Genetic analyses support the contribution of mRNA $N^6$-methyladenosine (m$^6$A) modification to human disease heritability

Zijie Zhang$^{1,2,9}$, Kaixuan Luo$^{3,9}$, Zhongyu Zou$^{1,2}$, Maguanyun Qiu$^{1,2}$, Jiakun Tian$^{1,2}$, Laura Sieh$^{1,2}$, Hailing Shi$^{1,2}$, Yuxin Zou$^4$, Gao Wang$^5$, Jean Morrison$^6$, Allen C. Zhu$^{1,2,5}$, Min Qiao$^6$, Zhongshan Li$^{3,7}$, Matthew Stephens$^{3,4,2}$, Xin He$^{3,2}$ and Chuan He$^{1,2,5,8}$

$m^6$A-modified sites play important roles in regulating messenger RNA processing. Despite rapid progress in this field, little is known about the genetic determinants of m$^6$A modification and their role in common diseases. In this study, we mapped the quantitative trait loci (QTLs) of m$^6$A peaks in 60 Yoruba (YRI) lymphoblastoid cell lines. We found that m$^6$A QTLs are largely independent of expression and splicing QTLs and are enriched with binding sites of RNA-binding proteins, RNA structure-changing variants and transcriptional features. Joint analysis of the QTLs of m$^6$A and related molecular traits suggests that the downstream effects of m$^6$A are heterogeneous and context dependent. We identified proteins that mediate m$^6$A effects on translation. Through integration with data from genome-wide association studies, we show that m$^6$A QTLs contribute to the heritability of various immune and blood-related traits at levels comparable to splicing QTLs and roughly half of expression QTLs. By leveraging m$^6$A QTLs in a transcriptome-wide association study framework, we identified putative risk genes of these traits.

We mapped m$^6$A QTLs using lymphoblastoid cell lines (LCLs), for which QTL data of multiple molecular traits are available$^{36,41,42}$. We found that the m$^6$A consensus motif (RRACH), while highly enriched, explains only a small fraction of m$^6$A QTLs. We observed that m$^6$A QTLs are enriched in RNA-binding protein (RBP) target sites, riboSNitches (variants affecting RNA secondary structure) and transcriptional features, suggesting that these factors are important regulators of m$^6$A installation. Through integration with other molecular QTL data, we found that the regulatory effects of m$^6$A on downstream traits, such as translation, may vary across m$^6$A sites in a context-dependent manner.

We conducted joint analysis of m$^6$A QTLs and genome-wide association study (GWAS) data. Current efforts to characterize GWAS variants have largely focused on transcriptional effects. However, recent studies, employing different approaches from colocalization to heritability analyses, estimate that eQTLs explain only 10–25% of GWAS signals$^{36,41,42}$. To fill this gap, researches have suggested other mechanisms, such as RNA splicing$^{43,44}$. In our analysis, we found that m$^6$A QTLs are enriched for risk variants of a range of complex traits, particularly autoimmune diseases and blood cell-related traits. The contribution of m$^6$A QTLs to heritability of these traits is roughly half of eQTLs and comparable to splicing QTLs (sQTLs). Treating m$^6$A levels as molecular traits, we performed a transcriptome-wide association study (TWAS) of these traits and identified a number of m$^6$A sites and related genes. Taken together, our results demonstrate that m$^6$A variation is an important link between genetic and phenotypic variation.

$^{1}$Department of Chemistry and Institute for Biophysical Dynamics, The University of Chicago, Chicago, IL, USA. $^{2}$Howard Hughes Medical Institute, The University of Chicago, Chicago, IL, USA. $^{3}$Department of Human Genetics, The University of Chicago, Chicago, IL, USA. $^{4}$Department of Biostatistics and Data Science, School of Public Health, The University of Texas Health Science Center at Houston, Houston, TX, USA. $^{5}$Institute of Genomic Medicine, Wenzhou Medical University, Wenzhou, China. $^{6}$Department of Biochemistry and Molecular Biology, The University of Chicago, Chicago, IL, USA. $^{7}$These authors contributed equally: Zijie Zhang, Kaixuan Luo. $^{50}$e-mail: mstephens@uchicago.edu; xinhe@uchicago.edu; chuanhe@uchicago.edu

https://doi.org/10.1038/s41588-020-0644-z
Results

Mapping cis-m6A QTLs. We used m6A sequencing (m6A-seq)\(^{44,45}\) to profile m6A levels across the transcriptome in LCLs derived from 60 YRI individuals. We obtained on average 60 million reads for unmodified (input) and immunoprecipitated mRNA libraries for each cell (Fig. 1a). We called peaks jointly on all samples (see Methods) and identified 20,044 peaks (Supplementary Table 1) located in the transcripts of 8,448 genes, with an average peak length of 351 base pairs (bp) (Extended Data Fig. 1 and Fig. 1a). Consistent with previous reports\(^{4,49}\), m6A peaks are enriched near start and stop codons (Extended Data Fig. 1b,c) and sequences within peaks are enriched in the RRACH motif (Extended Data Fig. 1d).

We tested the association between the m6A level of each peak and nearby genetic variants using a linear model, accounting for GC content and other covariates (Fig. 1b; see Methods and Supplementary Note). To determine a proper window size for cis-QTL mapping, we first tested all SNPs within 2 Mb of the m6A peaks (Extended Data Fig. 1e). Most SNPs strongly associated with m6A are within 100 kilobases (kb) of m6A peaks (Fig. 1c). Therefore, we restricted our cis-tests to SNP–peak pairs within 100 kb. The resulting \(P\) values show a strong deviation from the null expectation, while \(P\) values from permutations are consistent with the null expectation, indicating that the test is well-calibrated (Fig. 1d). We used a permutation scheme implemented in FastQTL\(^{60}\) to account for multiple genetic variants tested per peak. This resulted in 822 peaks with at least one significant cis-m6A QTL (denoted as ePeaks, following the literature of eQTLs), at a 10% false discovery rate (FDR)\(^{60}\) (Extended Data Fig. 1f). Most of these ePeaks (86%) have a single causal effect (Extended Data Fig. 1g), based on computational fine-mapping analysis\(^{41}\).

We quantified the contribution of genetic variation to interindividual variation of m6A levels by estimating the cis-heritability of each peak. Most peaks have low heritability values, with 918 peaks having heritability \(>0\) (Extended Data Fig. 2a). The heritability of ePeaks is higher with a median of 0.31 (Extended Data Fig. 2b).

We next examined the distribution of m6A QTLs relative to gene-based features, using the program Torus\(^{61}\). m6A QTLs are most enriched at the 3' UTR (log OR = 4.9, where OR is the odds ratio) followed by 5' UTR (log OR = 4.5) and coding sequence (CDS) (log OR = 3.8), but not in intergenic repressive regions as marked by H3K27me3 (Extended Data Fig. 2c).

Comparing m6A QTLs (derived from mRNA m6A peaks) to the eQTLs mapped in the same LCLs, we found that genes containing ePeaks and eQTLs are largely distinct (Extended Data Fig. 1e). In genes with both m6A and eQTLs, the lead SNPs of two types of QTLs are mostly >10 kb apart and in low linkage disequilibrium (LD) (Fig. 1f and Extended Data Fig. 2d). Comparison of m6A QTLs to sQTLs shows similar patterns (Extended Data Fig. 2e–g). These results suggest that m6A QTLs and eQTLs/sQTLs are distinct types of molecular QTLs.

m6A QTLs are enriched in RNA-related features. To understand which factors may determine m6A deposition, we analyzed the features of m6A QTLs and their surrounding sequences. We annotated SNPs using RNA-related features including the m6A consensus motif (RRACH) contained in m6A peaks, binding sites of 121 RBPs (ENCODEx enhanced crosslinking and immunoprecipitation sequencing (eCLIP-seq) peaks)\(^ {19}\), riboSNitches\(^ {11}\) (variant RNAs changing RNA secondary structure) and predicted microRNA (miRNA) binding sites\(^ {21}\). We used two approaches to test enrichment. In our primary analysis, we used Fisher's exact test, comparing possible causal variants of m6A from fine-mapping\(^ {18}\) and randomly sampled SNPs that matched the key properties of m6A QTLs (see Methods)\(^ {51}\). We found the m6A consensus motif in m6A peaks highly enriched in m6A QTLs with an OR of 685 (\(P = 1.0 \times 10^{-33}\)). However, only 12% of m6A QTLs contained in m6A peaks (most QTLs were outside peaks) disrupt the consensus motif. Despite this, we found that m6A QTLs tend to be located in proximity to the consensus motif since an additional 33% of m6A QTLs contained in m6A peaks are located within 50 bp of the motif and 46% within 100 bp, suggesting many m6A QTLs may indirectly affect the binding of methyltransferase, demethylase or reader proteins to the consensus motif. We also found enrichment in riboSNitches (OR = 6.2, \(P = 5.9 \times 10^{-4}\)) and RBP binding sites (OR = 2.5, \(P = 8.3 \times 10^{-7}\)) but not predicted miRNA targets\(^ {52}\) (Fig. 2a). In a secondary analysis, we tested enrichment using Torus, which accounts for uncertainty of causal variants due to LD. This analysis revealed similar results (Extended Data Fig. 3a).

We tested enrichment for each RBP and miRNA separately using Torus (Supplementary Table 2). Interestingly, several of the most enriched RBPs are known m6A-interacting proteins including FTO, an m6A demethylase\(^ {47,48}\), and IFG2BP2, an m6A reader protein that stabilizes nuclear RNA\(^ {19}\) (Fig. 2b). Analyses of individual miRNAs show enrichment of m6A QTLs at the binding sites of hsa-miR-582-5p and hsa-miR-331-3p. This finding is in line with previous reports that miRNA could affect m6A levels\(^ {39}\).

The enrichment of binding sites of an RBP in m6A QTLs could occur if the binding sites of an RBP co-occur with cis-elements regulating m6A, without the RBP playing a direct role in m6A deposition. We reason that if the RBP is causal, alterations in motif scores (disruption or creation) of SNPs should correlate with their effects on m6A deposition. We limited this correlation analysis to fine-mapped m6A QTLs that also have significant effects on motif score. As proof of principle, DNA variants creating a consensus motif were much more likely to be positively associated with m6A levels (Fig. 2c; see an example in Fig. 2d). We tested 29 RBPs with more than 5 data points and identified 3 RBPs with significant correlations at an FDR = 10% (Fig. 2e). Interestingly, SRSF1 is a known splicing factor\(^ {1}\), suggesting a possible connection of splicing with m6A deposition.

m6A modification is coupled with transcriptional processes. Recent studies suggest that the deposition of m6A may occur cotranscriptionally and be influenced by transcription processes\(^ {2,24}\). We used our m6A QTLs and ENCODE chromatin immunoprecipitation sequencing (ChIP–seq) data from LCLs to examine potential link between m6A and transcription\(^ {42}\). We observed significant enrichment (Fisher’s exact test) of fine-mapped m6A QTLs in RNA polymerase II, phospho-RNA polymerase II and transcription-factor-binding sites (TFBSs) as well as in histone marks of promoters (H3K4me3), enhancers (H3K4me1, H3K27ac) and active transcription (H3K36me3) (Fig. 3a). The enrichment of m6A QTLs in H3K36me3, the most enriched feature, remained strong when conditioned on other histone modifications using Torus (Extended Data Fig. 3b). A recent study\(^ {50}\) showed that H3K36me3 is recognized by the m6A writer protein METTL14 to facilitate m6A installation on mRNA, thus validating our finding.

We then compared the contributions of RNA-related and transcriptional features (TFBSs) to m6A QTLs. We used fine-mapping to quantify the probability of an SNP being a causal variant of m6A, known as posterior inclusion probability (PIP). We estimated the proportion of causal variants attributed to a feature by summing the PIPs of all variants located within that feature (see Methods). Using this approach, we found that TFBS and RBP binding sites make an about equal contribution to m6A QTLs (17.8 and 15.8%, respectively) and the RRACH motif contributes 1.95% (Extended Data Fig. 3c).

These findings support a tight connection between transcriptional processes and m6A installation. Two models have been suggested to explain this connection (Fig. 3b). In the first model (the transcription rate model), m6A installation is affected by the progression rate of RNA polymerase II, with fast progression associated with lower m6A methylation\(^ {11}\). In the second model (the transcription factor recruitment model), the methyltransferase complex
(MTC) is recruited to mRNA by transcription factors, for example, ZFP217 (ref. 58) and CEBPZ or adapter proteins, for example, SMAD2/3 (ref. 56).

If the transcription rate model is correct, we expect the correlation between variant effects on transcription rate and variant effects on m6A level in the matched transcript. To assess this, we ascertained the lead SNPs of transcription rate QTLs from the same LCL samples but found little correlation between transcription rates and m6A effect sizes (Fig. 3c). In a positive control, we observed a strong correlation of transcription rate QTL effects with eQTL effects.

**Fig. 1 | Mapping common genetic variants associated with m6A.**

a, Overall study design and workflow of m6A QTL mapping. The linear regression model for association testing, adjusting for guanine–cytosine (GC) content and immunoprecipitation efficiency. b, An example of m6A QTL. The left panel shows the box plot of m6A levels grouped by the genotype of the example m6A QTL (rs1045405). n = 60 biologically independent samples. The lower and upper hinges correspond to the first and third quartiles. The horizontal line indicates the median and the whiskers correspond to the value no further than 1.5× the interquartile range. The right panel shows the mean coverage of each genotype at the m6A peak. The m6A peak is shown by the blue track and the gene model by the gray track. The coverages of the input and immunoprecipitation libraries are shown in lines and shadows, respectively. c, Spatial distribution of m6A QTLs represented by the cumulative fraction of SNPs with increasing distance from m6A peaks at varying P value cutoffs of SNP-peak association. d, Quantile–quantile plot of P values. The cis-tests (n = 60 individuals) results are plotted in black and the results of five permutation tests are shown in different colors. e, Overlap between ePeak-harboring genes and eGenes (both at an FDR < 10%) mapped in the same cohort of YRI LCL samples. f, Distribution of the distance between the lead ePeak SNP and the eGene SNP in genes that have both ePeaks and eGenes mapped.
**Fig. 2** Functional features enriched in m^6^A QTLs. a, log_2 OR enrichment of fine-mapped m^6^A QTLs (SNP with the highest PIP in each credible set) versus random control SNPs (see Methods) in RNA features by Fisher’s exact test. The error bars represent the 95% CIs from two-tailed tests. b, Enrichment of m^6^A QTLs at the RBPs of individual RBPs using eCLIP-seq data from ENCODE^{39,41}. The red dashed line represents the Bonferroni-corrected P ≤ 0.05 threshold. c, Distribution of m^6^A QTL effect sizes between SNPs creating versus breaking the m^6^A consensus motif. P was computed using Welch’s test (n = 32 SNPs). The lower and upper hinges correspond to the first and third quartiles. The horizontal line indicates the median and the whiskers correspond to the value no further than 1.5 times the interquartile range. d, An m^6^A QTL example illustrating how a genetic variant disrupting an RRACH motif could lead to m^6^A variation. e, RBPs for which changes in binding affinities were significantly correlated with fine-mapped m^6^A QTL effect sizes (all SNPs with PIP > 0.5 and maximum PIP SNPs for ePeaks without SNP PIP > 0.5). Changes in binding affinity are represented by the alteration of motif match scores from the reference to the alternative allele. The shaded region and line show the 95% CI and fitted line from the linear model (n = 7, 5 and 5 SNPs for SRSF1, DDX55 and RPS3, respectively).

Effects (R^2^ = 0.69 and 0.65) and with protein QTL effects (R^2^ = 0.37 and 0.42) in the matched transcripts (Extended Data Figs. 3d,e). These data suggest that, overall, transcription rate may not determine m^6^A deposition in LCLs. It is possible that other mechanisms, such as RNA polymerase II pausing explain the observed correlation between m^6^A and transcription rates in an earlier study^{57}.

To examine the transcription factor recruitment model, we used Torus to assess the enrichment of m^6^A QTLs for binding sites of individual transcription factors while accounting for enrichment of m^6^A QTLs in H3K27ac, a general transcription marker. Fifty transcription factors are significantly enriched at a Bonferroni-corrected P ≤ 0.05 (Fig. 3d and Supplementary Table 2). We then selected a few of these based on a literature review and performed coimmunoprecipitation experiments. Two transcription factors robustly pulled down m^6^A methyltransferase components in LCLs, including RBBP5, a component of the COMPASS histone H3K4 methylase complex, and BACH1, a regulator of oxidative stress^{61,62} (Fig. 3e and associated source data), supporting the transcription factor recruitment model.

**Analysis of molecular QTLs suggests context-dependent effects of m^6^A.** It is generally believed that specific reader proteins...
recognize m^6A and mediate downstream effects^{1,2}. The best-studied readers are known to promote translation (for example, YTHDF1), mRNA degradation (for example, YTHDF2) or affect mRNA nuclear processing (for example, YTHDC1) (refs. 3,6–8). We used m^6A QTLs as natural perturbations of m^6A to explore its effects on five possible downstream traits: mRNA expression; ribosome binding; protein level; mRNA decay rate; and alternative polyadenylation (APA)^{29,33,35,40}.

We first ascertained the lead m^6A QTLs at different P thresholds and then estimated the percentage of m^6A QTLs that were also QTLs of other traits using Storey's π^1 method^{40,47}. We found that m^6A QTLs were more likely to be other QTLs than random SNP–gene pairs, with increased sharing at a more stringent P value threshold (Fig. 4a), suggesting functional connections between m^6A and other molecular phenotypes, as expected from earlier studies^{1,2}.

Fig. 3 | m^6A installation is coupled with transcriptional processes. a, Enrichment of fine-mapped m^6A QTLs (SNP with the highest PIP in each credible set) in chromatin features by two-sided Fisher's exact test comparing m^6A QTLs to control SNPs. The error bars represent the 95% CIs. b, Two possible models of m^6A regulation through transcription. TF, transcription factor. c, Effect sizes of ascertained transcription rate QTLs versus their effects on m^6A levels. The transcription rate was measured by 4SU-seq in an earlier study^{40}. 4SU-seq with 30 min 4SU labeling (upper, n = 698 SNPs) and 60 min 4SU labeling (lower, n = 688 SNPs) shows similar results. The shaded region and line show the 95% CI and fitted line from the linear model. d, Enrichment of m^6A QTL at the TFBSs of individual transcription factors conditioned on H3K27ac peaks by T orus analysis. The red dashed line shows the Bonferroni-corrected P ≤ 0.05 cutoff. e, Western blot of the transcription factor coimmunoprecipitation experiment. Ten percent of lysate was loaded as the ‘input’. The cropped blot of each transcription factor of interest is shown, as well as three m^6A MTC components—METTL3, WTAP and VIRMA. These experiments were repeated twice with similar results.
the cascade from transcription to protein\(^*\). One potential problem is that sharing of m\(^6\)A QTLs and other molecular QTLs may be confounded by eQTLs, since transcription may influence both m\(^6\)A and other traits. However, the majority of m\(^6\)A QTLs mapped here are not chromatin-associated eQTLs, suggesting that in practice, this is not a main concern (Extended Data Fig. 4a).

Based on our current understanding that m\(^6\)A function is mediated by reader proteins with certain downstream effects (for example,
increase of translation efficiency by YTHDF1), we hypothesized that m’A QTLs and other molecular QTLs would have consistent effect directions. To test this hypothesis, we first confirmed that molecular traits along the cascade from transcription to translation show high positive correlations in QTL effects (Extended Data Fig. 4b). Surprisingly, the effect sizes of m’A QTLs and other molecular traits are poorly correlated (Fig. 4b). This lack of overall correlation suggests that the effects of m’A on downstream molecular phenotypes may be heterogeneous, with mechanisms varying across transcripts.

The context dependency of m’A function may be partially mediated by RBPs bound near m’A peaks. For example, binding by different m’A readers may lead to different downstream effects. To examine this hypothesis, we stratified our effect size correlation analysis by m’A peaks bound by different RBPs (using eCLIP-seq data). In 8 RBPs, we observed significant correlations (FDR < 10%) between the effect sizes of m’A QTLs and at least 1 related molecular trait (Fig. 4c). This result suggests that m’A function may vary according to binding of specific RBPs.

**m’A affects translation efficiency in a context-dependent manner.** To further investigate the context-dependent effects of m’A, we made use of data from an earlier study of m’A effects on translation in HeLa cells. This study examined the impacts of m’A depletion (by METTL3 knockdown) and YTHDF1 (m’A reader) knockdown on the translation efficiency of all transcripts, measured by ribosome profiling. Across all m’A modified transcripts, the effects of m’A depletion on translation efficiency were heterogeneous, with similar numbers of upregulated and downregulated genes (Extended Data Fig. 4c). To assess the impact of the RBP context, we compared the effects of m’A depletion on translation efficiency of transcripts containing m’A sites targeted by an RBP versus transcripts not targeted. Among 121 tested RBPs, 33 show statistically significant differences in target sites versus nontargets (FDR < 10%) (Fig. 4d). Again, the effects are quite heterogeneous, with almost equal numbers of RBPs involved in upregulation and downregulation of translation efficiency on m’A depletion. This list includes all four RBPs (YBX3, GRWD1, HLTF and PPIG) we identified from m’A versus ribosome-some QTL effect correlation analysis (Fig. 4c). Furthermore, the effect directions were consistent between two studies: m’A depletion resulted in higher translation efficiency of the RBP targets, in agreement with negative correlations of m’A versus ribosome-QTL effects (Fig. 4c,d). These results provide independent support to the hypothesis that the effects of m’A on translation efficiency depend on contexts, in particular binding of specific RBPs.

Interestingly, even in transcripts targeted by the classical m’A reader YTHDF1, the effect of m’A may be more complex than previously thought. While depletion of YTHDF1 leads to an overall reduction of translation efficiency in transcripts harboring YTHDF1-bound m’A peaks, approximately 33% of YTHDF1 targets show the opposite effects (Extended Data Fig. 4d). This observation suggests the possibility that the action of reader proteins may be modulated by additional, yet to be identified factors, diversifying m’A effects.

We validated an m’A effector protein, YBX3, as a translation repressor of m’A-modified and YBX3-bound transcripts (Fig. 4e). We knocked down YBX3 in HeLa cells and performed polysome profiling followed by quantitative reverse transcription PCR (RT–qPCR). We found more RNAs in polysome-bound fractions in YBX3-depleted cells compared with controls (Extended Data Fig. 5a), suggesting YBX3 as a translation repressor. To further validate YBX3 function, we selected five transcripts harboring m’A peaks overlapping with YBX3-bound sites, all of which show elevated translation efficiency on METTL3 knockdown (Fig. 4e). We quantified these transcripts in three polysome-bound fractions using RT–qPCR on YBX3 knockdown. The translation efficiency of these target transcripts was elevated in YBX3-depleted cells compared with controls in all but one case (Extended Data Fig. 5b). As a negative control, three YTHDF1-targeted m’A transcripts did not show translation efficiency elevation on YBX3 depletion. These results suggest that YBX3 probably mediates the m’A effect by repressing translation of YBX3-bound m’A transcripts. This effect is opposite from the YTHDF1 effect (increasing translation), providing a partial explanation of why m’A downstream effects appear heterogeneous (Fig. 4b,d).

**m’A QTLs make a significant contribution to the genetics of complex traits.** To study the role of m’A QTLs in human phenotypic variation, we collected GWAS summary statistics of 45 complex traits with an emphasis on immune and blood-related traits. For comparison, we also included eQTLs and sQTLs from LCLs. All three types of QTLs showed an excess of low P values in the GWAS of these traits, for example, lymphocyte counts (Fig. 5a and Extended Data Fig. 6a). We used stratified LD score regression (S-LDSC) to formally test enrichment of GWAS variants in m’A QTLs. S-LDSC is a tool for assessing how the heritability of a complex trait is partitioned among functional features, while controlling for LD, allele frequency and other baseline features. Following a previously used strategy, we fine-mapped m’A QTLs and used the resulting PIPs as an annotation, representing probable causal m’A variants (known as quantitative trait nucleotides (QTNs)). We found 10–20-fold enrichment of heritability in m’A QTNs in several selected traits (Fig. 5b and Extended Data Fig. 6b). The enrichment increased to 15–35-fold (Extended Data Fig. 7a) when we used m’A-related annotations (Extended Data Fig. 7b), such as RBP binding, as priors to improve fine-mapping (see Methods). Including QTNs of expression and splicing in S-LDSC only modestly reduced the enrichment level (Fig. 5b). We note, however, that m’A may affect expression (for example, by changing mRNA stability) and pre-mRNA splicing; therefore, adjusting eQTLs and sQTLs probably leads to underestimation of m’A QTL enrichment.

Expanding the S-LDSC analysis to all 45 traits, we found that m’A PIPs are enriched in most immune and blood traits (Fig. 5c and Extended Data Fig. 7c) as well as a small number of other traits, such as coronary artery disease and age at menopause, where immune systems may play a significant role. Thus, these results support the specificity of our finding and are consistent with the known role of m’A in the immune system.

Using S-LDSC, we compared the relative contributions to trait heritability by m’A QTLs, eQTLs and sQTLs (FDR < 10%). For traits where m’A QTNs showed enrichment (Fig. 5c), m’A QTLs explain about 2–4% of heritability, comparable to sQTLs and roughly 50–100% of the heritability explained by eQTLs (Fig. 5d and Extended Data Fig. 8). These estimates are probably conservative, since many QTLs below the FDR cutoff may contribute to trait heritability. Nevertheless, given the established roles of eQTLs and sQTLs, this relative comparison suggests that m’A QTLs can make a significant contribution to the heritability of complex traits.

**TWAS using m’A QTLs.** To highlight the potential of using m’A QTLs to identify specific risk genes, we performed a TWAS using m’A as a molecular-level trait. We built prediction models of m’A levels using genetic variants as explanatory variables, then assessed if genetically determined m’A levels correlate with specific phenotypes by using TWAS/FUSION. We found a number of m’A peaks passing the Bonferroni threshold across a range of immune and blood-related traits (Fig. 6a) as well as several other phenotypes (Extended Data Fig. 9a). These results showed limited overlap, at the level of genes, with the TWAS results using eQTLs and sQTLs mapped in LCLs (Fig. 6b and Supplementary Table 3), suggesting that m’A variation represents a distinct path from genetic to phenotypic variation.
We performed an in-depth analysis of lymphocyte count. The m' A TWAS identified a total of 30 significant m' A peaks in 28 genes (Fig. 6c). Since TWAS associations can result from LD and/or pleiotropic effects, we conducted colocalization analysis to identify cases where a single causal variant drives both m' A QTL and GWAS association. Among 30 peaks, 10 have high colocalization probabilities (PP4 from Coloc > 0.5) (Supplementary Table 4). In one example, an m' A peak in the DDX55 gene shows high colocalization probability (PP4 = 0.929). The SNP driving the colocalization result, rs3204541, is the top SNP in both m' A QTL and GWAS (Fig. 6d). A conditional association test adjusting for m' A showed that the TWAS association almost fully explains the GWAS signal in the region (Fig. 6e). The same m' A peak in DDX55 is also found by the m' A TWAS in leukocyte counts (Extended Data Fig. 9b,c). DDX55 is a DEAD box helicase gene; its paralog gene, DDX10, is implicated in myelodysplastic syndrome, a disease with abnormal blood cell counts. Importantly, DDX55 is not found by expression or splicing TWAS (both P ≥ 0.1). Together, our TWAS results highlight the promise of using m' A QTLs to reveal mechanisms in GWAS loci where genetic effects are not mediated by expression or splicing.

**Discussion**

We report a systematic genetic analysis of the most abundant mRNA modification—m' A. Our analysis reveals insights into mRNA m' A regulation, highlighting the importance of both RNA features (for example, RBPs and secondary structure) and transcriptional regulation (for example, transcription factor binding). We find that the functional effects of m' A on downstream processes, in particular translation, can be highly heterogeneous and depend on binding of specific RBPs. Our integrated analysis of m' A QTLs with GWAS supports the role of m' A as an important link from genetic to phenotypic variation.

Using an analysis that correlates SNP effects on RBP motifs and m' A levels, we identified specific RBPs, such as SRSF1, that may be m' A regulators (Fig. 2e). However, this analysis has some limitations. It may not be able to distinguish RBPs from the same families that share similar motifs. Due to the small sample size of our study, it may also be underpowered to detect many more RBPs regulating m' A. The enrichment of m' A QTLs in transcription-related features supports an emerging connection between mRNA modification and transcriptional control. In support of the recruitment model (Fig. 3b), TFBSs are enriched in m' A QTLs and several transcription factors interact with the m' A MTC in LCLs. Given the additional transcription factor–methyltransferase interactions reported previously in pluripotent stem cells and acute myeloid leukemia, we think transcription factor–methyltransferase interactions may broadly exist and participate in cell type-specific m' A regulation.

Previous studies found that m' A promotes translation efficiency and mRNA decay via interactions with reader proteins. Our results add additional insights into this model, suggesting that m' A effects...
on downstream processes, for example, translation, are much more heterogeneous across transcripts than previously appreciated. We identified RBPs that may influence the effects of m6A, including some with reported functions in RNA processing (Fig. 4c,d), for example, YBX3 (ref. 75) and HNRNPA1 (ref. 76). The RBPs uncovered in this study provide a resource for future studies.

We hypothesize two potential mechanisms that may explain context-dependent m6A effects. First, there may be more m6A reader proteins, with potentially different effects, than are currently known; some could be readers that respond to m6A through a structure switching mechanism77. Alternatively, the functions of RNA regulators may depend on m6A, even if they do not directly bind and recognize m6A or respond through structure switching (and hence not readers). These proteins may bind the motif that harbors m6A or a motif nearby m6A sites and compete with bona fide reader proteins on the modified transcripts. Future studies are needed to

Fig. 6 | m6A TWAS and colocalization analysis. a, Number of significant m6A TWAS genes in selected immune and blood-related traits. b, Overlap between significant genes discovered by the TWAS analyses using m6A, expression and splicing as molecular-level phenotypes. c, Manhattan plot of m6A TWAS associations of lymphocyte count. The dashed line shows the Bonferroni-corrected P ≤ 0.05 threshold. Genes are colored by Coloc PP4 (posterior probability of GWAS traits and m6A QTLs sharing common genetic causal variants). Ten genes colored with Coloc PP4 > 0.5 are labeled. d, Aligned Manhattan plots of GWAS and m6A QTLs at the DDX55 locus generated by LocusCompare. SNPs are colored by LD (r2) with the lead m6A QTL (rs3204541). e, Manhattan plot of GWAS association signal of lymphocyte count at the DDX55 locus before (gray dots) and after (blue dots) conditioning on the TWAS-predicted m6A level. Top panel: all genes within 200 kb and the significant m6A peak (green) are labeled.
assess these RBPs and their interactions as well as competition in RNA binding.

Our integrated analysis of m6A QTLs and GWAS highlights the importance of m6A to the etiology of complex traits and adds to the growing evidence that posttranscriptional regulation plays a key role in common diseases. Genetic variants affecting RNA processing are almost as common as, and are largely independently from, those affecting transcription9. These variants have been implicated in a number of diseases including cyctic fibrosis, type 2 diabetes, Crohn’s disease and lung cancer80. However, identifying variants with posttranscriptional regulation effects is more challenging than transcriptional effects. Mapping m6A QTLs may be an effective strategy to address this challenge, given the central role of m6A modification in almost every step of RNA processing (Extended Data Fig. 10).

One potential caveat of our GWAS analysis is the mismatch between the population ancestries of QTL (African) and GWAS (mostly European) data. However, the impact of this mismatch is probably limited. Studies have suggested that associations with complex traits, especially causal variants, are broadly shared across populations9,10. A systematic study with multiple complex traits estimated that more than 80% of causal variants are shared between European and Asian ancestries81. In another study, a TWAS on asthma using eQTL models trained on data from European and African ancestries gave broadly similar results82. Given these findings, we think many m6A QTNs (causal variants) in YRI LCLs are probably shared in Europeans. Therefore, population mismatch probably has a small impact in our S-LDSC analysis, which used PIPs as SNP annotations, and in our TWAS, where results are often driven by single shared variants between molecular QTLs and GWAS82. Finally, we note that population mismatch will generally drive underestimation of enrichment in S-LDSC and false negatives in GWAS.82. Finally, we note that population mismatch will generally drive underestimation of enrichment in S-LDSC and false negatives in GWAS.

Moving forward, we think there are three main challenges and opportunities to leverage m6A QTLs to study disease genetics. First, more work needs to be done to characterize the possible mechanisms of how m6A QTLs influence phenotypes. Second, eQTLs or sQTLs are often cell type- and condition-specific83,84. For m6A, recent studies suggest that its effects on decay or translation are probably strongest in cells undergoing differentiation85,86 or stimulation87. Thus, a major future direction is to map m6A QTLs under various disease-related cellular and physiological contexts. Third, recent work has shown that chromosome-associated regulatory RNA (carRNA) m6A methylation regulates transcription88. QTL studies of carRNA m6A may reveal new insights into mammalian transcriptional regulation.

Online content
Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41588-020-0644-z.

Received: 13 March 2019; Accepted: 11 May 2020; Published online: 29 June 2020

References
1. Fu, Y., Dominissini, D., Rechavi, G. & He, C. Gene expression regulation mediated through reversible m6A RNA methylation. Nat. Rev. Genet. 15, 293–306 (2014).
2. Roundtree, I. A., Evans, M. E., Pan, T. & He, C. Dynamic RNA modifications in gene expression regulation. Cell 169, 1187–1200 (2017).
3. Xiao, W. et al. Nuclear m6A reader YTHDC1 regulates mRNA splicing. Mol. Cell 61, 507–519 (2016).
4. Kasowitz, S. D. et al. Nuclear m6A reader YTHDC1 regulates alternative polyadenylation and splicing during mouse oocyte development. PLoS Genet. 14, e1007412 (2018).
5. Louloudi, A., Ntini, E., Conrad, T. & Orom, U. A. Transient N-6-methyladenosine transcriptome sequencing reveals a regulatory role of m6A in splicing efficiency. Cell Rep. 23, 3429–3437 (2018).
6. Roundtree, I. A. et al. YTHDC1 mediates nuclear export of N6-methyladenosine methylated mRNAs. elf6 6, e31311 (2017).
7. Wang, X. et al. N6-methyladenosine-dependent regulation of messenger RNA stability. Nature 565, 117–120 (2014).
8. Wang, X. et al. N6-methyladenosine modules messenger RNA translation efficiency. Cell 161, 1388–1399 (2015).
9. Zhou, J. et al. Dynamic m6A mRNA methylation directs translational control of heat shock response. Nature 526, 591–594 (2015).
10. Shi, H. et al. m6A facilitates hippocampus-dependent learning and memory through YTHDF1. Nature 563, 249–253 (2018).
11. Li, A. et al. Cytoplasmic m6A reader YTHDF3 promotes mRNA translation. Cell Res. 27, 444–447 (2017).
12. Shi, H. et al. YTHDF3 facilitates translation and decay of N6-methyladenosine-modified RNA. Mol. Cell 71, 973–985.e5 (2018).
13. Liu, J. et al. A METTL3–METTL14 complex mediates mammalian nuclear RNA N6-adenosine methylation. Nat. Chem. Biol. 10, 93–95 (2013).
14. Wang, P., Dostádér, K. A. & Nam, Y. Structural basis for cooperative function of Mettl3 and Mettl4 methyltransferases. Mol. Cell 63, 306–317 (2016).
15. Zheng, G. et al. ALKBH5 is a mammalian RNA demethylase that impacts RNA metabolism and mouse fertility. Mol. Cell 49, 18–29 (2013).
16. Jia, G. et al. N6-methyladenosine in nuclear RNA is a major substrate of the obesity-associated FTO. Nat. Chem. Biol. 7, 885–887 (2011).
17. Wei, J. et al. Differential m6A, m6Am, and m1A demethylation mediated by FTO in the cell nucleus and cytoplasm. Mol. Cell 71, 973–985.e5 (2018).
18. Frye, M., Harada, B. T., Behm, M. & He, C. RNA modifications module gene expression during development. Science 361, 1346–1349 (2018).
19. Huang, H. et al. Recognition of RNA N6-methyladenosine by IGF2BP proteins enhances mRNA stability and translation. Nat. Cell Biol. 20, 285–295 (2018).
20. Edupuganti, R. R. et al. N6-methyladenosine (m6A) recruits and repels proteins to regulate mRNA homeostasis. Nat. Struct. Mol. Biol. 24, 870–878 (2017).
21. Liu, J. et al. m6A mRNA methylation regulates AKT activity to promote the proliferation and tumorigenicity of endometrial cancer. Nat. Cell Biol. 20, 1074–1083 (2018).
22. Deng, X. et al. RNA N6-methyladenosine modification in cancers: current status and perspectives. Cell Res. 28, 507–517 (2018).
23. Barbieri, I. et al. Promoter-bound METTL3 maintains myeloid leukemias by m6A-dependent translation control. Nature 552, 126–131 (2017).
24. Vu, L. P. et al. The N6-methyladenosine (m6A)-forming enzyme METTL3 controls myeloid differentiation of normal hematopoietic and leukemia cells. Nat. Med. 23, 1369–1376 (2017).
25. Li, Z. et al. FTO plays an oncogenic role in acute myeloid leukemia as a N6-methyladenosine RNA demethylase. Cancer Cell 31, 127–141 (2017).
26. Su, R. et al. R2HG exhibits anti-tumor activity by targeting FTO/m6A/mRNA axis. Cell Rep. 17, 90–105.e23 (2018).
27. Zhao, S. et al. Detailed modeling of positive selection improves detection of cancer driver genes. Nat. Commun. 10, 3399 (2019).
28. Banovich, N. E. et al. Methylation QTLs are associated with coordinated changes in transcription factor binding. Cell Rep. 27, 315–328 (2017).
29. Kayser, J. et al. The effect of regulatory variation from RNA to protein. Science 347, 285–295 (2015).
30. Ledbetter, S. & Ledbetter, D. Genomic imbalance in human cancer. Nat. Rev. Genet. 13, 444–452 (2012).
31. Pare, P. et al. Genetic determinants of co-accessible chromatin regions in activated T cells across humans. Nat. Genet. 50, 1140–1150 (2018).
32. Grubert, F. et al. Genetic control of chromatin states in humans involves local and distal chromosomal interactions. Cell 162, 1051–1065 (2015).
33. Pali, A. A. et al. The contribution of DNA methylation to low幫refmtic variation trait loci to inter-individual variation in steady-state gene expression levels. PLoS Genet. 8, e1003000 (2012).
34. Chen, L. et al. Genetic drivers of epigenetic and transcriptional variation in human immune cells. Cell 167, 1398–1414.e24 (2016).
35. Pickrell, J. K. et al. Understanding mechanisms underlying human gene expression variation with RNA sequencing. Nature 464, 768–772 (2010).
36. Hormozdian, F. et al. Leveraging molecular quantitative trait loci to understand the genetic architecture of diseases and complex traits. Nat. Genet. 50, 1041–1047 (2018).
37. Lee, M. N. et al. Common genetic variants modulate pathogen-sensing responses in human dendritic cells. Science 343, 1246980 (2014).
38. Nicolae, D. L. et al. Trait-associated SNPs are more likely to be eQTLs: annotation to enhance discovery from GWAS. PLoS Genet. 6, e1000888 (2010).

39. Wén, X., Pique-Regí, R. & Luca, F. Integrating molecular QTL data into genome-wide genetic association analysis: probabilistic assessment of enrichment and colocalization. PLoS Genet. 13, e1006646 (2017).

40. Li, Y. I. et al. RNA splicing is a primary link between genetic variation and disease. Science 352, 600–604 (2016).

41. Chua, S. et al. Limited statistical evidence for shared genetic effects of eQTLs and autoimmune-disease-associated loci in three major immune-cell types. Nat. Genet. 49, 600–605 (2017).

42. Yao, D. W., O’Connor, L. J., Price, A. L. & Gusev, A. Quantifying genetic effects on disease mediated by assayed gene expression levels. Nat. Genet. 52, 626–633 (2020).

43. Takata, A., Matsumoto, N. & Kato, T. Genome-wide identification of RNA-binding protein binding sites with enhanced CLIP (eCLIP). Proc. Natl. Acad. Sci. USA 100, 9440–9445 (2003).

44. Ongen, H., Buil, A., Brown, A. A., Dermitzakis, E. T. & Delaneau, O. Fast and efficient QTL mapper for thousands of molecular phenotypes. Bioinformatics 32, 1479–1485 (2016).

45. Stoeyn, J. D. & Tibshirani, R. Statistical significance for genomewide studies. Proc. Natl. Acad. Sci. USA 100, 1635–1636 (2012).

46. Wiel, C. et al. The SMAD2/3 interactome reveals that TGF-β stimulates zebrafish maternal-to-zygotic transition. Cell Stem Cell 18, 1494–1505 (2019).

47. Roy, R. et al. hnRNPA1 couples nuclear export and translation of specific mRNAs downstream of FGFR/S6K2 signalling. Nucleic Acids Res. 42, 12483–12497 (2014).

48. Liu, N. et al. N6-methyladenosine-dependent RNA structural switches regulate RNA–protein interactions. Nature 518, 560–564 (2015).

49. Manning, K. S. & Cooper, T. A. The roles of RNA processing in translating genotype to phenotype. Nat. Rev. Mol. Cell Biol. 18, 102–114 (2017).

50. Shih, H. Localizing components of shared transethnic genetic architecture of complex traits from GWAS summary data. Am. J. Hum. Genet. 106, 803–817 (2020).

51. Mogil, L. S. et al. Genetic architecture of gene expression traits across diverse populations. PLoS Genet. 14, e1007586 (2018).

52. Das, S. & Krainer, A. R. Emerging functions of SRSF1, splicing factor and oncoprotein, in RNA metabolism and cancer. Mol. Cancer Res. 12, 1195–1204 (2014).

53. Bertero, A. et al. The SMAD2/3 interactome reveals that TGFβ controls m6A RNA methylation in pluripotency. Nature 555, 256–259 (2018).

54. Slobodin, B. et al. Transcriptom-wide association study identifies novel targets of NMD-dependent RNA degradation. Cell 169, 325–337.e12 (2017).

55. Aguilo, F. et al. Coordination of m6A mRNA methylation and gene transcription by ZFP217 regulates pluripotency and reprogramming. Cell Stem Cell 17, 689–704 (2015).

56. Dunham, I. et al. An integrated encyclopedia of DNA elements in the human genome. Nature 489, 73–77 (2012).

57. Huang, H. et al. Histone H3 trimethylation at lysine 36 guides m6A RNA modification co-translationally. Nature 567, 414–419 (2019).

58. Lee, J. et al. Effective breast cancer combination therapy targeting BACH1 and mitochondrial metabolism. Nature 568, 254–258 (2019).

59. Wiel, C. et al. BACH1 stabilization by antioxidants stimulates lung cancer metastasis. Cell 178, 330–345.e22 (2019).

60. Wén, X. et al. Partitioning heritability by functional annotation using genome-wide association summary statistics. Nat. Genet. 47, 1228–1235 (2015).

61. Hansson, G. K. Inflammation, atherosclerosis, and coronary artery disease. N. Engl. J. Med. 352, 1685–1695 (2005).

62. Nath, A. F. et al. Multivariate genome-wide association analysis of a cytokines network reveals variants with widespread immune, haematological, and cardiometabolic pleiotropy. Am. J. Hum. Genet. 105, 1076–1090 (2019).

63. Stolk, L. et al. Meta-analyses identify 13 loci associated with age at menopause and highlight DNA repair and immune pathways. Nat. Genet. 44, 260–268 (2012).

64. Li, H.-B. et al. m6A mRNA methylation controls T cell homeostasis by targeting the IL-7/STAT5/SOCS pathways. Nature 548, 338–342 (2017).

65. Zheng, Q., Hou, J., Zhou, Y., Li, Z. & Cao, X. The RNA helicase DDX46 inhibits innate immunity by entrapping m6A-demethylated antiviral transcripts in the nucleus. Nat. Immunol. 18, 1094–1103 (2017).

66. Shih, H. Localizing components of shared transethnic genetic architecture of complex traits from GWAS summary data. Am. J. Hum. Genet. 106, 803–817 (2020).
**Methods**

**Cell culture.** Human LCLs of 60 YRI individuals were purchased from the Coriell Institute. These 60 individuals were chosen by the availability of other molecular QTL data in previous studies \(^{34,46}\). On receiving them, cells were split into flasks as technical replicates and processed independently thereafter. Cells were cultured and propagated in Roswell Park Memorial Institute 1640 medium with 15% FCS at 37°C and 5% CO₂, until collection.

**RNA extraction and m6A-seq.** Cells were collected by 1,000 g centrifugation. Total RNA was extracted using QiAamp RNA extraction kit (catalog no. R2072; Zymo Research) according to the manufacturer’s instructions. The RNA was further purified with the Dynabeads mRNA DIRECT purification kit (catalog no. 61011; Thermo Fisher Scientific). mRNA was adjusted to 15 ng µl⁻¹ in 100 µl and fragmented using the Bioruptor Ultrasonicator (Diagenode) with 30 s on/off for 30 cycles. Approximately 1 µg fragmented mRNA was saved as input sample and approximately 1,450 ng was subjected to m6A immunoprecipitation with the EpiMark N⁺-Methyladenosine Enrichment Kit (catalog no. E1610S; New England Biolabs). To minimize the variation due to the immunoprecipitation experiment, which is often a great source of technical noise in immunoprecipitation-based sequencing, m6A immunoprecipitation was performed by a robot (KingFisher Duo Prime System, Thermo Fisher) for 12 samples at a time. Although a monoclonal antibody was used, we further controlled for much of the variation by pooling the antibody before aliquoting to each of the 60 samples. RNA eluted from m6A immunoprecipitation was cleaned using the RNA Clean and Concentrator kit (catalog no. R1012) and immunoprecipitation samples were then used to prepare the libraries with the KAPA mRNA Hyper Kit (catalog no. KK8841; Roche). A total of 240 libraries (duplicates per individual, each with an input and immunoprecipitation) were constructed in three batches. All libraries were sequenced by the HiSeq 4000 platform on SE50 mode at the sequencing core facility of the University of Chicago. For each batch of library constructed, all libraries (with distinct index) were pooled and sequenced at a lane together for 3–5 repetitive lanes. This study design balanced the lane effect on each batch of libraries. In summary, about 30 million reads were obtained for each library and reads from technical replicates were pooled to result in 60 million reads for each input and immunoprecipitation sample per individual.

**m6A-seq data alignment.** For each dataset, the raw sequencing data were mapped to the hg19 reference genome by HISAT2 (v.2.1.0) with the parameter -k known-splice-site-infile <splice-file> extracted from Refseq hg19 GTF file > k 1. We used WASP (v.0.3.4)\(^{98}\) to control for alignment bias due to genetic variations. The BAM files obtained from the alignment were used as an input file for reads quantification.

**Joint m6A peak calling across samples.** The BAM files obtained from the alignment were used as an input file for reads quantification. Approximately 1,450 ng was subjected to m6A immunoprecipitation with the EpiMark N⁺-Methyladenosine Enrichment Kit (catalog no. E1610S; New England Biolabs). To minimize the variation due to the immunoprecipitation experiment, which is often a great source of technical noise in immunoprecipitation-based sequencing, m6A immunoprecipitation was performed by a robot (KingFisher Duo Prime System, Thermo Fisher) for 12 samples at a time. Although a monoclonal antibody was used, we further controlled for much of the variation by pooling the antibody before aliquoting to each of the 60 samples. RNA eluted from m6A immunoprecipitation was cleaned using the RNA Clean and Concentrator kit (catalog no. R1012) and immunoprecipitation samples were then used to prepare the libraries with the KAPA mRNA Hyper Kit (catalog no. KK8841; Roche). A total of 240 libraries (duplicates per individual, each with an input and immunoprecipitation) were constructed in three batches. All libraries were sequenced by the HiSeq 4000 platform on SE50 mode at the sequencing core facility of the University of Chicago. For each batch of library constructed, all libraries (with distinct index) were pooled and sequenced at a lane together for 3–5 repetitive lanes. This study design balanced the lane effect on each batch of libraries. In summary, about 30 million reads were obtained for each library and reads from technical replicates were pooled to result in 60 million reads for each input and immunoprecipitation sample per individual.

**m6A-seq data alignment.** For each dataset, the raw sequencing data were mapped to the hg19 reference genome by HISAT2 (v.2.1.0) with the parameter -k known-splice-site-infile <splice-file> extracted from Refseq hg19 GTF file > k 1. We used WASP (v.0.3.4)\(^{98}\) to control for alignment bias due to genetic variations. The BAM files obtained from the alignment were used as an input file for reads quantification.

**Joint m6A peak calling across samples.** Genes (concatenated exons) were first divided into 50-bp consecutive bins where the read counts of the input and immunoprecipitation sample were quantified. Second, we applied a two-tailed Fisher’s exact test to call peaks by significantly distinguishing the peak region from the input region. For each gene and an FDR 5% cutoff was used to call peaks. Third, to obtain a common set of peaks for all QTL pairs, we used motifbreakR (v.1.12.0)\(^{96}\) to find instances of m6A motifs overlapping with the RBP CLIP-seq peaks. For the RBP CLIP-seq peaks, we intersected the peaks with both ePeak and eGene mapped, we computed the pairwise distances and LD between the lead eSNP and the eGene SNPs. To compare m6A QTLs with eQTLs or sQTLs (both at an FDR < 10%), we compared the overlap of ePeak-harboring genes and eGenes in the same cohort of YRI LCL samples\(^{35}\). Then, for those genes with both ePeak and eGene mapped, we computed the pairwise distances and LD between the lead eSNP and the eGene SNPs. To compare m6A QTLs with sQTLs (both at an FDR < 10%), we used the sQTL data from a larger cohort of Genome Reference of YRI LCL samples (n = 87) (ref. \(^{40}\)). We mapped intron clusters with at least one significant sQTL (denoted as eSplicing intron clusters) and compared the overlap of ePeak-harboring genes and genes containing the eSplicing intron clusters. For genes with multiple ePeaks and/or multiple eSplicing intron clusters, we computed the pairwise distances and LD between all peaks within the lead ePeak SNPs and eSplicing SNPs.

**Functional annotations of m6A QTLs.** Our functional annotations included the m6A consensus motif (RRACH) in m6A peaks, RBP CLIP-seq peaks in KS62 and HepG2 cells from ENCODE\(^{39}\), transcription factor and histone modifications ChIP–seq peaks in LCLs from ENCODE\(^{39}\), experimentally determined riboSNitch\(^{51}\) and predicted miRNA expression in human and mouse cells from TargetScan (downloaded from the TargetScanHuman 7.0 database at http://www.targetscan.org/cgi-bin/targetscan_data_download.vert72.cgi\(^{41}\)).

To annotate m6A peaks by the m6A consensus motif (RRACH) in m6A peaks, we used motifbreakR (v.1.12.0)\(^{96}\) to find instances of m6A motifs overlapping with SNPs. We then intersected these motif matches with the joint peaks to obtain motifs in m6A peaks. For the RBP CLIP-seq peaks, we intersected the peaks of the two replicates and from the two cell lines to obtain a peak set that was consistent across replicates and cell lines. These peaks shared across cell lines are more likely to be functional in LCLs than those in single-cell lines. To define the ChIP–seq peaks for transcription factors and histone markers, we chose peaks that were ‘optimal irreproducible discovery rate peaks’ as defined by the ENCODE processing pipeline. We used the miRNA binding sites predicted by TargetScan\(^{39}\), limited to the sites targeted by the miRNA expressed in LCLs. miRNA expression data were obtained from the miRNA sequencing data of LCLs samples from Geuvadis\(^{36}\). We defined miRNA expressed in LCLs by requiring the median read count across individual samples to be at least five.

**Enrichment of functional annotation in m6A QTLs.** We took the independent SNPs from the fine-mapped m6A QTLs (see Fine-mapping m6A QTLs, eQTLs and sQTLs) by choosing SNPs with the maximum PIP per credible set. We then compared the number of QTLs versus the number of random control SNPs (see Pairwise eQTL and sQTL associations). To generate the control SNP set, we used a modified version of SNPpan\(^{2}\) to sample 100 sets of SNPs that matched the allele frequencies, number of SNPs in LD, distances to the nearest genes and gene density, as well as annotations about SNP locations relative to genes (5’-UTR, CDS, 3’-UTR, intron and intergenic regions). We also used Torus (last change 29 June 2017)\(^{32}\) as an alternative method to assess the enrichment of functional annotation in m6A QTLs. Torus fits a logistic regression model to estimate an enrichment parameter for each annotation, which enables joint analysis of multiple annotations.
Learning motifs of RBPs from CLIP-seq data. For each RBP, we took the top 3,000 peaks as ranked by the peak strength of each replicate and retained the ones that were consistent in both replicates. We then extended 5 bp at both sides for peaks that were shorter than 10 bp. The sequence of the resulting peaks served as the target sequence for de novo motif search with Homer2. To generate matched background peaks for each RBP, we first generated a large set of random 70 bp peaks (the mean width of CLIP-seq peaks) on the transcribed region including both exons and introns. Then, we annotated the genomic locations (5′ UTR, CDS, Intron, 3′ UTR, intergenic regions) of the top CLIP-seq peaks and drew random peaks with matched distribution of genomic annotation. At least 1 motif with \( P < 1 \times 10^{-6} \) was obtained for 113 RBPs. For each RBP, the top 2 motifs sorted by \( P \) value were used for the motif correlation analysis of RBP binding (Fig. 2e).

To visualize the motifs, we used the R package Loglas (v.1.6.0)\(^\text{98}\) to generate sequence logo plots that highlight both nucleotide conservation and depletion.

**Fine-mapping m\(\text{A} \)** QTLs, eQTLs and sQTLs. Many significant m\(\text{A} \) QTLs are probably not causal variants but target the causal SNPs. To better identify independent associations and likely causal variants, we performed fine-mapping of m\(\text{A} \) QTLs using the state-of-art tool Susi\(\text{E} \)\(^\text{10}\). We used the standard version of Susi\(\text{E} \), which takes individual-level phenotype and genotype data. For the Susi\(\text{E} \) parameters, the maximal causal variant per region was set to 3 and estimate_prior\_variance = TRUE.

We first fine-mapped m\(\text{A} \) QTLs with a uniform prior inclusion probability and applied this version to most of our analyses, including enrichment analysis comparisons. m\(\text{A} \) QTLs lead to thousands of control SNPs by the exact test, m\(\text{A} \) QTL motif break analysis and partition of GWAS trait heritability analysis. We also performed another version of fine-mapping that leveraged RNA annotations including riboSNIch and RBP binding sites by using informative priors in Susi\(\text{E} \) fine-mapping. For example, an SNP close to a peak and located in an RBP binding site would have a higher prior probability of being a causal variant. The informative prior probability used in fine-mapping was derived from the functional annotation enrichment analysis using Torus with flag: -dump_prior. We used the RNA features informed fine-mapping results in the S-LDSC analysis of the enrichment of GWAS variants in m\(\text{A} \) QTNs (Extended Data Fig. 7).

Similarly, we fine-mapped eQTLs and sQTLs using Susi\(\text{E} \) on individual-level expression and splicing data in Guettard YRI LCL samples with uniform prior inclusion and the same parameter settings used to fine-map m\(\text{A} \) QTLs.

**Evaluating the role of RBP binding in regulating m\(\text{A} \) levels**. We checked whether the impact of genetic variants on RBP binding was correlated with the effect on m\(\text{A} \) levels, as measured by m\(\text{A} \) QTL effect size, in a directionally consistent manner. To assess the effect of genetic variants on RBP binding affinity, we used motifBreakR\(^\text{11}\) to map SNPs that break an RBP motif. A cutoff of \( P < 1 \times 10^{-6} \) was used to filter the motif match result in the parameter setting. To enhance signals, we used fine-mapped m\(\text{A} \) QTLs as described in Fine-mapping m\(\text{A} \) QTLs using the state-of-art tool Susi\(\text{E} \). We then applied 5 rounds of stringent washes using dialysis buffer (20 mM of Tris-HCl, pH 7.5, 50 mM of KCl, 100 mM of NaCl, 2 mM of MgCl\(\text{2} \), 10% glycerol, 0.2 mM of EDTA, 0.2 mM of DTT, 10 mM of sodium butyrate, 1x protease and phosphatase inhibitor cocktail). Nuclei were homogenized by 150 strokes of the ‘tight’ pestle of a 2-ml Dounce homogenizer to obtain the nuclei suspension. We then applied 5 rounds of stringent washes using dialysis buffer (20 mM of Tris-HCl, pH 7.5, 50 mM of KCl, 100 mM of NaCl, 2 mM of MgCl\(\text{2} \), 10% glycerol, 0.2 mM of EDTA, 0.2 mM of DTT, 10 mM of sodium butyrate, 1x protease and phosphatase inhibitor cocktail) followed by elution in 1x Laemml SDS sample buffer.

**Molecular QTLs from earlier studies.** We collected the QTL data of multiple molecular traits in YRI LCLs from earlier studies, including transcription rate, mRNA levels, decay and splicing, ribosome loading and protein levels. We used processed phenotype data and YRI genotypes from Li et al.\(^\text{10} \) (downloaded from http://equil.ucchicago.edu/jointCLL). To map co-QTLs for these molecular phenotype traits, we chose SNPs with no interaction between SNPs and molecular traits in YRI LCLs from earlier studies, including transcription rate, mRNA levels, decay and splicing, ribosome loading and protein levels. We used processed phenotype data and YRI genotypes from Li et al.\(^\text{10} \) (downloaded from http://equil.ucchicago.edu/jointCLL). To map co-QTLs for these molecular phenotype traits, we chose SNPs with no interaction between SNPs and molecular traits in YRI LCLs from earlier studies, including transcription rate, mRNA levels, decay and splicing, ribosome loading and protein levels. We used processed phenotype data and YRI genotypes from Li et al.\(^\text{10} \) (downloaded from http://equil.ucchicago.edu/jointCLL).

**APA QTLs.** Using the RNA-seq data (input) we generated, we predicted and quantified APA events based on sequence features at the 3′ UTR region using a modified version of DaPars (v.0.9.1)\(^\text{99}\) as described in Li et al.\(^\text{10} \). We found 7,617 putative APA sites. Using the ratio of distal to proximal poly(A) site usage as a quantitative phenotype, we tested its association with genotypes within a 100 kb range by FastQTL\(^\text{10} \). Seven principal components were included to maximize QTL discovery. At an SNP-level FDR < 10% (Storey’s q-value method), we obtained 3 APA QTLs.

**Estimation of QTL sharing between m\(\text{A} \) and other molecular phenotypes.** Following the procedure outlined by Li et al.\(^\text{10} \), we first ascertained m\(\text{A} \) QTLs at a given \( P \) threshold. We then limited our analysis to the lead SNP per mRNA locus; thus, the SNPs we included are largely independent. We next assessed the proportion of the non-null expectation \( (\pi) \) from the \( P \) values of the ascertained SNPs in another molecular phenotype, using Storey’s q-value method in the qvalue R package (v.2.14.1)\(^\text{10} \). Eighty percent bootstrap CIs for the \( \pi \) estimates were empirically encoded by resampling \( P \) values with replacement 100 times. Control SNPs were randomly sampled across the genome.

**QTL effect size correlation analysis stratified by RBP binding.** We ascertained lead m\(\text{A} \) QTLs and grouped m\(\text{A} \) peaks bound by different RBPs, requiring the genomic intervals of RBP binding sites to be entirely within m\(\text{A} \) peaks. (The results are similar if we require only a 1-bp overlap.) Eighty-two RBPs with at least 50 data points (SNP–peak pairs) were selected. In each group of m\(\text{A} \) peaks bound by an RBP, we assessed the correlations of effect sizes between m\(\text{A} \) QTLs and the QTLs of other molecular phenotypes. When computing the effect sizes of the molecular QTLs, we used the slope of the linear regression as a measure of effect size and did not regress out principal components since that could modify the effect size estimates\(^\text{10} \). Significant correlations of effect sizes between m\(\text{A} \) QTLs and the QTLs of other molecular phenotypes were selected at the Benjamini–Hochberg FDR = 10% threshold.

**Reanalysis of ribosome profiling data of METTL3 and YTHDF1 knockdown in Hela cells.** To validate our finding of the heterogeneous effect of m\(\text{A} \) on downstream molecular traits, we used translation efficiency as an example and reanalyzed the ribosome profiling data of METTL3-depleted (m\(\text{A} \) methyltransferase) and YTHDF1-depleted (m\(\text{A} \) reader) Hela cells from a published study\(^\text{10} \). The detailed description is shown in the Supplementary Note.

**Validating YBX3 function in repressing translation efficiency.** We depleted YBX3 using small interfering RNA (catalog no. ST00355P01; QIAGEN) in Hela cells and performed polysome profiling as described previously to assess the effect on translation efficiency. We quantified the transcript levels of selected targets in three polysome-bound fractions: monosome: polysome 1 and polysome 2 (as in Supplementary Data Fig. 7). YBX3 knockdown reduced the m\(\text{A} \) methyltransferase and monosome peak fractions for further gene-specific analysis. We selected five YBX3 target genes from the analysis of ribosome profiling data in m\(\text{A} \) depleted cells; all five YBX3 targets had m\(\text{A} \)
peaks overlapping with YBX3 CLIP-seq peaks and showed increased translation efficiency on m^A depletion. As negative controls, we selected three YTHDF1 targets, which showed decreased translation efficiency on m^A depletion. For each gene, we normalized the monosome, polysome 1 and polysome 2 fractions by the nonpolysome-bound fraction to obtain a translation efficiency estimate.

GWAS summary statistics. We downloaded the summary statistics of 45 phenotypes from the UK Biobank\(^1\), the Price laboratory\(^2\) and the GWAS Catalog. The GWAS traits and corresponding references are listed in Supplementary Table 6.

Testing the enrichment of GWAS signals in m^A QTLs. We extracted GWAS SNPs that also belonged to m^A QTLs (association P < 10\(^{-10}\)) and plotted the quantile–quantile plot of the GWAS P values for those SNPs. Similarly, we plotted GWAS SNPs that were also eQTLs or sQTLs (association P < 10\(^{-10}\)) for comparison. Genome-wide GWAS P values were also plotted as a control.

Heritability and enrichment analysis of GWAS summary statistics using S-LDSC. We partitioned the heritability of complex traits and estimated the heritability enrichment of m^A QTLs, eQTLs and sQTLs. S-LDSC partitions the heritability of genomic annotations using GWAS summary statistics and estimates enrichment as a ratio of the proportion of heritability explained by an annotation divided by the proportion of SNPs in that annotation.

We then constructed a probabilistic (continuous-valued) annotation using the PIP estimates from SuSiE fine-mapping with RNA features informed prior (Extended Data Fig. 7). We used the posterior probability (Fig. 2d–f). We applied S-LDSC (v.1.0.1) to our QTL-based annotations using separate models for each QTL annotation and a joint model with all three types of QTL annotations together. In our S-LDSC analysis, we adjusted for various baseline annotations of SNPs using a baseline LD model\(^5\), including gene annotations (coding, UTRs, intron, promoter), minor allele frequency bins and LD-related annotations. We did not include functional annotations such as enhancer markers in our baseline model because these annotations probably correlate with the QTL features of interest (for example, enhancers are enriched in eQTLs); including them would have biased our estimated enrichment.

To estimate the heritability explained by molecular QTLs, we constructed a binary annotation containing all SNPs at given SNP-level FDR cutoffs. We repeated the analysis on m^A QTLs, eQTLs and sQTLs at thresholds of 20, 10 and 5% FDR (Extended Data Fig. 8). Our conclusion is robust at the varying thresholds.

TWAS and heritability analysis of m^A peaks. The TWAS was performed using the FUSION (2018/04/08 version)\(^3\) software. We trained regression models using our own m^A data in LCLs, published RNA-seq data in YRI LCLs\(^4\) and splicing data\(^6\) using the Geuvadis YRI LCL data\(^7\), and the corresponding YRI genotype data. In the m^A TWAS analysis, we computed weights for each m^A peak using the LASSO and Elastic Net models as well as a regression model with the most significantly associated SNPs (using the R function FUSION:compute_weights.R with the parameter --models lassoenet.top1). The best-performing model in cross-validation was selected for each peak to perform imputation. We used a 100-kb cis-window and restricted genotypes to the set of markers in the LD reference panel (1000 Genomes European samples) provided on the TWAS/ FUSION website (http://gusevlab.org/projects/fusion/), since we used the LD reference data for the GWAS-m^A association analysis; 19,130 m^A peaks had estimated heritability (Extended Data Fig. 2a). We obtained trained weights for 918 peaks with estimated heritability significantly greater than 0 (with the default parameter hsq P < 0.01). We then performed imputation of genetically determined m^A levels and estimated GWAS m^A associations. We selected genome-wide significant m^A peaks/genes at a Bonferroni-corrected P < 0.05. Similarly, we built a prediction model of gene expression as well as splicing (introns with missing values were ignored) and estimated GWAS-gene expression as well as GWAS-splicing associations using FUSION.

Colocalization of m^A QTLs and GWAS associations. Our colocalization analysis was performed using the approximate Bayes factor test implemented in the Coloc (v.3.2.1)\(^8\) package, which has been incorporated in the TWAS/FUSION pipeline. Coloc computes five posterior probabilities (PP0, PP1, PP2, PP3 and PP4), each corresponding to a hypothesis: H0, no association with either trait; H1, association with trait 1 but not with trait 2; H2, association with trait 2 but not with trait 1; H3, association with trait 1 and trait 2, two independent SNPs; H4, association with trait 1 and trait 2, one shared SNP. We ran Coloc incorporated in the FUSION pipeline with the default parameters for TWAS-significant associations (using the R function Fusion.assoc_test.R in FUSION with -coloc P_flag) and used PP4 to assess evidence of colocalization. We visualized the colocalization of m^A QTLs and GWAS associations using the LocusCompareR (v.1.0.0) package (https://github.com/boxianglu/locuscompare).

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability
The m^A profiles of the 60 YRI samples generated in this study have been deposited with the Gene Expression Omnibus repository under accession no. GSE125377. The summary statistics data and imputed genotype data are available at https://doi.org/10.5281/zenodo.3870952. The source data for Fig. 3e can be found in the Supplementary Information.

Code availability
The code used for m^A QTL data processing and analysis are available at https://scottzijiezheng.github.io/m6AQTL_reproducibleDocument/index.html. Our method for joint peak calling is implemented as the R package MeriRtools and is freely available at https://github.com/scottzijiezheng/MeriRtools.

References
89. van de Geijn, B., McVicker, G., Gilad, Y. & Pritchard, J. K. WASP: allele-specific software for robust molecular quantitative trait locus discovery. Nature Methods 12, 1061–1063 (2015).
90. Heinz, S. et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol. Cell 38, 576–589 (2010).
91. Cui, X. et al. Guitar: an R/Bioconductor package for gene annotation guided transcriptomic analysis of RNA-related genomic features. BioMed Res. Int. 2016, 8367534 (2016).
92. Auton, A. et al. A global reference for human genetic variation. Nature 526, 68–74 (2015).
93. Howie, B., Fuchsberger, C., Stephens, M., Marchini, J. & Abecasis, G. R. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. Nat. Genet. 44, 955–959 (2012).
94. Howie, B., Marchini, J. & Stephens, M. Genotype imputation with thousands of genomes. G3 (Bethesda) 1, 457–470 (2011).
95. Yu, G., Wang, L.-G. & He, Q. V. ChiPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. Bioinformatics 31, 2382–2383 (2015).
96. Coetze, S. G., Coetze, G. A. & Hazelett, D. J. motifbreakR: an R/Bioconductor package for predicting variant effects at transcription factor binding sites. Bioinformatics 31, 3847–3849 (2015).
97. Lappalainen, T. et al. Transcriptome and genome sequencing uncovers functional variation in humans. Nature 501, 506–511 (2013).
98. Dey, K. K., Xie, D. & Stephens, M. A new sequence logo plot to highlight enrichment and deletion. BMC Bioinformatics 19, 473 (2018).
99. Xia, Z. et al. Dynamic analyses of alternative polyadenylation from RNA-seq reveal a 3′-UTR landscape across seven tumour types. Nat. Commun. 5, 5274 (2014).
100. Sudlow, C. et al. UK Biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. PLoS Med. 12, e1001779 (2015).
101. Li, Y. I. et al. Annotation-free quantification of RNA splicing using LeafCutter. Nat. Genet. 50, 151–158 (2018).

Acknowledgements
We thank Y. Gilad, Y.I. Li, M. Chen and L. Barreiro for helpful discussions, and X. Wen for advice on computational analysis. The data on coronary artery disease have been contributed by CARDS6GRAMplus4CD investigators and have been downloaded from http://www.cardioangiplusid.org/. C.H. acknowledges support from National Institutes of Health (NIH) grant no. R11H008935. X.H. acknowledges support from NIH grant no. R01MH110531. M.S. acknowledges support from NIH grant no. HG002585.

Author contributions
Z. Zhang, K.L., M.S., X.H. and C.H. designed the study. Z. Zhang, K.L., M.S., X.H. and C.H. conducted and supervised the experiments. Z. Zhang, K.L., Y.Z., G.W. M. Qiao, Z.L., M.S. and X.H. conducted and supervised the analyses. Z. Zhang, K.L., L.S., J.M., M.S., X.H. and C.H. wrote the paper.

Competing interests
C.H. is a founder and scientific advisory board member of Accent Therapeutics and a shareholder of Epican Genetech.

Additional information
Extended data is available for this paper at https://doi.org/10.1038/s41588-020-0644-z. Supplementary information is available for this paper at https://doi.org/10.1038/s41588-020-0644-z.

Correspondence and requests for materials should be addressed to M.S., X.H. or C.H.

Reprints and permissions information is available at www.nature.com/reprints.
Extended Data Fig. 1 | See next page for caption.
Extended Data Fig. 1 | Joint m^6A peak calling and QTL mapping. a, Distribution of merged m^6A peak length. Dash line marks the mean peak width. 
b, Distribution of all m^6A peaks vs. ePeaks on a meta-gene. c, Proportion of all m^6A peaks vs. ePeaks in each genomic annotation. d, m^6A motif learned by Homer2, and visualized using EDlogo package. e, Spatial distribution of m^6A-QTLs illustrated by density plot of SNP to peak distances of m^6A-QTL with nominal P-value < 1x10^-4 in a 2 Mb window surrounding m^6A peaks. We also showed the significance by the -log_{10} P-value of the association tests in the blue dots. f, Volcano plot of overall statistics of m^6A-QTLs with peak-level FDR < 10% (ePeaks). g, Distribution of the number of causal effects of ePeaks (FDR < 10%) by SuSiE fine-mapping with uniform prior. We set SuSiE parameters L = 3 (assuming at most three causal effects) and coverage = 0.95 (95% coverage for credible sets).
Extended Data Fig. 2 | See next page for caption.
Extended Data Fig. 2 | Heritability of m^6A peaks and independence of m^6A-QTLs, eQTL and sQTLs. a, Distribution of estimated heritability of the 19,130 peaks included in the TWAS analysis, in which 918 peaks had estimated heritability significantly greater than 0 (minimum heritability P-value of 0.01). b, Distribution of estimated heritability of ePeaks (n = 822 peaks). c, Enrichment (log, odds ratio) of m^6A-QTLs in gene annotations. d, Distribution of the LD between the lead ePeak SNP and the eGene SNP in genes that have both ePeak and eGene mapped. e, Overlap between ePeak-harboring genes and eSplicing-harboring (splicing event with at least one significant sQTL) gene (both at FDR <10%) mapped in YRI LCL samples. f, Distribution of the distance between the lead ePeak SNP and the eSplicing SNP in genes that have both ePeak and eSplicing mapped. g, Distribution of the LD between the lead ePeak SNP and eSplicing SNP in genes that have both ePeak and eSplicing mapped.
Extended Data Fig. 3 | See next page for caption.
Extended Data Fig. 3 | Contribution of RNA features and transcriptional features to m\(^6\)A variation. a, Enrichment of m\(^6\)A-QTLs in RNA related features by Torus. Error bars represent the 95% confidence intervals. b, Enrichment of m\(^6\)A-QTLs in the binding sites of RNA polymerase2 subunit A (POLR2A), and phosphorylated POLR2A at two residues (S2 and S5) by Torus joint analysis of all annotations (upper panel), and enrichment of m\(^6\)A-QTLs in histone modifications from Torus joint analysis. Error bars indicate the 95% confidence intervals. c, Proportion of putative causal m\(^6\)A-QTNs in RNA features and transcription factor binding site annotations (see Methods). d-e, To confirm that transcription rate affects mRNA and protein level, we ascertained transcription rate QTLs (Txn-QTLs) and assessed the correlation between transcription rate (Txn)-QTL effect sizes (30 min and 60 min 4sU labelling, respectively) and eQTL effect size (panel d, n = 425 and 1,387 SNP-gene pairs), and protein-QTL effect sizes (panel e, n = 425 and 408 SNP-gene pairs). Correlation is computed using linear regression. Fitted lines and 95% confidence intervals are shown in blue lines and shades.
Extended Data Fig. 4 | Downstream effects of m^6A are context dependent. a, The number and fraction of m^6A-QTLs in chromatin related genomic regions (using the union of promoter and enhancer regions annotated by ChromHMM in GM12878 cell line), and in chromatin related eQTLs (eQTLs with nominal P-value < 0.05 and also in promoter and enhancer regions). b, High correlations of effect sizes between molecular QTLs along the cascade from transcription to translation. Correlation is computed using linear regression, in which fitted lines and 95% confidence intervals are shown in blue lines and shades. c, Log₂ fold change of translation efficiency of m^6A methylated transcripts in METTL3 knockdown vs. controls. d, Log₂ fold changes of translation efficiency upon YTHDF1 (m^6A reader protein) knockdown. Transcripts harboring YTHDF1-bound m^6A peaks are labeled in yellow and non-targets in blue.
Extended Data Fig. 5 | YBX3 mediates translation efficiency of m^6A modified transcripts. a, Sucrose gradient A260 absorbance profile from YBX3 knockdown and control Hela cells. The arrows indicate the fraction sampled for subsequent qPCR analysis of YBX3 target transcripts. This experiment is repeated 2 times. b. Translation efficiency of YBX3 targets in comparison with YTHDF1 targets. We accounted for mRNA level variation by dividing polysome-bound fraction by the non-polysome-bound fraction. Transcript levels are quantified using RT-qPCR. Three polysome-bound fractions, as indicated in panel a, are sampled from sucrose gradient fractionation. 2 technical replicates were measured to obtain the data. The lower and upper hinges correspond to the first and third quartiles. Horizontal line indicates median value, and whiskers correspond to the value no further than 1.5x inter-quartile range.
Extended Data Fig. 6 | Enrichment of GWAS signal in m^6A-QTLs. a, Quantile-quantile (QQ) plots of P-values from GWAS of selected traits. m^6A-QTLs, eQTLs and sQTLs are shown in comparison with genome wide SNPs. GWAS SNPs are binary annotated using m^6A-QTLs, eQTLs and sQTLs with P-value < 1x10^-4. b, Enrichment of GWAS trait heritability assessed by stratified LD-score regression (S-LDSC). Shown are the results of GWAS traits not reported in Fig. 5b. Posterior inclusion probability (PIPs) in this analysis are derived from SuSiE with default (uniform) priors. Error bars represent the 95% confidence intervals.
Extended Data Fig. 7 | Enrichment of complex trait heritability in m^6A-QTNs using RNA-features-informed priors. a, Enrichment of selected immune and blood GWAS trait heritability assessed by stratified LD-score regression (S-LDSC). PIPs of m^6A-QTLs are derived from SuSiE using RNA-features-informed priors. PIPs of eQTL and sQTL are based on uniform prior. Error bars represent 95% confidence intervals. b, Enrichment parameters of annotations that are used to derive RNA-features-informed priors (by Torus) for SuSiE fine-mapping. Error bars represent the 95% confidence intervals. c, Summary of GWAS traits heritability enrichment analysis using m^6A-QTL PIP (using RNA-feature informed priors) as annotation. The -log_{10} P-value is plotted against the enrichment of heritability. The dots are colored by disease category. The red dashed line shows FDR 5% threshold.
Extended Data Fig. 8 | Partitioning complex trait heritability by m^A-QTLs, eQTLs and sQTLs. Heritability is assessed by S-LDSC in which QTLs are binary annotated with varying SNP-level FDR thresholds of 5%, 10%, and 20%. Error bars represent standard errors.
Extended Data Fig. 9 | m⁶A-TWAS identifies putative risk genes in human complex traits. a, Number of significant m⁶A-TWAS genes in all 45 GWAS traits. Significance is defined by the Bonferroni corrected P-value 0.05. b, LocusCompare plot showing the scatter plot and aligned Manhattan plots of leukocyte count GWAS and m⁶A-QTL association signal at the DDX55 locus. c, Manhattan plot of GWAS association signals before and after conditioning on the TWAS-predicted m⁶A level (gray and blue dots, respectively) for the leukocyte count at the DDX55 locus.
Extended Data Fig. 10 | m^6A modification mediates the impact of genetic variation on human complex traits. Genetic variation exerts its impact on complex traits through various mechanisms. As one of these mechanisms, we propose that variation of m^6A modification may lead to variation of mRNA processing, including mRNA decay, splicing, APA, export and translation efficiency. These variations in turn may change protein levels and functions, and lead to phenotypic variations.
 Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

- n/a
- Confirmed

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
- Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F, t, r) with confidence intervals, effect sizes, degrees of freedom and P value noted
- Give P values as exact values whenever suitable.
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen’s d, Pearson’s r), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection
Sequencing data of m6A was collected by the University of Chicago Genomics Facility using the Illumina Hiseq4000 system. Raw data was processed with the alignment softwares: WASP (v0.3.4), Hisat2 (v2.1.0), Samtools (v1.9).

Data analysis
We have used the R package "MeRIPtools" to perform peak calling in current study, which is available at: https://github.com/scottzijiezhang/MeRIPtools. Other custom codes used for data analysis in current study are available at:https://scottzijiezhang.github.io/m6AQTL_reproducibleDocument/index.html.
Software used in data analysis include: Homer2 (v4.10.1), FastQTL (v2.0), DaPars (0.9.1), S-LDSC (1.0.1), FUSION (last change 2018/04/08); R packages: Guitar (v1.20.1), ChIPseeker (1.18.0), Logolas (1.6.0), SuSiE (0.6.1.03), MotifBreakR (1.12.0), Coloc (v3.2-1).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The m6A profiles of the 60 YRI samples generated in this study has been deposited in GEO repository: accession GSE125377.
Raw data associated with Fig. 3e is in supplementary data.
Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

**Sample size**
We used a sample size of 60, which is determined from previous studies. There are a number of studies (Degner et al, Nature, 2012; McVicker et al, Science, 2013; Battle et al, Science 2014; Grubert et al, Cell, 2015 and Li et al, Science, 2017) using similar number of YRI LCL samples for other types of QTL study. These studies mapped plausible number of QTLs (roughly 1000 independent associations) using this sample size and yielded interesting insight into regulatory variation. Thus we think this sample size is likely sufficient to generate a plausible number of QTLs associated with m6A modifications.

**Data exclusions**
No data were excluded from the analysis.

**Replication**
We performed two technical replicates from cell culturing to sequencing. The two technical replicates for each individual showed relatively low variation compare to cross sample variation. We merged the data of two technical replicates in the data analysis stage.

**Randomization**
Randomization is not relevant in current study because when testing the association between genotypes with m6A level, the group by genotype gets re-shuffled at each locus. In another word, randomization was performed naturally in molecular traits QTL mapping.

**Blinding**
All samples used in current study are non-related individuals from Hapmap and 1k genome project and all samples are treated as the same. Grouping of samples is based on genotype at each locus and varies from locus to locus. Therefore, no blinding is needed.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

### Materials & experimental systems

| n/a | Involved in the study |
|-----|-----------------------|
| ☒ | Antibodies |
| ☒ | Eukaryotic cell lines |
| ☒ | Palaeontology |
| ☒ | Animals and other organisms |
| ☒ | Human research participants |
| ☒ | Clinical data |

### Methods

| n/a | Involved in the study |
|-----|-----------------------|
| ☒ | ChIP-seq |
| ☒ | Flow cytometry |
| ☒ | MRI-based neuroimaging |

### Antibodies

**Antibodies used**
- anti-METTL3, Cell signaling, ca#86132, E3F2A, Lot1 , 1:1000.
- anti-WTAP, Cell signaling, ca#41934, E70SB, Lot1 , 1:1000.
- anti-VIRMA, Proteintech, cat#25714-1-AP, Lot00025060, 1:300.
- anti-RBBP5, Cell signaling, ca#13171, D3I6P, Lot1, 1:1000.
- anti-BACH1, Bethyl, ca#A303-058A, Lot1, 1:2000, 5ug/mg lysate.
- anti-RBBP5, Bethyl, ca#A300-109A, Lot2, 5ug/mg lysate.
- VeriBlot HRP, abcam, ca#ab131366, Lot GR3276535-2 , 1:300.

**Validation**
- anti-METTL3 validated on human for WB, IP and ChIP.
- anti-WTAP validated on human for WB and IP.
- anti-VIRMA validated on human for WB, IHC, IF, CoIP, ELISA.
- anti-RBBP5 validated on human for WB, IP and ChIP.
- anti-BACH1 validated on human for WB and IP.
- anti-RBBP5 validated on human for WB and IP.

### Eukaryotic cell lines

**Policy information about** cell lines

**Cell line source(s)**
- We purchased the lymphoblastoid cell line of YRI origin from the Coriell Institute.

**Authentication**
- None of the cell line used was authenticated.
Mycoplasma contamination

We tested Mycoplasma contamination by aligning the RNA-seq reads to the Mycoplasma reference genome and found <0.01% reads aligned to Mycoplasma genome.

Commonly misidentified lines
(See ICLAC register)

No commonly misidentified cell lines were used.