SUPPLEMENTAL INFORMATION

Supplemental Figure Legends

Figure S1. m^6A-Containing Transcripts are Found at Varying Levels Across Individual Cell Lines and are Not Detected in Poly(A) Tails, Related to Figures 2 and 3.

A. Cell line-specific changes in the levels of m^6A indicate its dynamic nature. RNA was isolated from various mammalian cell lines and subjected to immunoblot analysis with the anti-m^6A antibody. The levels of m^6A vary substantially across different cell lines. For instance, the cancer cell lines HEPG2 and MCF7 appear to have relatively high levels of m^6A, whereas the prostate cancer cell lines PC3 and PC9 have comparatively low levels. m6A immunoblotting of total mRNA often reveals two bands of low signal intensity that are approximately 1.9 kb and 5 kb in size, which correspond to the 18S and 28S rRNA species, respectively. This likely indicates low levels of m^6A within these rRNAs.

B. Developmentally regulated increases in m^6A abundance are observed in cultured neurons. RNA was collected from cultured neurons isolated from embryonic day 18 (E18) and postnatal day 3 (P3) rat brain and subjected to immunoblot analysis with the anti-m^6A antibody. A substantial increase in m^6A abundance is observed in P3 neurons compared to E18 neurons, despite loading of less RNA in the P3 sample (indicated by ethidium bromide staining of 28S rRNA bands, bottom panel). Additionally, compared to RNA isolated from adult rat brain tissue, the m^6A content of E18 or P3 cultured neuronal RNA is significantly enriched in lower molecular weight RNA species.

C. Probing for m^6A content in isolated poly(A) tails alone indicates an absence of m^6A. To further explore whether poly(A) tails contain m^6A residues, we purified poly(A) tails from poly(A) RNA, and directly assayed for m^6A. Poly(A) RNA was isolated from total brain RNA using oligo(dT) Dynabeads. To harvest the poly(A) tails, half the sample was digested with RNase A, which cleaves after C and U residues. Whereas poly(A) tails are resistant to this treatment, the remainder of the transcript is digested (Perry et al., 1975). Dot blot analysis of total poly(A) RNA and poly(A) tails alone using anti-m^6A immunoblotting shows negligible levels of m^6A in the poly(A) tail sample.

D. To confirm the successful removal of the poly(A) tail from the remainder of the transcript in **(C)**, 3'RACE and RT-PCR were used to detect β-actin. In this experiment, the transcript is amplified only if it contains a poly(A) tail, which is utilized in the 3'RACE procedure. As expected, the target amplicon is observed only in the total poly(A) RNA sample.

E. To confirm the presence of poly(A) tails in the poly(A) tail-only sample, 3'RACE and RT-PCR were performed using a $GGA_{(18)}$ forward primer. This primer binds to the 5' end of any poly(A) tails that contain two guanine residues immediately preceding the poly(A) segment, which are not cleaved off by RNase A. When this primer is coupled with a 3'RACE-specific reverse primer and used for RT-PCR, several products of predominantly low molecular weight

are observed in the poly(A) tail-only sample (arrow, middle lane), indicating that poly(A) tails are indeed present. This result is not due to nonspecific amplification, since a control PCR lacking template cDNA has no amplified product ("No cDNA" lane).

Figure S2. MeRIP Enriches m^{6}A-Containing Targets, Related to Figures 1 and 4.

A. Two rounds of MeRIP produce optimal enrichment for m^{6}A. MeRIP assays were performed using a 1:1 mixture of methylated (containing m^{6}A) and unmethylated DNA, and enrichment was measured using qRT-PCR (see Extended Experimental Procedures). Enrichment is represented as the fold change in methylated/unmethylated DNA relative to the input sample. Following a single round of MeRIP, m^{6}A-containing DNA was enriched approximately 70-fold relative to the input sample. This enrichment increased to over 130-fold after a second round of MeRIP. A third round of MeRIP failed to further enrich for m^{6}A (K.D.M., S.R.J., data not shown).

B. *In vitro* MeRIP assays demonstrate the ability of MeRIP to enrich for m^{6}A. MeRIP assays were performed using a 1:1 mixture of unmethylated and methylated DNA species, or of unmethylated and methylated RNA species as a source of m^{6}A. Levels of enrichment for m^{6}A-containing DNA or RNA were measured by qPCR as described in Extended Experimental Procedures. As a control for nonspecific RNA immunoprecipitation, MeRIP assays were also performed using rabbit IgG in place of the m^{6}A antibody. Enrichment is shown as the fold change in methylated/unmethylated species relative to the input sample. Values represent mean $\pm$ SEM (n = 16 for m^{6}A IP; n = 12 for IgG IP).

Figure S3. MeRIP-Seq Reads Cluster as Distinct Peaks Surrounding m^{6}A Sites, Related to Figures 4-6.

A. MeRIP-Seq replicates share common m^{6}A peaks. Venn diagram showing the number of m^{6}A peaks shared among MeRIP-Seq replicates. The 13,471 m^{6}A peaks shared by all samples corresponds to the peaks in our "high-confidence" set. Sample 1 was generated with the SySy antibody and run on the Illumina Genome Analyzer II. Sample 2 (SySy antibody) and Sample 3 (NEB antibody) were run on the Illumina HiSeq2000.

B-G. UCSC Genome Browser plots of representative RNAs demonstrate the specific, non-random pattern of MeRIP-Seq peaks. Shown are examples of sequencing read peaks in the 3' UTRs of several mRNAs: **(B)** *Dlg4* (encoding PSD-95), **(C)** *Myo5A*, **(D)** *Kcnc2*, **(E)** *Nova1*, **(F)**, *Creb1*, and **(G)** *Neo1*. Sequencing reads obtained from the MeRIP sample cluster as distinct peaks (pink tracks; bottom). This pattern is not observed in reads obtained from the non-immunoprecipitated control sample (brown tracks; top). Peak height is displayed as reads per base per million mapped reads (BPM). Plots displayed are screen shots from the UCSC Genome Browser (http://genome.ucsc.edu).

Figure S4. Specificity of an Additional m^{6}A Antibody Used for MeRIP-Seq, Related to Figure 1.

A. Comparison of both m^{6}A antibodies used for MeRIP-Seq. In order to validate our MeRIP-Seq data, we sought to determine if similar MeRIP-Seq data are obtained using a separate, independently generated m^{6}A-binding antibody (generated by New England Biolabs). To establish the specificity of the antibody, we used the same tests as described for the first m^{6}A antibody. In brief, increasing amounts of a 25 nt-long oligonucleotide containing either m^{6}A or an unmodified adenosine (A) at position 13 were spotted onto a nylon membrane (from left to right: 0.1 ng, 1 ng, 10 ng, 100 ng, 1 µg). The membrane was then probed with one of two anti-m^{6}A antibodies (top panels: New England Biolabs antibody ("NEB"); bottom panels: Synaptic Systems antibody ("SySy")). Both antibodies are highly specific for m^{6}A.

B. The NEB anti-m^{6}A antibody is specific for *N*^{6}-methyladenosine and does not cross-react with other forms of methylated adenosine. Competition dot blot assays were performed on membranes spotted with 100 ng of m^{6}A-containing oligonucleotide as in **(A)**. Prior to probing the membrane, the NEB anti-m^{6}A antibody was pre-incubated with increasing amounts of *N*^{6}-methyladenosine triphosphate (*N*^{6}-MeATP), adenosine triphosphate (ATP), *N*^{1}-methyladenosine triphosphate (*N*^{1}-MeATP), or 2'-*O*-methyladenosine triphosphate (2'-*O*-MeATP). Only *N*^{6}-MeATP is able to compete with antibody binding. Concentration of competitor nucleotide used (left to right): 0 µM, 1 µM, 2 µM, 4 µM.

C. m^{6}A antibodies do not cross-react with 2'-*O*-methyladenosine. Dot blot assays were performed to directly test potential cross-reactivity of m^{6}A antibodies with the naturally-occurring nucleoside 2'-*O*-methyladenosine. Equal amounts of *in vitro* transcribed RNAs containing either m^{6}A, 2'-*O*-methyladenosine, or unmodified adenosine were spotted onto a nylon membrane and probed with the NEB and SySy antibodies. Both antibodies bound strongly to m^{6}A-containing RNA but did not display cross-reactivity to 2'-*O*-methyladenosine or adenosine-containing RNA. Amount of RNA spotted (left to right): 1.0 ng, 10 ng, 50 ng, 100 ng.

Figure S5. Validation of MeRIP-Seq Target mRNAs, Related to Figure 5.

A. MeRIP-Seq identifies *Drd1a* as an mRNA containing m^{6}A. UCSC Genome Browser tracks (http://genome.ucsc.edu) displaying read clusters from a MeRIP-Seq sample (bottom) and a non-IP control sample (top). *Drd1a* exhibits distinct m^{6}A peaks in the MeRIP sample, whereas it lacks these peaks in the non-IP sample. Peak height is displayed as reads per base per million mapped reads (BPM).

B. Confirmation of the presence of m^{6}A in *Drd1a*, an mRNA identified with MeRIP-Seq. *Drd1a* mRNA was isolated from total mouse brain RNA using a biotinylated oligonucleotide probe in an RNA pull-down. Immunoblot analysis with the anti-m^{6}A antibody was then performed to confirm m^{6}A presence in *Drd1a*. A control sample using a probe of equal size that is not specific for any known mouse mRNA (Control Probe) was run in parallel. Total mouse brain RNA (Input) is also shown as a reference for m^{6}A labeling. Arrows indicate m^{6}A-

immunoreactive band following pull-down of *Drd1a*. The size of the band is consistent with known molecular weights of *Drd1a* transcript variants (Thierry-Mieg and Thierry-Mieg, 2006).

C. MeRIP-Seq identifies *Grm1* as an mRNA m^6A-containing mRNA. UCSC Genome Browser tracks (http://genome.ucsc.edu) displaying read clusters from a MeRIP-Seq sample (bottom) and the control sample, which comprised the RNA sample prior to immunoprecipitation with the anti- m^6A antibody ("non-IP," top). *Grm1* exhibits distinct m^6A peaks along its length in the MeRIP sample, whereas it lacks these peaks in the non-IP sample. Peak height is displayed as reads per base per million mapped reads (BPM).

D. Validation of the presence of m^6A in the *Grm1* mRNA. *Grm1* mRNA was isolated from total mouse brain RNA using a target-specific, biotinylated oligonucleotide probe as in (**B**). Immunoblot analysis with anti-m^6A was subsequently performed to confirm m^6A presence. Total mouse brain RNA (Input) is also shown, as is the results of a control sample using no probe (No Probe). An m^6A immunoreactive band is observed following pull-down of *Grm1* (arrow). The size of the band is consistent with known molecular weights of *Grm1* transcript variants (Thierry-Mieg and Thierry-Mieg, 2006).

E. Immunodepletion of m^6A-containing mRNAs from complex RNA samples following m^6A immunoprecipitation. RiboMinus-treated mouse brain RNA was fragmented and subjected to m^6A immunoprecipitation. Unbound RNAs were isolated, and the abundance of target RNAs was measured by qRT-PCR. All transcripts were normalized to the amount of *Rps14* mRNA within each sample. *Rps14* was chosen because it is an abundant transcript which does not have m^6A peaks. Compared to the input RNA, the levels of *Rps21* and *Ndel1*, two transcripts which lack m^6A peaks, show only slight decreases in the unbound sample (which might be due to non-specific binding of RNA to the magnetic beads used during immunoprecipitation). However, the levels of *Drd1a*, *Grm1*, *Ptpn4*, and *Tlr3*, all transcripts which contain m^6A peaks, are dramatically decreased in the unbound RNA fraction, indicating that the m^6A antibody selectively immunodepletes these methylated transcripts from the unbound RNA pool.

Figure S6. Distribution of m^6A Peaks in Mouse Brain and HEK293T Cell RNA, Related to Figures 5 and 6.

A. Many transcripts contain adjacent m^6A peaks. The number of transcriptome-wide m^6A peaks that are contiguous with neighboring m^6A peaks is shown. Many m^6A peaks occur singly (1 peak within a cluster), although the majority of peaks are part of adjacent peak pairs (2 peaks within a cluster) or contiguous peak triplets (3 peaks within a cluster). A small number of peaks are highly clustered (4 or more peaks within a cluster). These data suggest that some transcripts contain a single region of adenosine methylation, while other transcripts are multi-methylated on several adenosine residues which cluster in distinct regions of a transcript.

B. Number of motifs found in m^6A peaks. The percentage of m^6A peaks that contain various numbers of the motifs identified in **(Figure 6B)** was determined. Only 10% of m^6A peaks lack a motif, whereas the majority of peaks (57.3%) contain one or two motifs. 28.5% of m^6A peaks

have only a single motif within their sequence, suggesting that a single m^6A residue accounts for these peaks.

C. Relationship between m^6A peak enrichment and mRNA abundance in mouse brain. Plotted is the peak enrichment value (the ratio of MeRIP sample reads to non-IP sample reads within the area of a peak, each normalized to the number of reads within the sample) relative to the abundance of the transcript within the input RNA. The RPKM (reads per kilobase per million mapped reads) of mRNAs in the non-IP sample (y-axis) provides an estimate of transcript abundance within the input RNA. The most highly enriched m^6A peaks are often observed in transcripts of low abundance.

D. Relationship between m^6A peak enrichment and mRNA abundance in HEK293T cells. The enrichment of individual m^6A peaks is plotted relative to the abundance of the transcript in which the peak resides as in (**C**). As in the mouse brain dataset, there is a tendency for highly enriched m^6A peaks to occur in weakly expressed transcripts.

Figure S7. Features of Adenosine Methylation in the Mouse and Human Transcriptomes, Related to Figures 5 and 6.

A. Distribution of m^6A peaks surrounding the CDS start site. The distribution of m^6A peaks 1 kb upstream and downstream of the CDS start sites of known RefSeq genes is shown. A steady increase in the number of peaks is observed which plateaus approximately 500 nt after the CDS start site.

B. Distribution of m^6A peaks surrounding the CDS end site. The distribution of m^6A peaks 1 kb upstream and downstream of the CDS end sites of known RefSeq genes is shown. A strong and very distinct enrichment of m^6A peaks surrounding the stop codon is observed.

C. Distribution of m^6A enrichment along the length of mRNA transcripts in mouse. Peaks that fell within gene exons were mapped to percentile locations within the 5' UTR, CDS and 3' UTR segments of the mature transcript. Shown is the sum of the enrichments of the peaks that fell within each percentile bin. m^6A enrichment is particularly strong at the 3' end of the CDS and the 5' end of the 3' UTR.

D. Transcriptome-wide distribution of HEK293T m^6A peaks. A pie chart shows the percentage of m^6A peaks within distinct RNA sequence types. m^6A is highly enriched in 3' UTRs and CDSs, similar to the pattern observed in mouse brain RNA (**Figure 5C**).

TABLES

Table S1. Filtered Set of Mouse Brain m^{6}A Peaks, Related to Figure 5. (Excel File)

Listed are the set of filtered m^{6}A peaks, which is the total list of m^{6}A peaks identified in all MeRIP-Seq replicates. Information provided includes peak coordinates (Chr; Start; End), the transcript to which the peak maps (RefSeq Accession, Name), and the location of the peak within the transcript (Peak Location).

Table S2. High-Confidence Set of Mouse Brain m^{6}A Peaks, Related to Figure 5. (Excel File)

A. The list of high-confidence m^{6}A peaks, which are present in all MeRIP-Seq replicates. Shown are the peak coordinates (Chr; Start; End), the transcript to which the peak maps (RefSeq Accession, Name), and the location of the peak within the transcript (Peak Location).

B. The list of m^{6}A peaks which overlap with ncRNAs in the FANTOM3 dataset. Information is shown as in (**A**), with the addition of the transcription start and end sites for each ncRNA, the strand of the ncRNA, and the length of the ncRNA.

C. m^{6}A peaks mapping to lincRNAs. Information presented is as described in (**A**) and (**B**).

Chr	Start	End	RefSeq Accession	Name	# m6A Peaks
chr18	34380637	34481844	NM_007462	Apc	24
chr2	121115337	121136568	NM_032393	Mtap1a	23
chr9	15714636	16182675	NM_001080814	Fat3	21
chr11	55064111	55125759	NM_001029988	Fat2	19
chr13	55311142	55419686	NM_008739	Nsd1	19
chr15	72636029	72640754	NR_002864	Peg13	18
chr5	14514917	14863459	NM_011995	Pclo	17
chr5	42178777	42235554	NM_001081422	Bod1l	17
chr6	22825501	23002916	NM_001081306	Ptprz1	17
chr14	93412919	94287951	NM_001081377	Pcdh9	17
chr10	79764564	79781001	NM_011789	Apc2	17
chr19	9063773	9093685	NM_009643	Ahnak	17
chr7	17079821	17200342	NM_172739	Grlf1	16
chr1	30859186	30920101	NM_001081080	Phf3	16
chr8	124122602	124175833	NM_080855	Zcchc14	16
chr11	117515602	117624753	NM_198022	Tnrc6c	16
chr1	20880702	20928837	NM_028829	Paqr8	16
chr4	34750358	34830197	NM_013889	Zfp292	16
chr16	91648068	91663563	NM_019973	Son	16
chr5	107926763	107941575	NM_001007574	A830010M20Rik	16

Table S3. mRNAs Containing the Greatest Number of m^{6}A Peaks, Related to Figure 5.

Listed are the top 20 mouse brain mRNAs identified by MeRIP-Seq as having the greatest number of m^{6}A sites along their length.

Chr	Start	End	RefSeq Accession	Name	Length Spanned
chr13	98016375	98016739	NM_007930	Enc1	1014
chr9	58337925	58338199	NM_176921	6030419C18Rik	1024
chr1	193732829	193733854	NM_009579	Slc30a1	1025
chr2	160774888	160775913	NM_173368	Chd6	1025
chr3	32418578	32419603	NM_144519	Zfp639	1025
chr7	95279165	95280190	NM_001081414	Grm5	1025
chr9	111294275	111295300	NM_001164659	Trank1	1025
chrX	61523268	61524293	NM_178740	Slitrk4	1025
chr16	96223704	96224729	NM_001103179	Brwd1	1025
chr2	79294197	79295205	NM_010894	Neurod1	1030
chr3	16104675	16105424	NM_172677	Ythdf3	1049
chr3	82196199	82197249	NM_001081230	Mtap9	1050
chr3	107990483	107991533	NM_146137	Amigo1	1050
chr9	16179392	16180442	NM_001080814	Fat3	1050
chr9	101003350	101004400	NM_001100451	Msl2	1050
chr11	49074547	49075597	NM_001110148	Mgat1	1050
chr18	39645651	39646701	NM_008173	Nr3c1	1050
chr1	58959643	58960718	NM_172406	Trak2	1075
chr7	51716878	51717953	NM_198250	Lrrc4b	1075
chr8	124124352	124125427	NM_080855	Zcchc14	1075
chr9	20241361	20242436	NM_011753	Zfp26	1075
chr11	22735963	22737038	NM_016888	B3gnt2	1075
chr12	93046957	93048032	NM_001039089	Sel1l	1075
chr9	110148798	110149894	NM_013884	Cspg5	1096
chr1	136644451	136645551	NM_009307	Syt2	1100
chr8	75338622	75339722	NM_010687	Large	1100
chr11	60883429	60884529	NM_010603	Kcnj12	1100
chr19	23239297	23240397	NM_010638	Klf9	1100
chr2	28084698	28085823	NM_001038613	Olfm1	1125
chr2	28084698	28085823	NM_019498	Olfm1	1125
chr13	60862120	60863245	NM_029653	Dapk1	1125
chr11	79315458	79316529	NM_019409	Omg	1140
chr2	125565321	125566471	NM_177608	Secisbp2l	1150
chr17	5342062	5343212	NM_001085355	Arid1b	1150
chr8	34496846	34498001	NM_152821	Purg	1155
chr2	168007289	168008464	NM_009628	Adnp	1175
chr6	56728743	56729918	NM_145958	Kbtbd2	1175
chr8	63149995	63151170	NM_027756	Mfap3l	1175
chrX	163911865	163913040	NM_001033330	Frmpd4	1175
chr11	20625687	20626862	NM_181411	Aftph	1175
chr3	27140122	27141322	NM_178772	Nceh1	1200
chr10	34002445	34003645	NM_009433	Tspyl1	1200
chr12	112948203	112949403	NM_027404	Bag5	1200
chr17	32910441	32911641	NM_172458	Zfp871	1200
chr4	68422805	68424030	NM_019967	Dbc1	1225
chr7	71035186	71036436	NM_021366	Klf13	1250

chr4	49598302	49599577	NM_025944	2810432L12Rik	1275
chr7	13395301	13396576	NM_026046	Zfp329	1275
chr6	8900268	8900307	NM_008751	Nxph1	1289
chr18	37304447	37305772	NM_001003672	Pcdhac2	1325
chr1	20926262	20927612	NM_028829	Paqr8	1350
chr2	67955794	67957144	NM_020283	B3galt1	1350
chr15	100870221	100871571	NM_001077499	Scn8a	1350
chr5	82223313	82224688	NM_198702	Lphn3	1375
chr9	56466831	56468206	NM_181074	Lingo1	1375
chr18	46664659	46666043	NM_173423	Fem1c	1384
chr2	140485333	140486733	NM_001172160	Flrt3	1400
chr12	42179629	42181029	NM_010733	Lrrn3	1400
chr4	11887956	11889380	NM_001098231	Pdp1	1424
chr2	83719391	83720816	NM_175514	Fam171b	1425
chr9	111174998	111176523	NM_175266	Epm2aip1	1525
chr5	58111684	58113234	NM_001122758	Pcdh7	1550
chr10	112364133	112365733	NM_001033474	Atxn7l3b	1600
chr6	77193821	77195446	NM_028880	Lrrtm1	1625
chr15	80709571	80711271	NM_144812	Tnrc6b	1700
chr2	97469495	97471220	NM_178725	Lrrc4c	1725
chr11	117582888	117584738	NM_198022	Tnrc6c	1850
chr15	72636954	72640029	NR_002864	Peg13	3075

Table S4. Transcripts with Multiple Adjacent m^{6}A Peaks, Related to Figure 5.

Shown are the 68 RNAs in mouse brain that have long ($\geq$ 1 kb) stretches of contiguous m^{6}A peaks. The genomic coordinates for each region of adjacent peaks are given (Chr;Start;End), as well as the RefSeq accession number and gene symbol for the transcript that contains each cluster of peaks. The distance that each set of contiguous peaks spans is also provided (Length Spanned). These sites of multiple contiguous m^{6}A peaks are likely to represent regions of highly clustered m^{6}A residues.

Table S5. Gene Ontology Analysis of Genes Encoding m^{6}A-Containing RNAs, Related to Figure 5. (Excel File)

Gene ontology (GO) analysis of genes encoding m^{6}A-containing RNAs was performed for biological process, cellular component, and molecular function. Shown are the GO term (Term), the number of genes that fall within each category (Count), the percentage of total genes encoding m^{6}A-containing mRNAs that fall within each category (%), the p-values obtained from GO enrichment analysis (P-Value), the RefSeq IDs of the m^{6}A-containing RNAs of each term (RefSeq), the Bonferroni corrected-p-values (Bonferroni), and the false discovery rate (FDR). Panels A-C were performed using all genes expressed in the non-IP and MeRIP samples combined as the background list. Panels D-F were performed using a list of genes randomly sampled from the mouse genome as the background list. Panels G-I were performed using the list of genes encoding m^{6}A-containing RNAs in the HEK293T dataset, with genes from the human transcriptome used as the background list.

Table S6. High-Confidence Set of HEK293T cell m^{6}A Peaks, Related to Figure 5. (Excel File)

A. The list of high-confidence m^{6}A peaks in HEK293T cells. Shown are the peak coordinates (Chr; Start; End), the transcript to which the peak maps (RefSeq Accession, Name), and the location of the peak within the transcript (Peak Location).

B. Gene symbols for the list of overlapping and unique m^{6}A peak-containing transcripts in the mouse brain and HEK293T datasets.

EXTENDED EXPERIMENTAL PROCEDURES

RNA Isolation

Adult C57BL/6 mice (age 6-16 weeks) were sacrificed by CO_2 inhalation and cervical dislocation in accordance with the Weill Cornell Medical College Institutional Animal Care and Use Committee (IACUC). RNA from various tissues was immediately isolated using TRIzol according to the manufacturer's instructions (Invitrogen). For isolation of embryonic rat brain RNA, timed-pregnant female dams were sacrificed by CO_2 inhalation and cervical dislocation, and embryos were immediately removed and their brains isolated for RNA collection. Postnatal and adult rats were sacrificed as described for adult mice. HEK293T cell RNA was isolated from 10 cm dishes of HEK293T cells using TRIzol as above. For isolation of poly(A) RNA, total mouse brain RNA was subjected to two rounds of purification using oligo(dT)-coupled magnetic beads according to the manufacturer's instructions (Invitrogen). The uncaptured RNA after two rounds was designated as the "unhybridized RNA." Lastly, for MeRIP-Seq experiments, total RNA was isolated as above and subjected to RiboMinus treatment (Invitrogen) prior to immunoprecipitation. Notably, although these samples likely contain more mature, spliced mRNA than pre-mRNA, we observed that the percentage of intronic reads in the input RNA samples is consistent with that of rRNA-depleted samples (51% of total intronic and exonic sequences were introns), indicating that intronic sequences are indeed present within the input RNA.

Antibodies

Two independently-derived antibodies generated against m^6A were used in these studies. One is a rabbit polyclonal antibody originally developed by a research group in Germany (Munns et al., 1977) and now commercially available through Synaptic Systems (SySy; Germany). The other is a separate rabbit polyclonal antibody of independent origin which was developed by a research group at New England Biolabs (NEB) (Kong et al., 2000).

Cell Culture

Cortical neurons were isolated from embryonic day 18 (E18) or postnatal day 3 (P3) rats and cultured according to established methods (Cohen et al., 2011). Briefly, cells were plated in Culture Media (Neurobasal supplemented with 1% penicillin/streptomycin, 1% Glutamax, and 2% NS21 (Chen et al., 2008)). After 3 DIV, half the media was replaced with Culture Media + 20 μM 5'-fluoro-2'-deoxyuridine (FdU) to eliminate dividing glial cells and obtain neuron-enriched cultures, and on every third day thereafter, one-third of the media was replaced with fresh Culture Media. Neurons were grown for 8-10 DIV and RNA isolated as described above. Immortalized cell lines were cultured in appropriate media (American Type Culture Collection; ATCC) and RNA was isolated as described above.

FTO Overexpression

Overexpression experiments were carried out by infecting HEK293T cells with FLAG-tagged human FTO lentivirus or no virus control. Cells were cultured for 48hr, and total RNA was isolated using TRIzol as above and then subjected to m^6A immunoblotting. Additionally, heterologous overexpression of Flag-tagged human METTL3 was carried out in HEK293T cells using Lipofectamine 2000 according to the manufacturer's instructions (Invitrogen). RNA was isolated after 24 or 48 hr as above; however, increases in m^6A levels detected by m^6A immunoblotting were inconsistent and therefore not shown.

m^6A Immunoblotting

RNA samples were quantified using UV spectrophotometry, and equal amounts were mixed 1:1 with glyoxal loading dye (Ambion) and denatured for 20 min at 50°C. Samples were then run on a 1% agarose gel for 1 h at 70 V and transferred to a nylon membrane for 2-3 h using the NorthernMax-Gly kit according to the manufacturer's instructions (Ambion). RNA was UV crosslinked to the membrane, and membranes were blocked for 1 h in 5% nonfat dry milk in 0.1% PBST (0.1% Tween-20 in 1x PBS, pH 7.4) (Blocking Buffer). Rabbit anti-m^6A antibody (SySy or NEB) or was diluted 1:1000 in 0.1% PBST and incubated on the membranes for 1 h (25°C) to overnight (4°C). Following extensive washing with 0.1% PBST, HRP-conjugated donkey anti-rabbit IgG (GE Healthcare) was diluted 1:2500 in Blocking Buffer and added to the membranes for 1 h at 25°C. Membranes were washed again in 0.1% PBST and developed with enhanced chemiluminescence (ECL; GE Healthcare).

Dot Blot Assays

Dot blots were performed essentially as described for m^6A immunoblotting (above). Two 25 nt-long oligonucleotides (Midland) were designed to contain either m^6A or A at a single internal position (5' AGTCGTTCATCT**A**GTTGCGGTGTAC 3') and were spotted onto a nylon membrane (GE Healthcare). The membrane was then UV crosslinked, blocked, and exposed to rabbit anti-m^6A antibody as described above. For competition assays, rabbit anti-m^6A antibody was pre-mixed with competitor RNA or competitor NTP for 30 min at 25°C. Competitor NTPs used were: N^6-methyladenosine triphosphate, adenosine triphosphate, N^1-methyladenosine triphosphate, and 2'-*O*-methyladenosine triphosphate (TriLink). For experiments using competitor RNA, a PCR product was amplified from rat brain cDNA using the following primers: 5' *TAATACGACTCACTATAGG*GTGTCACCAACTGGGA 3' Fwd; T7 promoter sequence is italicized, 5' ACCCTCATAGATGGGCACAG 3' Rev. RNA was *in vitro* transcribed from this PCR product using the AmpliScribe T7-Flash kit (Epicentre). NTPs were added individually to the *in vitro* transcription reaction and included GTP, CTP, UTP and either N^6-methyladenosine triphosphate (N^6-MeATP) or adenosine triphosphate (ATP). Following pre-incubation with competitor NTPs or RNA, the m^6A antibody was incubated on the membranes for 1 h at 25°C. Membranes were then washed in 0.1% PBST, followed by incubation in secondary antibody (HRP-conjugated donkey anti-rabbit IgG, diluted 1:2500 in block) for 1 h at 25°C. Membranes were again washed in 0.1% PBST and developed with ECL.

m^6A DNA Immunoblotting

Cells from from *dam*+ (DH5alpha; Invitrogen) or *dam*- (K12 ER2925; New England Biolabs) *E. coli* strains were grown overnight at 37°C in 5 ml cultures of Luria-Bertani (LB) broth. Genomic DNA was isolated from each cell line and randomly sheared using a Branson sonicator (5 rounds of 7 sec pulses at 20% amplitude with 1 sec intervals) DNA was then treated with RNase A for 30 min at 37°C to remove any RNA and quantified with UV spectrophotometry. 1 µg of *dam*+ DNA and 1.5 µg of *dam*- DNA were loaded into a 1% agarose gel and electrophoresed for 40 min at 100 V. DNA was passively transferred to a nylon membrane (GE Healthcare) for ~2 h with 20x SSC. Membrane was then UV crosslinked and blotted with anti-m^6A as described above.

Densitometry

Densitometry analysis of the relative abundance of m^6A in immunoblot experiments was performed using the ImageJ software (Abramoff, 2004). The levels of m^6A across various samples were determined relative to the levels of ethidium bromide staining of the corresponding 28S rRNA band (images of 28S rRNA band intensity were acquired prior to transferring the RNA from the gel to the membrane during immunoblotting).

m^6A Immunoprecipitation

For immunoprecipitation of RNA for MeRIP-Seq experiments, 12 µl rabbit anti-m^6A antibody (Synaptic Systems or New England Biolabs) was coupled to sheep anti-rabbit Dynabeads (Invitrogen) in 300 µl 1 M IP Buffer (1 M NaCl, 10 mM sodium phosphate, 0.05% Triton-X) for 2 h at 4°C. Beads were then washed 3 times in 300 µl 140 mM IP Buffer (140 mM NaCl, 10 mM sodium phosphate, 0.05% Triton-X). Fragmented RNA was denatured 5 min at 75°C, cooled on ice 2-3 min, and bound to antibody-coupled beads in 300 µl of 140 mM IP Buffer (2 h at 4°C). Beads were treated with 300 µl Elution Buffer (5 mM Tris-HCL pH 7.5, 1 mM EDTA pH 8.0, 0.05% SDS, 4.2 µl Proteinase K (20 mg/ml)) for 1.5 h at 50°C, and RNA was recovered with phenol:chloroform extraction followed by ethanol precipitation.

To demonstrate the ability of the m^6A antibody to immunoprecipitate targets that contain m^6A (**Figure S2**), we performed *in vitro* immunoprecipitation experiments. We mixed m^6A-containing DNA or RNA of known sequence with unmethylated DNA to achieve a heterogeneous population of both methylated and unmethylated targets, analogous to the mixture of both methylated and unmethylated RNA fragments that are used in MeRIP-Seq. The source of unmethylated DNA was a PCR product generated from rat brain cDNA using the following primers targeting β-actin: 5' TGTCACCAACTGGGACGATA 3' Fwd, 5' ACCCTCATAGATGGGCACAG 3' Rev. The source of methylated DNA was a 3 kb fragment of the pcDNA3.1 vector (Invitrogen) obtained by restriction enzyme digest with PvuII. The digested vector was grown in *dam*+ cells (DH5alpha, Invitrogen), and the presence of m^6A was confirmed with *dam*-sensitive restriction enzyme digest. For experiments using methylated RNA as a source of m^6A, total mouse brain RNA was used in the input sample. For all assays,

unmethylated DNA and methylated DNA/RNA were mixed 1:1 and used as the input sample. Immunoprecipitation using the m^6A antibody was performed as described for MeRIP-Seq (above), and m^6A target enrichment was measured with real-time quantitative PCR (qPCR; see below). For experiments using RNA as the source of m^6A, cDNA was generated from both input and immunoprecipitated samples using random hexamers and Superscript III reverse transcriptase according to the manufacturer's instructions (Invitrogen) and then subjected to real-time qPCR analysis. For control experiments, MeRIP was performed using rabbit IgG in place of the m^6A antibody. All other experimental parameters were kept the same.

Real-Time Quantitative PCR

Real-time qPCR reactions were performed using iQ SYBR Green Supermix (Bio-Rad) and an Eppendorf Mastercycler ep realplex thermocycler. The ratio of methylated DNA (pcDNA3.1 or U2snRNA cDNA) to unmethylated DNA (β-actin) was determined for each sample, and enrichment for m^6A targets was calculated relative to the input sample. All samples were run in duplicate. Primers used to amplify each target are as follows:

β-actin: 5' TGTCACCAACTGGGACGATA 3' Fwd
5' ACCCTCATAGATGGGCACAG 3' Rev

pcDNA3.1: 5' TGTAGGCGGTGCTACAGAGTTCTT 3' Fwd
5' TTTCTGCGCGTAATCTGCTGCTTG 3' Rev

U2 snRNA: 5' GGCCTTTTGGCTAAGATCAAGTGT 3' Fwd
5' GGACGGAGCAAGCTCCTATTCCAA 3' Rev

Rps14: 5' ACCTGGAGCCCAGTCAGCCC 3' Fwd
5' CACAGACGGCGACCACGACG 3' Rev

Rps21: 5' CTGCGGAGGCACGAGCTACT 3' Fwd
5' TTCCGCGGCACGTACAGGTC 3' Rev

Ndel1: 5' TGTCACCAACTGGGACGATA 3' Fwd
5' ACCCTCATAGATGGGCACAG 3' Rev

Ptpn4: 5' CCTCCCATCCCGGTCTCCACC 3' Fwd
5' GGCTGCCCATCTTCAGGGGT 3' Rev

Grm1: 5' GCCTCAGTGTGACGGTGGCC3' Fwd
5' AGCTTGCCGTCACCGACGTG 3' Rev

Drd1a: 5' TGTGTGGTTTGGCTGGGCGA3' Fwd
5' TGGAGATGGAGCCTCGGGGC 3' Rev

Tlr3: 5' TGCTCAGGAGGGTGGCCCTT 3' Fwd

5’ CGGGGTTTGCGCGTTTCCAG 3’ Rev

3’ Rapid Amplification of cDNA Ends (RACE)

RACE-ready cDNA was generated from equal amounts of total poly(A) RNA, poly(A) tail-depleted RNA, or poly(A) tail-only RNA using a 1:1 ratio of the GeneRacer oligo(dT) primer (Invitrogen) and random hexamers according to the manufacturer’s instructions (Invitrogen). PCR was carried out using Phusion High-Fidelity PCR Master Mix (Finnzymes).

RNA Pull-Down

In experiments designed to validate the presence of m^6A in specific transcripts, RNAs were selectively pulled down using biotinylated (attached at the 3’ or 5’ end) DNA probes. These probes were 50 nt in length, and were complementary to target mRNAs. The probes (300 pg) were mixed with 10 µg total mouse brain RNA in 50 µl total volume of hybridization buffer (2x SSC, 40 U/ml RNaseOUT, 300 ng/ml salmon sperm DNA). Samples were denatured 3 min at 75°C and hybridized 30 min at 37°C with occasional mixing. Meanwhile, 100 µl of MyOne T1 streptavidin-coupled Dynabeads (Invitrogen) were equilibrated by washing twice in 100 µl Binding/Washing (B&W) Buffer (5 mM Tris-HCl, 0.5 mM EDTA, 1 M NaCl), once in 0.1 M NaOH, and once in 0.1 M NaCl. Beads were then resuspended in 100 µl B&W Buffer + 1 µg yeast tRNA (Roche) and incubated at room temperature for approximately 15 min. Buffer was then replaced with 50 µl B&W Buffer, and beads were added to the probe/RNA hybridization mix and incubated 10 min at room temperature with gentle rotation. Following 3 washes in B&W Buffer, 1 wash each in 0.5x SDS, 1x SSC, and 0.2x SSC, beads were resuspended in 100 µl Elution Buffer (10 mM EDTA, pH 8.0, 95% formamide), and biotinylated probe:RNA complexes were eluted by heating for 2 min at 90°C. Eluate was then digested with RNase H for 20 min at 37°C and ethanol precipitated.

Unique Properties of MeRIP-Seq Data

Although MeRIP-Seq is conceptually similar to CLIP-Seq, there are important technical differences between the two methods which necessitate the use of unique strategies for analyzing MeRIP-Seq data. In CLIP-Seq, endogenous RNA binding protein (RBP)-RNA interactions are stabilized by UV crosslinking and then partially digested with ribonucleases, leaving a unique “footprint” of RNA that is protected from digestion by the presence of the bound protein. The protein of interest is then immunoprecipitated (along with any bound RNAs). These RNA-RBP complexes are then end-labeled, separated by SDS-PAGE, and transferred to nitrocellulose. The short RNA fragments, or “tags,” are then isolated in accordance with the predicted size of the RBP, allowing them to be separated from any non-RBP bound RNA. Finally, these short RNA tags are then subjected to high-throughput sequencing, and the location of RBP binding sites can be determined by identifying regions of the transcriptome that contain unique overlapping tags (Ule et al., 2005).

MeRIP-Seq, however, seeks to identify the location of a unique methylated base throughout the transcriptome. Instead of immunoprecipitating full-length RNA directly and then digesting it to produce protected fragments (as in CLIP-Seq), MeRIP-Seq digests the RNA *first*, then immunoprecipitates m^6A-containing RNA fragments with an antibody that recognizes m^6A. These fragments are then subjected to high-throughput sequencing, and are analogous to the tags sequenced in CLIP-Seq studies. The method for RNA fragmentation in MeRIP-Seq follows the recommendations of the sequencing platform used (Illumina) and digests RNAs to a range of sizes approximately 100 nt long. Therefore, one or more m^6A residues could lie within each 100 nt-long tag that is sequenced. In contrast, CLIP-Seq tags are defined by the small regions in RNA that are protected from digestion by a bound RBP and thus identify a narrow region of RNA in which a given RBP binds.

This difference in the way CLIP-Seq and MeRIP-Seq tags are generated is important, because it translates to the use of different methods for identifying areas of CLIP-Seq and MeRIP-Seq tag clustering throughout the transcriptome. Both techniques are susceptible to non-specific immunoprecipitation of RNA, which can occur when RNA is binds to beads or other surfaces used in immunoprecipitation protocols. These non-specifically bound RNAs are presumably degraded by the ribonuclease digestion step in CLIP-Seq and are further excluded following the SDS-PAGE and subsequent size selection steps. However, such a step does not exist in the MeRIP-Seq protocol, as the addition of ribonucleases would lead to the degradation of m^6A-containing RNAs as well. Thus, MeRIP-Seq data requires the pool of RNA prior to m^6A immunoprecipitation to be sequenced in parallel. This population of RNA provides a measure of the abundance of each individual transcript prior to immunoprecipitation; thus, a comparison of tags from m^6A immunoprecipitation (the MeRIP sample) to those from RNA prior to immunoprecipitation (the non-IP sample) is necessary to distinguish the MeRIP tags that are significantly enriched due to recognition by the m^6A antibody from those that are randomly immunoprecipitated. In addition, because MeRIP-Seq seeks to uncover information regarding the frequency of adenosine methylation (see below), it is necessary to take into account the abundance of individual transcripts prior to immunoprecipitation. Because m^6A immunoprecipitation enriches for m^6A-containing transcripts, and thus changes the abundance of individual RNA fragments in the MeRIP sample, analyzing non-IP sample tags is necessary to determine the abundance of individual transcripts.

Identification of m^6A Peaks Genome-Wide

In order to identify regions that contain m^6A, we developed a method for detecting MeRIP-Seq read peaks according to their genomic annotations. To do this, we divided the entire genome (mm9 and hg19 builds) into 25 nt-wide discrete, non-overlapping windows and compared the number of reads in the MeRIP sample to the number of reads in the non-IP (control) sample. Since the non-IP sample was generated from the same initial pool of RNA as the MeRIP sample, it serves as a measure of the abundance of individual transcripts that were in the MeRIP sample prior to immunoprecipitation. Thus, by comparing the number of reads in the MeRIP sample to those in the non-IP sample, we minimized any biases that might be caused by non-random fragmentation of the RNA or by variability in the abundance of individual transcripts.

We compared the number of reads that mapped to a given window for the MeRIP sample and the non-IP sample to the total number of reads in each, and used Fisher's exact test to determine the probability of observing this under the null hypothesis for each window (the p-value). Fisher's exact test is non-parametric and makes no assumptions about the model underlying the data. To account for the multiple testing hypothesis with such a large number of independent statistical tests (i.e., the large number of individual windows), we used Benjamini-Hochberg to adjust the p-values to reduce our false discovery rate to 5%. We defined a window as significant when the adjusted p-value was ≤ 0.05 for each replicate. We then used Fisher's Method to combine p-values across replicates to calculate a final p-value for the window. This strategy allowed us to identify regions of MeRIP read clusters at high resolution (25 nt), while simultaneously filtering out those windows that reach significance by chance, because of artifacts in the data or because of sequencing errors. Our analysis revealed 93,074 significant 25 bp windows for all three replicates in the mouse tissue samples and 440,910 in the HEK293T tissue samples.

In order to identify m^6A peaks throughout the genome, we next had to determine the sites at which these significant 25 nt-wide windows clustered together to form distinct peaks. The size of the individual RNA fragments from which all samples were prepared was approximately 100 nt. For each immunoprecipitated fragment, an m^6A residue could technically exist at the 5'-most base of the fragment or at the 3'-most base of the fragment. Therefore, a single m^6A residue could be part of immunoprecipitated RNA fragments that at their extremes contain bases 100 nt upstream or 100 nt downstream of the actual m^6A site. Thus, when identifying m^6A peaks, we predicted that they would be approximately 200 nt wide at their base (**Figure 4**).

To determine the 200 nt-wide regions of significant m^6A enrichment, we concatenated adjacent significant windows from our 25 nt analysis and filtered out concatenated windows that were smaller than 100 bp wide. We reasoned that because the RNA in each sample was sheared to approximately 100 nt-long fragments, MeRIP reads which clustered around m^6A sites would be at the highest density within the central 100 nt-wide region of the peak. Thus, we chose 100 nt as our minimum size of the concatenated windows required for peak definition. In some cases, the length of the concatenated windows spanned >200 nt. In such cases, these regions were considered to have *n* m^6A peaks, where *n* corresponds to the minimum number of m^6A peaks that could result in a concatenated window of that length. Using this method for peak calling, we identified 41,072 peaks total across all MeRIP-Seq mouse samples and 57,236 peaks total across HEK293T samples. Of these peaks, 13,471 m^6A were significant in all three mouse brain samples and 18,756 in all three HEK293T samples.

The stringent filtering criteria outlined above allowed us to minimize the false discovery rate of our method for identifying m^6A sites. The drawback to this approach, of course, is that the final list of high-confidence m^6A sites we identified is likely to be an underestimate of the true number of m^6A sites throughout the genome. The 41,072 significant windows identified from our 25 nt analysis (above) likely contain many valid m^6A sites, and we have provided the list of coordinates for each of these significant windows (**Table S1**).

Determining m^6A Peaks Across the Transcriptome

Determining m^6A peaks across the genome (above) was necessary for identifying regions of m^6A localization within intronic and intergenic regions. However, to determine the enrichment and clustering of m^6A peaks within individual transcripts, we also identified m^6A peaks across the transcriptome. To do this, we split each RefSeq exon into windows approximately 25 nt in size, using all known annotated transcript forms of each gene. The actual size of the window was computed by counting how many 25 nt windows would be needed to span each exon, and then distributing the base pairs evenly across those windows, to avoid the creation of small windows at the end of each exon. We then used TopHat (Trapnell et al., 2009) to align the original sequenced data to the genome, with RefSeq exon-exon splice junctions used for the known exon junctions parameter. The split size was also set to half of the sequence length of Sample 1 (16 nt), which had a smaller sequenced length than the other samples. The default split size (25 nt) was used for the other replicates.

We then determined the number of reads within each replicate that mapped to each window by using BEDTools' intersectBed. As with our genome-wide peak identification (above), we then compared the number of reads in the MeRIP sample to the number in the non-IP sample within each window and used Fisher's exact test to compute p-values for the windows of each replicate. These p-values were then adjusted using Benjamini-Hochberg, and only those windows with p-values less than or equal to 0.05 in all samples were kept. Next, windows were concatenated, and those that did not join to span contiguous regions at least 100 bp in length across mature, spliced transcripts were filtered out. The remaining windows were split into peaks between 100-200 nt in size. This method identified 23,924 peaks across the transcriptome, which overlapped with 93% of the peaks in our high-confidence set (above).

We then used the mergeBed program from BEDTools (Quinlan and Hall, 2010) to join these peaks into contiguous regions across the genome and then re-split them into individual 100-200 nt peaks. This method allowed us to remove redundant peaks which mapped to the same area of two or more transcript variants of a given gene, and it resulted in a total of 17,830 m^6A peaks. To determine the overlap between these transcriptomics peaks and genomically-defined peaks in our "high-confidence" set (above), we used BEDTools' intersectBed. We required that each peak must overlap at least 50% with another peak, setting the –f parameter to 0.5.

Estimation of False Discovery Rate for Identification of m^6A Peaks

Analyses of both the BWA-aligned sequences across the genome and the TopHat-aligned sequences across the transcriptome are susceptible to the multiple testing problem, which is caused by the large number of 25 nt windows being independently tested in each analysis. To account for this, we used the Benjamini-Hochberg method for error correction, which seeks to estimate the threshold at which a certain false discovery rate (FDR) is achieved. We chose to use an FDR of 0.05, and the p-values were adjusted accordingly per sample. However, it is likely that the FDR for our high-confidence set of m^6A peaks is actually lower that 0.05, because of the numerous filtration steps used to obtain this list of m^6A peaks (above). First, a window was only considered significant if the adjusted p-value was less than or equal to 0.05 in all three replicates. Second, we only used significant windows that continuously spanned a region of 100 nt or more

when joined. Thus, while the FDR is estimated to be around 0.05 *per sample*, it is likely less for our list of high-confidence peaks.

Annotation of m^{6}A Peaks

Peaks were annotated by applying BEDTools' intersectBed in a tiered fashion. First, RefSeq gene annotations were split into two subsets, those for protein coding genes and those for non-protein coding genes. The tiered system first mapped peaks to protein coding gene coding sequences, UTRs, exons, introns, and finally full genes, in that order. Those that mapped 90% were given priority, then 50%, and finally any bp overlap. Then, the same was performed on non-protein coding gene annotations, in the same order. Duplicate mappings were removed such that a peak was mapped only once to any given RefSeq gene annotation. Because individual m^{6}A peaks often mapped to multiple transcript variants of the same RNA, we used only one transcript variant and RefSeq accession number per gene when generating our list of 4,654 unique genes in the high-confidence set of m^{6}A peaks. Additionally, we reported only one transcript variant per gene in Table 1 and Table S3, which list the m^{6}A peaks from genes with the greatest enrichment and the greatest number of m^{6}A peaks, respectively.

Distribution of m^{6}A Peaks and Samples

The peak annotations from above were then compiled into the pie chart distributions for the mouse brain peaks (**Figure 5C**) and the HEK293T peaks (**Figure S7D**). The distribution for the control data sets was computed in a similar tiered fashion, but by comparing RefSeq annotations against the original control datasets. These percentages were then averaged across the replicate controls.

Analysis of m^{6}A Peak Distribution Along an mRNA

First, a subset of the RefSeq gene annotations was derived by taking only one transcript variant of each gene. Next overlapping transcript variants were removed from the set, to reduce any ambiguity in determining which transcript a peak is from. Peaks were then mapped into this single-transcript-variant non-overlapping RefSeq subset. If the peak fell within a gene exon, then its position within the mature transcript was calculated using the exon lengths. This was then converted to a position within the 5' UTR, the coding sequence, or the 3' UTR segments, and divided by the length of that region and multiplied by 100 to determine a percentile for where this peak fell. The percentile bin that the peak fell into was then incremented, and the bins were plotted as a percentage of the total number of peaks in the dataset.

For plotting m^{6}A enrichment (**Figure S7C**), BEDTools' intersectBed was used to first calculate the number of reads that mapped to each peak, for each sample and replicate, which was then compiled into a single file to store peak read counts. A similar procedure was then performed on all RefSeq gene exons, and then tabulated by gene to get read counts from all samples for mature transcripts. The peak enrichment was computed for each peak by dividing the number of MeRIP reads by the number of control reads that mapped to that peak, each normalized for the total number of reads that were mapped, for each replicate, and then averaged across the three

replicates. Peaks that had control RPKMs of less than 1 or that were in genes that had control RPKMs of less than 1 were filtered out. These are still peaks, given their high number of reads in the MeRIP samples, but the lack of reads in the control skews the enrichment score. The peaks were then mapped to percentile bins for the 5' UTR, CDS, and 3' UTR regions as above. The purpose of this plot was to show the distribution of potential m^6A sites and their cumulative enrichment, and so the sum of the enrichment scores at bin was used to accurately both the peak enrichment and the number of peaks in each bin.

Determining the Most Frequently Methylated m^6A Peaks

To determine the frequency of methylation at individual m^6A peaks, we calculated the ratio of the number of reads in the MeRIP sample within the region defined by each m^6A peak, normalized by the total number of reads mapped to the genome in that sample, to the RPKM of the gene that the peak resides in. This ratio was averaged across all three replicates, and then shown on a $\log_2$ scale. This method allowed us to determine the relative frequency of methylation at a given m^6A peak. However, m^6A peaks may be due to the presence one or more m^6A residues. Therefore, our determination of the m^6A peaks with high degrees of methylation could reflect either the stoichiometry of a single m^6A residue, or a cluster of highly adjacent m^6A residues, each with potentially low or varying stoichiometry. In many cases, single MeRIP-Seq peaks contained only one m^6A consensus motif, suggesting a single methylation site (**Figure S6B**); however until m^6A sites can be determined transcriptome-wide with single nucleotide resolution, it is impossible to know for sure whether an m^6A peak corresponds to a single or multiple m^6A residues

m^6A Enrichment vs RPKM

To compute the enrichment of human and mouse m^6A peaks, we divided the normalized number of MeRIP reads by the normalized number of control reads that mapped to each peak, averaged across replicates. The RPKM was computed for the gene transcript that the peak fell into, averaged across control replicates.

Distribution of m^6A Surrounding CDS Start and End Sites

Using the filtered subset of RefSeq annotations that had only one transcript variant per gene and no overlapping regions, a .bed file of 10-bp windows 1kb upstream and downstream of both the coding sequence start and end sites was created. Windows were generated with the transcriptome coordinates of the specific transcript, taking into consideration the length limitations of each transcript. For example, the coding sequence start windows would stop at the beginning of the 5' UTR and the end of the coding sequence. These transcriptome windows were then translated into genomic coordinates and the peaks were translated as single bp points at the center of each peak. BEDTools' intersectBed was used to count the number of peaks that fell into each window. The peak counts were then tabulated across all genes into two sets of 200 bins that represented 1kbp upstream and downstream of both the CDSs and the CDSe. The bins were plotted with 100 bins on each side of the CDSs or CDSe, and the center point was computed as the average of the adjacent bins.

MeRIP-Seq Gene Ontology

Gene ontology (GO) analysis was performed using the DAVID bioinformatics database (Huang da et al., 2009a, b). GO classification for cellular component, biological process, and molecular function were performed at default settings. To provide additional validation of the results, two separate analyses were performed using two different lists of genes as background for the mouse brain dataset: 1) the list of genes expressed in all MeRIP and non-IP samples combined and 2) a list of random genes taken from the mouse transcriptome.

Evolutionary Conservation and Motif Statistics

Analysis of phylogenetic conservation was done by comparing PhyloP (Pollard et al., 2010) scores of m^6A peaks to those same peaks randomly shuffled within gene exons using BEDTools (Quinlan and Hall, 2010). PhyloP scores were computed for each using completeMOTIFs (http://cmotifs.tchlab.org/), which uses the phastCons scores from vertebrates. Significant differences in the distributions of the PhyloP scores were determined with the Kolmogorov-Smirnov (K-S) test in the R programming environment, using the stats library package. Motif analysis was done using FIRE (Elemento et al., 2007) with default RNA analysis parameters. For motif analysis, the sequence under the peaks located in RefSeq mRNAs (downloaded June 2010) were extracted and converted to the appropriate strand. MicroRNA analyses were performed using custom scripts and TargetScan miRNA predictions.

Analysis of m^6A Localization to Splice Junctions

The number of m^6A peaks found at exon-exon junctions was determined by overlapping the set of 14,416 transcriptome-wide m^6A peaks that fall within CDSs with exon-exon junctions compiled from known RefSeq exons. We also generated four different sets of control "peaks" which were used to establish a background level of overlap with exon-exon junctions. These control sets included 1) randomly generated peaks 2) upstream adjacent region peaks 3) downstream adjacent region peaks and 4) mixed adjacent region peaks.

The set of random control peaks was generated with BEDTools' shuffleBed program to randomly shuffle regions of the same size as the m^6A CDS peaks throughout coding sequence exons. Peaks were shuffled only to exons on the same chromosome, and the shuffleBed program was modified so that it would retain the transcript of the new exon to which it was mapped. By default, shuffleBed allows the new random peak to extend beyond the end of the exon; therefore, we further modified the code to allow it to map peaks up to 50 nt upstream of the start of an exon. To make this a fair comparison, the code was also modified to allow it to map peaks up to 50 nt upstream of the start of the exon. Our peaks are between 100-200 nt in size, so 50 nt is enough to allow a peak to cross the 5' junction but not so much that a peak would end up being mapped completely out of the exon. These shuffled peaks were then mapped to transcriptome coordinates, (the coordinates of mature transcripts for individual transcript variants of a gene).

The adjacent region control sets of peaks were generated by taking the regions immediately 5' (the upstream adjacent regions set), 3' (the downstream adjacent regions set), or either 5' or 3'

(the mixed adjacent regions set) to each m^6A peak within a CDS. The size of each control peak matched that of the adjacent m^6A peak which was located either up- or downstream. Additionally, if the region adjacent to an m^6A peak contained another m^6A peak, the next available adjacent region which did not contain an m^6A peak was used. This step ensured that the control peaks were adjacent to, but not overlapping with, m^6A peaks,

After generating all four sets of control peaks, the number of exon-exon junctions that overlapped with the peaks within each set was determined as above.

Poly(A) Site Analysis

To determine the degree of overlap between poly(A) cleavage sites and m^6A peaks within 3'UTRs, we used a list of known poly(A) cleavage sites (Brockman et al., 2005) and examined whether 50 nt regions upstream of each cleavage site overlapped with the regions of m^6A peaks in 3'UTRs. We also examined the overlap between these poly(A) cleavage sites and randomly generated regions in the same 3' UTRs. These random regions were generated by using the BEDTools shuffleBed program with RefSeq 3' UTR as the inclusion regions and the chromosome flag set. Using shuffleBed, 100 different sets of random peaks were generated and the average of the number of intersections with polyA sites. BEDTools' intersectBed was used to determine the overlapping regions, with the –f flag set to 0.2 to require that at least one fifth of each window (20-40 nt) overlap with a polyA site. The peaks were shuffled a total of 100 times, and the average of the total number of overlaps with polyA sites was used for the random counts.

MicroRNA Expression Analysis

A wildtype mouse brain miRNA expression profile was downloaded from http://www.microrna.org/. Mouse miRNA TargetScan target predictions were downloaded from http://www.targetscan.org/mmu_60/ and mapped to RefSeq transcript 3' UTRs. m^6A peaks were mapped to the same RefSeq transcript 3' UTRs. For each miRNA, we determined the number of target transcripts, i.e. the number of 3' UTRs with at least 1 miRNA target. For each miRNA, we also determined the number of m^6A peaks found in all 3' UTRs targeted by the miRNA. We then calculated the ratio between the number of m^6A peaks and the number of target transcripts for each miRNA, so as to obtain an average number of m^6A peaks per target 3' UTR. Using this miRNA expression profile, we then identified the 25 least and 25 most expressed miRNAs in mouse brain and compared average numbers of m^6A peaks per target 3' UTR for these two miRNA groups using Wilcoxon tests and boxplots.

Comparison of Mouse Brain and HEK293T Datasets

RefSeq gene annotations were used to compare the peaks found in the mouse brain tissue samples to those found in the HEK293T tissue cell line. The peaks were matched with RefSeq annotations, with priority given to coding sequences, then the UTRs, exons, introns, and lastly, full genes, using BEDTools' intersectBed. If a peak mapped to more than one CDS exon, for

example, all matches were kept. Using Microsoft Excel, the official gene symbols present in each peak set were tabulated and compared to determine the gene symbols present in both datasets.

Analysis of m^6A-containing Transcripts

The pattern of m^6A peaks gives hints as to the stage at which the RNA can become methylated. We observe a small percentage of m^6A peaks within intronic regions, which suggests that at least some mRNAs are methylated as immature pre-mRNAs within the nucleus. The methylation of pre-mRNAs is consistent with the nuclear localization of MT-A70, the adenosine methyltransferase (Bokar et al., 1997). However, we also find that 5% of m^6A peaks throughout the transcriptome contain reads that span an exon-exon junction, indicating that mature, spliced transcripts contain m^6A and suggesting that m^6A might have roles in mRNA processing events that occur both within and outside of the nucleus.

A-to-I Editing Site Analysis

We sought to explore whether adenosine methylation serves to regulate the conversion of adenosine to inosine, which is mediated by ADAR enzymes. ADARs exhibit markedly reduced activity towards m^6A compared to adenosine (Veliz et al., 2003), raising the possibility that adenosine methylation may act as a regulatory mechanism to control editing. We compared m^6A peaks to a list of 2,545 total A-to-I editing sites identified in mouse and human (Bahn et al., 2011; Enstero et al., 2010; Li et al., 2011; Maas et al., 2011; Neeman et al., 2006; Sakurai et al., 2010; Wahlstedt et al., 2009). Human A-to-I editing sites were converted to mm9 genomic coordinates using LiftOver from the UCSC Genome Browser (Rhead et al., 2010). We used genomic regions defined by m^6A peaks in our high-confidence set to look for overlaps with A-to-I sites. The random regions used as a reference for background overlap were generated using BEDTools' shuffleBed program (with the –incl flag set to known RefSeq genes). Random regions were set to be the same size as the regions of m^6A peaks and were investigated for overlaps with A-to-I sites. This process was repeated for 100 permutations of random regions, and the average number of overlaps with A-to-I sites was determined.

This analysis revealed that only 10 of these A-to-I sites overlapped with m^6A peaks, compared to an average of 8.25 overlaps the control regions, indicating that m^6A peaks are not significantly overrepresented at A-to-I editing sites. ($p=0.54$; chi-square test).

Although the presence of m^6A would technically inhibit A-to-I editing, and therefore potentially explain this lack of association, both A-to-I editing and m^6A peaks exhibit substoichiometric modification (Iwamoto et al., 2005; Narayan and Rottman, 1988; Rana and Tuck, 1990). Therefore, the absence of a correlation between these two modifications is unlikely to reflect a complete inhibition of A-to-I editing by adenosine methylation. Nevertheless, because a given adenosine could be methylated or deaminated at very low levels, we cannot rule out the possibility that m^6A inhibits A-to-I editing at some sites.

SUPPLEMENTAL REFERENCES

Abramoff, M.D., Magelhaes, P.J., Ram, S.J. (2004). Image Processing with ImageJ. Biophotonics International *11*, 36-42.

Bahn, J.H., Lee, J.H., Li, G., Greer, C., Peng, G., and Xiao, X. (2011). Accurate identification of A-to-I RNA editing in human by transcriptome sequencing. Genome Res.

Bokar, J.A., Shambaugh, M.E., Polayes, D., Matera, A.G., and Rottman, F.M. (1997). Purification and cDNA cloning of the AdoMet-binding subunit of the human mRNA (N6-adenosine)-methyltransferase. RNA *3*, 1233-1247.

Brockman, J.M., Singh, P., Liu, D., Quinlan, S., Salisbury, J., and Graber, J.H. (2005). PACdb: PolyA Cleavage Site and 3'-UTR Database. Bioinformatics *21*, 3691-3693.

Chen, Y., Stevens, B., Chang, J., Milbrandt, J., Barres, B.A., and Hell, J.W. (2008). NS21: re-defined and modified supplement B27 for neuronal cultures. J Neurosci Methods *171*, 239-247.

Cohen, M.S., Bas Orth, C., Kim, H.J., Jeon, N.L., and Jaffrey, S.R. (2011). Neurotrophin-mediated dendrite-to-nucleus signaling revealed by microfluidic compartmentalization of dendrites. Proc Natl Acad Sci U S A *108*, 11246-11251.

Elemento, O., Slonim, N., and Tavazoie, S. (2007). A universal framework for regulatory element discovery across all genomes and data types. Mol Cell *28*, 337-350.

Enstero, M., Akerborg, O., Lundin, D., Wang, B., Furey, T.S., Ohman, M., and Lagergren, J. (2010). A computational screen for site selective A-to-I editing detects novel sites in neuron specific Hu proteins. BMC Bioinformatics *11*, 6.

Huang da, W., Sherman, B.T., and Lempicki, R.A. (2009a). Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res *37*, 1-13.

Huang da, W., Sherman, B.T., and Lempicki, R.A. (2009b). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nature protocols *4*, 44-57.

Iwamoto, K., Bundo, M., and Kato, T. (2005). Estimating RNA editing efficiency of five editing sites in the serotonin 2C receptor by pyrosequencing. RNA *11*, 1596-1603.

Kong, H., Lin, L.F., Porter, N., Stickel, S., Byrd, D., Posfai, J., and Roberts, R.J. (2000). Functional analysis of putative restriction-modification system genes in the Helicobacter pylori J99 genome. Nucleic Acids Res *28*, 3216-3223.

Li, M., Wang, I.X., Li, Y., Bruzel, A., Richards, A.L., Toung, J.M., and Cheung, V.G. (2011). Widespread RNA and DNA Sequence Differences in the Human Transcriptome. Science.

Maas, S., Godfried Sie, C.P., Stoev, I., Dupuis, D.E., Latona, J., Porman, A.M., Evans, B., Rekawek, P., Kluempers, V., Mutter, M., *et al.* (2011). Genome-wide evaluation and discovery of vertebrate A-to-I RNA editing sites. Biochem Biophys Res Commun *412*, 407-412.

Munns, T.W., Liszewski, M.K., and Sims, H.F. (1977). Characterization of antibodies specific for N6-methyladenosine and for 7-methylguanosine. Biochemistry *16*, 2163-2168.

Narayan, P., and Rottman, F.M. (1988). An in vitro system for accurate methylation of internal adenosine residues in messenger RNA. Science *242*, 1159-1162.

Neeman, Y., Levanon, E.Y., Jantsch, M.F., and Eisenberg, E. (2006). RNA editing level in the mouse is determined by the genomic repeat repertoire. RNA *12*, 1802-1809.

Perry, R.P., Kelley, D.E., Friderici, K., and Rottman, F. (1975). The methylated constituents of L cell messenger RNA: evidence for an unusual cluster at the 5' terminus. Cell *4*, 387-394.

Pollard, K.S., Hubisz, M.J., Rosenbloom, K.R., and Siepel, A. (2010). Detection of nonneutral substitution rates on mammalian phylogenies. Genome Res *20*, 110-121.

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841-842.

Rana, A.P., and Tuck, M.T. (1990). Analysis and in vitro localization of internal methylated adenine residues in dihydrofolate reductase mRNA. Nucleic Acids Res *18*, 4803-4808.

Rhead, B., Karolchik, D., Kuhn, R.M., Hinrichs, A.S., Zweig, A.S., Fujita, P.A., Diekhans, M., Smith, K.E., Rosenbloom, K.R., Raney, B.J., *et al.* (2010). The UCSC Genome Browser database: update 2010. Nucleic Acids Res *38*, D613-619.

Sakurai, M., Yano, T., Kawabata, H., Ueda, H., and Suzuki, T. (2010). Inosine cyanoethylation identifies A-to-I RNA editing sites in the human transcriptome. Nat Chem Biol *6*, 733-740.

Thierry-Mieg, D., and Thierry-Mieg, J. (2006). AceView: a comprehensive cDNA-supported gene and transcripts annotation. Genome Biol *7 Suppl 1*, S12 11-14.

Trapnell, C., Pachter, L., and Salzberg, S.L. (2009). TopHat: discovering splice junctions with RNA-Seq. Bioinformatics *25*, 1105-1111.

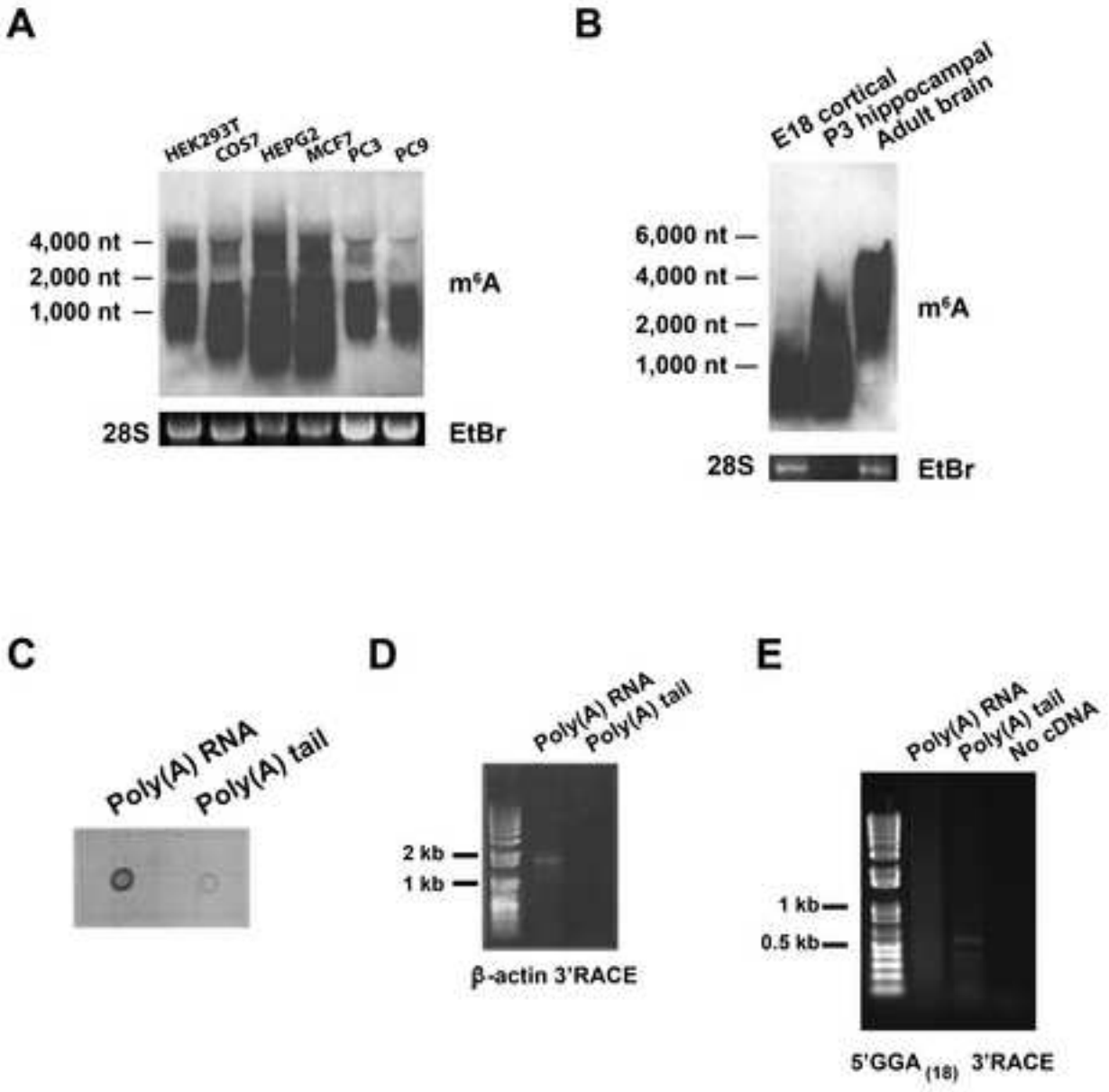
Ule, J., Jensen, K., Mele, A., and Darnell, R.B. (2005). CLIP: a method for identifying protein-RNA interaction sites in living cells. Methods *37*, 376-386.

Veliz, E.A., Easterwood, L.M., and Beal, P.A. (2003). Substrate analogues for an RNA-editing adenosine deaminase: mechanistic investigation and inhibitor design. J Am Chem Soc *125*, 10867-10876.

Wahlstedt, H., Daniel, C., Enstero, M., and Ohman, M. (2009). Large-scale mRNA sequencing determines global regulation of RNA editing during brain development. Genome Res *19*, 978-986.

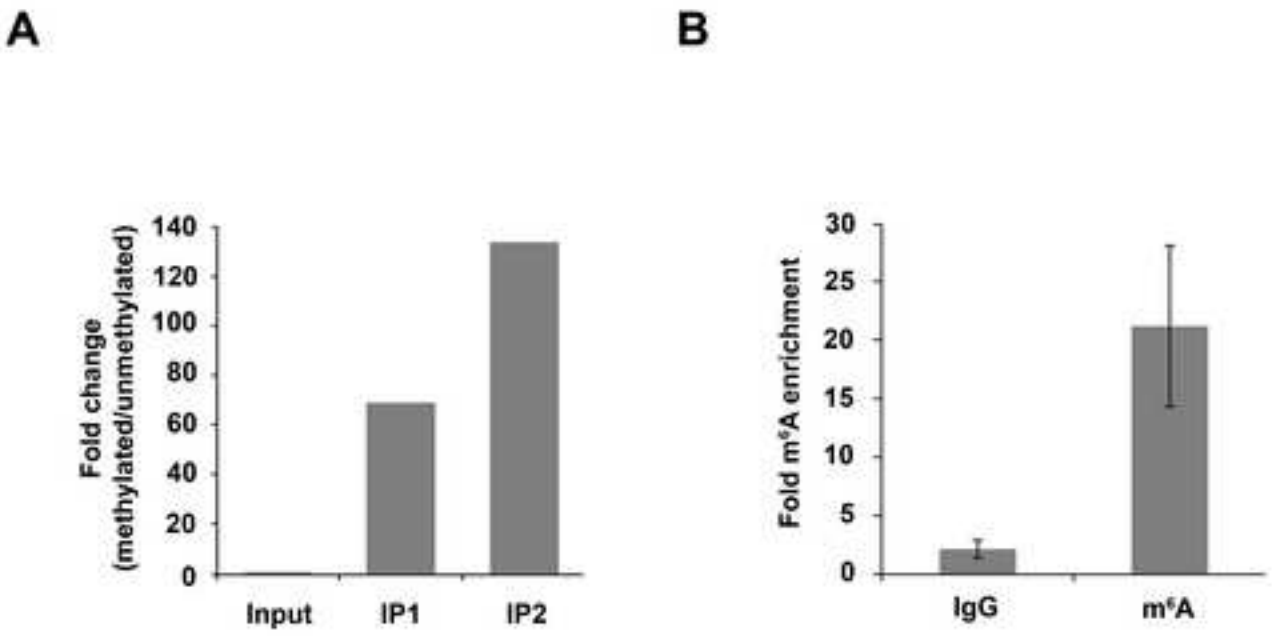
Supplemental Figure 1

Figure S1



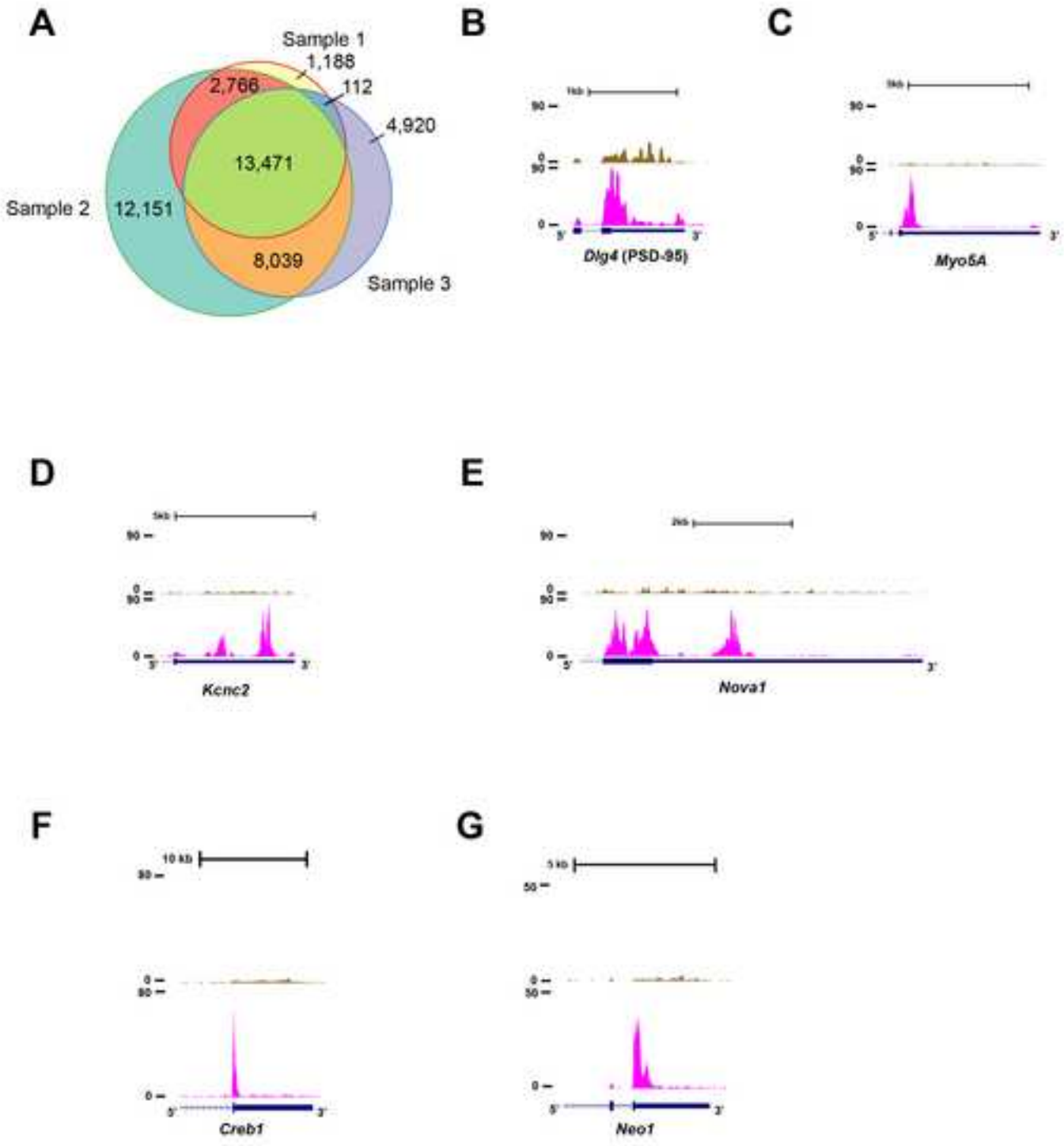
Supplemental Figure 2

Figure S2



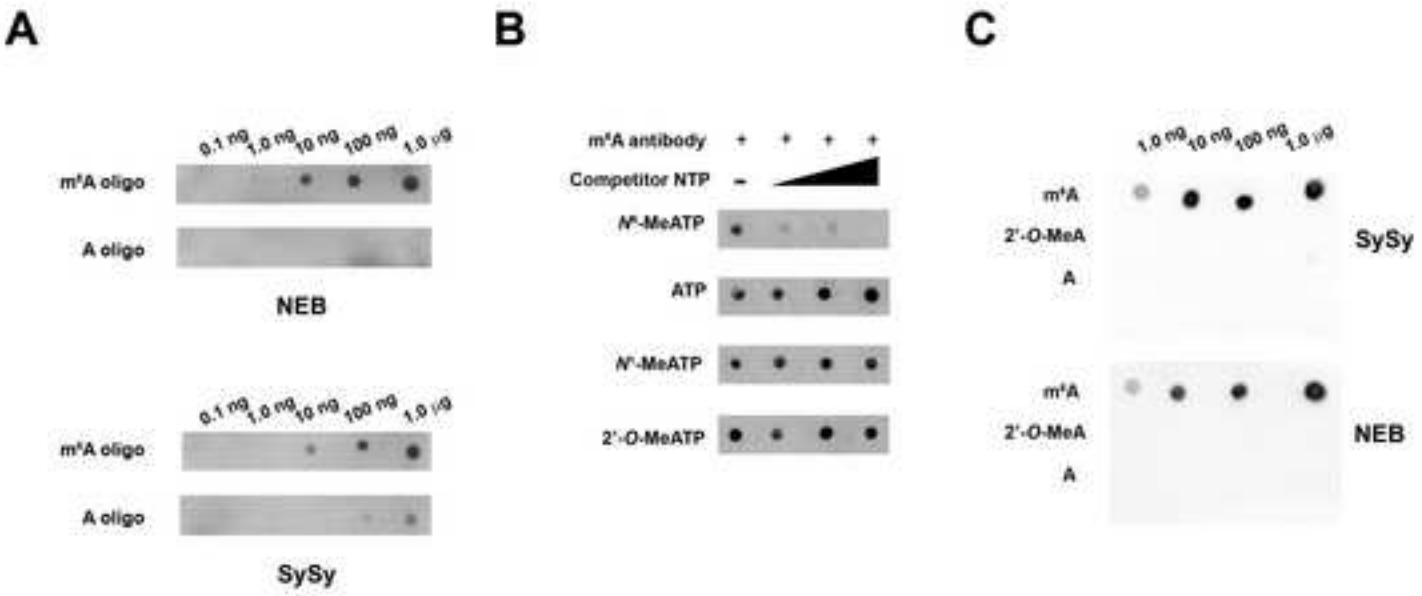
Supplemental Figure 3

Figure S3



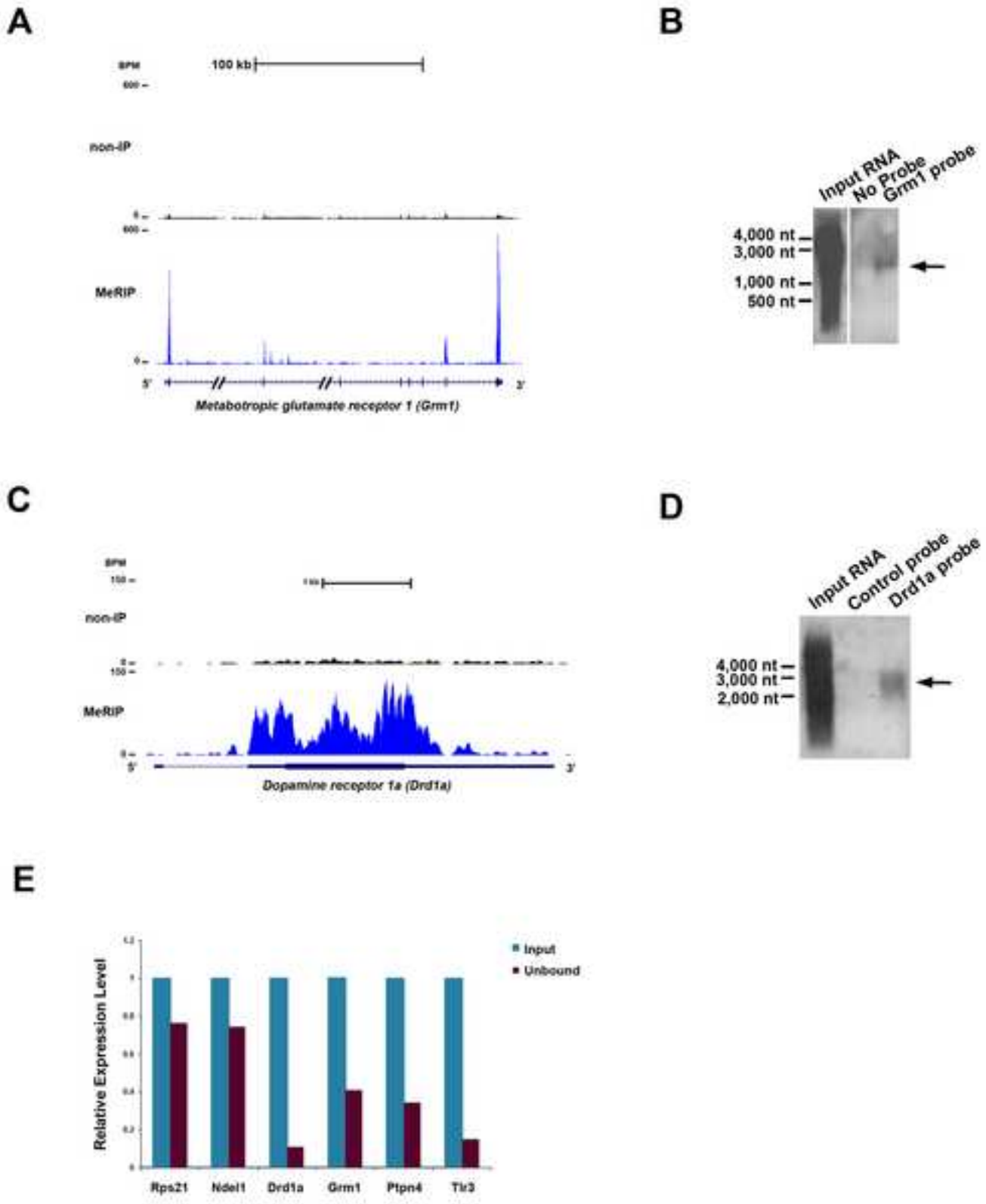
Supplemental Figure 4

Figure S4



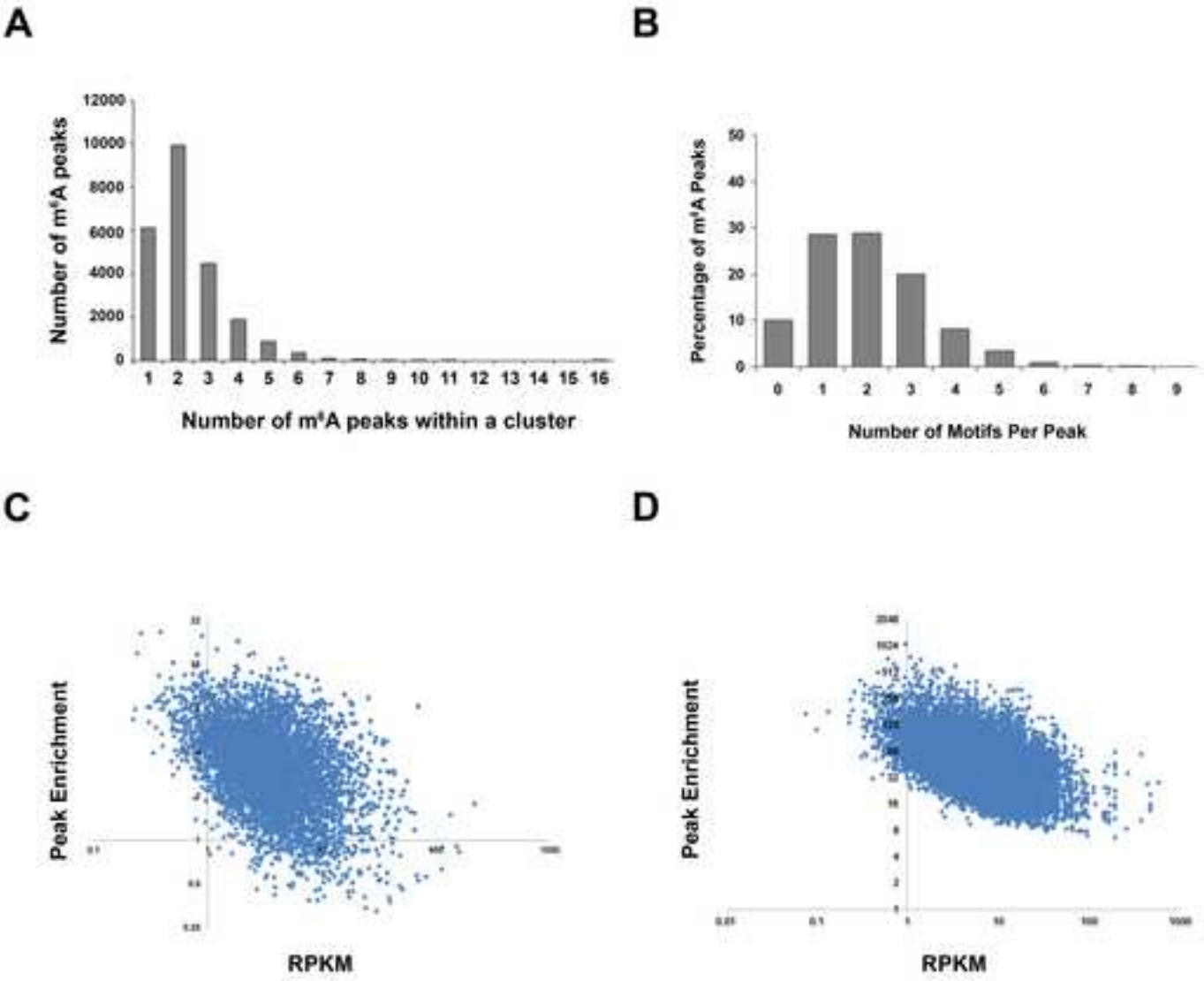
Supplemental Figure 5

Figure S5



Supplemental Figure 6

Figure S6



Supplemental Figure 7

Figure S7