs. The structural m⁶A sites around stop codons are probably regulated mainly in *cis* and directly mediated by m⁶A methyltransferase complex. This idea is supported by the previous report that VIRMA and ZC3H13, which are important components of methyltransferase complex, specifically deposit m⁶A around the stop codon of mRNA (4,5). On the other hand, the m⁶A sites away from stop codons, such as those within coding regions, tend to display cell-specificities, thus could be considered as ‘dynamic m⁶A sites’. They are precisely and dynamically regulated through a number of regulators expressed with spatial and temporal specificities, providing a novel mechanism for genes to play diverse roles in different cells. As previously reported, transcription factor CEBPZ induces the m⁶A specifically within the coding region of its associated mRNAs through co-transcriptional recruitment of METTL3 at the promoters (19). In this study, we found TRA2A and CAPRIN1 also selectively modulate the m⁶A sites within the coding regions (Supplementary Figure S5C and D). These results further support that dynamic m⁶A sites are away from stop codons.

Our study provided a new perspective on how the m⁶A sites were regulated. It is well known that m⁶As are modified co-transcriptionally (17,67), thus m⁶A can be regulated through co-transcriptional mechanisms. Two well-known m⁶A regulators SMAD2/3 and CEBPZ are both transcription factors and regulate m⁶A co-transcriptionally (19,20). Moreover, a recent study reported that H3K36me3, a histone marker for transcription elongation, could guide the installation of m⁶A modifications with classic enrichment near stop codons through direct recruitment of METTL14 (18). In this study, besides transcription factors, we also found classic RBPs worked as regulators that selectively regulated subsets of m⁶A sites through direct recruitments of methyltransferase complex, suggesting that various RBPs and transcription factors work together to modulate the precise levels of specific m⁶A sites. In contrast to transcription factors, which always bind to the promoters, the RBPs confer the m⁶A specificity by their RNA binding specificities. Moreover, the m⁶A sites can be controlled precisely through the modulations of multiple regulators. Therefore, it is very likely that m⁶A RNA methylation is precisely controlled in a similar manner as alternative splicing, which is regulated complicatedly by histone modifications co-transcriptionally as well as a variety of splicing factors that bind the *cis*-regulatory elements of splicing (68).

Technically, reliable quantification of m⁶A sites on mRNAs is still of great challenges. Therefore, we made multiple methodological improvements in order to mitigate the impacts of technical biases of m⁶A-seq data, which were important for the successful identification of m⁶A regulators. First of all, we compared our winscore-based method with exomePeak (60) and MeTPeak (69) using one of the HepG2 m⁶A-seq dataset. Though our winscore-based peaks had a similar number and distribution across 5'UTR, CDS and 3'UTR as exomePeak and MeTPeak, the density of m⁶A motifs (number of RRAC motifs in 100 bp of peak) of

winscore-based peaks were more than 2-fold higher than exomePeak and MeTPeak called peaks (Supplementary Figure S7A and B), suggesting our peaks are more centralized to real m⁶A sites. This should be important for the quantification of m⁶A peaks, because non-centralized long peaks may dilute the signals of m⁶A differences. Another important technical detail was that we defined the m⁶A ratio of each merge peaks with multiple sliding windows as the maximum m⁶A ratios of all windows for each sample, respectively. The exact locations and widths of m⁶A peaks may be biased by the RNA fragment lengths, sequencing read lengths and *et al.*, it is of great advantage to use the peak summits of each library for the overlapped peaks other than the exact same region when comparing the m⁶A ratios using m⁶A-seq data from different labs. Third, to calculate the m⁶A ratios, we required the input RPKM > 5 to deal with the unreliable m⁶A ratios with low read coverage. Fourth, we used quantile normalization to normalize the m⁶A ratios across all samples, thus the bias caused by different immunoprecipitation efficiencies across the libraries were minimized. Last but the most important, we identified the m⁶A regulators based on m⁶A modules other than single m⁶A sites, which greatly minimized the impact of using the noisy m⁶A quantifications of single m⁶A peaks.

DATA AVAILABILITY

The raw data of the low-input m⁶A-seq data have been deposited in the Sequence Read Archive (SRA) database (<https://dataview.ncbi.nlm.nih.gov/>) under the accession number SRP211943.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank Yi Xing and Rui Zhang for the suggestions in preparing the manuscript.

FUNDING

National Key R&D Program of China [2018YFA0107200 to J.W.]; National Natural Science Foundation of China [31771446 to J.W.]; Guangzhou Science and Technology Program [201904010181 to J.W.]; Fundamental Research Funds for the Central Universities [18ykzd12 to J.W.]. Funding for open access charge: National Key R&D Program of China [2018YFA0107200].

Conflict of interest statement. None declared.

REFERENCES

- 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.
- Dominissini, D., Moshitch-Moshkovitz, S., Schwartz, S., Salmon-Divon, M., Ungar, L., Osenberg, S., Cesarkas, K., Jacob-Hirsch, J., Amariglio, N., Kupiec, M. *et al.* (2012) Topology of the human and mouse m⁶A RNA methylomes revealed by m⁶A-seq. *Nature*, **485**, 201–206.

3. Meyer, K.D. and Jaffrey, S.R. (2017) Rethinking m(6)A readers, writers, and erasers. *Annu. Rev. Cell Dev. Biol.*, **33**, 319–342.
4. 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 Discov.*, **4**, 10.
5. Wen, J., Lv, R., Ma, H., Shen, H., He, C., Wang, J., Jiao, F., Liu, H., Yang, P., Tan, L. *et al.* (2018) Zc3h13 regulates nuclear RNA m(6)A methylation and mouse embryonic stem cell self-renewal. *Mol. Cell*, **69**, 1028–1038.
6. 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.
7. Harcourt, E.M., Kietrys, A.M. and Kool, E.T. (2017) Chemical and structural effects of base modifications in messenger RNA. *Nature*, **541**, 339–346.
8. Nachtergaele, S. and He, C. (2018) Chemical modifications in the life of an mRNA transcript. *Annu. Rev. Genet.*, **52**, 349–372.
9. Frye, M., Harada, B.T., Behm, M. and He, C. (2018) RNA modifications modulate gene expression during development. *Science*, **361**, 1346–1349.
10. Roundtree, I.A., Evans, M.E., Pan, T. and He, C. (2017) Dynamic RNA modifications in gene expression regulation. *Cell*, **169**, 1187–1200.
11. Delaunay, S. and Frye, M. (2019) RNA modifications regulating cell fate in cancer. *Nat. Cell Biol.*, **21**, 552–559.
12. Wang, S., Chai, P., Jia, R. and Jia, R. (2018) Novel insights on m(6)A RNA methylation in tumorigenesis: a double-edged sword. *Mol. Cancer*, **17**, 101.
13. Lian, H., Wang, Q.H., Zhu, C.B., Ma, J. and Jin, W.L. (2018) Deciphering the Epitranscriptome in Cancer. *Trends Cancer*, **4**, 207–221.
14. Deng, X., Su, R., Weng, H., Huang, H., Li, Z. and Chen, J. (2018) RNA N(6)-methyladenosine modification in cancers: current status and perspectives. *Cell Res.*, **28**, 507–517.
15. Jonkhout, N., Tran, J., Smith, M.A., Schonrock, N., Mattick, J.S. and Novoa, E.M. (2017) The RNA modification landscape in human disease. *RNA*, **23**, 1754–1769.
16. Garcia-Campos, M.A., Edelheit, S., Toth, U., Safra, M., Shachar, R., Viukov, S., Winkler, R., Nir, R., Lasman, L., Brandis, A. *et al.* (2019) Deciphering the 'm(6)A Code' via antibody-independent quantitative profiling. *Cell*, **178**, 731–747.
17. Ke, S., Pandya-Jones, A., Saito, Y., Fak, J.J., Vagbo, C.B., Geula, S., Hanna, J.H., Black, D.L., Darnell, J.E. Jr. and Darnell, R.B. (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.
18. Huang, H., Weng, H., Zhou, K., Wu, T., Zhao, B.S., Sun, M., Chen, Z., Deng, X., Xiao, G., Auer, F. *et al.* (2019) Histone H3 trimethylation at lysine 36 guides m(6)A RNA modification co-transcriptionally. *Nature*, **567**, 414–419.
19. Barbieri, I., Tzelepis, K., Pandolfini, L., Shi, J., Millan-Zambrano, G., Robson, S.C., Aspris, D., Migliori, V., Bannister, A.J., Han, N. *et al.* (2017) Promoter-bound METTL3 maintains myeloid leukaemia by m(6)A-dependent translation control. *Nature*, **552**, 126–131.
20. Bertero, A., Brown, S., Madrigal, P., Osnato, A., Ortman, D., Yiangou, L., Kadiwala, J., Hubner, N.C., de Los Mozos, I.R., Sadee, C. *et al.* (2018) The SMAD2/3 interactome reveals that TGFbeta controls m(6)A mRNA methylation in pluripotency. *Nature*, **555**, 256–259.
21. Wang, L., Wen, M. and Cao, X. (2019) Nuclear hnRNPA2B1 initiates and amplifies the innate immune response to DNA viruses. *Science*, **365**, eaav0758.
22. Zheng, Q., Hou, J., Zhou, Y., Li, Z. and 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.
23. Batista, P.J., Molinier, B., Wang, J., Qu, K., Zhang, J., Li, L., Bouley, D.M., 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.
24. 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.
25. Schwartz, S., Mumbach, M.R., Jovanovic, M., Wang, T., Maciag, K., Bushkin, G.G., Mertins, P., Ter-Ovanesyan, D., Habib, N., Cacchiarelli, D. *et al.* (2014) Perturbation of m6A writers reveals two distinct classes of mRNA methylation at internal and 5' sites. *Cell Rep.*, **8**, 284–296.
26. Lichinchi, G., Gao, S., Saletore, Y., Gonzalez, G.M., Bansal, V., Wang, Y., Mason, C.E. and Rana, T.M. (2016) Dynamics of the human and viral m(6)A RNA methylomes during HIV-1 infection of T cells. *Nat. Microbiol.*, **1**, 16011.
27. Lin, S., Choe, J., Du, P., Triboulet, R. and Gregory, R.I. (2016) The m(6)A Methyltransferase METTL3 promotes translation in human cancer cells. *Mol. Cell*, **62**, 335–345.
28. Cui, Q., Shi, H., Ye, P., Li, L., Qu, Q., Sun, G., Lu, Z., Huang, Y., Yang, C.G. *et al.* (2017) m(6)A RNA methylation regulates the Self-Renewal and tumorigenesis of glioblastoma stem cells. *Cell Rep.*, **18**, 2622–2634.
29. Li, Z., Weng, H., Su, R., Weng, X., Zuo, Z., Li, C., Huang, H., Nachtergaele, S., Dong, L., Hu, C. *et al.* (2017) FTO plays an oncogenic role in acute myeloid leukemia as a N(6)-Methyladenosine RNA Demethylase. *Cancer Cell*, **31**, 127–141.
30. Shi, H., Wang, X., Lu, Z., Zhao, B.S., Ma, H., Hsu, P.J., Liu, C. and He, C. (2017) YTHDF3 facilitates translation and decay of N(6)-methyladenosine-modified RNA. *Cell Res.*, **27**, 315–328.
31. Xiang, Y., Laurent, B., Hsu, C.H., 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.
32. Zhang, S., Zhao, B.S., Zhou, A., Lin, K., Zheng, S., Lu, Z., Chen, Y., Sulman, E.P., Xie, K., Bogler, O. *et al.* (2017) m(6)A demethylase ALKBH5 maintains tumorigenicity of glioblastoma stem-like cells by sustaining FOXM1 expression and cell proliferation program. *Cancer Cell*, **31**, 591–606.
33. Liu, J., Eckert, M.A., Harada, B.T., Liu, S.M., Lu, Z., Yu, K., Tienda, S.M., Chryplewicz, A., Zhu, A.C., Yang, Y. *et al.* (2018) m(6)A mRNA methylation regulates AKT activity to promote the proliferation and tumorigenicity of endometrial cancer. *Nat. Cell Biol.*, **20**, 1074–1083.
34. Rubio, R.M., Depledge, D.P., Bianco, C., Thompson, L. and Mohr, I. (2018) RNA m(6)A modification enzymes shape innate responses to DNA by regulating interferon beta. *Genes Dev.*, **32**, 1472–1484.
35. Tan, B., Liu, H., Zhang, S., da Silva, S.R., Zhang, L., Meng, J., Cui, X., Yuan, H., Sorel, O., Zhang, S.W. *et al.* (2018) Viral and cellular N(6)-methyladenosine and N(6),2'-O-dimethyladenosine epitranscriptomes in the KSHV life cycle. *Nat. Microbiol.*, **3**, 108–120.
36. Weng, H., Huang, H., Wu, H., Qin, X., Zhao, B.S., Dong, L., Shi, H., Skibbe, J., Shen, C., Hu, C. *et al.* (2018) METTL14 inhibits hematopoietic stem/progenitor differentiation and promotes leukemogenesis via mRNA m(6)A modification. *Cell Stem Cell*, **22**, 191–205.
37. Zhong, X., Yu, J., Frazier, K., Weng, X., Li, Y., Cham, C.M., Dolan, K., Zhu, X., Hubert, N., Tao, Y. *et al.* (2018) Circadian clock regulation of hepatic lipid metabolism by modulation of m(6)A mRNA methylation. *Cell Rep.*, **25**, 1816–1828.
38. Kim, D., Langmead, B. and Salzberg, S.L. (2015) HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods*, **12**, 357–360.
39. Pertea, M., Pertea, G.M., Antonescu, C.M., Chang, T.C., Mendell, J.T. and Salzberg, S.L. (2015) StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.*, **33**, 290–295.
40. Galili, T. (2015) dendextend: an R package for visualizing, adjusting and comparing trees of hierarchical clustering. *Bioinformatics*, **31**, 3718–3720.
41. Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H. and Glass, C.K. (2010) Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell*, **38**, 576–589.
42. Linder, B., Grozhik, A.V., Orlangerin-George, A.O., Meydan, C., Mason, C.E. and Jaffrey, S.R. (2015) Single-nucleotide-resolution mapping of m6A and m6Am throughout the transcriptome. *Nat. Methods*, **12**, 767–772.
43. Langfelder, P. and Horvath, S. (2008) WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*, **9**, 559.

44. Huang da,W., Sherman,B.T. and Lempicki,R.A. (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.*, **4**, 44–57.
45. Cerami,E., Gao,J., Dogrusoz,U., Gross,B.E., Sumer,S.O., Aksoy,B.A., Jacobsen,A., Byrne,C.J., Heuer,M.L., Larsson,E. *et al.* (2012) The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov.*, **2**, 401–404.
46. Gao,J., Aksoy,B.A., Dogrusoz,U., Dresdner,G., Gross,B., Sumer,S.O., Sun,Y., Jacobsen,A., Sinha,R., Larsson,E. *et al.* (2013) Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci. Signal.*, **6**, pii.
47. The Gene Ontology, C. (2019) The Gene Ontology Resource: 20 years and still GOing strong. *Nucleic Acids Res.*, **47**, D330–D338.
48. Li,J.H., Liu,S., Zhou,H., Qu,L.H. and Yang,J.H. (2014) starBase v2.0: decoding miRNA-ceRNA, miRNA-ncRNA and protein-RNA interaction networks from large-scale CLIP-Seq data. *Nucleic Acids Res.*, **42**, D92–D97.
49. Sloan,C.A., Chan,E.T., Davidson,J.M., Malladi,V.S., Strattan,J.S., Hitz,B.C., Gabdank,I., Narayanan,A.K., Ho,M., Lee,B.T. *et al.* (2016) ENCODE data at the ENCODE portal. *Nucleic Acids Res.*, **44**, D726–D732.
50. Yang,J.H., Li,J.H., Shao,P., Zhou,H., Chen,Y.Q. and Qu,L.H. (2011) starBase: a database for exploring microRNA-mRNA interaction maps from Argonaute CLIP-Seq and Degradome-Seq data. *Nucleic Acids Res.*, **39**, D202–D209.
51. Polymenidou,M., Lagier-Tourenne,C., Hutt,K.R., Huelga,S.C., Moran,J., Liang,T.Y., Ling,S.C., Sun,E., Wancewicz,E., Mazur,C. *et al.* (2011) Long pre-mRNA depletion and RNA missplicing contribute to neuronal vulnerability from loss of TDP-43. *Nat. Neurosci.*, **14**, 459–468.
52. Ray,D., Kazan,H., Cook,K.B., Weirauch,M.T., Najafabadi,H.S., Li,X., Gueroussov,S., Albu,M., Zheng,H., Yang,A. *et al.* (2013) A compendium of RNA-binding motifs for decoding gene regulation. *Nature*, **499**, 172–177.
53. Ule,J., Stefani,G., Mele,A., Ruggiu,M., Wang,X., Taneri,B., Gaasterland,T., Blencowe,B.J. and Darnell,R.B. (2006) An RNA map predicting Nova-dependent splicing regulation. *Nature*, **444**, 580–586.
54. Dittmar,K.A., Jiang,P., Park,J.W., Amirkian,K., Wan,J., Shen,S., Xing,Y. and Carstens,R.P. (2012) Genome-wide determination of a broad ESRP-regulated posttranscriptional network by high-throughput sequencing. *Mol. Cell Biol.*, **32**, 1468–1482.
55. Anderson,E.S., Lin,C.H., Xiao,X., Stoilov,P., Burge,C.B. and Black,D.L. (2012) The cardiotoxic steroid digitoxin regulates alternative splicing through depletion of the splicing factors SRSF3 and TRA2B. *RNA*, **18**, 1041–1049.
56. Arguello,A.E., DeLiberto,A.N. and Kleiner,R.E. (2017) RNA Chemical Proteomics Reveals the N-6-Methyladenosine (m(6)A)-Regulated Protein-RNA Interactome. *J. Am. Chem. Soc.*, **139**, 17249–17252.
57. Huang,H., Weng,H., Sun,W., Qin,X., Shi,H., Wu,H., Zhao,B.S., Mesquita,A., Liu,C., Yuan,C.L. *et al.* (2018) Recognition of RNA N(6)-methyladenosine by IGF2BP proteins enhances mRNA stability and translation. *Nat. Cell Biol.*, **20**, 285–295.
58. Edupuganti,R.R., Geiger,S., Lindeboom,R.G.H., Shi,H., Hsu,P.J., Lu,Z., Wang,S.Y., Baltissen,M.P.A., Jansen,P., Rossa,M. *et al.* (2017) N(6)-methyladenosine (m(6)A) recruits and repels proteins to regulate mRNA homeostasis. *Nat. Struct. Mol. Biol.*, **24**, 870–878.
59. Wan,C., Borgeson,B., Phanse,S., Tu,F., Drew,K., Clark,G., Xiong,X., Kagan,O., Kwan,J., Bezginov,A. *et al.* (2015) Panorama of ancient metazoan macromolecular complexes. *Nature*, **525**, 339–344.
60. Meng,J., Lu,Z., Liu,H., Zhang,L., Zhang,S., Chen,Y., Rao,M.K. and Huang,Y. (2014) A protocol for RNA methylation differential analysis with MeRIP-Seq data and exomePeak R/Bioconductor package. *Methods*, **69**, 274–281.
61. Thorvaldsdottir,H., Robinson,J.T. and Mesirov,J.P. (2013) Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief. Bioinform.*, **14**, 178–192.
62. Love,M.I., Huber,W. and Anders,S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.*, **15**, 550.
63. Anders,S., Pyl,P.T. and Huber,W. (2015) HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics*, **31**, 166–169.
64. Zeng,Y., Wang,S., Gao,S., Soares,F., Ahmed,M., Guo,H., Wang,M., Hua,J.T., Guan,J., Moran,M.F. *et al.* (2018) Refined RIP-seq protocol for epitranscriptome analysis with low input materials. *PLoS Biol.*, **16**, e2006092.
65. Sundararaman,B., Zhan,L., Blue,S.M., Stanton,R., Elkins,K., Olson,S., Wei,X., Van Nostrand,E.L., Pratt,G.A., Huelga,S.C. *et al.* (2016) Resources for the comprehensive discovery of functional RNA elements. *Mol. Cell*, **61**, 903–913.
66. Horiuchi,K., Kawamura,T., Iwanari,H., Ohashi,R., Naito,M., Kodama,T. and Hamakubo,T. (2013) Identification of Wilms’ tumor 1-associating protein complex and its role in alternative splicing and the cell cycle. *J. Biol. Chem.*, **288**, 33292–33302.
67. Knuckles,P., Carl,S.H., Musheev,M., Niehrs,C., Wenger,A. and Buhler,M. (2017) RNA fate determination through cotranscriptional adenosine methylation and microprocessor binding. *Nat. Struct. Mol. Biol.*, **24**, 561–569.
68. Lee,Y. and Rio,D.C. (2015) Mechanisms and regulation of alternative Pre-mRNA splicing. *Annu. Rev. Biochem.*, **84**, 291–323.
69. 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.