Genetic drivers of m6A methylation in human brain, lung, heart and muscle

HHS Public Access

Author manuscript

Nat Genet. Author manuscript; available in PMC 2022 May 17.

Published in final edited form as:

Nat Genet. 2021 August ; 53(8): 1156–1165. doi:10.1038/s41588-021-00890-3.

Genetic drivers of m^{6}A methylation in human brain, lung, heart and muscle

Xushen Xiong[1,2,3,#], **Lei Hou**[1,2,#], **Yongjin Park**[1,2], **Benoit Molinie**[2], **GTEx Consortium**, **Richard I. Gregory**[3], **Manolis Kellis**[1,2,*]

[1.] Computer Science and Artificial Intelligence Lab, Massachusetts Institute of Technology, 32 Vassar St, Cambridge, Massachusetts 02139, USA.

[2.] The Broad Institute of Harvard and MIT, 415 Main Street, Cambridge, Massachusetts 02142, USA.

[3.] Stem Cell Program, Division of Hematology/Oncology, Boston Children's Hospital, Boston, MA, USA.

Abstract

The most prevalent post-transcriptional mRNA modification, *N*^{6}-Methyladenosine (m^{6}A), plays diverse RNA-regulatory roles, but its genetic control in human tissues remains uncharted. Here, we report 129 transcriptome-wide m^{6}A profiles, covering 91 individuals and 4 tissues (brain, lung, muscle and heart) from GTEx/eGTEx. We integrate these with inter-individual genetic and expression variation, revealing 8,843 tissue-specific and 469 tissue-shared m^{6}A-QTLs, which are modestly enriched in but mostly orthogonal to eQTLs. We integrate m^{6}A-QTLs with disease genetics, identifying 184 GWAS-colocalized m^{6}A-QTL loci, including brain m^{6}A-QTLs underlying neuroticism, depression, schizophrenia, and anxiety; lung m^{6}A-QTLs underlying expiratory flow and asthma; and heart/muscle m^{6}A-QTLs underlying coronary artery disease. Lastly, we predict novel m^{6}A regulators that show preferential binding in m^{6}A-QTLs, protein interactions with known m^{6}A regulators, and expression correlation with m^{6}A levels of their targets. Our results provide important insights and resources for understanding both *cis* and *trans* regulation of epitranscriptomic modifications, their inter-individual variation, and their roles in human disease.

Editor summary:

[*] Correspondence to: manoli@mit.edu.
[#]These authors contributed equally

Author Contributions

This study was designed by X.X., L.H., and M.K., and directed and coordinated by M.K.. B.M. performed the m^{6}A profiling, X.X. and L.H. performed the bioinformatic analysis with the help from Y.P., R.I.G. and under the supervision of M.K. All authors participated in the discussion of the project. X.X., L.H. and M.K. wrote the manuscript.

Competing interests

The authors declare no competing interests.

GTEx Consortium Authors

The Broad Institute of Harvard and MIT, 415 Main Street, Cambridge, Massachusetts 02142, USA
Kristin G. Ardlie & François Aguet

Peer review Information:

Nature Genetics thanks the anonymous reviewers for their contribution to the peer review of this work.

Analysis of 129 N6-methyladenosine (m6A) profiles across 4 tissues (brain, lung, muscle and heart) identifies 8,843 tissue-specific and 469 tissue-shared m6A quantitative trait loci (QTLs). 184 m6A-QTLs colocalize with GWAS signals.

Introduction

Genome-wide association studies (GWAS) identified >100,000 genetic loci associated with complex traits and diseases, but 93% do not affect protein-coding regions directly and remain largely uninterpreted, hindering the search for deciphering the molecular basis of human disease[1,2]. To bridge this gap between genetic variation and disease phenotypes, many studies profiled diverse molecular processes affected by non-coding variants, revealing thousands of genetic variants impacting gene expression (expression Quantitative Trait Loci, eQTLs), mRNA splicing (sQTLs), DNA methylation (meQTL), histone acetylation (haQTLs), RNA editing (edQTL), and protein levels (pQTLs)[3–11], and providing insights into increasing numbers of GWAS loci. However, the role of genetic variants affecting post-transcriptional mRNA modifications in human tissues and complex diseases remains uncharacterized.

More than 150 types of post-transcriptional RNA modifications regulate non-coding RNAs and protein-coding messenger RNA (mRNA) transcripts, similar to epigenomic modifications of DNA and histone modifications[12,13]. Among many such mRNA "epitranscriptomic" marks, N^6-methyladenosine (m^6A) is the most prevalent, and most frequently associated with development and disease[12,14–18]. m^6A modulates mRNA maturation, degradation, export, and translation efficiency, and is involved in spermatogenesis, stem cell differentiation, immune response, neurodevelopment, tumorigenesis, and other processes[14,16,17,19–21]. It is deposited by m^6A "writer" methyltransferases (METTL3, METTL14) and adaptors (WTAP, VIRMA, ZC3H13, HAKAI, RBM15, RBM15B)[12,22] at preferential sequence motifs[12], removed by m^6A "eraser" demethylases (FTO/ALKBH9, ALKBH5), and recognized by m^6A "readers" (YT521-B Homology, YTH family) and other direct or indirect binders (IGF2BP1–3, FMR1, HNRNPC/G, HNRNPA2B1), with different "readers" leading to decay (e.g. YTHDF2) or stabilization (IGF2BPs). However, many m^6A regulators remain uncharacterized[23–25].

Diverse diseases are linked to m^6A dysregulation, including glioblastoma, lung cancer, and acute myeloid leukemia[16,20], and m^6A loci are enriched in disease-associated variants indicating possibly causal roles[26,27]. Genetic loci controlling m^6A level as a quantitative trait (m^6A Quantitative Trait Loci, m^6A-QTLs) were identified in lymphoblastoid cell lines derived from 60 individuals of YRI ancestry (Yoruban from Nigeria)[28], but m^6A-QTLs remain uncharacterized in primary human tissues or in other genetic backgrounds.

Here, we report a collection of transcriptome-wide post-transcriptional modifications in human primary tissues, spanning 176 samples across 107 individuals in brain, lung, heart and muscle. We discover 9,312 m^6A-QTLs in brain, lung, heart, and muscle targeting 1,270 m^6A sites (g-m^6As). Most m^6A-QTLs are distinct from eQTLs, but the small subset of eQTL-colocalized m6A-QTLs are preferentially degradation-associated. m^6A-QTLs enrich

in disease-associated loci. Several m^{6}A-QTLs help elucidate GWAS loci, including brain m^{6}A-QTLs for psychiatric disorders, lung m^{6}A-QTLs for expiratory flow and asthma, and heart/muscle m^{6}A-QTLs for heart diseases and blood pressure loci. We predict new m^{6}A-regulating RNA-binding proteins (RBPs) with preferential binding in m^{6}A-QTL single-nucleotide polymorphisms (SNPs), experimentally confirmed interactions with known m^{6}A regulators, and significant expression correlation with m^{6}A levels of their targets.

Results

m^{6}A variation across tissues and individuals

We generated 176 transcriptome-wide m^{6}A profiles across 107 individuals in human brain, lung, heart and muscle (Supplementary Table 1), as part of the enhancing GTEx (eGTEx) consortium[29]. We selected individuals from the Genotype-Tissue Expression (GTEx) project[4,8,30], enabling us to directly compare our discovered m^{6}A-QTLs with mRNA expression eQTLs and other pre- and post-transcriptional processes.

We extended methylated-RNA immunoprecipitation sequencing (meRIP-seq)[18,31,32] protocols for m^{6}A identification and quantification, which we optimized for 400-fold-reduced starting material from GTEx compared to previous studies[31], and used matched RNA-seq as background (Fig. 1a). Our aggregated meRIP-seq signal showed strong enrichment surrounding the known GGACH m^{6}A deposition sequence motif (centered on the modified A base) (Extended Data Fig. 1a). After stringent quality controls using peak calling, positional enrichment, and consensus motif enrichment, we focused on a subset of 129 high-quality m^{6}A maps across 91 individuals that also have genetic information, including 53 brain, 12 heart, 32 muscle and 32 lung samples (Fig. 1a, Extended Data Fig. 1a–c).

Across all samples, we found >278,000 multiply detected m^{6}A sites (detected in ≥2 individuals), with an average of ~20,000 sites per sample (Extended Data Fig. 1d). Our profiling greatly expanded the previously known set of m^{6}A methylated sites, with 77% of our multiply detected sites in brain, 60% in heart sites, 58% in muscle, and 72% in lung not previously reported[33] (Extended Data Fig. 1e). Our m^{6}A sites were enriched for the known m^{6}A consensus motif, and showed enrichment near stop codons, consistent with previous studies[31,32,34] (Extended Data Fig. 1b–c). Previously undetected m^{6}A sites showed equally strong positional (Extended Data Fig. 1f) and motif enrichment (P adj = 5.7×10^{-26}, Fisher test) with previously detected sites, confirming their high quality.

Clustering all m^{6}A profiles by their global similarity, we found that the tissue profiled was the primary driver of variation (Fig. 1b), with brain samples most distinct, heart and muscle clustering first together and then with lung, consistent with tissue-specific biology[26,27,35] and with RNA-based tissue similarity (Extended Data Fig. 1g). Genes harboring tissue-specific m^{6}As showed tissue-relevant functional enrichments (Extended Data Fig. 1h), including synaptic and neuronal pathways in brain, and cardiomyopathy and muscle-related functions in muscle and heart. Noteworthy examples of tissue-specific methylation include brain-specific m^{6}A in *POU3F2* (Fig. 1c), where m^{6}A regulates glioblastoma[36], and lung-specific m^{6}A for *EGFR* (Fig. 1d), where m^{6}A promotes lung cancer[35]. While genes with

tissue-shared m^{6}As show little expression difference across tissues (Fig. 1e), tissue-specific m^{6}As do not primary stem from mRNA tissue-specific expression: for 72.8% of tissue-specific m^{6}A cases, the corresponding transcript was broadly expressed in the other tissues but m^{6}A modifications were not detected in those tissues (Fig. 1f), indicating that m^{6}A can lead to tissue-specific functions for otherwise broadly expressed transcripts.

m^{6}A genetic driver discovery and validation

We next sought to recognize m^{6}A-QTLs and their target m^{6}A sites (genetically driven m^{6}A sites, g-m^{6}As). Given the smaller sample size in heart, we called m^{6}A-QTLs in heart and muscle jointly to gain power, as heart and muscle were previously shown to share many eQTLs[8], and co-cluster for both GTEx mRNA data (Extended Data Fig. 1g) and our m6A data (Fig. 1b).

To recognize and remove “unwanted” global variation that confounds and hides subtle effects of *cis*-acting genetic variants[8,37], we used a Bayesian factor based tool (PEER[37]). We removed the top 7 factors in brain, top 5 in muscle/heart, and top 2 in lung, as inclusion of additional factors led to rapid decrease in the number of g-m^{6}As (Fig. 2a), and additional factors were highly correlated with each other, indicating signal saturation (Fig. 2b, Extended Data Fig. 2a). The removed factors correlated with batch, sex, age, ethnicity, as expected, and also immune disease, psychiatric traits, and medication history (Fig. 2c, Extended Data Fig. 2b), which are expected confounders acting in non-local ways whose removal decreases noise during QTL analyses.

We used FastQTL to perform a permutation-based m^{6}A-QTL search for each m^{6}A site, calculating the empirical *P* value for the SNP with the strongest genetic effect for each m^{6}A site. As m^{6}A modifications are deposited both co-transcriptionally and post-transcriptionally, we searched for m^{6}A-SNPs in promoter regions (where they can act at the DNA level), in introns (where they can act both at the DNA level and at the RNA level prior to splicing), and in exons (including UTRs and coding regions, where they act at the DNA level or RNA level, either pre- or post-splicing).

This resulted in 9,312 m^{6}A-QTLs targeting 1,058 genes across the four tissues, of which ~5,200 m^{6}A-QTLs act in brain (472 genes), ~3,100 in lung (401 genes), and ~2,100 in muscle/heart (279 genes), after correcting for multiple-SNP testing in each locus (empirical $P < 0.005$) (Fig. 2d, Extended Data Fig. 2c–d, Supplementary Table 2). The higher sample size in brain allowed us to also report ~1,300 higher-confidence brain m^{6}A-QTLs (targeting 94 genes) (Supplementary Table 3), using two rounds of multiple-testing correction for both multiple SNPs and multiple m^{6}A targets (see Methods). These m^{6}A-QTLs showed substantial genotype-driven differences in m^{6}A levels between individuals, and often affected m^{6}A levels of biologically important genes in relevant tissues, including neuronal-function genes *PADI2, MOBP, DLG1* in brain, lung-function and respiratory genes *CFLAR, SPTBN1, SLIT2* in lung, and skeletal/heart-muscle-function genes *PDIA6, SSR1, HSPA8* in muscle/heart (Fig. 2e, Extended Data Fig. 2c–d).

To validate our m^{6}A-QTL results (in our “discovery study”), we used m^{6}A levels measured in muscle, heart, and lung from two individuals[26] (“validation study”). As

two individuals are insufficient for m^{6}A-QTL discovery, we tested whether effect size direction matched between the two studies, focusing on N = 281 m^{6}A-QTL SNPs (5.6%) for which the two individuals from the validation study have different genotypes, and separating our m^{6}A-QTLs into positive-effect (minor allele shows significantly higher m^{6}A level), negative-effect (minor allele significantly lower), and no-effect (no significant m^{6}A difference). Despite using only two individuals, and despite potential population-specific effects (between our European-ancestry discovery study and the Asian-ancestry validation) that may reduce agreement, we found a consistent and significant agreement between our discovery study and the validation study (Fig. 2f): our positive-effect m^{6}A-QTLs showed 28% higher median m^{6}A level in the validation study ($P = 0.04$, Wilcoxon test), our negative-effect m^{6}A-QTLs showed 36% lower median m^{6}A levels in validation ($P = 5.5 \times 10^{-5}$, Wilcoxon test), and our no-effect SNPs showed only 1.007-fold difference between alleles (used as the baseline to compute the aforementioned *P* values).

Tissue specificity of m^{6}A-QTLs

Only 5.0% of m^{6}A-QTLs (6.8% of g-m^{6}As) were shared between any pair of tissues (Fig. 3a,b, Extended Data Fig. 3a), even though 46% of m^{6}A peaks were shared (Fig. 1f–g), with tissue-specific m^{6}A-QTLs enriched in distinct functional pathways (Extended Data Fig. 3b), including synaptic function and signalling in brain.

To evaluate whether m^{6}A-QTL specificity stemmed from biological factors or lack of discovery power, we tested whether the 8,800 tissue-specific m^{6}A-QTLs (~95% of all m^{6}A-QTLs) showed consistent directionality in the tissues where they were not discovered as m^{6}A-QTLs. We reasoned that if the m^{6}A-QTL effects were shared in those tissues, we would see common directionality, despite having limited power to detect them, with complete sharing resulting in 100% effect directionality agreement, and complete tissue specificity resulting in 50% effect directionality agreement expected by chance. Indeed, we found a gradual transition from 100% consistency to near-50% consistency (Fig. 3c,d), at decreasing *P* value thresholds for the second tissue: (i) the 469 shared m^{6}A-QTLs (5%) that were significant in both tissues showed 100% directionality consistency, providing strong confirmation of our results based on independent discovery in multiple tissues (red in Fig. 3d); (ii) an additional 592 m^{6}A-QTLs (6.4%) that were significant in one tissue ($P < 0.005$) and near-threshold *P* values in another tissue ($0.005 < P < 0.05$) showed 99% consistency, indicating that our current threshold for m^{6}A-QTL discovery is quite stringent, and that even looser thresholds for m^{6}A-QTL discovery would still result in additional meaningful m^{6}A-QTLs; (iii) another ~1,680 m^{6}A-QTLs (~18%) that were significant in one tissue and sub-threshold in the second tissue ($0.05 < P < 0.5$) showed 92% consistency, indicating increasing numbers of false positives at those thresholds; (iv) the remaining ~6,600 m^{6}A-QTLs (~71%) that were significant in one tissue and showed $P > 0.5$ in the second tissue showed only 60% consistency in m^{6}A genetic effect directionality (near the 50% expected by chance), indicating that the vast majority of calls would be false positives at such thresholds.

Leveraging the fact that our m^{6}A profiling was done in individuals from the GTEx cohort, we also directly compared the tissue sharing of eQTLs and the tissue sharing of m^{6}A-QTLs

in the exact same cohort of individuals, by subsampling the transcriptional data from GTEx to the specific individuals studied here, thus matching discovery power, allele frequencies, and potential *trans*-acting or environmental effects. In the most stringent eQTL threshold (two-step correction, FDR = 0.1), we found that 14% of eQTLs were independently discovered in multiple tissues (Extended Data Fig. 3c), substantially more than the 5% sharing found for m^{6}A-QTLs, suggesting that m^{6}A-QTLs are approximately 3-fold more tissue-specific than eQTLs for matched cohorts. Repeating our directionality consistency analysis for eQTLs across both the full GTEx cohort and a sample-size-matched subsampled GTEx dataset, we found that directionality consistency is robust to sample size (90% for full dataset vs. 89% for subsampled eQTLs, Extended Data Fig. 3d–e), and is substantially higher for eQTLs than for m^{6}A-QTLs (89% for eQTLs vs. 71% for m^{6}A-QTLs, compared to 50% expected by chance, Extended Data Fig. 3e, Fig. 3d), indicative of approximately twice as much eQTL tissue sharing than m^{6}A-QTL tissue sharing. We also used simulations with matching effect size, minor allele frequency, and sample size under the assumption of 100% tissue-sharing, and found much higher overlap and directionality consistency between tissues than observed in our measurements (38%−90% vs. ~5%, Extended Data Fig. 3f–i), indicating that the observed tissue specificity is not simply stemming from discovery power.

We next asked whether the tissue specificity of m^{6}A-QTLs stems from strong differences in gene expression levels. We first evaluated whether genes with tissue-specific m^{6}A-QTLs in a given tissue show higher relative expression levels than genes with tissue-shared m^{6}A-QTLs in the same tissue (Extended Data Fig. 3j), and found that in all cases there was no substantial difference in expression level between tissue-specific and tissue-shared m^{6}A-QTLs. We next evaluated whether genes with tissue-specific m^{6}A-QTLs in a given tissue show higher expression levels in that tissue than in tissues where they lack m^{6}A-QTLs (Extended Data Fig. 3k), and found that in 4 out of 6 pairwise comparisons, the differences in expression are not significant, and in the last two comparisons, gene expression levels are within 50% of each other. These results indicate that the observed m^{6}A-QTL tissue specificity is not simply stemming from differences in expression levels.

We next asked whether the primary tissue m^{6}A-QTLs discovered here were already captured in the m^{6}A-QTLs previously reported[28] in immortalized lymphoblastoid cell lines (LCLs), derived from peripheral B lymphocytes by Epstein-Barr virus (EBV) transformation in 60 Geuvadis YRI samples[30,38]. As a benchmark, we first evaluated the sharing of eQTLs from the four primary tissues profiled here in the full GTEx cohort (205 Brain, 515 Lung, 386 Heart, and 706 Muscle samples) and eQTLs from YRI LCLs in the full Geuvadis cohort (89 LCL samples)[30,38]. This comparison showed strong eQTL sharing between primary tissues in mostly EUR samples (R^2 = 0.66–0.80) but very low sharing with YRI LCLs (R^2 = 0.06–0.12) (Extended Data Fig. 3l), as expected given the substantially different biology of immortalized cell lines and primary tissues, compounded with potential differences between EUR and YRI ancestry groups, and indicating that the two capture different parts of biological diversity. Consistent with these eQTL results, we found our m^{6}A-QTLs in primarily EUR-ancestry tissues showed were largely distinct from m^{6}A-QTLs of YRI-ancestry immortalized cell lines, with only 0.8% shared m^{6}A-QTLs (Extended Data Fig. 3m), very low correlation (R^2 = 0.0007–0.021, Extended Data Fig. 3n), and only 51.25% directionality agreement (Extended Data Fig. 3o), which is close to the 50% expected by

chance ($P = 0.31$) and substantially lower than the 71% consistency found between tissues ($P < 2.2 \times 10^{-16}$). However, the 0.8% of m^{6}A-QTLs that were shared (N = 26) showed 100% consistent directionality (green dots in Extended Data Fig. 3n), providing a form of additional validation for the validity of both studies. These results highlight the importance of also profiling m^{6}A-QTLs in primary human tissues of multiple ancestry backgrounds, as they capture additional biological information not captured in immortalized cell lines from a single ancestry background.

m^{6}A-QTLs and eQTLs sometimes overlap but are mostly independent

As m^{6}A can increase mRNA stability (e.g. via IGF2BP[39]) or degradation (e.g. via YTHDF2[40]), m^{6}A-QTL effects may be positively or negatively correlated with m^{6}A target gene expression. Across all SNPs, m^{6}A-QTLs showed a mild but significant enrichment in GTEx matched-tissue eQTLs, across brain (1.5-fold, $P = 8.3 \times 10^{-7}$, prefrontal cortex BA9 eQTLs), muscle/heart (2-fold, $P = 0.0026$, muscle eQTLs; 1.6-fold, $P = 0.0050$, left ventricle eQTLs), and lung (1.6-fold, $P = 0.0030$, lung eQTLs) (Fig. 4a). Fold-enrichments were similar when exonic and intronic QTLs were considered separately, although significance levels were lower due to fewer data points (Extended Data Fig. 4a–b).

Among m^{6}A-QTLs co-localized with eQTLs in matched individuals (after PEER factor removal) (Extended Data Fig. 4c), 23% were positively correlated with gene expression, reflecting stability-increasing effects (Fig. 4b–c, teal points), and 77% were negatively correlated, reflecting degradation effects (Fig. 4b–c, red points). Repressive-effect m^{6}A-QTLs were significantly higher (OR = 2.5, $P = 2.3 \times 10^{-14}$, Fisher test) indicating more widespread degradation effects in the tissues studied here. Degradation-mediating g-m^{6}As preferentially localized in CDS and 3'-UTRs, and stabilization-mediating g-m^{6}As preferentially localized in lncRNAs and 5'-UTRs, although these differences were not significant due to small counts (Extended Data Fig. 4d).

Despite these enrichments, 88% of m^{6}A-QTLs showed no significant eQTL effect and 94% of eQTLs showed no significant m^{6}A-QTL effect in matched samples (no P value < 0.005, no effect size > 0.5, Fig. 4c–d), and m^{6}A-QTL target genes (gmGenes, for genetically controlled m^{6}A-levels genes) and eQTL target genes (eGenes) showed very little overlap (Extended Data Fig. 4e, $P = 0.21$), with 93% of gmGenes lacking eQTLs and 96% of eGenes lacking m^{6}A-QTLs, indicating that m^{6}A-QTL effects are not simply a consequence of changes in expression. Even across the full GTEx cohort for matching tissues, 79% of m^{6}A-QTLs do not show any eQTL effect (Extended Data Fig. 4f), and conversely 99% of eQTLs do not show any m^{6}A-QTL effect (Extended Data Fig. 4g). Even when a gmGene was also an eGene, their lead QTL SNPs differed greatly (~10-kb average distance, Fig. 4e), indicating that m^{6}A-QTLs and eQTLs constitute largely independent gene-regulatory paths.

m^{6}A-QTLs help interpret GWAS loci

Given the known roles of m^{6}A in multiple human diseases[16,17,20,21], including cancer, immune disease, and brain disorders, we next asked whether our discovered m^{6}A-QTLs can help shed light into potential mechanisms for currently unexplained genome-wide significant and sub-threshold GWAS loci, and possibly implicate m^{6}A function in additional disorders.

We found 179 m^{6}A-QTLs colocalized with 80 GWAS traits (Supplementary Table 4), with 88% (N = 157) of lead SNPs outside protein-coding exons, and 82% (N = 147) distinct from eQTLs in the full GTEx cohort for our tissues, thus providing new tissue-specific mechanistic hypotheses for these loci, including candidate causal variants, target genes, and tissues of action.

For brain, 71 m^{6}A-QTLs colocalize with GWAS variants. Neuroticism-associated rs12471193 is a brain m^{6}A-QTL for potassium channel gene KCNJ3, a neuronal excitability regulator implicated in multiple brain disorders[41] (Fig. 5a, Supplementary Table 4). Depression-associated rs1827603 is a brain m^{6}A-QTL for postsynaptic receptor GRM5[42] (Fig. 5a–b). Schizophrenia-associated rs7285557 is a brain m^{6}A-QTL for a brain-enriched lincRNA, LINC00634, down-regulated in schizophrenia[43]. Anxiety-associated rs1541627 is a brain m^{6}A-QTL for synaptic plasticity regulator ANKS1B[44].

For lung, 62 m^{6}A-QTLs colocalize with GWAS variants. Peak expiratory flow (PEF)-associated intronic rs682164 is a lung m^{6}A-QTL for retinoid-responsive CXXC5 that attenuates lung fibrosis in mice[45]. PEF-associated rs79966207 is a lung m^{6}A-QTL for Plexin-B2 (PLXNB2) acting in adult lung bronchial epithelium[46] (Fig. 5c, Extended Data Fig. 5a). Asthma-associated rs3194051 is a lung m^{6}A-QTL for immune-related IL-7 that contributes to atopic asthma, acting in bronchoalveolar lavage fluid, and regulating airway eosinophilia[47]. Forced expiratory volume (FEV1)-associated rs35956171 is a lung m^{6}A-QTL for alveolar fatty acid oxidation regulator CPT1A, implicated in acute lung injury[48] (Extended Data Fig. 5a, Supplementary Table 4).

In heart/muscle, 50 m^{6}A-QTLs colocalize with GWAS variants (Extended Data Fig. 5b). Coronary-artery-disease (CAD)-associated rs888298 is a muscle/heart m^{6}A-QTL targeting cardiac myocyte mitochondrial oxidation regulator WIPI1 involved in signaling and autophagy[49]. Heart-pulse-rate-associated rs6791834 is a muscle/heart m^{6}A-QTL for myocyte microtubule differentiation regulator MAP4 involved in heart development[50]. High-blood-pressure-associated rs56104944 is a muscle/heart m^{6}A-QTL for heat shock protein HSPA4 involved in cardiac hypertrophy and fibrosis[51] (Extended Data Fig. 5b, Supplementary Table 4).

Beyond GWAS SNPs, m^{6}A-QTLs helped interpret 3’-UTR and intronic SNPs from ClinVar (Supplementary Table 5). Congenital-cataract-associated rs13069079 is an intronic brain m^{6}A-QTL for eye segment morphology regulator FYCO1[52] (Extended Data Fig. 5c). Nemaline-myopathy and familial-restrictive-cardiomyopathy-associated rs605430 is a 5’-UTR muscle/heart m^{6}A-QTL for actin isoform ACTA1, essential for muscle contraction in sarcomeric thin filaments of skeletal muscle[53] (Extended Data Fig. 5d).

Tissue-specific m^{6}A-QTL-GWAS enrichments

We next assessed global tissue-specific m^{6}A-QTL enrichments for GWAS variants by heritability partitioning[54,55] after correcting for eQTLs, and found 27 significantly enriched traits ($P \leq 0.05$) (Fig. 5d), indicating m^{6}A-QTLs capture residual heritability beyond eQTLs. These results held even after correcting for coding regions, UTRs, promoters, and conserved

regions (Extended Data Fig. 5e), indicating they are not simply driven by fortuitous overlaps.

Brain m^{6}A-QTLs enriched for 6 traits, including ALS, tense mood, schizophrenia, and Alzheimer's disease (AD) (Fig. 5d), consistent m^{6}A roles in brain disorders[17,21]. Interestingly, AD GWAS variants enrich for GTEx eQTLs[56] and our m^{6}A-QTLs in bulk brain samples (composed primarily of neurons), while epigenomic enrichments implicate immune and microglia cells[57–59], indicating both neurons and immune cells may mediate AD genetic effects.

Lung m^{6}A-QTLs enriched for 19 traits, including both lung-related traits of forced vital capacity (FVC), forced expiratory volume in one second (FEV1), and asthma (Fig. 5d), and blood- and immune-related traits, that may reflect immune roles of lung tissue[60–62].

Muscle/heart m^{6}A-QTLs enriched for 8 traits, including hypertension, high blood pressure, and gout (Fig. 5d), reinforcing reports of m^{6}A regulators (FTO, METTL3, ALKBH5) on cardiovascular diseases[63], and suggesting potential interactions between *trans*-regulator effects and *cis*-m^{6}A-QTL effects.

Novel m^{6}A regulators prediction

As only a small number of m^{6}A regulators are known, accounting for only a small fraction of the observed complexity of m^{6}A dynamics[22], we next used our m^{6}A-QTLs to help reveal candidate novel m^{6}A regulators, by searching for RNA-binding-proteins (RBPs) that preferentially bind m^{6}A-QTL regions using CLIP-seq data for 171 RBPs across ~34M target sites[64].

We predicted 69 candidate m^{6}A regulators in all three tissues, of which 26 are shared by at least two tissues (Bonferroni-adj. $P = 0.05$, Fig. 6a, Extended Data Fig. 6a, Supplementary Table 6), including one m^{6}A writer adapter (RBM15), four known m^{6}A readers (YTHDF2, YTHDC1, FMR1, HNRNPC), and four previously proposed candidate m^{6}A readers using preferential binding of m^{6}A-modified oligonucleotides[23,24] (CPSF6, NUDT21, TARDBP, PRPF8), consistent with evidence that m^{6}A readers can impact m^{6}A level by protecting m^{6}A from demethylation by m^{6}A erasers[65,66].

Our candidates formed a tight protein-protein interactions (PPI) network with several known m^{6}A readers/writers/erasers, suggesting m^{6}A cofactor interactions[22] reminiscent of pre-transcriptional and splicing regulators. For example, DD3X3 showed experimentally validated interactions with known m^{6}A eraser ALKBH5 (Fig. 6b), with which it was shown to modulate mRNA demethylation[67]. An additional five RBPs (ATXN2, EFTUD2, UPF1, NCBP2, LARP4; Fig. 6b, dotted circles) showed multiple interactions with both m^{6}A writers (purple) and readers (green), indicating they may function as adapter proteins between them.

We also found several cases where allele-specific RBP binding[68] overlapped our m^{6}A-QTLs, including: SRSF1 and PRPF8 for brain m^{6}A-QTLs; NCBP2, TARDBP, UCHL5 and ZNF622 for lung m^{6}A-QTLs; and RBM15 for muscle/heart m^{6}A-QTLs (Supplementary Table 7).

Seven of our 26 candidates were also supported by significant correlations (across individuals, adjusted $q < 0.1$) between RBP expression levels (measured by GTEx RNA-seq) and m^{6}A level of RBP-bound m^{6}A-QTL targets (measured by our eGTEx m^{6}A levels) (Fig. 6a, ii). For example, known m^{6}A writer RBM15 and predicted regulator FIP1L1 (which directly interacts with known writer ZC3H13), and 5 additional RBPs (FMR1, NUDT21, STAU1, UCHL5, ZNF622) were all positively correlated with m^{6}A methylation level of their CLIP-inferred bound targets, consistent with contributions to m^{6}A deposition or blocking demethylation (Fig. 6a,c, Extended Data Fig. 6c–d). By contrast, TARDBP, CSTF2T, CPSF7, and GEMIN5 were negatively correlated with m^{6}A methylation level of their CLIP-inferred bound targets, consistent with contributions to m^{6}A demethylation or blocking deposition (Fig. 6a,d).

Discussion

Our study reports epitranscriptomic inter-individual variation in post-transcriptional m^{6}A mRNA modifications across multiple primary human tissues, and its integration with genetic variation, expression variation, disease-associated loci, and RBP binding, to recognize *cis*- and *trans*-acting drivers of mRNA modifications and their roles in human disease.

We generated transcriptome-wide post-transcriptional modification profiles in human, spanning 176 experiments across 107 individuals in brain, lung, heart and muscle, filtered to 129 high-quality experiments using stop-codon and m^{6}A-consensus-motif enrichment and other quality control metrics. We report 278k multiply detected m^{6}A sites, greatly expanding known m^{6}A-modified loci, and implicating ~43% of all protein-coding genes.

We defined m^{6}A-QTLs at both a stringent threshold for pinpointing individual m^{6}A-QTLs and a more inclusive threshold for recognizing functional properties of m^{6}A-QTLs, as biologically meaningful GWAS enrichments continue well past stringent thresholds to even nominally significant thresholds[69,70]. We show that m^{6}A-QTLs at both thresholds are highly reliable, with ~100% directionality consistency between tissues, and significant validation in an independent cohort. Even near-threshold and sub-threshold m^{6}A-QTLs showed >90% consistency, indicating that even looser thresholds may capture additional reliable and biologically meaningful m^{6}A-QTLs.

We found that m^{6}A-QTLs are approximately 3-fold more tissue-specific than eQTLs for matched cohorts, which may stem from tissue-specific co-transcriptional and post-transcriptional m^{6}A regulators[12,71], and selective pressures against variants with multi-tissue effects, similar to GWAS variants that primarily localize in tissue-specific enhancers rather than tissue-shared promoters. However, the ~470 tissue-shared m^{6}A-QTLs were >99% consistent in effect directionality between tissues, and even sub-threshold m^{6}A-QTLs were >90% consistent in directionality, indicating that m^{6}A *trans*-acting regulators rarely change direction of effect, even though their activity level for different targets may vary across tissues.

We provided several lines of evidence that the observed m^{6}A-QTL tissue-specificity is biological rather than technical, including: (a) using simulations to assess the m^{6}A-

QTL calling power, and confirming that our observed data are more tissue-specific than expected even at matching MAF, effect size, and sample size; (b) partitioning cross-tissue comparisons of m^{6}A-QTLs across different *P* value thresholds, and showing that "no-effect" m^{6}A-QTLs show near-random directionality, indicating that lack of m^{6}A-QTL effect is not stemming from threshold effects; and (c) showing that m^{6}A-QTLs are substantially more tissue-specific than eQTLs, by subsampling GTEx eQTLs to matched individual donors, thus controlling for sample size, discovery power, allele frequency, and potential *trans*-acting effects.

We note however that as mRNA expression measurements are more robust and biologically more stable, while m^{6}A may be more variable due to both biological and technical reasons, which may partly account for the observed increased tissue-specificity of m^{6}A-QTLs. Moreover, we expect the fraction of tissue-sharing to increase for m^{6}A-QTLs with increased discovery power at larger sample sizes, as previously observed for eQTLs[4,8,30]. However, our directionality consistency analysis indicates that the large majority of m^{6}A-QTLs are tissue-specific irrespective of discovery power, as GTEx eQTL directionality shows near-perfect agreement at ~20-fold reduced sample sizes (e.g. from 706 individuals to 38 for muscle eQTLs), thus enabling us to estimate the fraction of true m^{6}A-QTL effects even at much lower discovery thresholds.

Our results shed light on ~400 intronic and exonic eQTLs whose mechanism of action was not previously characterized, by showing that they act as tissue-specific m^{6}A-QTLs, which may mediate their effect on expression levels through mRNA degradation or stability, as previously shown for the YTHDF2 and IGFBP family regulators respectively. The two directionalities of effect showed uneven proportions among m^{6}A-QTLs that also impact expression: only one quarter of increased-m^{6}A alleles showed increased expression indicating roles in mRNA stability, while three quarters showed decreased expression indicating roles in mRNA degradation. In the vast majority of cases however, m^{6}A-QTLs acted through distinct sets of SNPs from eQTLs, thus expanding the set of genetic variants known to affect molecular phenotypes of protein-coding genes, and helping shed function on potential disease-associated variants.

Our results also revealed the widespread role of m^{6}A-QTLs in human disease. At the genome-wide level, we found that m^{6}A-QTLs were significantly enriched for disease-associated genetic loci and showed compelling tissue-specific enrichments, with psychiatric and neurodegenerative traits enriched in brain m^{6}A-QTLs, respiratory traits enriched in lung m^{6}A-QTLs, blood pressure traits enriched in muscle/heart m^{6}A-QTLs. At the single-locus level, we also found multiple tissue-specific examples of m^{6}A-QTLs affecting biologically meaningful target genes through colocalization analysis with GWAS SNPs, thus providing new insights and candidate mechanistic hypotheses for GWAS hits that were previously uncharacterized. While experimental validation of these hypotheses through CRISPR-Cas9 genome editing requires tissue systems and animal models of human disease that are not yet developed and will take years to complete for heart, muscle, lung, and brain, they can help guide future experiments and pre-clinical studies by expanding the diversity of mechanistic hypotheses underlying human disease genetics to include epitranscriptomic

molecular phenotypes, and they demonstrate the broad utility of our results for the field of human genetics and GWAS interpretation.

Our results also revealed 26 RNA-binding proteins as m^{6}A regulator candidates, greatly expanding the set of factors and co-factors currently implicated in m^{6}A regulation, which can help guide the systematic dissection of m^{6}A regulatory circuitry. All 26 RBPs showed preferential mRNA binding in m^{6}A-QTL loci, 20 also showed protein-protein interactions with known m^{6}A regulators, 7 showed allele-specific RBP binding overlapping our m^{6}A-QTLs, and 11 also showed significant expression correlation with m^{6}A levels of their targets. Five of the putative novel regulators (UPF1, NCBP2, LARP4, ATXN2, EFTUD2) were particularly intriguing, as they showed multiple interactions with both m^{6}A writers and readers, indicating possible functions as adapter proteins between them. The tight network of our predicted m^{6}A regulators and known m^{6}A regulators suggests extensive cooperation of multiple co-factors guiding m^{6}A methylation, similar to current pre-transcriptional and splicing regulatory models.

Our study has several limitations. First, our study is still confined to a relatively small number of tissues and a small number of individuals, due to the difficulty and sample requirements of profiling m^{6}A in primary human tissues. Thus our study only identified the strongest m^{6}A-QTLs, and we expect future studies to reveal many more m^{6}A-QTLs, which we estimate will show a modestly higher overlap between tissues, although still substantially lower than for eQTLs. Second, our antibody-based m^{6}A profiling does not provide single-nucleotide resolution and is not fully quantitative. We expect future studies to overcome these limitations, although existing methods that rely on restriction enzymes are confined to specific sequence contexts[72–76] and can only detect a subset of m^{6}A sites. Third, similar to GTEx, our study only sampled m^{6}A levels at bulk-tissue resolution, without distinguishing between different cell types within them. We expect that single-cell profiling of m^{6}A modifications, when such technologies become available, will help reveal the specific cell types where our m^{6}A-QTLs act, and to also reveal additional single-cell m^{6}A-QTLs that are not captured at the bulk level. Lastly, we focused on polyadenylated protein-coding mRNAs, while m^{6}A methylation can affect chromatin-associated RNAs and other non-coding RNAs, whose inter-individual m^{6}A variation may be highly informative in gene regulation and disease studies.

Overall, our datasets and analyses provide a foundation to bridge genetic variation with epitranscriptomic regulation and human disease in human primary tissues. This is particularly important as m^{6}A is increasingly recognized to play important roles in human disease, and many disease-associated variants remain uninterpreted. The resulting m^{6}A-QTLs, target genes, upstream regulators, and biological insights can help further our understanding of epitranscriptomic gene-regulatory control, and help pave the path for new therapeutic targets in human disease.

Methods

m^{6}A RIP-seq across human tissues

Samples collection by GTEx consortium, with donor enrollment and consent, histopathological review procedures, biospecimen procurement methods and fixation, and informed consent approval as previously described[77]. m^{6}A profiling was performed across four eGTEx tissues, including Brain cortex (n = 53), Lung (n = 32), Muscle - Skeletal (n = 32) and Heart - Left Ventricle (n = 12), using an optimized version of methylated RNA immunoprecipitation sequencing (meRIP-seq). Figure 1a created with BioRender.com. These four tissue types were chosen as they are highly disease-relevant and are representative for different germ layers (brain: ectoderm; muscle/heart: mesoderm; lung: endoderm).To minimize potential confounding factors introduced by experimental differences, the samples from each tissue were balanced across the three experimental batches (Supplementary Table 1). We adopted the protocols from Batista *et al.* (2014)[78], adapted to enable 400-fold less starting material available for the eGTEx samples. We calibrated the m^{6}A-Dynabeads complex by coupling the synaptic system anti m^{6}A-antibody to magnetic beads at a ratio of 5 μg of antibody for 1 mg of Dynabeads, and adjusting the reaction conditions and volume of m^{6}A/Dynabeads complex to 25 μl of m^{6}A/Dynabeads complex in a total volume of 50 μl IP volume. The new wash steps were set at 100 μl volume, and the new elution steps were established at 4 consecutive elution steps each using 50 μl of elution buffer at 52°C for 5 minutes under gentle rotation. The combined 240 μl elution was then precipitated by addition of 400 mM NaCl, 2 μl of ultra-pure Glycogen at 20 mg/ml and 2.5 times the volume of 200° proof ethanol at −20°C for one hour or overnight. The RNA was precipitated at 4°C for 25 minutes by centrifugation at 13,000g. The pellet was then washed twice in 70% ethanol, dried at 25°C for 15 minutes, and re-suspended in 7 μl of ultrapure H_2O prior library construction. We used the SMARTer Stranded RNA-Seq Kit from Clontech/Takada which is optimized to work with 100 pg of starting material. The libraries were sent for 2×45-bp pair-end sequencing.

m^{6}A sites identification and quantification

The RNA-seq (without immunoprecipitation) data were taken from the GTEx portal (V8; https://gtexportal.org/home/datasets) as input for m^{6}A peak calling and quantification. Reads from the RNA-seq and meRIP-seq were first aligned against tRNA (Downloaded from UCSC Table Browser) and rRNA (downloaded from NCBI nucleotides) using bowtie2 (v2.3.4.3)[79], with the unmapped reads kept for further analysis. The remaining reads were mapped to hg38 human genome using hisat2 (v2.1.0)[80], with GENCODE (v26; downloaded from https://gtexportal.org/home/datasets) as annotation, and with the following parameters: -k 1 --no-discordant. Peak calling was carried out using MACS2 (version 2.1.1), with the parameters of -- nomodel and --extsize 100. Peak processing performed using bedtools (v2.28.0). The high-confidence peaks were selected by requiring q value ≤ 10^{-5} and fold-change ≥ 3. The raw methylation level for each m^{6}A site was quantified as[28]:

$$\log\left[\left(\mathrm{Reads}_{\mathrm{Peak,\ MeRIP}}/\mathrm{Reads}_{\mathrm{Peak,\ RNA\text{-}seq}}\right)/\left(\mathrm{Reads}_{\mathrm{Total,\ MeRIP}}/\mathrm{Reads}_{\mathrm{Total,\ RNA\text{-}seq}}\right)\right]$$

The m^{6}A level matrix across samples and peaks was further normalized prior to m^{6}A-QTL calling (see the "m^{6}A normalization" section below).

m^{6}A sites quality control metrics

To evaluate if the m^{6}A immunoprecipitation was successful, we used deeptools (v3.0.2) to examine the reads density of meRIP sequencing surrounding the GGACH motif, comparing to other control sequence motifs, including the reverse complement sequence (DGTCC), reverse sequence (HCAGG) and complement sequence (CCTGD). We then evaluated the quality of the identified m^{6}A peaks using the two well-recognized m^{6}A metrics: m^{6}A consensus sequence and distribution along mRNA. For the distribution evaluation, we first converted the genomic m^{6}A positions to transcriptomic coordinate, and then calculated the density of m^{6}A along the mRNA structure. For each individual sample, we performed motif analysis using AME (v5.0.3) from the MEME toolset by examining the enrichment of m^{6}A motif (GGACH) over the randomly shuffled sequences with the nucleotide content preserved. The enrichment *P* value of motif was calculated by AME based on Fisher's exact test, and multiple test correction was performed using Bonferroni.

Tissue-specificity of m^{6}A sites

We performed hierarchical clustering with the global sequencing signals by segmenting the genome into 200-kb bins. For the individual m^{6}A, we defined an m^{6}A site as tissue-specific m^{6}A if it occurred in only one of the four tissues. We then further categorized the tissue-specific m^{6}A into two groups: the expression specificity induced tissue-specific m^{6}A and epitranscriptome-layer tissue-specific m^{6}A, depending on whether the tissue specificity mainly comes from the gene expression level. A tissue-specific m^{6}A is considered as expression specificity induced tissue-specific m^{6}A when the expression of the host gene in the tissue is 10 times higher than in any of the other tissues, whereas the rest were defined as epitranscriptomic-layer tissue-specific m^{6}A.

m^{6}A and SNPs pre-filtration

For each tissue, we filtered out the variants with less than 8 minor allele individuals (minor allele homozygotes + heterozygotes) to avoid false signals caused by a small number of individuals and to lower the multi-test correction burden. Additionally, we envision that m^{6}A regulation may take place a) at DNA level, where a SNP in DNA may affect m^{6}A given that the installation process happens co-transcriptionally; b) at RNA level, where a SNP in a transcript can modulate the m^{6}A level in *cis* in the same RNA molecule in a post transcriptional fashion. Therefore, to ensure both the sensitivity and the biological rationality, we specified the m^{6}A-QTL searching regions to be the promoter plus the gene body for each m^{6}A in a gene.

For m^{6}A-QTL calling, we selected the high-confidence m^{6}A sites by requiring either 1) the m^{6}A peak was previously curated[33] and meanwhile was captured in a tissue for no less than 2 individuals; or 2) the m^{6}A peak was captured across 20% of the samples in a tissue.

m^{6}A normalization

The m^{6}A signal matrix of one tissue was further processed and normalized prior to QTL identification: 1) the m^{6}A with more than 20% of the missing values (caused by the undetectable RNA expression) across individuals were excluded; 2) the IP efficiency differences between samples and the GC content across peaks were corrected following a well-established procedures[28]; 3) the IP- and GC-corrected m^{6}A matrix was further standardized by subtracting the mean plus dividing by the s.d. for each peak across all the samples in that tissue; 4) lastly, the matrix was quantile-normalized across all the peaks. The resulting m^{6}A matrix was used for latent factor identification and m^{6}A-QTL calling.

Latent factor identification and removal

To regress out the unwanted variance stemming from non-genetic factors, we employed PEER[37] (R package 1.0) to identify the latent covariates for the samples in each tissue. We chose the PEER factors for QTL calling based on: a) maximizing the number of g-m^{6}A identified (Fig. 2b); b) the intra PEER factor correlation (Fig. 2a, Extended Data Fig. 2a–b). We included top 7, 2, 5 PEER factors for brain, lung, muscle/heart, respectively, plus experimental batch and top three genotype PCs as covariates for m^{6}A-QTL calling.

m^{6}A-QTL identification

We used FastQTL[81] (v2.0) for m^{6}A-QTL identification with the PEER factors included as covariates as described above. The searching window was set to the gene body plus the promoter (1 kb upstream of the transcription start site). To identify g-m^{6}A, the nominal *P* value of the lead SNP of each m^{6}A peak was corrected to get an empirical *P* value that accounts for the multiple variants tested. Empirical *P* value was extrapolated based on beta distribution fitting to the permutation results by FastQTL (parameter --permute 1000). We applied a threshold of empirical *P* value < 0.005 on the lead variant to identify g-m^{6}A, and then for each g-m^{6}A we applied the nominal *P* value threshold that corresponds to the empirical *P* value of 0.005 for each locus to identify m^{6}A-QTL (Supplementary Table 2). Particularly given larger sample size in brain, we provided a higher-confidence sets of g-m^{6}As and m^{6}A-QTLs by further correcting the empirical *P* value with Benjamini-Hochberg procedure and applying a threshold of FDR < 0.2 (Supplementary Table 3).

m^{6}A-QTL validation

To validate the m^{6}A-QTLs we identified in an unbiased way, we utilized an independent dataset published by another study[26], from which we got the m^{6}A profiles of heart, muscle and lung, each with two adult post-mortem individuals. Although brain samples are available from this study, the region they used was from cerebrum, which is much larger and more complex than the cortex region used by our m^{6}A-QTL analysis. Given the complexity of the human brain, we did not include the brain samples in the validation. To do the validation, we first grouped the m^{6}A-QTLs from our study based on the direction of effect size, where "+" represent the m^{6}A level increases from major to minor alleles, and the "-" represents the opposite, together with a "control" group of SNPs which are not m^{6}A-QTL. We then used samtools (v1.9) to call SNPs with RNA-seq data of each individual from the validation study. The SNPs with different genotypes in the two individuals were kept for further usage.

For each SNP, the individual with more minor alleles were regarded as “Minor” while the other one regarded as “Major”. The m^{6}A fold change of the “Minor” individual over the “Major” one was calculated for each variant in the validation cohort, and compared across the “−”, “+” and “control” groups defined from this study by Wilcoxon rank-sum test.

Simulation analysis for m^{6}A-QTL calling power evaluation

We carried out simulations to assess the power of m^{6}A-QTL calling, and thus to assess the extent of m^{6}A-QTL tissue-specificity driven by the current sample sizes, following the simulation strategy from an eQTL power analysis paper[82]. Specifically, we used the effect size, MAF, and sample size parameters observed from the lead m^{6}A-QTLs we identified, and performed the same number of simulations as the number of g-m^{6}As identified in each tissue following a simple linear regression: $m6A = \beta g + \in$, *where* $\in \sim N(0,1)$; *g denotes genotypes*, assuming each g-m^{6}A site is only regulated by one single SNP. We also simulated the same number of controls with the effect size set to 0. For m^{6}A-QTLs identified in each tissue, we simulated a corresponding set of m^{6}As with the sample sizes of other tissues. m^{6}A-QTL was mapped with linear regression in the simulation dataset. The m^{6}A-QTL overlap rate and directionality consistency were calculated by comparing the simulated dataset with the real dataset from which the m^{6}A-QTL parameters were used for simulation.

Enrichment testing and overlap analysis between m^{6}A-QTL and eQTL

The enrichment analysis of m^{6}A-QTLs versus eQTLs were carried out using GARFIELD[83] (version 2). The most updated version of eQTLs were used (V8), which are available in the GTEx portal (https://gtexportal.org/home/). The enrichment of m^{6}A-QTLs in eQTLs was quantified with odd ratios and significance was calculated using a generalized linear model, with minor allele frequency, distance to TSS and number of LD proxies ($r^2 > 0.8$) accounted for (see Garfield manual for details). To test if the overlap between eGenes and gmGenes is more than expected by chance (i.e. Fig. 4e), we randomly sampled the same number of eGenes and gmGenes from the expressed gene pool in our data (n = 26,359) for 10,000 times, the *P* value was calculated as the percentage of cases where the overlap between the randomly sampled genes was greater than the observed overlap.

To enable a more accurate comparison between the effect of m^{6}A-QTLs on m^{6}A versus the effect of eQTLs on the corresponding gene expression, we re-performed eQTL analysis in each tissue using the same cohort as used in m^{6}A-QTL calling. Similar to m^{6}A-QTL analysis, we first used PEER to capture the unwanted variance and included the first 6, 3 and 4 PEER factors for brain, lung and muscle/heart as covariates during eQTL calling. The slopes of eQTLs and m^{6}A-QTLs were then compared. Correlation of the slopes between the overlapped m6A-QTLs and eQTLs (separated by effect directionality) was calculated using a two-sided pearson correlation test, *P* value was calculated with linear regression.

GWAS heritability partition of m^{6}A-QTL

Enrichment analysis of m^{6}A-QTLs in GWAS was performed using stratified LD score regression (S-LDSC, v1.0.1) based on the tutorial[54]. The enrichment was calculated as the proportion of heritability over the proportion of SNPs, then the standard error was estimated

and used for *P* value calculation. We used two baseline models: 1) baseline v1.2 (provided by LDSC) plus eQTLs from the same tissue (**Fig. s5e**); 2) baseline v1.2 with the gene-body related annotations excluded (coding region, UTRs, promoter, conserved region), and with the eQTLs from the same tissue included (Fig. 5d). We used two sets of m6A-QTLs with different thresholds, which were $P < 10^{-3}$ and $P < 10^{-4}$. respectively. For brain m^{6}A-QTL heritability analysis, we further included the high confidence m^{6}A-QTLs that were identified with two rounds of multiple-test correction (labeled as "2StepAdj." in the figure). The GWAS summary statistics files were downloaded from a) the files curated by the Alkes group (https://data.broadinstitute.org/alkesgroup/LDSCORE/independent_sumstats/); b) UK Biobank; and c) respiratory traits from GWAS Atlas. Disease relevant traits were selected.

Colocalization analysis between m^{6}A-QTLs and GWAS loci

Colocalization analysis between m^{6}A-QTLs and GWAS variants was carried out using a Bayes Factor colocalization framework by Coloc[84]. GWAS loci (filtered by $P < 10^{-4}$) that overlap with g-m^{6}As were used for colocalization tests. The ratio between PP3 (posterior probability of both traits are associated with different causal variants) and PP4 (posterior probability of both traits are associated with the same causal variant) was calculated, and those loci with PP3/PP4 < 0.05 and PP4 > 0.1 were selected as the GWAS-m^{6}A-QTL colocalized events.

Enrichment analysis of m^{6}A-QTLs in RBP binding

To predict potential indirect m^{6}A regulators that function through binding to m^{6}A-QTLs, we carried out enrichment analysis between m^{6}A-QTLs and RBP binding sites. We downloaded the binding sites of 171 RBPs curated by POSTAR2 database[64], and then performed enrichment analysis against m^{6}A-QTLs using GARFIELD[83]. The enrichment of m^{6}A-QTLs in RBP binding sites was quantified as odd ratios, and significance was calculated using the generalized linear model, with minor allele frequency, distance to TSS and number of LD proxies ($r^2 > 0.8$) accounted for. The RBPs with less than 100 baseline SNPs overlapping binding sites (NAnnot < 100) were excluded to ensure the detection reliability. RBPs showing a Bonferroni adjusted *P* value < 0.05 were selected as potential m^{6}A regulators.

To further prioritize the enriched RBPs, we carried out a RBP-expression vs. m^{6}A level correlation analysis. For each enriched RBP, we first calculated an "aggregated" m^{6}A score for each individual, where the methylation levels of the m^{6}A sites whose corresponding m^{6}A-QTLs are within the RBP binding sites were aggregated, and then performed regression between the aggregated m^{6}A level and the gene expression level of the RBP across each individual. *P* values were calculated based on linear regression, and multiple test correction was performed by Benjamini & Hochberg (BH) correction. FDR of adjusted *P* value < 0.1 was used as the threshold.

Protein-Protein-Interaction (PPI) analysis

We examined the protein-protein interaction network composed of the enriched RBPs and the known m^{6}A regulators (writer/reader/eraser) using the "STRINGdb" R package (v1.24.0)[85]. Only the interactions with the source of "Experiments" were considered.

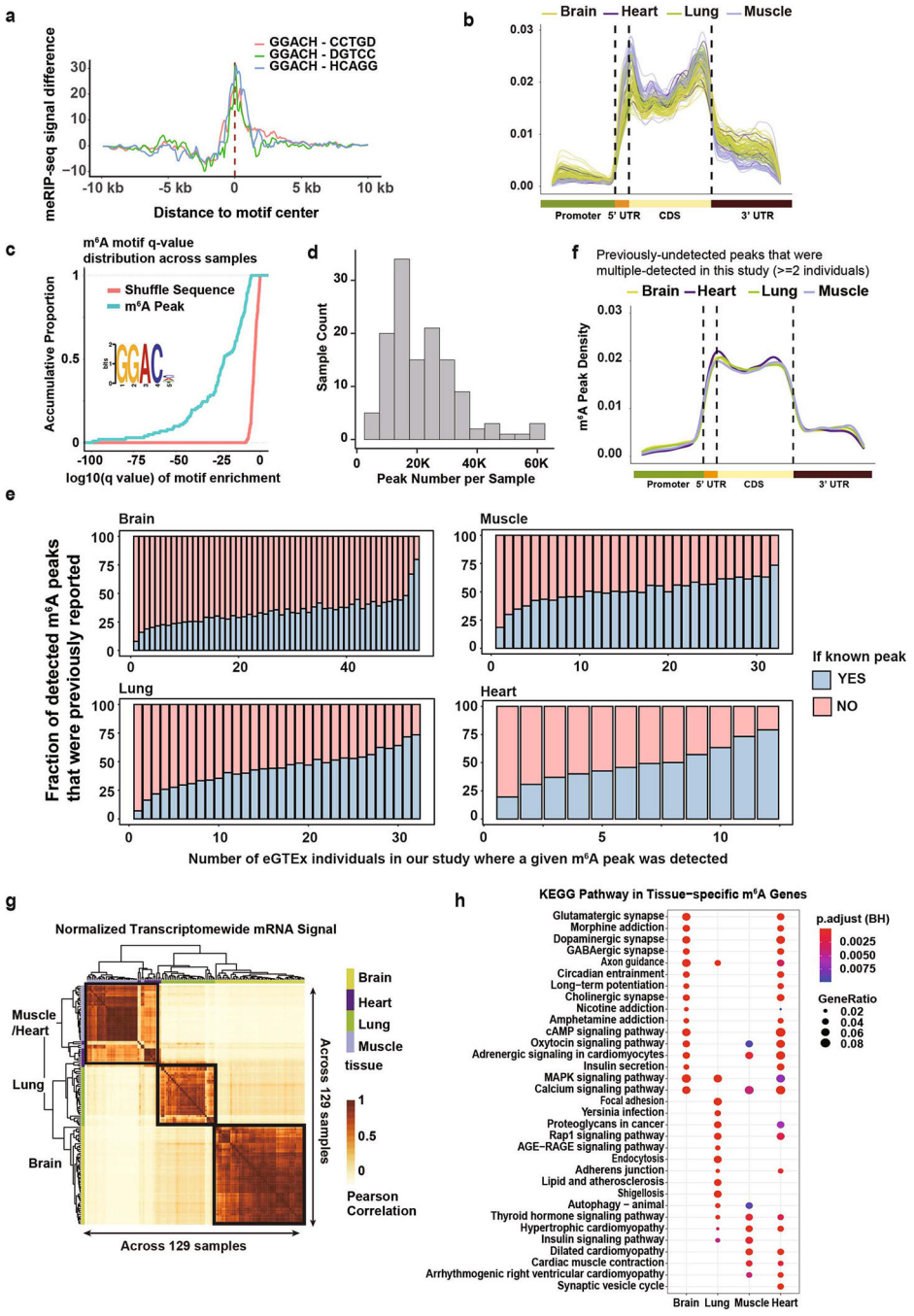
Data availability

All eGTEx protected data, including m^6A sequencing reads and matched RNA-seq data are available on dbGaP with accession number phs000424.v8.p2. Additionally, the data can be accessed via AnVIL with authentication: https://anvil.terra.bio/#workspaces/anvil-datastorage/AnVIL_GTEx_V8_hg38. Since the raw sequencing data with genetic information are protected, application and authentication are needed before accessing the data. All non-protected data of m^6A can be visualized via the GTEx Portal (www.gtexportal.org) as part of eGTEx v8. The m^6A-QTLs identified in each tissue can be downloaded from the supplementary tables. The eQTL datasets are from GTEx v8, which can be accessed at https://gtexportal.org/home/datasets. The LCL m^6A-QTL datasets from Zhang et al. can be downloaded from https://doi.org/10.5281/zenodo.3870952. The previously curated m^6A sites can be downloaded from RMBase (http://rna.sysu.edu.cn/rmbase/). The RNA binding sites can be downloaded from http://lulab.life.tsinghua.edu.cn/postar/.

Code availability

Code for m^6A data processing, m^6A-QTL calling, relevant functional analyses, and additional supplementary information can be found in: http://compbio.mit.edu/m6AQTLs/ and also Zenodo with DOI: 10.5281/zenodo.4764136.

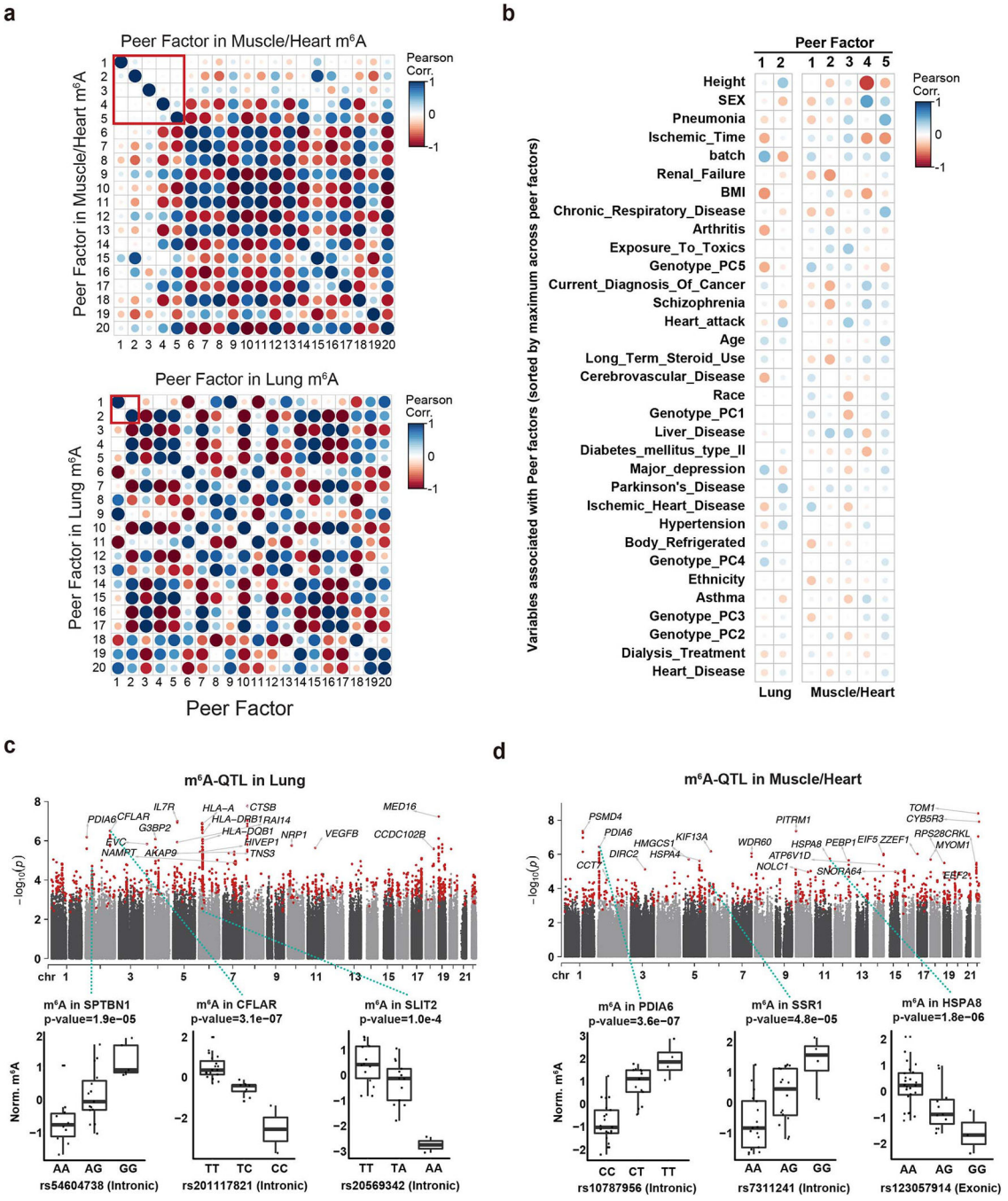
Extended Data



Extended Data Fig. 1. m^{6}A landscape across tissues.

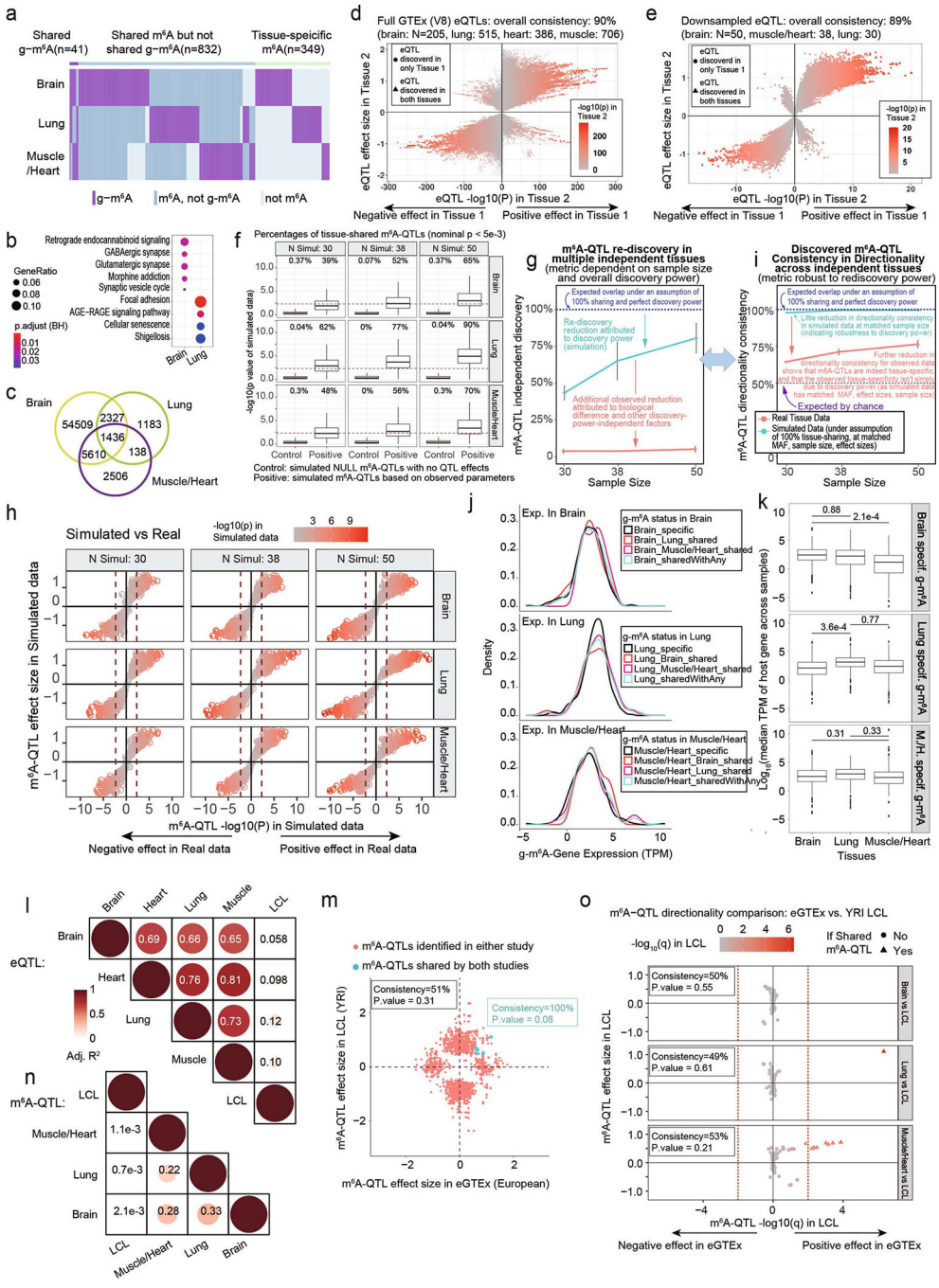
a. Positional enrichment for aggregated meRIP-seq signal surrounding the known m^{6}A motif (GGACH) vs. three control sequences (colors). b. m^{6}A peak positional density (y-axis) along the gene structure (x-axis) in brain, lung, heart and muscle. c. Cumulative distribution (y-axis) of GGACH motif enrichment (-log10q-value, x-axis) across samples (teal) vs. shuffled controls (salmon). d. Peak Count per sample. e. Peaks shared by

fewer individuals (left) are more likely previously-undetected (red). f. Similar to b for for only previously-unreported m^{6}A peaks found in >=2 individuals in our study. g. Pearson correlation (heatmap) and hierarchical clustering (tree) of mRNA-seq profiles across the 129 GTEx samples (rows/columns) that match the individuals profiled here shows tissue-specific clustering (colors) and co-clustering of heart and muscle, as shown for m^{6}A profiles in Fig. 1b. h. KEGG pathway enrichments for the genes harboring tissue-specific m^{6}A.



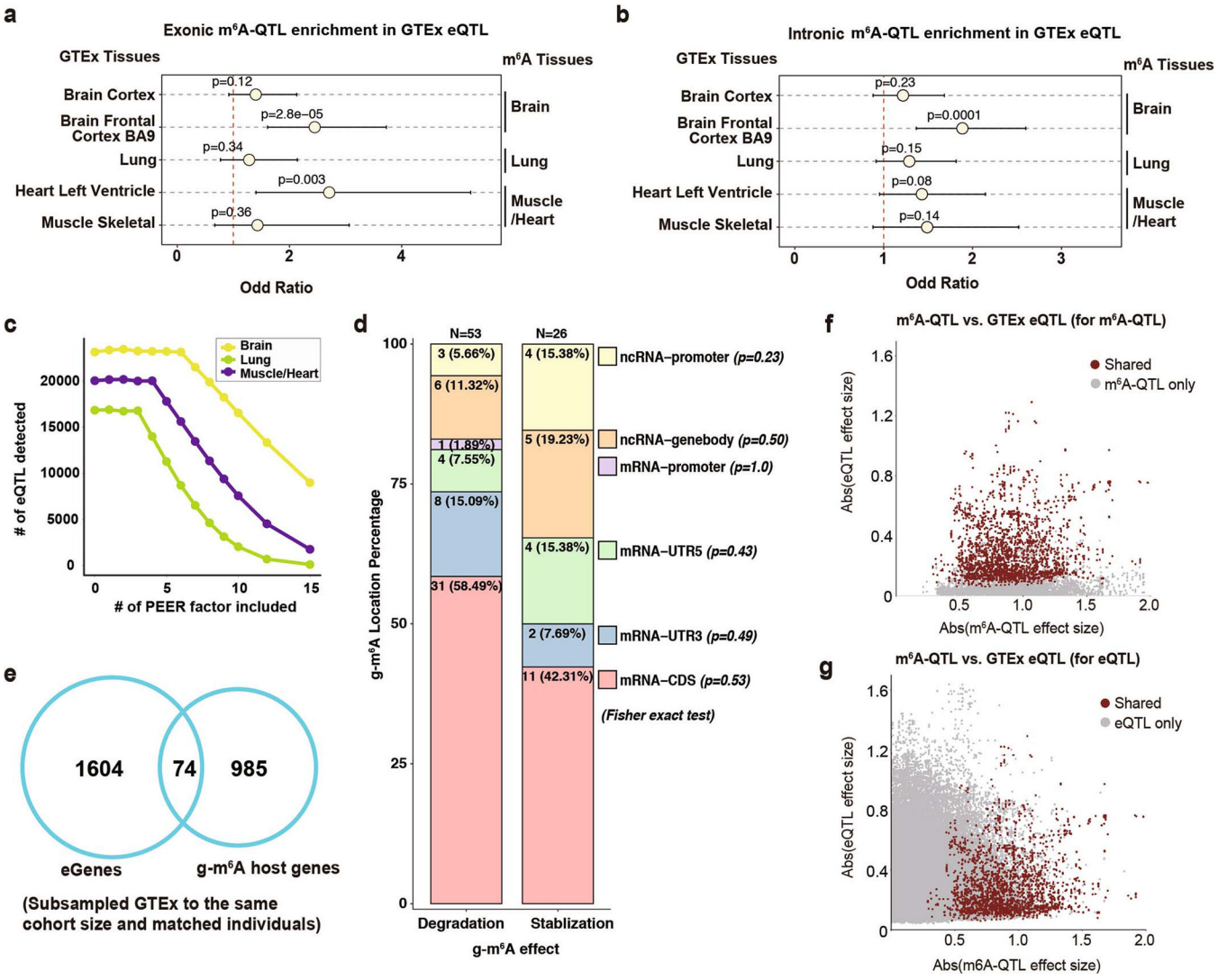
Extended Data Fig. 2. Identification of genetically-driven m^{6}A.

a. Pearson correlation (color) for PEER factors shows saturation (correlated factors) after first 2 factors (red box) in lung (top) and after first 5 factors (red box) in muscle/heart (bottom). b. Pearson correlation between PEER factors (columns) and donor phenotypic measurements (rows) helps interpret factors in lung and muscle/heart. c-d. Manhattan plot of m^{6}A-QTLs in lung (c) and muscle/heart (d), and m^{6}A-QTL examples. Boxes=25%−75% percentile; line=median; whiskers=1.5 IQR; p-values=FastQTL linear regressions.



Extended Data Fig. 3. Tissue specificity of m^{6}A-QTL.

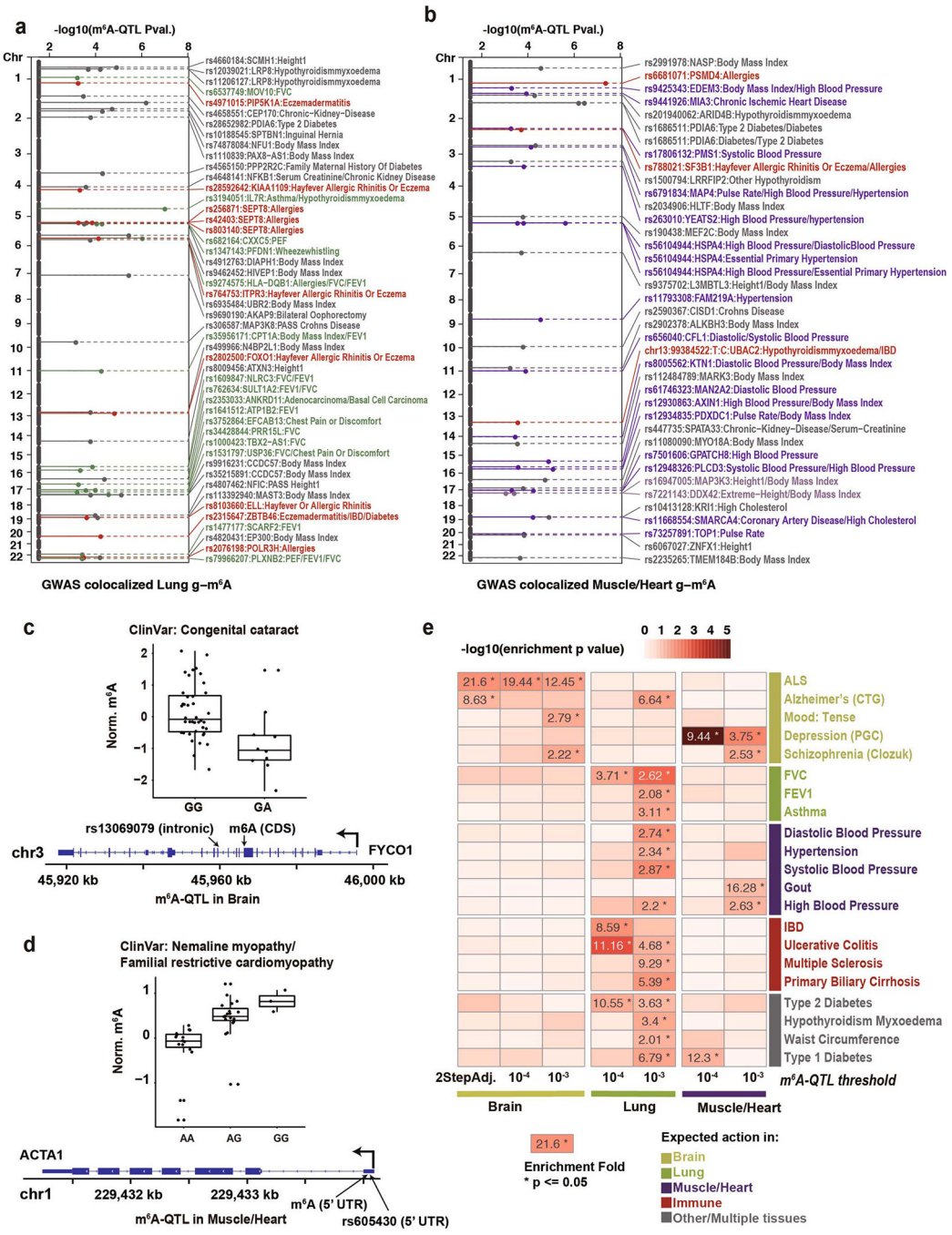
a. Summary of shared/specific g-m^{6}As and m^{6}A across tissues. b. KEGG pathway enrichments for tissue-specific gmGenes. c. Tissue-intersections of the eQTLs identified from the same samples for m^{6}A-QTL calling. d-e. Correlation between p-value (x-axis) and eQTL effect (y-axis) between tissues, with positive- and negative-effect eQTLs separated for full GTEx-V8 (d)), and subsampled to individuals used here (e;). f. Nominal p-values (y-axis) of simulated m^{6}A-QTLs (Positive), and simulated NULL m^{6}A-QTLs controls without QTL effects (Control). Boxes=25%−75% percentile; line=median; whiskers=1.5 IQR. g. m^{6}A-QTL overlaps between tissues in simulated data show much higher tissue-sharing (teal curve) than in observed real data (peach curve). h. Effect sizes directionality between m^{6}A-QTL from real tissue data and simulated data are almost 100% consistent. i. Effect directionality consistency when m^{6}A-QTLs were identified with 50, 38 and 30 samples. j. Gene expression distribution of the tissue-specific vs. tissue-shared m^{6}A-QTLs in each tissue. k. Gene expression in tissues for tissue-specific gmGenes in each tissue. Statistical test was carried out by two-sided paired Wilcoxon test. l. Correlation (adjusted R^2) of eQTLs between GTEx primary tissues and YRI LCL cells. Boxes=25%−75% percentile; line=median; whiskers=1.5 IQR. m. Comparison of m^{6}A-QTL effects size between this study (eGTEx tissues, x-axis) and the other m^{6}A-QTL study (YRI LCLs, y-axis). Green dots represent the m^{6}A-QTLs shared by the two studies. Directionality consistency and corresponding p-value (vs. the 50% expected by chance) calculated using one-sided Fisher exact test (inset box). n. Correlation of m^{6}A-QTLs between eGTEx primary tissues and YRI LCLs. o. Correlation between p-value and m^{6}A-QTL effect in LCL cell lines for m^{6}A-QTLs identified in eGTEx tissues, with positive- and negative-effect eGTEx m^{6}A-QTLs separated. Directionality consistency and corresponding p-value (vs. the 50% expected by chance) calculated using one-sided Fisher exact test.



Extended Data Fig. 4. Comparison between eQTL and m^{6}A-QTL.

a-b. m^{6}A-QTLs show a modest but significant enrichment for eQTLs in the matching tissues, with m^{6}A-QTLs separated into exonic (b) and intronic (c). NAnnot=21811, 17687, 40078, 24317, 39687 for Brain Cortex, Brain Frontal Cortex (BA9), Lung, Heart Left Ventricle, Muscle Skeletal (see Methods) for exonic m^{6}A-QTLs. NAnnot=23091, 18724, 41807, 25927, 42425 for Brain Cortex, Brain Frontal Cortex (BA9), Lung, Heart Left Ventricle, Muscle Skeletal (see Methods) for intronic m^{6}A-QTLs. Error bars denote the upper bound and the lower bound for the 95% CI of effect size. P-values are calculated by Garfield using a logistic regression model with 'feature matching'. c. Number of eQTLs identified (y-axis) for increasing number of PEER factors removed (x-axis) shows inflection-point for each tissue (colors). d. Genomic region distribution for g-m^{6}As mediating stabilization vs. degradation (p-values: Fisher exact test). e. Overlap between gmGenes and eGenes identified from the matching GTEx individuals. f. Effect size comparison between m^{6}A-QTL and GTEx eQTL for the m^{6}A-QTLs identified in this study.

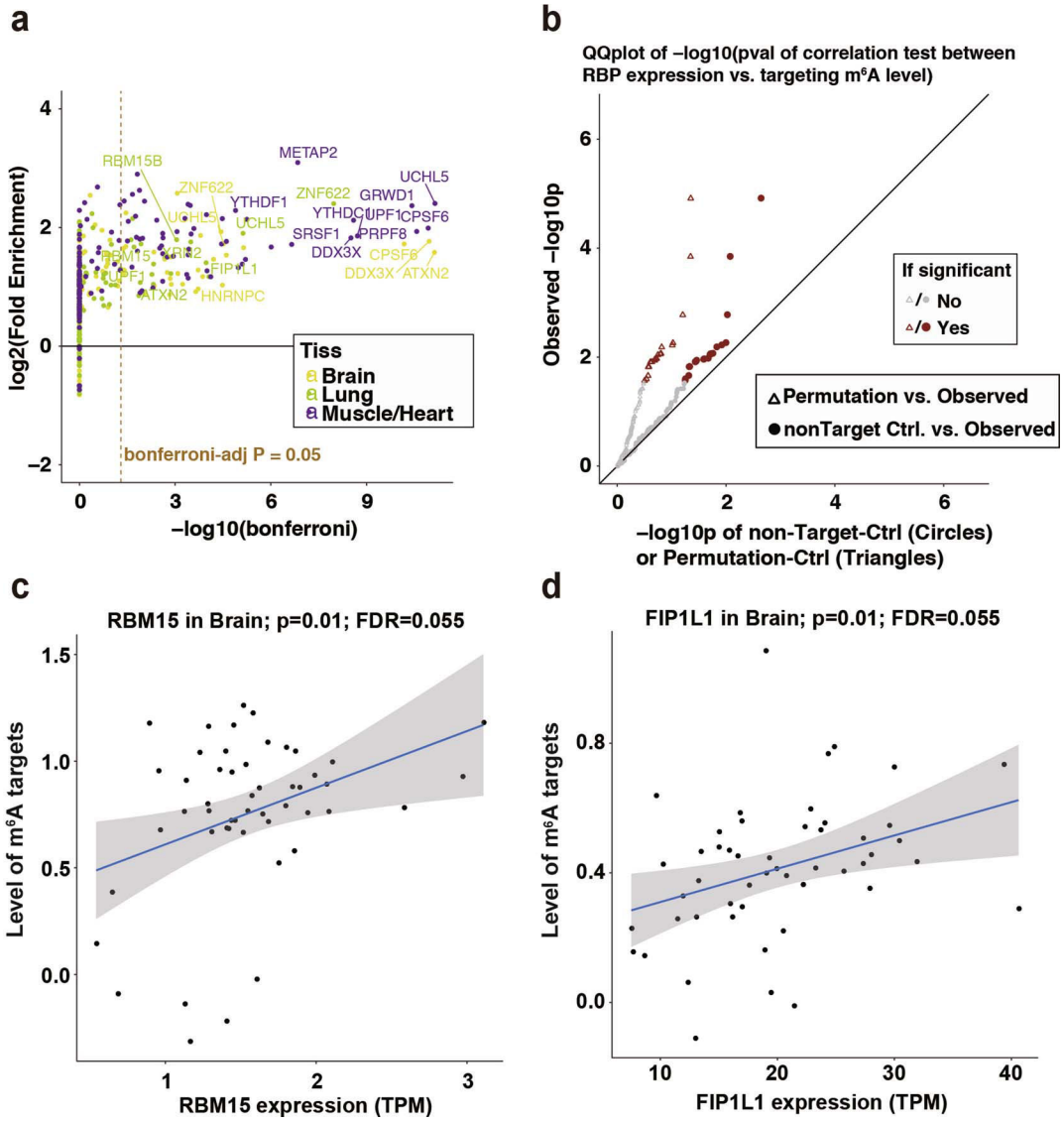
g. Effect size comparison between m^{6}A-QTL and GTEx eQTL for the eQTLs identified by GTEx V8.



Extended Data Fig. 5. Overlap between m^{6}A-QTL and disease GWAS hits.

a-b. Overlaid Manhattan plots showing genomic position (y-axis) and m^{6}A-QTL P-value (x-axis) for lead SNPs (points) across traits (colors) that show colocalization between GWAS variants and m^{6}A-QTLs in lung (a) and muscle/heart (b). c. Illustrative example showing a brain intronic m^{6}A-QTL that is overlapped with a ClinVar-curated variant related

to Congenital cataract. d. Illustrative example showing a muscle/heart 5'-UTR m^{6}A-QTL that is overlapped with a ClinVar-curated variant related to Nemaline myopathy. Boxes=25% −75% percentile; line=median; whiskers=1.5 IQR. e. Same plot as Fig. 5d, but shown using S-LDSC Baseline v1.2, which corrects for coding region, UTR, intron, promoter, enhancer, multiple histone marks, and eQTLs, shows robustness of results to this correction.



Extended Data Fig. 6. Identification of m^{6}A regulator candidates.

a. Enrichment (y axis, log) and corresponding p-value (x-axis, Bonferroni-corrected) between m^{6}A-QTL SNPs and RNA binding protein (RBP) binding sites, for each tissue (color), highlighting 10 most enriched RBPs in each tissue (labels). b. Quantile-Quantile plot showing the p-value distribution observed in the correlation test between RBP expression vs. m^{6}A levels (y-axis), compared to the non-target (circle) or permutation controls (triangle) (x-axis). Significant RBPs (FDR < 0.1) are shown in red. P-values are calculated by two-sided Pearson correlation tests. c. Correlation of predicted m^{6}Aregulator RBP mRNA

expression level (x-axis) vs. methylation level of its m^6A targets (y-axis) for ZNF622 in muscle/heart. Grey shadow denotes the 95% confidence region for the regression fit.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We thank Carles Boix for helping with figure visualization, analysis suggestion and discussion, and thank Kai Kang and other Kellis laboratory members for discussions and suggestions. We thank GTEx consortium for providing the samples for m^6A profiling. We thank Kristin G. Ardlie and Francois Aguet for helping with GTEx data deposition and providing helpful suggestions. This work was supported by NIH grants NIH HG007610, HG008155, HG009446, MH109978, AG054012, AG058002, AG062377, NS110453, NS115064, AG067151, AG062335 and MH119509 to M.K., and CA232115, and CA233671 to R.I.G.

REFERENCES

1. Ward LD & Kellis M Interpreting noncoding genetic variation in complex traits and human disease. Nat. Biotechnol. 30, 1095–1106 (2012). [PubMed: 23138309]
2. Tak YG & Farnham PJ Making sense of GWAS: using epigenomics and genome engineering to understand the functional relevance of SNPs in non-coding regions of the human genome. Epigenetics Chromatin 8, 57 (2015). [PubMed: 26719772]
3. Grundberg E et al. Mapping cis- and trans-regulatory effects across multiple tissues in twins. Nat. Genet. 44, 1084–1089 (2012). [PubMed: 22941192]
4. GTEx Consortium. The Genotype-Tissue Expression (GTEx) project. Nat. Genet. 45, 580–585 (2013). [PubMed: 23715323]
5. Sun W et al. Histone Acetylome-wide Association Study of Autism Spectrum Disorder. Cell 167, 1385–1397.e11 (2016). [PubMed: 27863250]
6. Wainberg M et al. Opportunities and challenges for transcriptome-wide association studies. Nat. Genet. 51, 592–599 (2019). [PubMed: 30926968]
7. Huan T et al. Genome-wide identification of DNA methylation QTLs in whole blood highlights pathways for cardiovascular disease. Nat. Commun. 10, 4267 (2019). [PubMed: 31537805]
8. Consortium GTEx et al. Genetic effects on gene expression across human tissues. Nature 550, 204–213 (2017). [PubMed: 29022597]
9. W u L et al. Variation and genetic control of protein abundance in humans. Nature 499, 79–82 (2013). [PubMed: 23676674]
10. Li YI et al. RNA splicing is a primary link between genetic variation and disease. Science 352, 600–604 (2016). [PubMed: 27126046]
11. Park E et al. Population and allelic variation of A-to-I RNA editing in human transcriptomes. Genome Biol. 18, 143 (2017). [PubMed: 28754146]
12. Roundtree IA, Evans ME, Pan T & He C Dynamic RNA Modifications in Gene Expression Regulation. Cell 169, 1187–1200 (2017). [PubMed: 28622506]
13. Frye M, Jaffrey SR, Pan T, Rechavi G & Suzuki T RNA modifications: what have we learned and where are we headed? Nat. Rev. Genet. 17, 365–372 (2016). [PubMed: 27140282]
14. Zaccara S, Ries RJ & Jaffrey SR Reading, writing and erasing mRNA methylation. Nat. Rev. Mol. Cell Biol. 20, 608–624 (2019). [PubMed: 31520073]
15. Zhao BS, Roundtree IA & He C Post-transcriptional gene regulation by mRNA modifications. Nat. Rev. Mol. Cell Biol. 18, 31–42 (2017). [PubMed: 27808276]
16. Barbieri I & Kouzarides T Role of RNA modifications in cancer. Nat. Rev. Cancer (2020) doi:10.1038/s41568-020-0253-2.
17. Yao B et al. Epigenetic mechanisms in neurogenesis. Nat. Rev. Neurosci. 17, 537–549 (2016). [PubMed: 27334043]

18. Li X, Xiong X & Yi C Epitranscriptome sequencing technologies: decoding RNA modifications. Nat. Methods 14, 23–31 (2016). [PubMed: 28032622]
19. Frye M, Harada BT, Behm M & He C RNA modifications modulate gene expression during development. Science 361, 1346–1349 (2018). [PubMed: 30262497]
20. Huang H, Weng H & Chen J mA Modification in Coding and Non-coding RNAs: Roles and Therapeutic Implications in Cancer. Cancer Cell 37, 270–288 (2020). [PubMed: 32183948]
21. Livneh I, Moshitch-Moshkovitz S, Amariglio N, Rechavi G & Dominissini D The mA epitranscriptome: transcriptome plasticity in brain development and function. Nat. Rev. Neurosci. 21, 36–51 (2020). [PubMed: 31804615]
22. Shi H, Wei J & He C Where, When, and How: Context-Dependent Functions of RNA Methylation Writers, Readers, and Erasers. Mol. Cell 74, 640–650 (2019). [PubMed: 31100245]
23. Arguello AE, DeLiberto AN & Kleiner RE RNA Chemical Proteomics Reveals the N-Methyladenosine (mA)-Regulated Protein-RNA Interactome. J. Am. Chem. Soc. 139, 17249–17252 (2017). [PubMed: 29140688]
24. Edupuganti RR et al. N-methyladenosine (mA) recruits and repels proteins to regulate mRNA homeostasis. Nat. Struct. Mol. Biol. 24, 870–878 (2017). [PubMed: 28869609]
25. Patil DP, Pickering BF & Jaffrey SR Reading m6A in the Transcriptome: m6A-Binding Proteins. Trends Cell Biol. 28, 113–127 (2018). [PubMed: 29103884]
26. Liu J. ‘e et al. Landscape and Regulation of m6A and m6Am Methylome across Human and Mouse Tissues. Mol. Cell 77, 426–440.e6 (2020). [PubMed: 31676230]
27. Xiao S et al. The RNA N6-methyladenosine modification landscape of human fetal tissues. Nat. Cell Biol. 21, 651–661 (2019). [PubMed: 31036937]
28. Zhang Z et al. Genetic analyses support the contribution of mRNA N-methyladenosine (mA) modification to human disease heritability. Nat. Genet. 52, 939–949 (2020). [PubMed: 32601472]
29. eGTEx Project. Enhancing GTEx by bridging the gaps between genotype, gene expression, and disease. Nat. Genet 49, 1664–1670 (2017). [PubMed: 29019975]
30. GTEx Consortium. The GTEx Consortium atlas of genetic regulatory effects across human tissues. Science 369, 1318–1330 (2020). [PubMed: 32913098]
31. Dominissini D et al. Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. Nature 485, 201–206 (2012). [PubMed: 22575960]
32. Schwartz S et al. High-resolution mapping reveals a conserved, widespread, dynamic mRNA methylation program in yeast meiosis. Cell 155, 1409–1421 (2013). [PubMed: 24269006]
33. Xuan J-J et al. RMBase v2.0: deciphering the map of RNA modifications from epitranscriptome sequencing data. Nucleic Acids Res. 46, D327–D334 (2018). [PubMed: 29040692]
34. Meyer KD et al. Comprehensive analysis of mRNA methylation reveals enrichment in 3’ UTRs and near stop codons. Cell 149, 1635–1646 (2012). [PubMed: 22608085]
35. Lin S, Choe J, Du P, Triboulet R & Gregory RI The m(6)A Methyltransferase METTL3 Promotes Translation in Human Cancer Cells. Mol. Cell 62, 335–345 (2016). [PubMed: 27117702]
36. Visvanathan A et al. Essential role of METTL3-mediated mA modification in glioma stem-like cells maintenance and radioresistance. Oncogene 37, 522–533 (2018). [PubMed: 28991227]
37. Stegle O, Parts L, Durbin R & Winn J A Bayesian framework to account for complex non-genetic factors in gene expression levels greatly increases power in eQTL studies. PLoS Comput. Biol. 6, e1000770 (2010). [PubMed: 20463871]
38. Lappalainen T et al. Transcriptome and genome sequencing uncovers functional variation in humans. Nature 501, 506–511 (2013). [PubMed: 24037378]
39. Huang H et al. Recognition of RNA N-methyladenosine by IGF2BP proteins enhances mRNA stability and translation. Nat. Cell Biol. 20, 285–295 (2018). [PubMed: 29476152]
40. Wang X et al. N6-methyladenosine-dependent regulation of messenger RNA stability. Nature 505, 117–120 (2014). [PubMed: 24284625]
41. Amaral AC et al. Prenatal protein malnutrition decreases KCNJ3 and 2DG activity in rat prefrontal cortex. Neuroscience 286, 79–86 (2015). [PubMed: 25446346]
42. Doré AS et al. Structure of class C GPCR metabotropic glutamate receptor 5 transmembrane domain. Nature 511, 557–562 (2014). [PubMed: 25042998]

43. Gandal MJ et al. Transcriptome-wide isoform-level dysregulation in ASD, schizophrenia, and bipolar disorder. Science 362, (2018).
44. Tindi JO et al. ANKS1B Gene Product AIDA-1 Controls Hippocampal Synaptic Transmission by Regulating GluN2B Subunit Localization. J. Neurosci. 35, 8986–8996 (2015). [PubMed: 26085624]
45. Cheng W, Wang F, Feng A, Li X & Yu W CXXC5 Attenuates Pulmonary Fibrosis in a Bleomycin-Induced Mouse Model and MLFs by Suppression of the CD40/CD40L Pathway. Biomed Res. Int. 2020, 7840652 (2020). [PubMed: 32337277]
46. Zielonka M, Xia J, Friedel RH, Offermanns S & Worzfeld T A systematic expression analysis implicates Plexin-B2 and its ligand Sema4C in the regulation of the vascular and endocrine system. Exp. Cell Res. 316, 2477–2486 (2010). [PubMed: 20478304]
47. Kelly EAB et al. Potential contribution of IL-7 to allergen-induced eosinophilic airway inflammation in asthma. J. Immunol. 182, 1404–1410 (2009). [PubMed: 19155487]
48. Cui H et al. Impairment of Fatty Acid Oxidation in Alveolar Epithelial Cells Mediates Acute Lung Injury. Am. J. Respir. Cell Mol. Biol. 60, 167–178 (2019). [PubMed: 30183330]
49. Tzimas C et al. WIPI1 is a conserved mediator of right ventricular failure. JCI Insight 5, (2019).
50. Webster DR & Bratcher JM Developmental regulation of cardiac MAP4 protein expression. Cell Motil. Cytoskeleton 63, 512–522 (2006). [PubMed: 16767748]
51. Mohamed BA et al. Targeted disruption of Hspa4 gene leads to cardiac hypertrophy and fibrosis. J. Mol. Cell. Cardiol. 53, 459–468 (2012). [PubMed: 22884543]
52. Hernandez-Zimbron LF et al. Molecular Age-Related Changes in the Anterior Segment of the Eye. J. Ophthalmol. 2017, 1295132 (2017). [PubMed: 29147580]
53. Laing NG et al. Mutations and polymorphisms of the skeletal muscle alpha-actin gene (ACTA1). Hum. Mutat. 30, 1267–1277 (2009). [PubMed: 19562689]
54. Finucane HK et al. Partitioning heritability by functional annotation using genome-wide association summary statistics. Nat. Genet. 47, 1228–1235 (2015). [PubMed: 26414678]
55. Andersson R et al. An atlas of active enhancers across human cell types and tissues. Nature 507, 455–461 (2014). [PubMed: 24670763]
56. Gamazon ER et al. Using an atlas of gene regulation across 44 human tissues to inform complex disease- and trait-associated variation. Nat. Genet. 50, 956–967 (2018). [PubMed: 29955180]
57. Roadmap Epigenomics Consortium et al. Integrative analysis of 111 reference human epigenomes. Nature 518, 317–330 (2015). [PubMed: 25693563]
58. Gjoneska E et al. Conserved epigenomic signals in mice and humans reveal immune basis of Alzheimer's disease. Nature 518, 365–369 (2015). [PubMed: 25693568]
59. Corces MR et al. Single-cell epigenomic analyses implicate candidate causal variants at inherited risk loci for Alzheimer's and Parkinson's diseases. Nat. Genet. 52, 1158–1168 (2020). [PubMed: 33106633]
60. Ardain A, Marakalala MJ & Leslie A Tissue-resident innate immunity in the lung. Immunology 159, 245–256 (2020). [PubMed: 31670391]
61. Whitsett JA & Alenghat T Respiratory epithelial cells orchestrate pulmonary innate immunity. Nat. Immunol. 16, 27–35 (2015). [PubMed: 25521682]
62. Chiu C & Openshaw PJ Antiviral B cell and T cell immunity in the lungs. Nat. Immunol. 16, 18–26 (2015). [PubMed: 25521681]
63. Zheng N, Su J, Hu H, Wang J & Chen X Research Progress of N6-Methyladenosine in the Cardiovascular System. Med. Sci. Monit. 26, e921742 (2020). [PubMed: 32350237]
64. Zhu Y et al. POSTAR2: deciphering the post-transcriptional regulatory logics. Nucleic Acids Res. 47, D203–D211 (2019). [PubMed: 30239819]
65. Zhou J et al. Dynamic m(6)A mRNA methylation directs translational control of heat shock response. Nature 526, 591–594 (2015). [PubMed: 26458103]
66. Yang Y, Hsu PJ, Chen Y-S & Yang Y-G Dynamic transcriptomic mA decoration: writers, erasers, readers and functions in RNA metabolism. Cell Res. 28, 616–624 (2018). [PubMed: 29789545]
67. Shah A et al. The DEAD-Box RNA Helicase DDX3 Interacts with mA RNA Demethylase ALKBH5. Stem Cells Int. 2017, 8596135 (2017). [PubMed: 29333169]

68. Yang E-W et al. Allele-specific binding of RNA-binding proteins reveals functional genetic variants in the RNA. Nat. Commun. 10, 1338 (2019). [PubMed: 30902979]
69. Wang X et al. Discovery and validation of sub-threshold genome-wide association study loci using epigenomic signatures. Elife 5, (2016).
70. Boyle EA, Li YI & Pritchard JK An Expanded View of Complex Traits: From Polygenic to Omnigenic. Cell 169, 1177–1186 (2017). [PubMed: 28622505]
71. Zhou KI et al. Regulation of Co-transcriptional Pre-mRNA Splicing by mA through the Low-Complexity Protein hnRNPG. Mol. Cell 76, 70–81.e9 (2019). [PubMed: 31445886]
72. Garcia-Campos MA et al. Deciphering the 'mA Code' via Antibody-Independent Quantitative Profiling. Cell 178, 731–747.e16 (2019). [PubMed: 31257032]
73. Wang Y, Xiao Y, Dong S, Yu Q & Jia G Antibody-free enzyme-assisted chemical approach for detection of N-methyladenosine. Nat. Chem. Biol. (2020) doi:10.1038/s41589-020-0525-x.
74. Zhang Z et al. Single-base mapping of mA by an antibody-independent method. Sci Adv 5, eaax0250 (2019). [PubMed: 31281898]
75. Shu X et al. A metabolic labeling method detects mA transcriptome-wide at single base resolution. Nat. Chem. Biol. (2020) doi:10.1038/s41589-020-0526-9.
76. Liu H et al. Accurate detection of mA RNA modifications in native RNA sequences. Nat. Commun. 10, 4079 (2019). [PubMed: 31501426]
77. Carithers LJ et al. A Novel Approach to High-Quality Postmortem Tissue Procurement: The GTEx Project. Biopreserv. Biobank 13, 311–319 (2015). [PubMed: 26484571]
78. Batista PJ et al. m(6)A RNA modification controls cell fate transition in mammalian embryonic stem cells. Cell Stem Cell 15, 707–719 (2014). [PubMed: 25456834]
79. Langmead B & Salzberg SL Fast gapped-read alignment with Bowtie 2. Nat. Methods 9, 357–359 (2012). [PubMed: 22388286]
80. Kim D, Langmead B & Salzberg SL HISAT: a fast spliced aligner with low memory requirements. Nat. Methods 12, 357–360 (2015). [PubMed: 25751142]
81. Ongen H, Buil A, Brown AA, Dermitzakis ET & Delaneau O Fast and efficient QTL mapper for thousands of molecular phenotypes. Bioinformatics 32, 1479–1485 (2016). [PubMed: 26708335]
82. Huang QQ, Ritchie SC, Brozynska M & Inouye M Power, false discovery rate and Winner's Curse in eQTL studies. Nucleic Acids Res. 46, e133 (2018). [PubMed: 30189032]
83. Iotchkova V et al. GARFIELD classifies disease-relevant genomic features through integration of functional annotations with association signals. Nat. Genet. 51, 343–353 (2019). [PubMed: 30692680]
84. Giambartolomei C et al. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. PLoS Genet. 10, e1004383 (2014). [PubMed: 24830394]
85. Szklarczyk D et al. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Res. 47, D607–D613 (2019). [PubMed: 30476243]

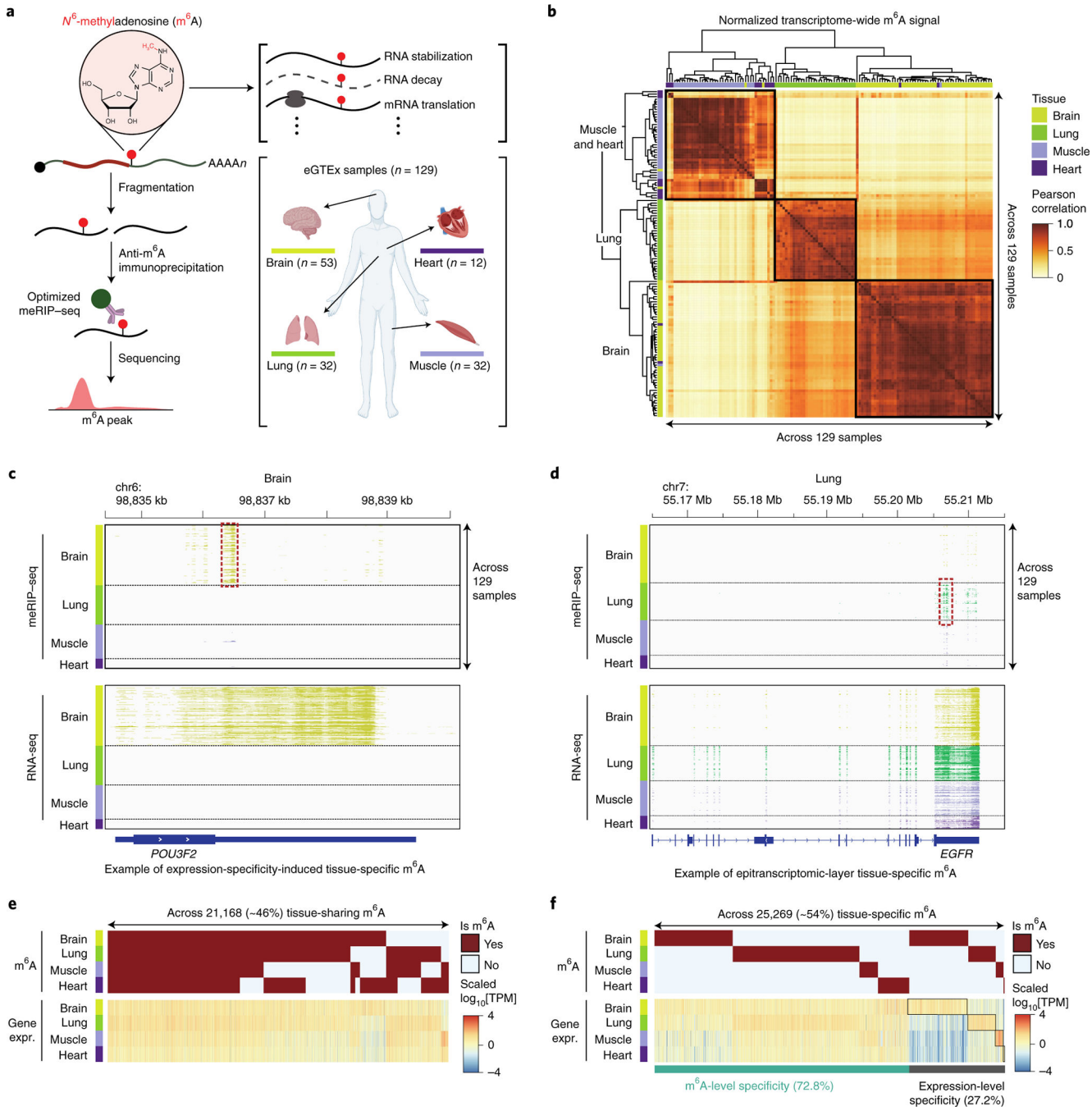


Figure 1. Study design and m^{6}A landscape across tissues.

a. Overview of m^{6}A profiling method and samples. **b.** Pearson correlation (heatmap) and hierarchical clustering (tree) of m^{6}A profiles across 129 samples (rows/columns) shows tissue-specific clustering (colors). **c.** Example of brain-specific m^{6}A (top) in *POU3F2* 3'-UTR with brain-specific RNA-seq expression (bottom). **d.** Example of lung-specific m^{6}A (top) in *EGFR* 3'-UTR despite tissue-shared expression (bottom). **d.** Tissue-shared m^{6}A (top) and corresponding expression (bottom) across peaks (columns) and tissues (rows). **f.** Tissue-specific m^{6}A (top) and corresponding expression (bottom) across peaks (columns)

and tissues (rows), showing 72.8% of tissue-specific m^6As lack tissue-specific expression (e.g. panel d) and 27.2% also show tissue-specific expression (e.g. panel c).

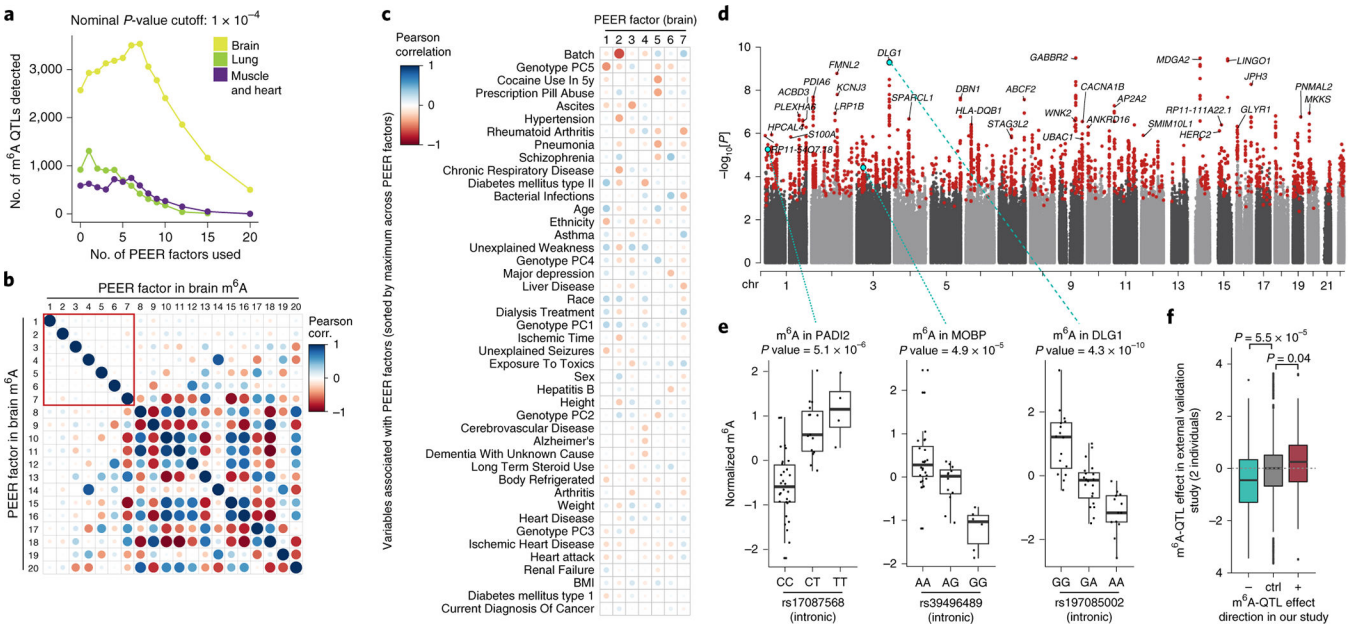


Figure 2. Genetically driven m^{6}A across tissues.

a. Number of genetically driven m^{6}A sites (nominal $P < 10^{-4}$, P value by FastQTL linear regression) identified (y-axis) for increasing number of PEER factors removed (x-axis) shows inflection-point for each tissue (colors). **b.** Pearson correlation (color) for PEER factors in brain shows saturation (correlated factors) after first 7 factors (red box). **c.** Pearson correlation between PEER factors (columns) and donor phenotypic measurements (rows) helps interpret factors. **d,e.** Overlaid Manhattan plot showing genomic position (x-axis) and association nominal P value (y-axis) for all m^{6}A-QTLs in brain, highlighting three examples. Boxes = 25%−75% percentile (i.e. inter-quartile range; IQR); line = median; whiskers = 1.5 IQR. P value by FastQTL linear regression. **f.** Minor-allele effect size in validation cohort (y-axis) for increased-effect m^{6}A-QTLs (+) and decreased-effect m^{6}A-QTLs (−) relative to non-m^{6}A-QTL (SNPs tested but not significant during m^{6}A-QTL calling, ctrl). Data from Liu et al.[26], and merged m^{6}A profiles from 2 samples of muscle, lung and heart, each. P value using two-sided Wilcoxon test. Boxes = 25%−75% percentile; line = median; whiskers = 1.5 IQR.

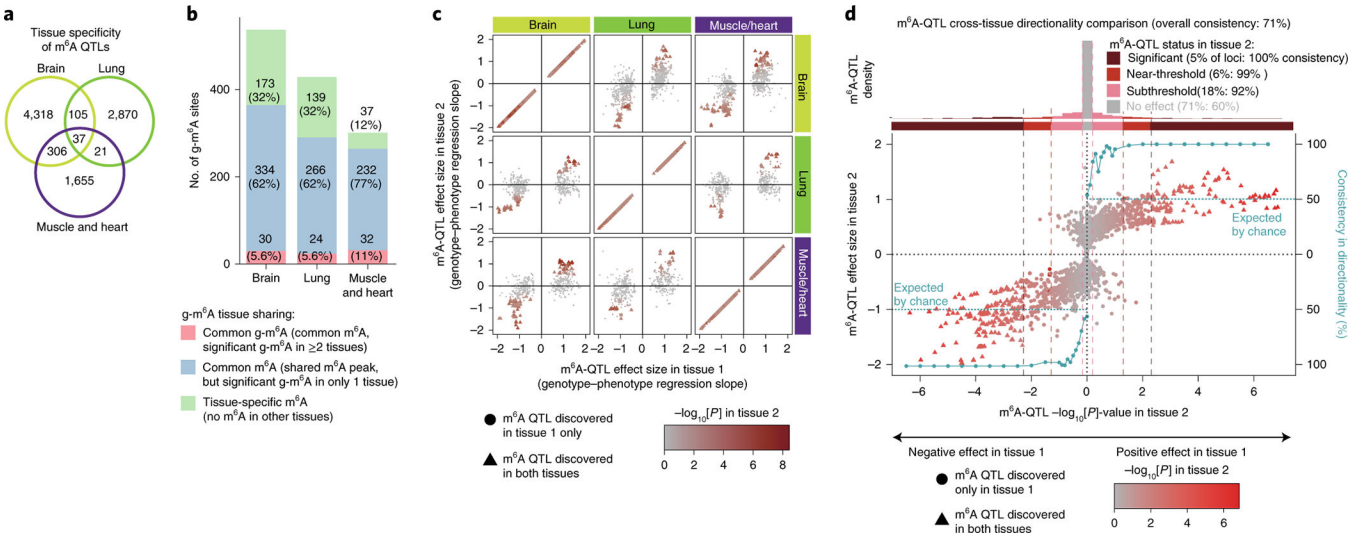


Figure 3. m^{6}A-QTL tissue specificity.

a. m^{6}A-QTLs are highly tissue-specific. **b.** Tissue-specificity of m^{6}A-QTL target peaks: among m^{6}A peaks that are m^{6}A-QTL targets (i.e. g-m^{6}A gene) 12%−32% are tissue-specific peaks (green), but even among tissue-shared peaks, only 5.0%−6.3% have shared m^{6}A-QTLs (red). **c.** m^{6}A-QTL effect size consistency: tissue-shared m^{6}A-QTLs (red triangles) show 100% consistent effect sizes between tissues (top-right and bottom-left quadrants), and even tissue-specific m6A-QTLs (circles) that are near-threshold in a second tissue (red shading) show highly consistent effects (99% agreement for $P < 10^{-2}$) for all pairs of tissues (9 panels). *P* value by FastQTL linear regression. **d.** Between-tissue consistency (teal y-axis) of m^{6}A-QTL effect size (black y-axis) increases with the significance of Tissue-2 m^{6}A-QTL *P* value (x-axis) for both positive-effect Tissue-1 m^{6}A-QTLs (right half-plane) and negative-effect Tissue-1 m^{6}A-QTLs (left half-plane). The top histogram shows the distribution of cases in each *P* value; most m^{6}A-QTLs in Tissue-1 show no effect in the second tissue (the grey bar in the middle). Percentage of loci in each group and corresponding directionality consistency in parentheses. *P* value by FastQTL linear regression.

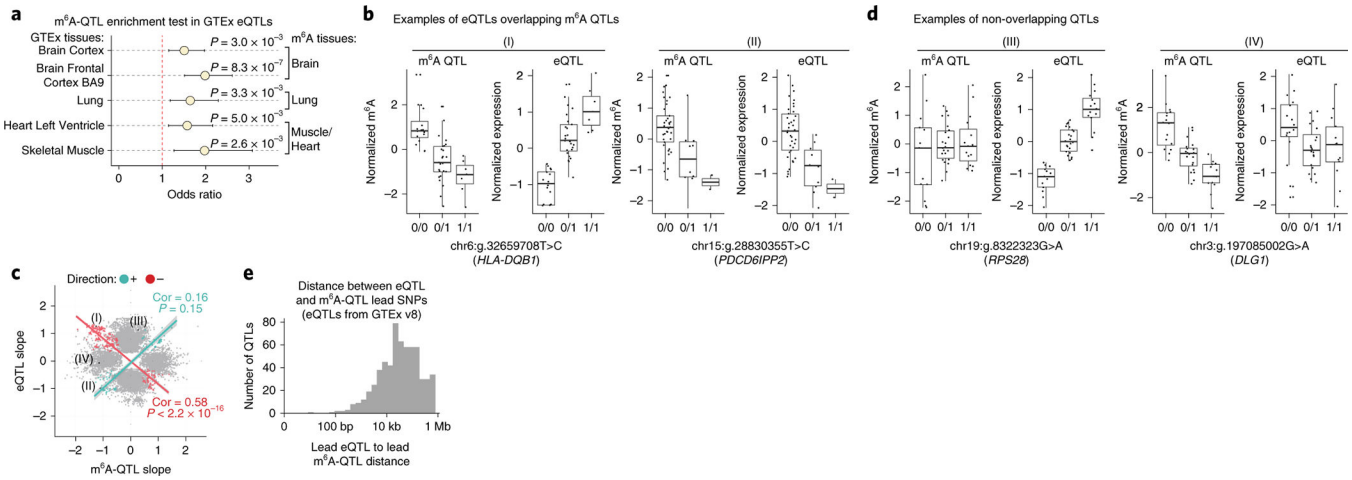


Figure 4. m^{6}A-QTL vs. eQTL comparison.

a. m^{6}A-QTLs show a modest but significant enrichment for eQTLs in matched tissues. Enrichment *P* values calculated using Garfield. NAnnot = 23,231, 18,663, 43,208, 27,398, 44,602 for Brain Cortex, Brain Frontal Cortex (BA9), Lung, Heart Left Ventricle, Muscle Skeletal (see Methods). Error bars denote upper bound and lower bound for 95% CI of effect size. *P* values by Garfield logistic regression with 'feature matching'. **b.** Examples of m^{6}A-QTLs (left) that are also eQTLs (right), with negative (top) or positive (bottom) effects on expression, indicating potential degradation or stabilization effects of m^{6}A, respectively. Boxes = 25%−75% percentile; line = median; whiskers = 1.5 IQR. **c.** Across all SNPs (dots), only a minority (colored) affect both m^{6}A (x-axis) and expression (y-axis), while most QTLs have independent m^{6}A-vs.-expression effects (grey points). **d.** Examples of an eQTL with no effect on m^{6}A (top) and an m^{6}A-QTL with no effect on expression (bottom). Boxes = 25%−75% percentile; line = median; whiskers = 1.5 IQR. **e.** Large distance between lead m^{6}A-QTL SNP and lead eQTL SNP (GTEx v8) for matching target genes indicates distinct mechanisms of action even when both QTL types are identified.

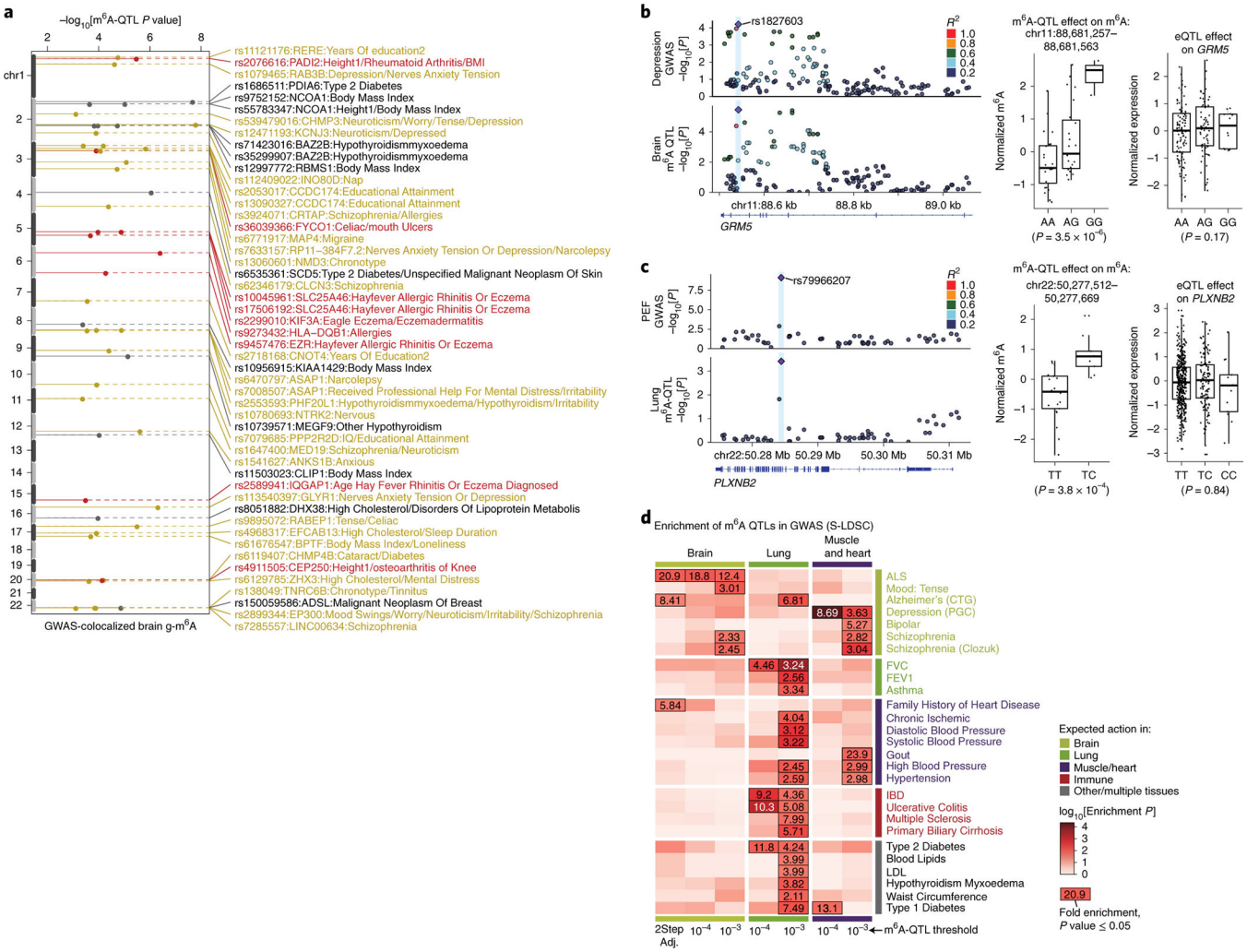


Figure 5. GWAS effects of m^{6}A-QTL.

a. Overlaid Manhattan plot showing genomic position (y-axis) and m^{6}A-QTL *P* value (x-axis) for lead SNPs (points) across traits (colors) showing colocalization of GWAS variants and m^{6}A-QTLs in brain**. b.** Example GWAS-m^{6}A-QTL co-localization for depression-associated A-to-G rs1827603, increasing m^{6}A level for *GRM5* in brain, with no effect on *GRM5* expression. Boxes = 25%−75% percentile; line = median; whiskers = 1.5 IQR. **c.** Example GWAS-m^{6}A-QTL co-localization for PEF-related T-to-C rs79966207, increasing m^{6}A level for *PLXNB2* in lung, with no effect on *PLXNB2* expression. Boxes = 25% −75% percentile; line = median; whiskers = 1.5 IQR. **d.** GWAS traits (rows) showing significant enrichment (heatmap) for m^{6}A-QTLs across tissues (columns) by stratified LD score regression (S-LDSC) grouped by enriched tissue, and colored by expected tissue of action. Enrichment *P* value is shown in a color scale, the enrichment folds are shown for those traits with $P < 0.05$. Enrichment *P* value reported by S-LDSC via z-score calculation, not adjusted for multiple tests.

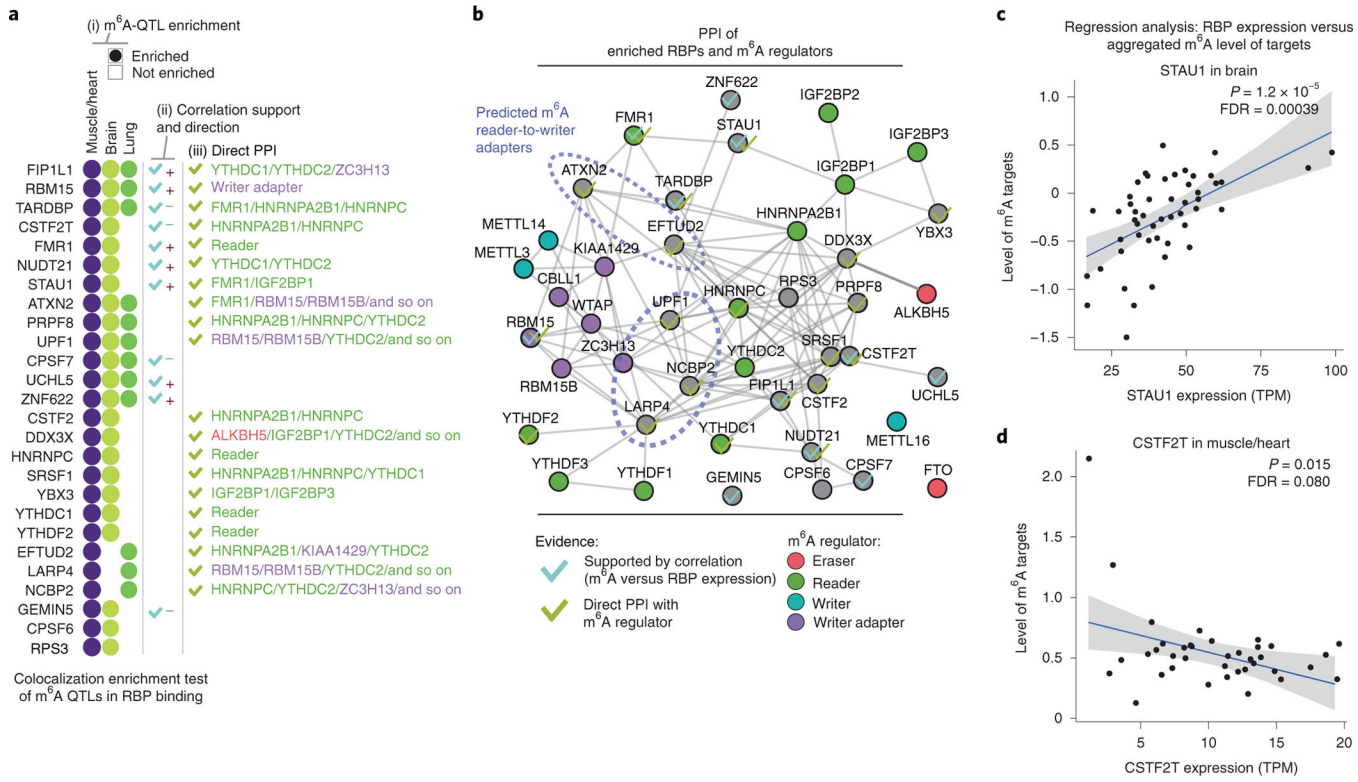


Figure 6. Predicted m^{6}A regulators.

a. Predicted m^{6}A regulators (rows) supported by three lines of evidence (checkmarks): (i) enrichment of RNA binding protein (RBP) regulator-bound sites for m^{6}A-QTLs in brain (yellow), lung (green) or muscle/heart (purple); (ii) significantly positive (red plus) or negative (green minus) correlation between RBP expression and methylation level its m^{6}A targets; (iii) experimentally validated protein-protein interaction with known m^{6}A regulators. **b.** Experimentally determined protein-protein interactions between predicted m^{6}A regulators (grey) and known m^{6}A regulators (colors) supported by enriched RBP binding in m^{6}A-QTLs (grey), RBP-expression vs. m^{6}A target level (blue checkmark), or direct interaction with known m^{6}A regulators (green check). Highlighted RBPs (dashed circle) have multiple interactions with both writer adaptors and readers suggesting potential adapter roles. **c.** Examples showing correlation of RBP expression (x-axis) vs. m^{6}A target methylation (y-axis) for predicted regulators STAU1 in brain. Grey shadows denote the 95% confidence region for the regression fit. **d.** Examples showing correlation of RBP expression (x-axis) vs. m^{6}A target methylation (y-axis) for predicted regulators CSTF2T in muscle/heart. Grey shadows denote the 95% confidence region for the regression fit.