OPEN ACCESS
CellPress

Cell Reports
Resource

Perturbation of m6A Writers Reveals Two Distinct Classes of mRNA Methylation at Internal and 5′ Sites

Schraga Schwartz,[1,12] Maxwell R. Mumbach,[1,12] Marko Jovanovic,[1] Tim Wang,[1,2,3] Karolina Maciag,[1,4] G. Guy Bushkin,[2] Philipp Mertins,[1] Dmitry Ter-Ovanesyan,[1] Naomi Habib,[1] Davide Cacchiarelli,[1,5] Neville E. Sanjana,[1] Elizaveta Freinkman,[2] Michael E. Pacold,[2,6] Rahul Satija,[1] Tarjei S. Mikkelsen,[1,5] Nir Hacohen,[1,7] Feng Zhang,[1,8,9] Steven A. Carr,[1] Eric S. Lander,[1,3,10,*] and Aviv Regev[1,3,11,*]

[1]Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA
[2]Whitehead Institute for Biomedical Research, Cambridge, MA 02142, USA
[3]Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[4]Graduate Program in Immunology, Division of Medical Sciences, Harvard Medical School, Boston, MA 02115, USA
[5]Harvard Stem Cell Institute and Department of Stem Cell and Regenerative Biology, Harvard University, Cambridge, MA 02138, USA
[6]Department of Radiation Oncology, Dana-Farber Cancer Institute, 450 Brookline Avenue, Boston, MA 02215, USA
[7]Center for Immunology and Inflammatory Diseases, Massachusetts General Hospital, Charlestown, MA 02129, USA
[8]McGovern Institute for Brain Research, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[9]Department of Brain and Cognitive Sciences and Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[10]Department of Systems Biology, Harvard Medical School, Boston, MA 02114, USA
[11]Howard Hughes Medical Institute, 4000 Jones Bridge Road, Chevy Chase, MD 20815, USA
[12]Co-first author
*Correspondence: lander@broadinstitute.org (E.S.L.), aregev@broad.mit.edu (A.R.)
http://dx.doi.org/10.1016/j.celrep.2014.05.048

SUMMARY

N6-methyladenosine (m6A) is a common modification of mRNA with potential roles in fine-tuning the RNA life cycle. Here, we identify a dense network of proteins interacting with METTL3, a component of the methyltransferase complex, and show that three of them (WTAP, METTL14, and KIAA1429) are required for methylation. Monitoring m6A levels upon WTAP depletion allowed the definition of accurate and near single-nucleotide resolution methylation maps and their classification into WTAP-dependent and -independent sites. WTAP-dependent sites are located at internal positions in transcripts, topologically static across a variety of systems we surveyed, and inversely correlated with mRNA stability, consistent with a role in establishing "basal" degradation rates. WTAP-independent sites form at the first transcribed base as part of the cap structure and are present at thousands of sites, forming a previously unappreciated layer of transcriptome complexity. Our data shed light on the proteomic and transcriptional underpinnings of this RNA modification.

INTRODUCTION

DNA, RNA, and proteins are all subjected to biochemical modifications following synthesis, which can alter and fine-tune their function by diverse regulatory mechanisms. N6-methyladenosine (m6A) is a highly prevalent base modification occurring on mammalian mRNA. Recent studies used immunoprecipitation of methylated RNA fragments followed by sequencing (m6A-seq) to globally map transcript regions enriched in m6A in mammalian cells, finding that it is strongly enriched near stop codons and in long exons (Dominissini et al., 2012; Meyer et al., 2012).

Conceptually, m6A in mammals has the potential of fine-tuning RNA function in different ways. One possibility is that genes are subjected to methylation only under specific conditions or in specific tissues ("condition specific methylation"), as is the case in yeast meiosis (Agarwala et al., 2012; Clancy et al., 2002; Schwartz et al., 2013). A nonmutually exclusive scenario is that m6A may mark and regulate a specific sets of transcripts ("transcript specific methylation"), for instance, by affecting their stability (Wang et al., 2014a). To explore potential roles for m6A, it is necessary to investigate the extent to which m6A varies across physiologically relevant conditions. To date, mammalian methylated sites have been mapped and characterized in only a small number of mammalian cell lines/tissues (Dominissini et al., 2012; Meyer et al., 2012; Wang et al., 2014a) limiting the ability to evaluate methylation dynamics. Moreover, the resolution of these maps was limited, with sites typically being >20 nt away from the nearest consensus signal, potentially reflecting a nonnegligible amount of false-positives.

Obtaining accurate maps of mRNA methylation requires identification of the proteins involved in catalyzing them. We recently found in yeast that available protocols for m6a-seq identified both true methylated sites and false-positive sites, and that

these two classes could be distinguished by mapping methylations that remain after knockout of the methyltransferases (Schwartz et al., 2013). Until recently, only one protein—METTL3 (Methyl-transferase-like 3)—was implicated in m6A methylation of mammalian mRNA (Bokar et al., 1997). However, it had been recognized that additional components were crucial for methylation (Bokar et al., 1997). While this manuscript was under preparation, two additional proteins, WTAP and METTL14, were identified as required for methylation (Liu et al., 2014; Ping et al., 2014; Wang et al., 2014b). These studies, however, did not study the extent to which individual sites were dependent on these proteins, which is important both to eliminate false-positives (Schwartz et al., 2013) and to identify sites that are methylated using orthogonal pathways.

Here, we have employed an unbiased proteomic approach to characterize the components of the methyltransferase complex, allowing us to identify and validate known and unknown components required for methylation. By mapping sites upon experimental depletion of these components, we were able to classify and characterize methylated sites based on their dependency on these proteins. Our analyses provide important resources at the proteomic and transcriptomic levels toward understanding the regulators ("who") and targets ("where") of RNA methylations, two crucial milestones toward addressing the function ("why") of this epitranscriptomic modification.

RESULTS

Proteomic Screens Identify Components of the Methyltransferase Complex

To identify components of the human m6A methyltransferase complex, we performed coimmunoprecipitation (co-IP) experiments with an overexpressed C terminus HIS-tagged METTL3 in HEK293 cells, followed by LC-MS/MS (Experimental Procedures). After filtering out background contaminants using the CRAPOME database (Mellacheruvu et al., 2013) (Experimental Procedures), one of the most enriched proteins was WTAP, the human ortholog of Mum2, a crucial component of the yeast methyltransferase complex (Agarwala et al., 2012; Schwartz et al., 2013) (Figure 1A). Proteins interacting with WTAP have recently been characterized in a proteomic screen (Horiuchi et al., 2013). Analysis of this data revealed reciprocal binding of WTAP to both METTL3 and to METTL14, a close paralog of METTL3 (Bujnicki et al., 2002). Confirming this association, both METTL3 and WTAP were enriched when we performed mass-spectrometry following pulling down on a 3′ terminus V5-tagged version of METTL14 (Figure 1B). The physical association of METTL14 and WTAP with METTL3, and our previous findings that all three proteins have tightly coevolved (Schwartz et al., 2013), thus strongly implicated them in mRNA methylation.

To explore additional proteins involved in the methylation program, we used in vitro transcribed biotinylated baits, which were either methylated or nonmethylated at a single position, to capture proteins from HEK293 protein lysates, followed by quantitative LC-MS/MS (Figure 1C). YTHDF1, YTHDF2, and YTHDF3, three proteins from the YTH domain family, were the top three enriched proteins in this screen. This enhances our previous results (Dominissini et al., 2012), where two of these three proteins were identified as m6A binders (Dominissini et al., 2012), and is consistent with recent biochemical results showing that these three proteins all selectively bind m6A (Wang et al., 2014a). Interestingly, the fourth most enriched protein in this assay was WTAP, suggesting that WTAP may be involved not only in mediating methylation, but also in binding it.

Finally, we used a V5-tagged version of each of three YTH proteins in affinity proteomics. These experiments highlighted potential associations between different YTH proteins (e.g., between YTHDF1 and YTHDF2) or between the YTH proteins and other components of the methyltransferase machinery identified here (e.g., YTHDF2 and WTAP, YTHDF1, and METTL3). The full proteomics data sets are available in Table S1.

Integrating the proteomic data into a network (Figure 1D; Experimental Procedures) highlights the centrality of WTAP and METTL14 in this complex, consistent with the three recent reports (Liu et al., 2014; Ping et al., 2014; Wang et al., 2014b). The density and topology of the network suggest that the conceptual distinction between m6A "writers" and "readers" may not be clear cut, because the putative writers (METTL3, WTAP, METTL14) and readers (YTH proteins) may physically interact with each other. Finally, 13 proteins with diverse functions interact with two or more of the seven core components studied here and may therefore form part of the methyltransferase complex as well. However, we cannot rule out that some of these associations may be spurious, and each of these candidates will therefore have to be individually tested. Below we validate three of these proteins.

WTAP Is Necessary for m6A mRNA Methylation

To determine whether WTAP is required for mRNA methylation in vivo, we depleted it in $p53^{-/-}$ mouse embryonic fibroblasts (MEFs) using WTAP-targeting small hairpin RNAs (shRNAs) to 10%–15% of WT levels (Figures S1A–S1D). We used anti-hnRPDL and anti-GFP shRNAs as negative controls. We immunoprecipitated and sequenced methylated RNA from these samples using an m6a-seq procedure, optimized for enhanced resolution and scalability using decreased starting material (Schwartz et al., 2013). Analyzing these profiles (Experimental Procedures), we identified 16,487 putative m6A sites, present in at least two of the eight profiled conditions.

Most sites (10,609/16,487, 64.3%) were WTAP dependent ("WTAP-dependent" cluster), showing dramatically decreased methylation Peak Over Input (POI) scores compared to any of the controls ($p < 2.2 \times 10^{-16}$ for all comparisons, Mann-Whitney) following knockdown with either of two shRNAs targeting different regions of WTAP (Figure 2A). WTAP-dependent sites were strongly enriched for hallmarks of m6A methylation: sites were within a median distance of 5 nt from the nearest consensus site (Figures 2B, S1E, and S1F); 34.6% of sites were within 200 nt of the stop codon (Figure 2C); and median internal exons harboring sites in this cluster tended to be ~7-fold longer than nonmethylated exons (Figure 2D).

The remaining sites formed two distinct clusters (Figure 2A), neither of which was enriched for the methylation consensus signal (median distances to consensus motif were 27 and 38 nt, respectively), or for any of the other hallmarks of m6A methylation (Figures 2B–2D). One of these two clusters showed

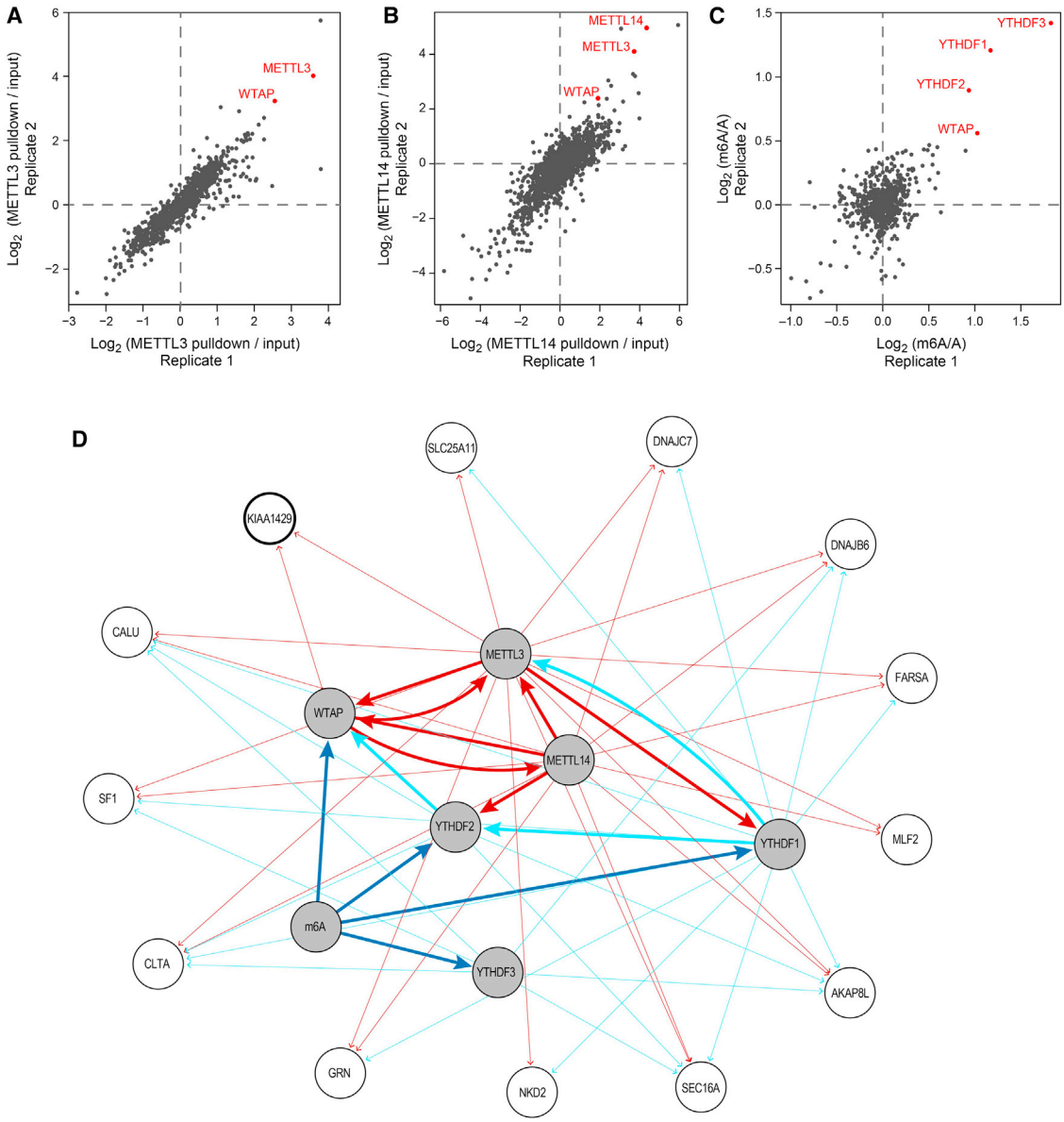


Figure 1. Proteomic Identification of Methyltransferase Complex Components

(A and B) Proteins associated with HIS-tagged METTL3 (A) and V5-tagged METTL14 (B). Fold changes quantifying enrichment versus control (Experimental Procedures) across two biological replicates.

(C) Proteins associated with biotinylated, methylated RNA baits, compared to biotinylated, nonmethylated counterparts, in two biological replicates.

(D) Network of associations with the bait proteins (shaded gray nodes) identified in the different mass-spec experiments. An edge from bait A to target B indicates that B was enriched (fold change >1.5) when performing IPs on A. Only target nodes with an incoming degree ≥ 2 are displayed. Edge width distinguishes association between baits (thick lines) and with nonbaits (thin lines). Edges are colored in dark red if they were enriched by a protein bait involving an m6a "writer" (METTL3/METTL14/WTAP), cyan if they were enriched by a "reader" (YTHDF1/YTHDF2/YTHDF3), or dark blue if they were enriched with a methylated bait. The node for KIAA1429, further analyzed functionally, is marked by a thicker border.

dramatic enrichment for sites near the TSS, with ~50% of the sites within 200 nt of the TSS ("TSS-enriched cluster"); we show below that the sites in this cluster represent methylations on the first transcribed nucleotide as part of the cap structure. The other cluster did not exhibit such dramatic enrichment for the TSS segment but was pervasive across all sampled conditions and showed no enrichment for the m6A consensus motif. These properties were reminiscent of a set of false-positive sites that we found in yeast when performing the same m6a-seq protocol to strains lacking a functional methyltransferase (Schwartz et al., 2013). Indeed, sequence analysis revealed that the sites in the cluster ("false-positive cluster") were highly enriched for the same degenerate purine-rich motif ("AGAAGAA") that we found at the yeast false-positive sites (Figures S1G and S1H). In yeast,

we have previously shown that these sites were false-positive sites enriched in a nonspecific manner during the immunoprecipitation process, because they are enriched even when we perform m6A-seq with in vitro synthesized sites without any methylations (Schwartz et al., 2013); thus, we considered it highly likely that these sets of sites are false-positives and conservatively opted to either analyze them separately, or eliminate them altogether, in the subsequent analyses.

We confirmed these results in human A549 cell lines where WTAP was perturbed using either shRNAs (Figure 2E) or siRNAs (Figure 2F). We observed a similarly strong reduction of methylation upon WTAP depletion in both sets of experiments, and clustering of the detected sites yielded very similar clusters to those obtained in the mouse studies above (Figures 2E and 2F), with similar characteristics (Figures S1I–S1K).

KIAA1429 Is Required for mRNA Methylation

Of the other 13 candidates associating with methyltransferase components in our proteomics screen (Figure 1D), we focused our attention on KIAA1429 because its *Drosophila* ortholog was biochemically shown to interact with *Drosophila* WTAP in the context of sex-specific splicing (Ortega et al., 2003). We used siRNAs to deplete KIAA1429 in human A549 cells to ~6% of WT levels (Figure S1C) and performed m6A-seq on the resulting cells. We observed a median ~4-fold decrease (Mann-Whitney, $p < 2.2 \times 10^{-16}$) in peak scores compared to cells treated with nontargeting siRNAs (Figure 2F). Although these decreases were less dramatic than those observed upon knockdown of WTAP (median decrease: 6.25-fold), they were substantially and significantly (Mann-Whitney, $p < 2.2 \times 10^{-16}$) more prominent than observed upon knockdown of either METTL3 or METTL14 (see below), demonstrating that KIAA1429 is required for the full methylation program in mammals.

Knockdown of METTL3 and/or METTL14 Leads to Milder Decreases in Methylation Levels

We next used shRNAs and siRNA to test the effect of depleting METTL3 and METTL14 in mouse fibroblasts (METTL3, Figure 2A) and in human A549 cells (both, Figures 2E and 2F). In mouse, we observed a moderate depletion of m6A levels for one of the hairpins (shMETTL3-2, where METTL3 transcript was reduced to 10% of WT levels), but not for another (shMETTL3-3, where knockdown efficiency was lower) (Figure S1A). In human, shRNA-mediated knockdown of METTL3 levels to ~10% of WT levels did not reveal a discernible effect on m6A methylation, but use of siRNAs, which enabled even higher knockdowns (~5% of WT levels, Figures S1C and S1D), yielded POI scores that were reduced to 60% of control (Mann-Whitney, $p < 2.2 \times 10^{-16}$, Figure 2F). Similarly, for METTL14 knockdown, we failed to observe an effect when knocking it down using shRNAs to ~20% of WT levels (Figure S1B), but upon knocking it down to ~6% with siRNAs (Figure S1C) we observed reductions to similar levels as for METTL3 (median decrease: 51% of control). Similar levels were achieved also with dual knockdown of METTL3 and METTL14 (median decrease: 53% of control, Figure S1C). Our results, which are consistent with Liu et al. (2014), demonstrate a strong threshold dependence for both METTL3 and METTL14, which is likely due to adequacy of even low levels of METTL3 and METTL14 to mediate mRNA methylations.

Bulk Measurements of m6A Levels Highly Consistent with m6A-Seq

To further confirm our results, we used HPLC-M/S to measure bulk levels of m6A in poly(A) mRNA from A549 cells upon siRNA-mediated knockdown of the above genes (Figure 2G). Consistently, highest levels of depletion in m6A content in the oligo-dT selected poly(A) fraction were obtained for WTAP (5.6-fold depletion), followed by KIAA1429 (~3-fold), followed by single or combined knockdown of METTL3 and METTL14 (1.9- to 2.5-fold). M6A levels in the flowthrough fraction were roughly equal for all samples (Figure 2G). These results were expected, because the flowthrough is dominated by rRNA, which harbors m6A at nonconsensus positions (Liu et al., 2013; Machnicka et al., 2013) that are likely deposited by proteins distinct from the ones identified here.

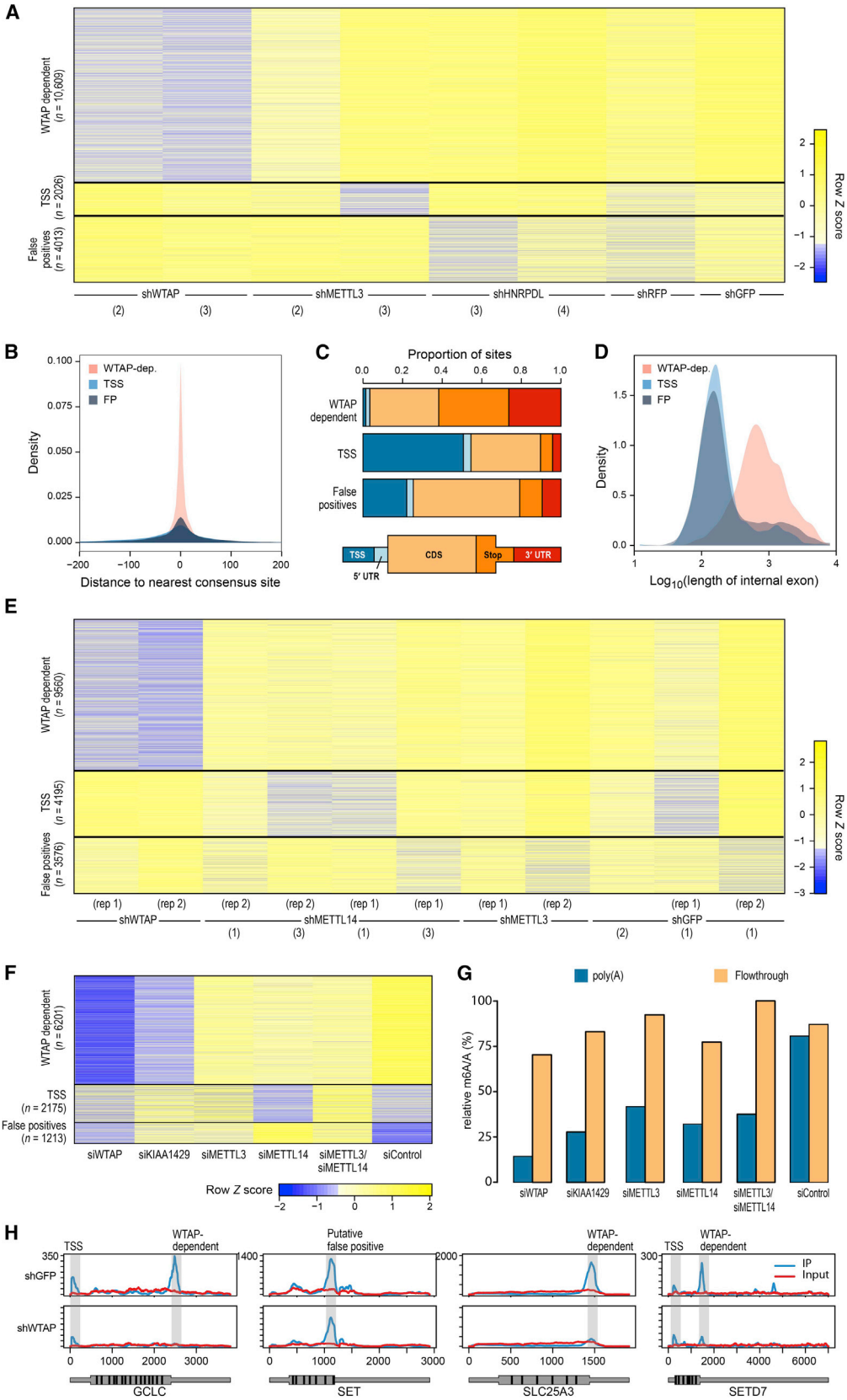
A High-Quality, High-Resolution Catalog of the m6A Methylome across Four Dynamic Systems

To date, mammalian methylated sites have been mapped and characterized in only a limited number of cell lines and tissues. To determine the extent to which m6A methylation may be modulated in physiologically relevant contexts, in both mitotic and postmitotic cells, we mapped methylation across four distinct dynamic systems in human and mouse: (1) mouse bone-marrow-derived dendritic cells (BMDCs) responding to lipopolysaccharide (LPS), (2) mouse embryonic (e16) and adult brains, a system chosen based on previous reports of lower levels of bulk m6A in embryonic brain (Meyer et al., 2012), (3) human fibroblasts undergoing reprogramming into induced pluripotent stem cells (iPSC) following doxycycline-induced expression of polycistronic OCT4-KLF4-MYC-SOX2, and (4) human embryonic stem cells undergoing differentiation into NPCs (Figure 3A).

Integrating all m6A profiles yielded 40,742 sites in human and 31,423 in mouse, present in at least two conditions, tripling the reported number of detected putative methylation sites compared to previous studies. We classified each site as high, intermediate, or low confidence based on a linear combination of three features: (1) number of samples in which peak was identified, (2) extent of dependency on WTAP, and (3) maximal POI score observed across any of the samples. The features were selected as informative based on distance to nearest consensus site (Figure S2A; Experimental Procedures), and the resulting ranking was independently consistent with enrichment of sites in long exons (Figure S2B) and near stop codons (Figure S2C). The sites identified here are at a dramatically increased resolution, compared with previous studies, with 50% of the top 15,000 sites being within 5–6 nt from the nearest consensus site, in both data sets. The full set of sites is provided (Table S2) allowing custom filtration of the data based on user-defined criteria.

WTAP-Dependent Sites Have a Static Topology across Multiple Dynamic Systems

Although initial examination suggested the presence of many condition-specific methylations, closer examination indicated



(legend on next page)

that observed differences were largely due to changes in underlying gene expression. For example, in the mouse data set, 3,629 sites classified as "high" or "intermediate" are present only in the two brain samples, and 1,312 such sites are present exclusively in the DCs. However, these differences clearly reflected the expression levels of the genes harboring the sites (Figures S2D and S2E), which indeed were enriched for genes involved in neural processes and immune response, respectively (Figures S2F and S2G).

To reliably compare sites across samples despite different expression levels, we therefore conservatively limited our analysis to a subset of WTAP-dependent sites within genes expressed above the 60th percentile in all samples, resulting in 11,247 sites in human (Figure 3B) and 8,456 sites in mouse (Figure 3C). Dividing these into three equally sized sets based on their maximal POI scores across all samples, we found that among the highest scoring sites >90% of the sites in human (Figure 3D) and >98% of the sites in mouse (Figure 3E) were detected as enriched (POI >4) across all samples. Among lower scoring sites, cases of peaks being called in one condition but not in another were more common (Figures 3D and 3E); however, upon manual inspection of such putative differential sites, it was difficult to identify convincing sites present in one condition and absent from another. Rather, such differences were typically due to subthreshold peaks not being called in a particular sample or insufficient coverage in a particular region. Indeed, the variability of POI scores across the different dynamic sets of conditions was similar to the variability across different experiments performed in the same cell line (MEFs in mouse, A549 in human) (Figures S2H and S2I). Thus, overall methylation profiles appear to be similar across diverse conditions, at least under our conservative criteria, suggesting that in mammals m6A plays a basal, ubiquitous role, shared across different cell types and systems. However, because m6A-seq cannot directly quantify methylation levels, our approach may be blind to quantitative differences in methylation levels between samples (see Discussion).

WTAP-Dependent Methylation Inversely Correlates with mRNA Stability and Is Depleted in Housekeeping Genes

Because we found little evidence for condition-specific differences in methylation, we next sought to analyze the potential of a transcript-specific role. Examining the relationship between gene-level attributes and methylation densities (number of WTAP-dependent methylation sites normalized by gene length) in mouse, we noted that the top 20% of expressed genes were significantly less likely to be methylated compared to more lowly expressed genes (Figure 4A). Indeed, genes completely lacking methylation sites are highly enriched for those involved in "housekeeping" cellular processes like translation, mitochondrial-related processes, chromatin regulation, and splicing. To study this further, we examined the proportion of genes lacking methylations across all Gene Ontology (GO) categories in human, mouse, and yeast. This analysis revealed that the functional group "structural constituents of ribosome" ranked highest in proportion of genes lacking methylation in both human and mouse, and second highest in yeast (Figure 4B). Other groups of housekeeping genes ranking high in both yeast and mammals include splicing and GTPase activity (Figure 4B). Thus, lack of methylations among ribosomal proteins, in particular, and specific sets of housekeeping proteins, in general, is conserved between yeast and mammals.

Because housekeeping genes generally have shorter mRNAs with shorter CDSs and UTRs and fewer exons (Eisenberg and Levanon, 2003) and longer RNA half-lives (Schwanhäusser et al., 2011), we directly examined the correlation between each of these variables and methylation densities (Figures 4C and 4D). Strikingly, the variable correlating most strongly with methylation density was mRNA half-life (Figures 4C–4E), as previously estimated in mouse fibroblasts (Schwanhäusser et al., 2011) (Spearman $\rho = -0.27$, $p = 1.2 \times 10^{-69}$), and this correlation remained of similar magnitude ($\rho = -0.22$) when performing this analysis after eliminating the top 10% or top 20% of genes, ranked by expression levels. Consistently, examining the difference in proportion of variance explained (R^2) when predicting methylation densities from all these variables, compared to exclusion of any single one, revealed that mRNA half-life accounted for the greatest difference (Figure 4D). We obtained similar results when performing these analyses for human transcripts using transcript stability estimates in human lymphoblastoid cell lines (Duan et al., 2013) (Figures S3A and S3B), with half-life exhibiting the strongest correlation with methylation density (Figure S3C), albeit at somewhat reduced levels ($\rho = -0.14$, $p = 3.4 \times 10^{-34}$). Our results are consistent with reports implicating methylations in mRNA degradation (Wang et al., 2014a, 2014b) and suggest that m6A methylation may

Figure 2. Perturbation of Methyltransferase Complex Components Identifies Complex-Dependent and -Independent Methylation Sites

(A) K-means clustering of log2 transformed peak over input scores (yellow, high; blue, low) across 16,487 sites (rows) detected in mouse embryonic fibroblasts (MEFs). Only sites detected in two or more of the presented samples (columns), with a maximal Peak Over Input score greater than 8, and within genes expressed above the 60th percentile are shown. Sites were initially clustered into five clusters using k-means, and three of the clusters were subsequently manually merged with existing clusters, based on similarity of characteristics, to yield three clusters (separated by black lines). For visualization purposes, the $\log_2$(POI) scores were capped at 4.5 and then transformed into *Z* scores. The internal shRNA identifiers are indicated in parentheses.

(B) Distributions of distance from the nearest consensus site of the sites in the three clusters in (A).

(C) Distribution of the sites in (A) grouped by the clusters in (A) across five gene segments, defined as in Dominissini et al. (2012).

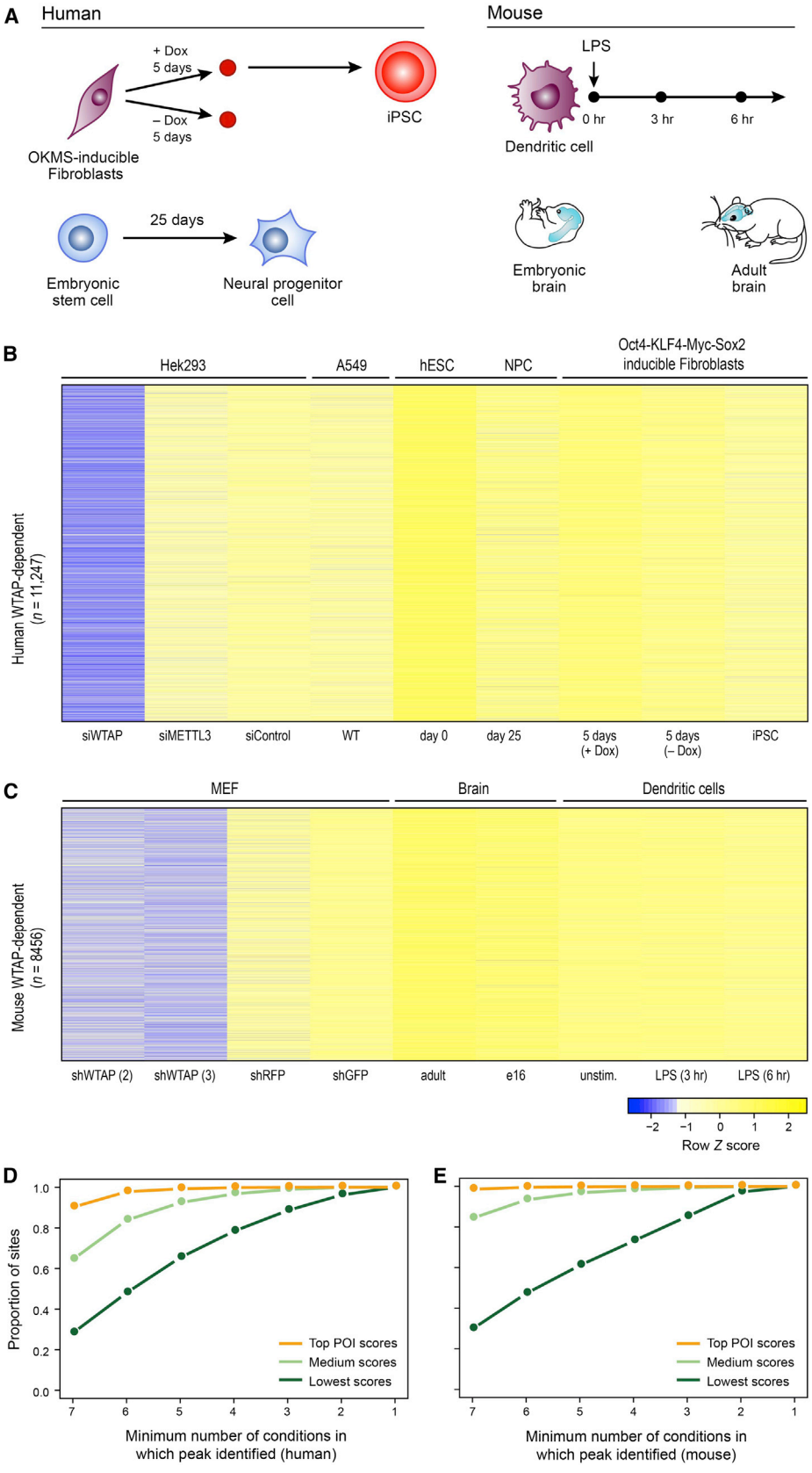
(D) Distribution of length of internal exons across the clusters identified in (A).

(E) K-means clustering of log2 transformed Peak Over Input scores across 17,331 sites (rows) detected in human A549 cells, visualized as in (A).

(F) K-means clustering of log2 transformed Peak Over Input scores for set of experiments involving perturbations of A549 cells with siRNAs, as indicated. Plotted are WTAP-dependent sites (rows) within genes expressed above the 60th percentile across all conditions.

(G) Relative m6A/A ratios across A549 cells treated with siRNA as indicated, as measured by HPLC-M/S. Values are plotted for both the poly(A) fraction (blue) and the flowthrough (orange).

(H) Examples of coverage from IP (blue) and input (red) experiments across selected genes in A549 cells treated with shGFP (top row) or shWTAP (bottom row).



(legend on next page)

help set a basal degradation program, and that highly abundant messages might have evolved to maximize their stability by avoiding mRNA methylation.

WTAP-Independent m6A Methylome at the mRNA Cap

Studies in the mid-1970s involving bulk analysis of mRNA caps revealed that when the first nucleotide of a transcript is an adenosine this base can be methylated at the N6 position (Furuichi et al., 1975; Keith et al., 1978; Wei et al., 1975). However, no study to date has been able to resolve which mRNA caps within which messages are methylated. We hypothesized that the consistently observed WTAP-independent peaks enriched near the TSS (Figures 2A, 2E, and 2F) reflect methylations occurring at the cap. Corroborating this hypothesis, we found that 30.5% of the mouse genes expressed above the 60th percentile with a "TSS peak" (i.e., an m6A peak within the first 200 nt) are annotated to begin with an "A", compared to only 23% of the genes lacking a peak (Fisher's exact test, $p = 2.8 \times 10^{-13}$), consistent with analyses performed in Dominissini et al. (2012). Although this enrichment was significant, it did not explain the peaks observed in the remaining ~70% of the genes.

We therefore considered the ~70% of genes whose RNA transcripts harbored a TSS peak but whose annotated TSS did not contain an "A" at the first base. We suspected that many of these cases might reflect limitations of the genomic annotation, as transcription initiation often occurs from a variety of closely spaced positions (Carninci et al., 2006; Ni et al., 2010; Plessy et al., 2010). We therefore leveraged the fact that our library construction method relies on ligation of adapters to both ends of the captured RNA fragment, such that although each transcript is fragmented "randomly," every transcript should yield some fragment beginning at its 5′ terminus, resulting in a "pileup" of reads beginning at the 5′ terminus of genes. Thus, these pileups harbor information on the precise transcript initiation sites. Moreover, methylated transcription start sites (mTSSs) should have a high ratio between the size of these pileups in the IP sample compared to the input sample; sites with high ratios should therefore be highly enriched toward harboring an adenosine, as opposed to typical transcription start sites that tend to begin with a guanosine (Carninci et al., 2006).

To detect methylated transcription start sites (mTSSs), we compared the number of reads stacks beginning at each of the first 50 annotated positions in the transcript in the IP and input samples (Experimental Procedures). We assembled a catalog of 33,714 sites in mouse with evidence of greater than ten reads beginning at a specific site (across all conditions) and assigned each site a fold change corresponding to its enrichment in IP over input. As expected, we found that sites with stronger fold changes were also dramatically more likely to harbor an adenosine either at the detected site or at the position immediately preceding it, with >80% of the sites harboring an adenosine in the bin of most highly enriched sites (Figure 5A). (The tendency for adenosines at the position preceding the pileups is likely due to reverse transcriptase drop-off one base prior to the TSS, as supported by several lines of evidence [Supplemental Discussion; Figures S4A and S4B; and below].) Further corroboration of the validity of the detected sites was obtained from an observation of a strong bias toward pyrimidines at position −1 and a discernible TATA box ~30 bp upstream of the detected sites, two hallmarks of transcription initiation sites (Carninci et al., 2006) (Figure 5B).

Our predicted mTSSs also agree well with the results of cap analysis of gene expression (CAGE) data. We performed the above analyses in human (obtaining highly similar results; Figure S4C) and focused on a subset of 9,757 putatively methylated TSSs (mTSSs) that were independently detected in at least two experiments in A549 cells, exhibiting at least 4-fold enrichment in IP over input and harboring an adenosine at either the position harboring the read stacks or the one preceding it. We compared these sites to CAGE data in A549 cells, obtained from the ENCODE project (Bernstein et al., 2012). Of the 9,757 sites, 7,255 (74%) were supported by at least two CAGE tags, compared to only 34.6% among an equally sized set of random controls (Fisher's exact test, $p < 2.2 \times 10^{-16}$). To confirm that the positions harboring the adenosines ("adenosine positions") were the true TSS even when they preceded the observed stacks by one base ("stack position") (a phenomenon we attribute to RT drop-off), we examined a subset of 4,879 sites in which the adenosine positions preceded the stack positions. Although 80.5% of the "adenosine positions" were supported by at least two CAGE tags, only 36.8% of the "stack positions" were supported by CAGE data, similar to the 34.5% overlap obtained in the random data set (Figure 5C). Moreover, there was a >6-fold increase in the mean number of CAGE tags supporting the "adenosine positions," compared to either the "stack positions" or the random positions (Figure 5D). Finally, when restricting this analysis to positions other than the first annotated position, these results were essentially identical (data not shown). Together, this confirms that our approach can identify transcription initiation sites at single nucleotide resolution and that mTSSs can be reliably detected also within a large number of positions other than the annotated TSS.

Based on these analyses, we compiled a catalog of putatively methylated transcription start sites (mTSS), comprising all unique adenosine-containing sites enriched >4-fold in at least one condition, encompassing 15,961 sites from 6,454 mouse genes and 12,601 sites from 5,774 human genes (Table S4).

Figure 3. A Human and Mouse Catalog of Methylation Dynamics

(A) Scheme depicting the four experimental systems profiled in this study. Top left: doxycycline-inducible OKMS (OCT4, KLF4, MYC, and SOX2) fibroblast reprogramming to iPSCs. Bottom left: human embryonic stem cells undergoing differentiation to neural progenitor cells. Top right: bone-marrow-derived dendritic cells stimulated by lipopolysaccharide (LPS). Bottom right: embryonic and adult mice brain.

(B and C) Heatmaps of POI scores (visualized as in Figure 1A) of all WTAP-dependent sites (rows) identified in one or more sample (columns), within genes expressed above the 60th percentile across all conditions (Experimental Procedures) in human (B) or mouse (C).

(D and E) Number of conditions (x axis, of a total of seven, as in [B] and [C], respectively, excluding any genetic perturbations) in human (D) and mouse (E), in which sites are identified as present (POI >4). Sites were grouped into three equally sized groups, based on POI score.

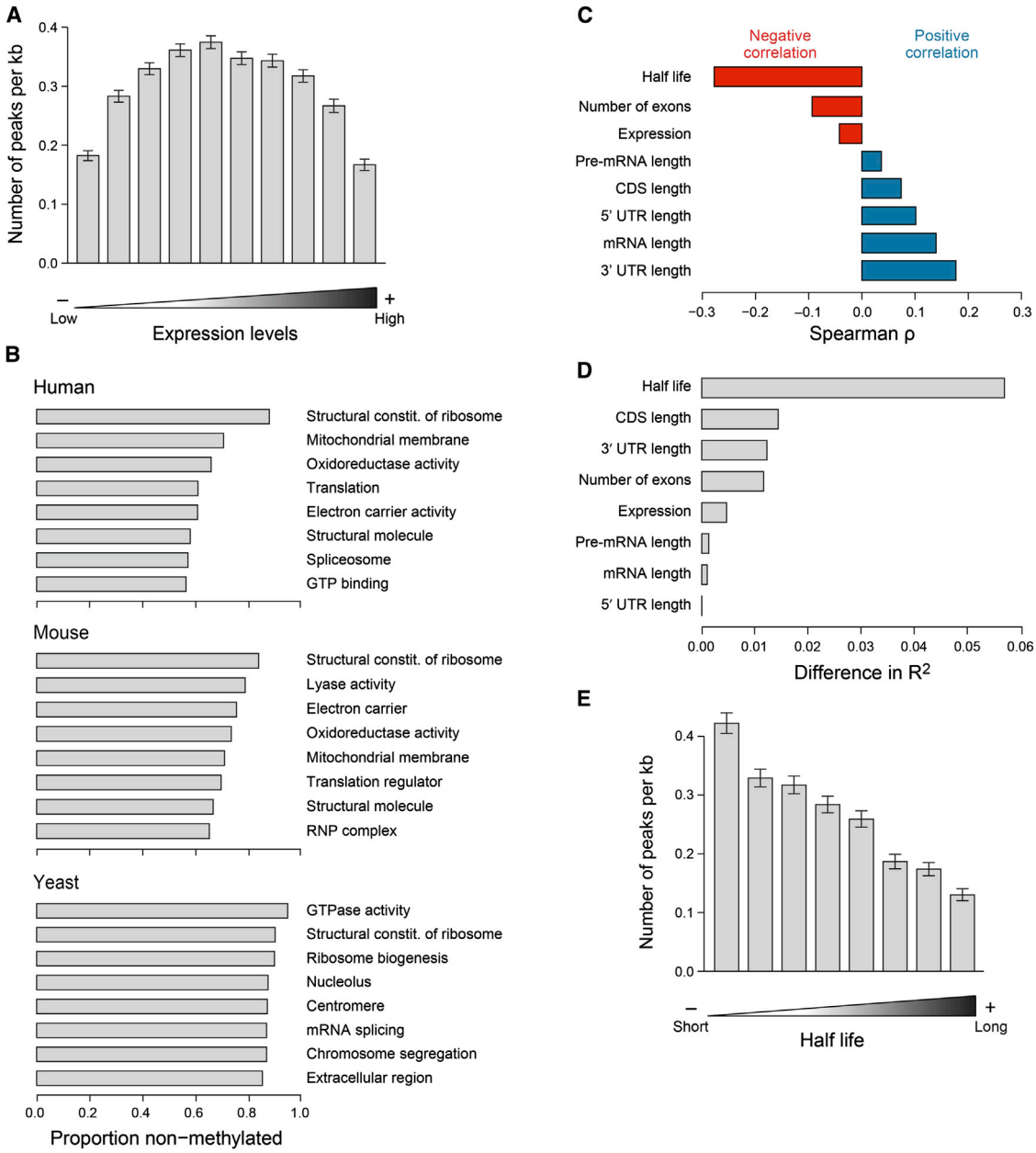


Figure 4. Methylations Are Depleted from Housekeeping Genes and Associated with Decreased Transcript Stability

(A) Association between gene expression and methylation density in MEFs.

(B) Top GO categories harboring nonmethylated genes in human (top), mouse (middle), and yeast (yeast). Shown is the proportion of nonmethylated genes in each category. Only genes expressed >60th percentile and longer than 500 nt and only GO categories harboring >30 matches genes were considered for this analysis.

(C) Correlations (Spearman ρ, x axis) between different gene attributes (y axis) and methylation density. Transcript half-lives were obtained from Schwanhäusser et al. (2011). We used only sites defined in A549 cells (Figure 2E), in protein coding genes longer than 500 nt and expressed above the 60th quantile, with a maximum POI score >8, with at least 50% increased POI score in nonperturbed cells compared to WTAP-depleted cells, and which were in segments other than the TSS segment.

(D) Analysis examining the difference in variability explained (R^2, x axis) upon elimination of each of the indicated variables from a linear regression model predicting methylation density on the basis of all of them.

(E) Association between mRNA half-lives (Schwanhäusser et al., 2011) and methylation densities. Half-lives were binned into ten equally sized bins, and mean methylation densities for each bin were calculated. Error bars represent the SEM.

Many sites—and particularly sites with stronger associated fold changes—were shared across multiple conditions (Figures S4D–S4F), and 51% of the above genes in mouse (23% in human) had more than one detected mTSS across all conditions. Limiting this analysis to a single condition, on average 13% of the transcripts had more than one mTSS in both mouse and human (e.g.,

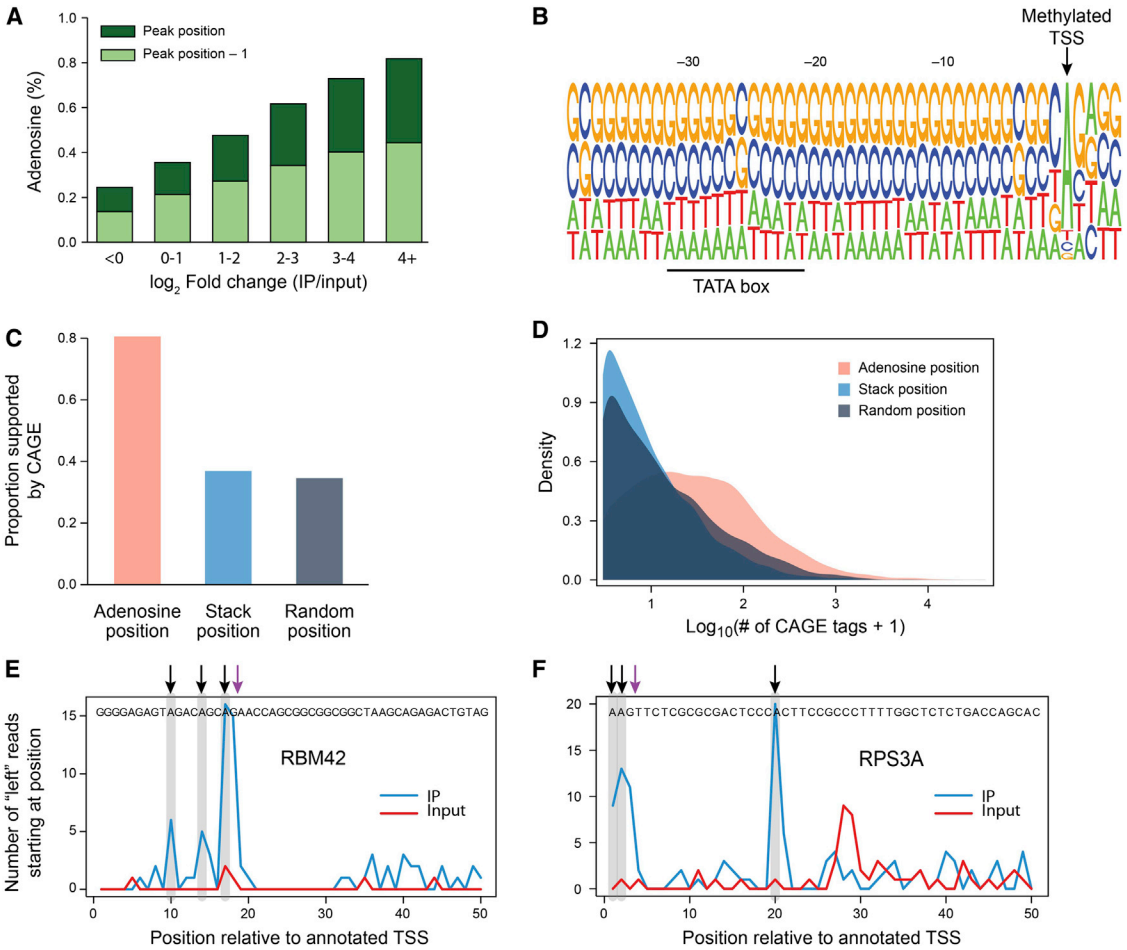


Figure 5. Identification of Methylated Transcription Start Sites

(A) Proportion of sites harboring an adenosine (y axis) at the detected position (dark green bars) or the position immediately upstream (light green bars) in 33,714 sites with greater than ten reads within 50 nt of annotated TSSs. Sites were binned based on fold changes (IP over input; x axis), using thresholds as indicated.

(B) Sequence logo for a set of 3,445 putative mTSSs with a log2 fold change exceeding 4.5. The position of the putative methylated mTSS is indicated, and of a TATA box signal located ~30 bp upstream.

(C) Proportion of putative mTSSs supported by two or more CAGE tags (y axis) for a subset of 4,879 sites, in which the "adenosine" position precedes the "Stack" position. Proportions are shown both for the adenosine position (pink bar), the stack position (cyan), and 4,879 positions randomly assigned to the same set of genes within the first 50 nt (gray bar).

(D) Distributions of the $\log_{10}$ transformed number of CAGE tags observed for the three groups defined in (C).

(E and F) Examples of genes with multiple internal mTSSs in human iPSCs. Number of reads beginning at each of the first 50 annotated positions in the IP (blue) and input (red) sample are shown. Detected mTSSs are highlighted in gray and marked with a black arrow; putative penultimate-base RT drop-offs are indicated with a purple arrow.

Figures 5E and 5F). Evaluation of the number of mTSSs found as a function of gene expression revealed that mTSS detection in our data has not reached saturation (Figures S4G and S4H), and thus the numbers reported here most likely reflect a lower bound.

To gain insight into the potential role(s) of mTSSs, we examined the extent to which mTSS presence in a gene was associated with gene structure, RNA stability, and translational efficiency features. Because our ability to detect mTSSs is biased toward more highly expressed genes, we used a randomly sampled control data set matching the expression patterns of the mTSS harboring genes (but lacking mTSSs). We found that the presence of mTSS is significantly inversely correlated with 5′ UTR length in both mouse and human (Figures S5A and S5E) and positively correlated with translation efficiency (Figures S5A, S5B, and S5F–S5H). However, in contrast to internal methylation, mTSSs were not associated with transcript stability (Figures S5A and S5B; see also Supplemental Discussion). These results may suggest that mTSSs and internal methylation sites may differ both in the factor(s) mediating their catalysis, and in their function, with internal sites potentially affecting stability and mTSSs potentially affecting translation. However, because these results are based on comparison between different transcripts, which may vary from each other in other respects that could potentially correlate with mTSS state, they must be interpreted with care. The presence and abundance of the mTSS

identified here add an additional unappreciated layer of complexity to the mammalian transcriptome, suggesting that a substantial fraction of transcripts exist in distinct isoforms, differing not only in the location of the TSS but also in its methylation state. It will be crucial to develop experimental methodologies to specifically perturb the methylation state of individual transcripts, and to monitor the impact of such perturbations on various stages of an RNA's life cycle, to conclusively understand the role of mTSSs.

DISCUSSION

The m6A Methylation Landscape

RNA methylations offer the potential of dynamically modulating characteristics of an mRNA's life cycle. Such regulation may potentially be applied either in specific conditions, or to specific transcripts, two scenarios that we set out to explore. By mapping methylations across various dynamic systems in human and mouse, we were able to more than triple the number of sites previously identified for either human or mouse (Dominissini et al., 2012; Meyer et al., 2012), while simultaneously enhancing the resolution of the detected sites, to nearly single-nucleotide level. We found that across the monitored conditions methylations appeared to be largely static, suggesting that the role of m6A in mammals is to a large extent condition independent. These results should best be interpreted with caution, because it is possible that m6a-seq is unable to capture subtle, quantitative differences in the proportion of transcripts methylated at a specific position, which could still be of functional consequence. To better assess the extent to which m6A is quantitative, measurements by orthogonal methods, such as the recently developed SCARLET (Liu et al., 2013), will be required for a large number of sites.

Although the topology of WTAP-dependent methylations is relatively static across mRNAs, we find that their distribution across genes is not uniform, and that methylations are depleted from abundantly transcribed messages. We moreover find that methylation densities correlate inversely with mRNA half-life. These findings are consistent with recent findings that YTHDF2 binds to methylated mRNAs and decreases messenger stability (Wang et al., 2014a), and with the fact that transcript half-lives are generally constant across different conditions. Indeed, in DCs, we do not see any change in m6A densities for those ~15% of transcripts where degradation rates change dynamically in response to LPS (data not shown). Collectively, our results suggest that internal m6A sites may play a ''basal'' role in controlling the half-lives of the methylated transcripts, and that this role is relatively constant across the different surveyed conditions.

The Multicomponent Methyltransferase Complex

Initial characterization of the mammalian methyltransferase complex revealed that at least two separable complexes (875 and 200 kDa in size) were required for restoring full m6A activity in vitro (Bokar et al., 1994). Subsequent analysis identified METTL3 as one of the components in the smaller complex (Bokar et al., 1997). Here, we show that WTAP, KIAA1429, and METTL14 are required for full m6A activity in vivo. The multiunit structure of the m6A methyltransferase is unique with respect to other characterized nucleic acid methyltransferases (Bokar et al., 1994), and offers the potential for complex regulation. Such is the case in yeast where onset and offset of methylations in meiosis were governed by up/downregulation of different members of this complex at different time points in meiosis (Schwartz et al., 2013).

Our results regarding WTAP and METTL14 are consistent with recently published results (Liu et al., 2014; Ping et al., 2014; Wang et al., 2014b), whereas KIAA1429 was not identified by these studies. The physical association of KIAA1429 with the methyltransferase complex is supported by the physical association of *Drosophila* homologs of KIAA1429 (virilizer) and WTAP (fl[2]d) in the context of female-specific alternative splicing (Granadino et al., 1990). Moreover, KIAA1429 in human localizes to the nuclear speckles (Horiuchi et al., 2013), as do METTL3 (Bokar, 2005) and WTAP (Horiuchi et al., 2013; Little et al., 2000), further supporting their association with each other. We identified another 13 proteins interacting with two or more of the above components. This collection may contain false-positives, and it will therefore be important to experimentally validate them via similar strategies.

The expression patterns of the different methyltransferase components offer important clues about the processes in which they are implicated. The yeast WTAP homolog, Mum2 is expressed in a meiosis-specific manner (Agarwala et al., 2012; Schwartz et al., 2013) and required for proper progression of meiosis (Davis et al., 2001). In *Drosophila*, fl(2)d and virilizer are both expressed at highest levels in the ovaries and brain (Robinson et al., 2013). Analysis of METTL14 expression levels across the Illumina Human Body Map (Farrell et al., 2014) also reveals particularly high levels of expression in testis and ovaries. These observations, combined with previous findings that ALKBH5, an m6A demethylase, is expressed at highest levels in the testis and required for spermatogenesis (Zheng et al., 2013), suggest that the methyltransferase complex may have an evolutionary conserved role in mammalian gametogenesis. Nonetheless, the ubiquitous expression of mammalian WTAP and KIAA1429 across tissues and ubiquitous presence of m6A suggests that this modification has acquired additional roles in mammals.

Rich Landscape of m6A at the mRNA Cap

In higher eukaryotes, but not in yeast, the 5′ cap structure contains 2′-*O*-ribose methylations at either the first nucleotide or the first and second nucleotide (Wei et al., 1975). When the first nucleotide of the transcript is an adenosine, this base can be methylated at the N6 position (Furuichi et al., 1975; Keith et al., 1978; Wei et al., 1975). However, no studies to date have resolved the methylation states of the caps of specific transcripts. Here, we simultaneously measured transcription initiation sites and their methylation status at thousands of sites and found that the transcription initiation landscape is highly complex, with transcript isoforms differing from each other in their methylation states. Our findings that mTSS state at the cap is correlated negatively with 5′ UTR length and positively with translation efficiencies suggest that the process most likely to be affected by mTSSs is translation. To investigate this more thoroughly, it will be necessary to better understand the

enzymatic components involved in cap methylation (Keith et al., 1978), and to develop targeted methodologies for monitoring and perturbing methylation status at the cap.

The ubiquity of m6A across dramatically different cell types and systems, the conservation of the different components involved in mediating and binding it across most studied systems, and the phenotypes associated with its depletion all suggest a fundamental role in mammalian cell biology. The rich proteomic and transcriptomic resources we provide in this study will help advance our understanding on the function of this epitranscriptomic modification.

EXPERIMENTAL PROCEDURES

Cell Culture

Human embryonic kidney 293T cells were transfected with plasmids encoding 3′ tagged proteins (Table S4) using Lipofectamine 2000 (Life Technologies). Pull-down was performed with anti-HIS (GenScript) or Anti-V5 (Life Technologies) antibodies. siRNAs (Thermo Scientific) were delivered using Lipofectamine RNAiMAX (Life Technologies); shRNAs (Broad RNAi Consortium) were delivered in polybrene supplemented media and selected for 72–96 hr. Details regarding derivation/culturing of HUES9, neural progenitors, mice brain samples, BMDCs, and hTERT immortalized fibroblasts reprogrammed into iPSCs are in Supplemental Experimental Procedures.

Mass Spectrometry Data Analysis

Details pertaining mass spectrometry data acquisition are in Supplemental Experimental Procedures. All mass spectra were processed using the Spectrum Mill software package v4.0 beta (Agilent Technologies) according to (Mertins et al., 2012). For each peptide, a log2 fold change was calculated between its intensities in the pull-down sample compared to the control. The peptides were mapped to genes based on the Uniprot database, and the median fold change was assigned to each gene. We then subtracted the median of the distribution of the log2 transformed values (across all genes) from the individual fold changes of each gene, to center the fold change distribution around 0. To filter out background contaminants, we used the CRAPOME database (Mellacheruvu et al., 2013), which summarizes the number of peptides identified for each protein across 343 control experiments. A protein was considered a contaminant if it was present (based on two or more peptides) across $\geq$20 control experiments, and filtered from subsequent analysis. We supplemented our data with published proteomics data of WTAP pull-down, obtained from Horiuchi et al. (2013). Then, we generated a network, whereby a directed edge existed between bait A (METTL3, WTAP, METTL14, YTHDF1, YTHDF2, YTHDF3, or m6A bait) and target B if the interaction between the two, in any of the experiments, was associated with a fold change >1.5.

M6A-Seq and Analysis of Internal m6A Methylation Sites

Isolation of total RNA, preparation of poly(A) RNA, the m6A pull-down procedure, and library preparation were performed as detailed in Schwartz et al. (2013). The computational analysis for internal sites is based on Schwartz et al. (2013), with some modifications (Supplemental Experimental Procedures).

Analysis of Putative Sites in the TSS

To identify putative mTSSs, we first counted the number of "left" reads originating at the first 50 bp of all annotated transcripts and recorded all positions in which a stack of five or more reads originated. An initial collection of all sites matching these criteria, and identified across any of the perturbations or dynamic systems, was generated. To account for the presumed RT drop-off, each site was assigned an "effective position," which corresponded to the stack position, unless that position did not harbor an adenosine and the position immediately preceding it did, in which case it was assigned the position preceding it. Sites were then aggregated based on genes and effective positions, and for each position we summarized (1) total number of reads originating from all input samples at each effective site, (2) total number of reads from all IP samples, and (3) a fold change between (1) and (2).

Statistical Analysis

All statistical analyses and visualizations were performed in R. Sequence logos were prepared using the SeqLogo package, and heatmaps were generated with the gplots package.

ACCESSION NUMBERS

Sequencing data have been deposited into the NCBI Gene Expression Omnibus under accession number GSE54365.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Discussion, Supplemental Experimental Procedures, five figures, and five tables and can be found with this article online at http://dx.doi.org/10.1016/j.celrep.2014.05.048.

AUTHOR CONTRIBUTIONS

S.S., M.R.M., A.R., and E.S.L. conceived most aspects of the study, acquired all m6A data, analyzed the data, and wrote the paper. T.W., K.M., T.S.M., D.T., R.S., and N. Hacohen provided important technical assistance. M.J., P.M., and S.A.C. performed the protein mass spectrometry. G.G.B., E.F., and M.E.P. performed nucleic acid mass spectrometry. D.C. performed iPSC experiments, N. Habib, N.E.S., and F.Z. performed experiments differentiating ES cell into NPCs.

ACKNOWLEDGMENTS

This work was supported by NHGRI CEGS P50 HG006193 (A.R.), U54 HG003067 (E.S.L), and Broad Institute Funds. A.R. was supported by an NIH Pioneer Award and HHMI. S.S. was supported by a European Molecular Biology Organization fellowship, and S.S. and D.C. were supported by Human Frontier Science Program fellowships. M.J. was supported by fellowships of the Swiss National Science Foundation for advanced researchers (SNF) and the Marie Sklodowska-Curie IOF.

Received: January 28, 2014
Revised: April 7, 2014
Accepted: May 27, 2014
Published: June 26, 2014

REFERENCES

Agarwala, S.D., Blitzblau, H.G., Hochwagen, A., and Fink, G.R. (2012). RNA methylation by the MIS complex regulates a cell fate decision in yeast. PLoS Genet. *8*, e1002732.

Bernstein, B.E., Birney, E., Dunham, I., Green, E.D., Gunter, C., and Snyder, M.; ENCODE Project Consortium (2012). An integrated encyclopedia of DNA elements in the human genome. Nature *489*, 57–74.

Bokar, J.A. (2005). The biosynthesis and functional roles of methylated nucleosides in eukaryotic mRNA. In Fine-Tuning of RNA Functions by Modification and Editing, H. Grosjean, ed. (New York: Springer), pp. 141–177.

Bokar, J.A., Rath-Shambaugh, M.E., Ludwiczak, R., Narayan, P., and Rottman, F. (1994). Characterization and partial purification of mRNA N6-adenosine methyltransferase from HeLa cell nuclei. Internal mRNA methylation requires a multisubunit complex. J. Biol. Chem. *269*, 17697–17704.

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.

Bujnicki, J.M., Feder, M., Radlinska, M., and Blumenthal, R.M. (2002). Structure prediction and phylogenetic analysis of a functionally diverse family of proteins homologous to the MT-A70 subunit of the human mRNA:m(6)A methyltransferase. J. Mol. Evol. *55*, 431–444.

Carninci, P., Sandelin, A., Lenhard, B., Katayama, S., Shimokawa, K., Ponjavic, J., Semple, C.A., Taylor, M.S., Engström, P.G., Frith, M.C., et al. (2006). Genome-wide analysis of mammalian promoter architecture and evolution. Nat. Genet. *38*, 626–635.

Clancy, M.J., Shambaugh, M.E., Timpte, C.S., and Bokar, J.A. (2002). Induction of sporulation in Saccharomyces cerevisiae leads to the formation of N6-methyladenosine in mRNA: a potential mechanism for the activity of the IME4 gene. Nucleic Acids Res. *30*, 4509–4518.

Davis, L., Barbera, M., McDonnell, A., McIntyre, K., Sternglanz, R., Jin, Q., Loidl, J., and Engebrecht, J. (2001). The Saccharomyces cerevisiae MUM2 gene interacts with the DNA replication machinery and is required for meiotic levels of double strand breaks. Genetics *157*, 1179–1189.

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

Duan, J., Shi, J., Ge, X., Dölken, L., Moy, W., He, D., Shi, S., Sanders, A.R., Ross, J., and Gejman, P.V. (2013). Genome-wide survey of interindividual differences of RNA stability in human lymphoblastoid cell lines. Sci Rep *3*, 1318.

Eisenberg, E., and Levanon, E.Y. (2003). Human housekeeping genes are compact. Trends Genet. *19*, 362–365.

Farrell, C.M., O'Leary, N.A., Harte, R.A., Loveland, J.E., Wilming, L.G., Wallin, C., Diekhans, M., Barrell, D., Searle, S.M., Aken, B., et al. (2014). Current status and new features of the Consensus Coding Sequence database. Nucleic Acids Res. *42* (Database issue), D865–D872.

Furuichi, Y., Morgan, M., Shatkin, A.J., Jelinek, W., Salditt-Georgieff, M., and Darnell, J.E. (1975). Methylated, blocked 5 termini in HeLa cell mRNA. Proc. Natl. Acad. Sci. USA *72*, 1904–1908.

Granadino, B., Campuzano, S., and Sánchez, L. (1990). The Drosophila melanogaster fl(2)d gene is needed for the female-specific splicing of Sex-lethal RNA. EMBO J. *9*, 2597–2602.

Horiuchi, K., Kawamura, T., Iwanari, H., Ohashi, R., Naito, M., Kodama, T., and Hamakubo, T. (2013). Identification of Wilms' tumor 1-associating protein complex and its role in alternative splicing and the cell cycle. J. Biol. Chem. *288*, 33292–33302.

Keith, J.M., Ensinger, M.J., and Mose, B. (1978). HeLa cell RNA (2′-O-methyladenosine-N6-)-methyltransferase specific for the capped 5′-end of messenger RNA. J. Biol. Chem. *253*, 5033–5039.

Little, N.A., Hastie, N.D., and Davies, R.C. (2000). Identification of WTAP, a novel Wilms' tumour 1-associating protein. Hum. Mol. Genet. *9*, 2231–2239.

Liu, N., Parisien, M., Dai, Q., Zheng, G., He, C., and Pan, T. (2013). Probing N6-methyladenosine RNA modification status at single nucleotide resolution in mRNA and long noncoding RNA. RNA *19*, 1848–1856.

Liu, J., Yue, Y., Han, D., Wang, X., Fu, Y., Zhang, L., Jia, G., Yu, M., Lu, Z., Deng, X., et al. (2014). A METTL3-METTL14 complex mediates mammalian nuclear RNA N6-adenosine methylation. Nat. Chem. Biol. *10*, 93–95.

Machnicka, M.A., Milanowska, K., Osman Oglou, O., Purta, E., Kurkowska, M., Olchowik, A., Januszewski, W., Kalinowski, S., Dunin-Horkawicz, S., Rother, K.M., et al. (2013). MODOMICS: a database of RNA modification pathways—2013 update. Nucleic Acids Res. *41* (Database issue), D262–D267.

Mellacheruvu, D., Wright, Z., Couzens, A.L., Lambert, J.P., St-Denis, N.A., Li, T., Miteva, Y.V., Hauri, S., Sardiu, M.E., Low, T.Y., et al. (2013). The CRAPome: a contaminant repository for affinity purification-mass spectrometry data. Nat. Methods *10*, 730–736.

Mertins, P., Udeshi, N.D., Clauser, K.R., Mani, D.R., Patel, J., Ong, S.E., Jaffe, J.D., and Carr, S.A. (2012). iTRAQ labeling is superior to mTRAQ for quantitative global proteomics and phosphoproteomics. Mol. Cell. Proteomics *11*, 014423.

Meyer, K.D., Saletore, Y., Zumbo, P., Elemento, O., Mason, C.E., and Jaffrey, S.R. (2012). Comprehensive analysis of mRNA methylation reveals enrichment in 3′ UTRs and near stop codons. Cell *149*, 1635–1646.

Ni, T., Corcoran, D.L., Rach, E.A., Song, S., Spana, E.P., Gao, Y., Ohler, U., and Zhu, J. (2010). A paired-end sequencing strategy to map the complex landscape of transcription initiation. Nat. Methods *7*, 521–527.

Ortega, A., Niksic, M., Bachi, A., Wilm, M., Sánchez, L., Hastie, N., and Valcárcel, J. (2003). Biochemical function of female-lethal (2)D/Wilms' tumor suppressor-1-associated proteins in alternative pre-mRNA splicing. J. Biol. Chem. *278*, 3040–3047.

Ping, X.L., Sun, B.F., Wang, L., Xiao, W., Yang, X., Wang, W.J., Adhikari, S., Shi, Y., Lv, Y., Chen, Y.S., et al. (2014). Mammalian WTAP is a regulatory subunit of the RNA N6-methyladenosine methyltransferase. Cell Res. *24*, 177–189.

Plessy, C., Bertin, N., Takahashi, H., Simone, R., Salimullah, M., Lassmann, T., Vitezic, M., Severin, J., Olivarius, S., Lazarevic, D., et al. (2010). Linking promoters to functional transcripts in small samples with nanoCAGE and CAGEscan. Nat. Methods *7*, 528–534.

Robinson, S.W., Herzyk, P., Dow, J.A., and Leader, D.P. (2013). FlyAtlas: database of gene expression in the tissues of Drosophila melanogaster. Nucleic Acids Res. *41* (Database issue), D744–D750.

Schwanhäusser, B., Busse, D., Li, N., Dittmar, G., Schuchhardt, J., Wolf, J., Chen, W., and Selbach, M. (2011). Global quantification of mammalian gene expression control. Nature *473*, 337–342.

Schwartz, S., Agarwala, S.D., Mumbach, M.R., Jovanovic, M., Mertins, P., Shishkin, A., Tabach, Y., Mikkelsen, T.S., Satija, R., Ruvkun, G., et al. (2013). High-resolution mapping reveals a conserved, widespread, dynamic mRNA methylation program in yeast meiosis. Cell *155*, 1409–1421.

Wang, X., Lu, Z., Gomez, A., Hon, G.C., Yue, Y., Han, D., Fu, Y., Parisien, M., Dai, Q., Jia, G., et al. (2014a). N6-methyladenosine-dependent regulation of messenger RNA stability. Nature *505*, 117–120.

Wang, Y., Li, Y., Toth, J.I., Petroski, M.D., Zhang, Z., and Zhao, J.C. (2014b). N6-methyladenosine modification destabilizes developmental regulators in embryonic stem cells. Nat. Cell Biol. *16*, 191–198.

Wei, C.M., Gershowitz, A., and Moss, B. (1975). Methylated nucleotides block 5′ terminus of HeLa cell messenger RNA. Cell *4*, 379–386.

Zheng, G., Dahl, J.A., Niu, Y., Fedorcsak, P., Huang, C.M., Li, C.J., Vågbø, C.B., Shi, Y., Wang, W.L., Song, S.H., et al. (2013). ALKBH5 is a mammalian RNA demethylase that impacts RNA metabolism and mouse fertility. Mol. Cell *49*, 18–29.

Cell Reports, Volume *8*

Supplemental Information

Perturbation of m6A Writers Reveals Two Distinct Classes of mRNA Methylation at Internal and 5′ Sites

Schraga Schwartz, Maxwell R. Mumbach, Marko Jovanovic, Tim Wang, Karolina Maciag, G. Guy Bushkin, Philipp Mertins, Dmitry Ter-Ovanesyan, Naomi Habib, Davide Cacchiarelli, Neville E. Sanjana, Elizaveta Freinkman, Michael E. Pacold, Rahul Satija, Tarjei S. Mikkelsen, Nir Hacohen, Feng Zhang, Steven A. Carr, Eric S. Lander, and Aviv Regev

Supplemental Discussion

RT-dropoff one nucleotide prior to the TSS

To investigate the extent to which transcription start sites, in general, and methylated TSS, in particular, could be precisely identified based on our data, we first examined the distribution of >30,000 sites in mouse with >10 reads beginning at a specific site (across all conditions). We noted that this catalogue is enriched towards the first (annotated) position, as expected, but to an equal extent also at the second position, which was unexpected **(Fig. S4A)**. To further address this, we analyzed the proportion of sites harboring an adenosine in the first or the second position, as a function of fold change over input. At the first position, we found a strong bias towards adenosine, as anticipated (**Fig. S4B**). However, at the second position, no such bias was noted. Instead, even when reads stacked at Position 2, the bias towards adenine was still at Position 1, and the magnitude of this bias was comparable to the one obtained when the stacks were at Position 1. Thus, the stacks at Positions 1 and 2 both appear to reflect methylations occurring at Position 1. The stacks at positions 2 likely reflect drop-off of the RT polymerase one nucleotide prior to reaching the 5' terminus, possibly due to steric obstruction of the cap structure. Virtually identical results were obtained when performing the same analysis in human (data not shown).

As an independent source of validation that the sites preceding the stacks are the true TSSs, in cases in which the former lack an adenosine, note also the analysis presented in **Fig. 5C-D**, assessing correspondence with a CAGE dataset.

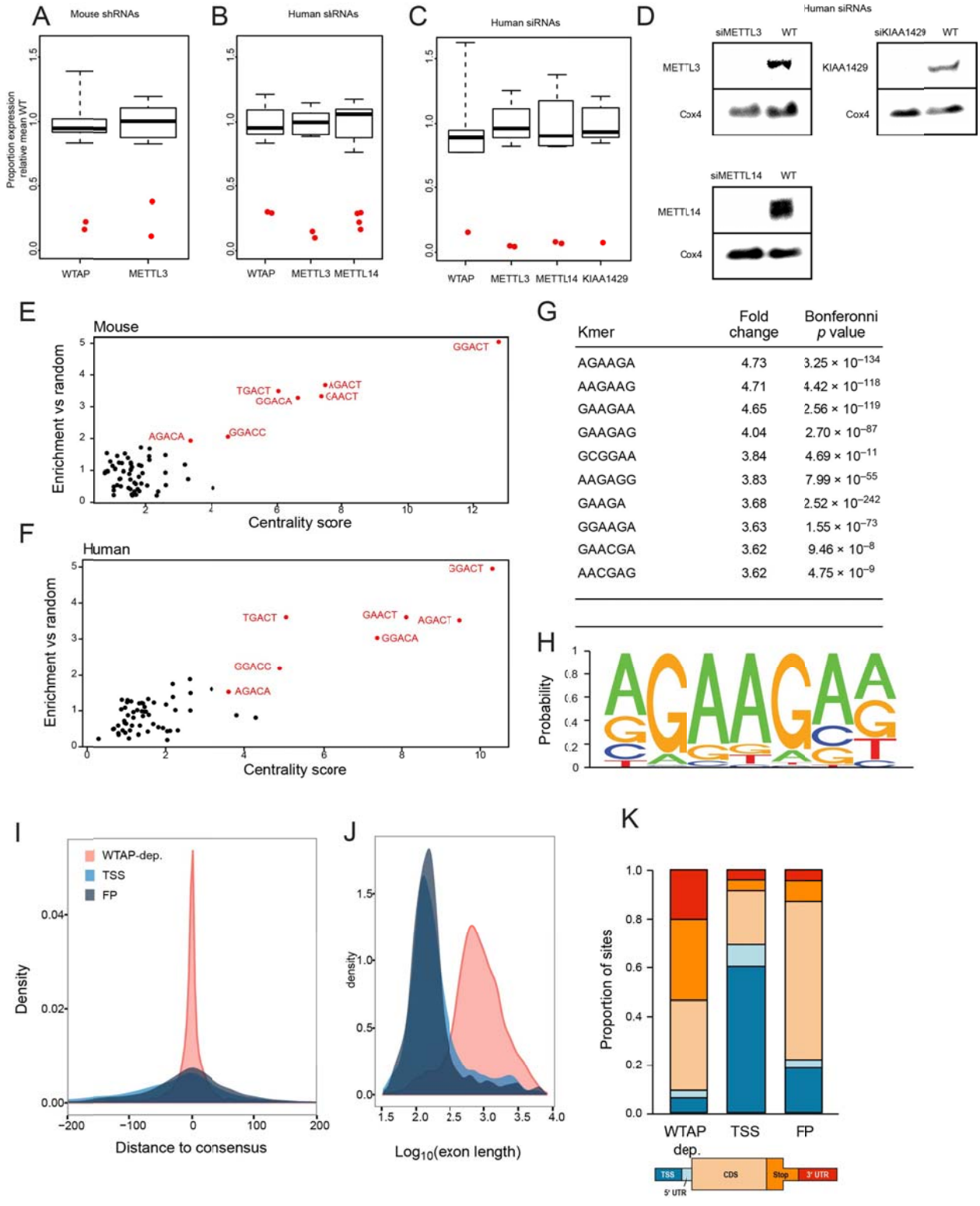
Correlation of mTSS state with transcript structure, stability and translation rates

To gain further insight into the role of mTSSs, we examined the extent to which mTSS presence in a gene was associated with gene structure, RNA stability and translational efficiency. As our ability to detect mTSSs is biased towards more highly expressed genes, we used a randomly sampled control dataset matching the expression patterns of the mTSS harboring genes (but lacking mTSSs). Specifically, we compared sets of 2,891 human genes and 3,459 mouse genes containing an mTSS in four or more of the samples, to a roughly equal number of expression-matched controls.

Across the vast majority of datasets, mTSS presence was most strongly associated with 5' UTR length, with genes harboring mTSSs having significantly shorter 5' UTRs compared to their non-methylated expression matched counterparts (**Figs. S5A-C**). This was observed using the full human and mouse datasets defined above (**Fig. S5A-C**) and when analyzing mTSSs identified in a specific condition (data not shown). Furthermore, a negative association with 5' UTR length was observed when binning genes into 5 equally sized bins based on number of samples in which an mTSS was observed (a variable capturing both our confidence in the sites, and potentially the stoichiometry of modification at site) (**Fig. S5D-E**).

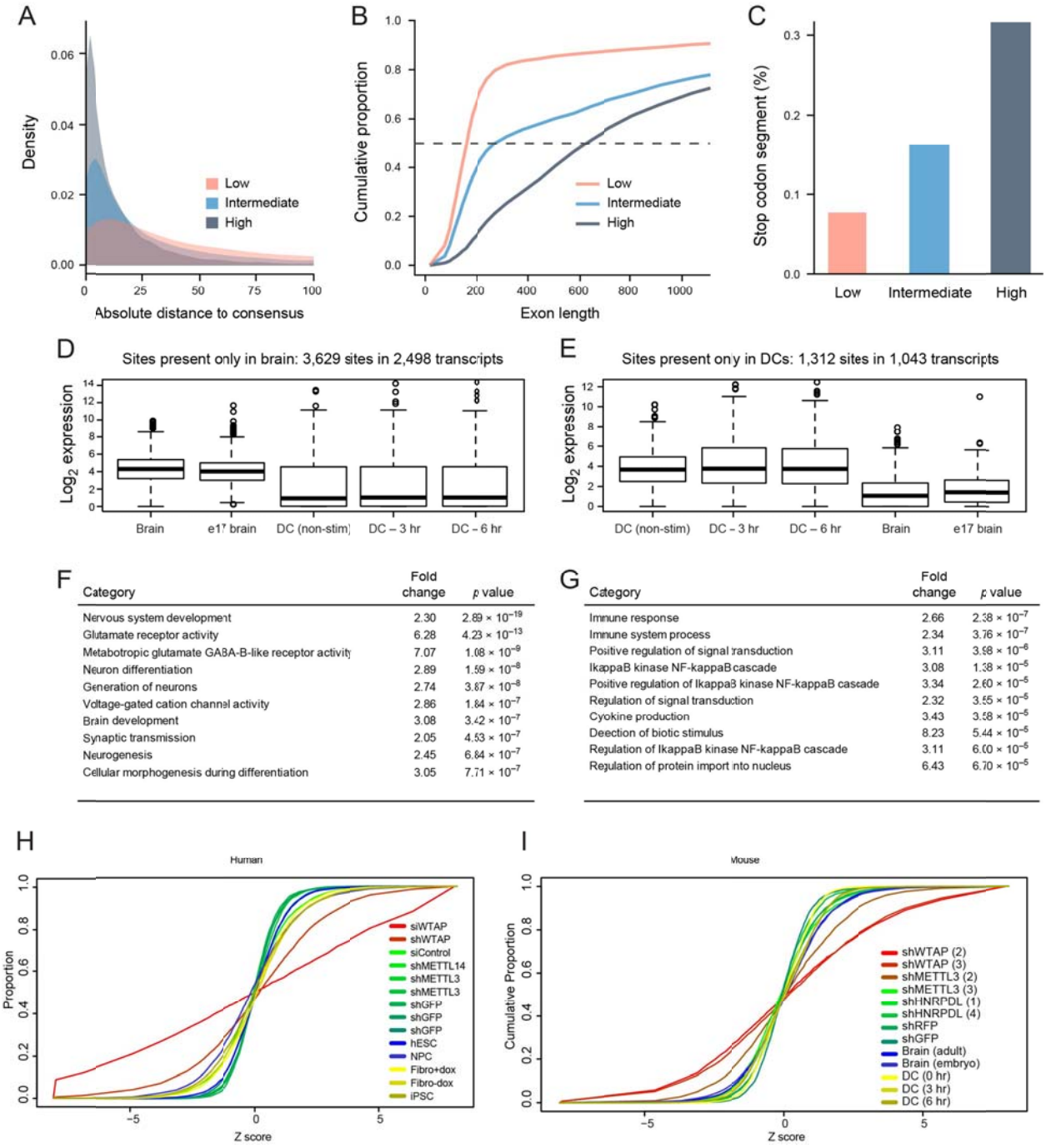
We also found a strong positive association between the presence of mTSS and translational efficiency (**Fig. S5A-B**), with genes harboring mTSSs exhibiting increased translational efficiencies compared to the expression matched controls (**Fig. S5F-G**). These effects were particularly strong when comparing our mouse dataset with translational efficiencies measured in mESC (Ingolia et al., 2011), but also observed in the human dataset when comparing with efficiencies measured in HeLa and HEK293 cells (Subtelny et al., 2014). We also observed a quantitative relationship between the number of conditions with mTSSs and translational efficiencies when using the binning approach described above (**Fig. S5H** and data not shown). As expression level, number of conditions in which an mTSS was observed, and 5' UTR length are all correlated with each other, we used a linear model to predict translational efficiencies based on these three variables. In mouse, all variables significantly contributed to the model in this joint context ($P<10^{-4}$ in all cases), whereas in human, the number of conditions in which mTSSs were observed ceased to remain significant. Finally, we note that in these analyses we noted only a negligible impact on RNA stability (**Fig. S5A-B**).

Together, our data strongly supports an inverse association between 5’ UTR length and presence of mTSS, and moreover provides indications that mTSSs in genes with shorter 5’ UTR are associated with increased translation efficiencies.



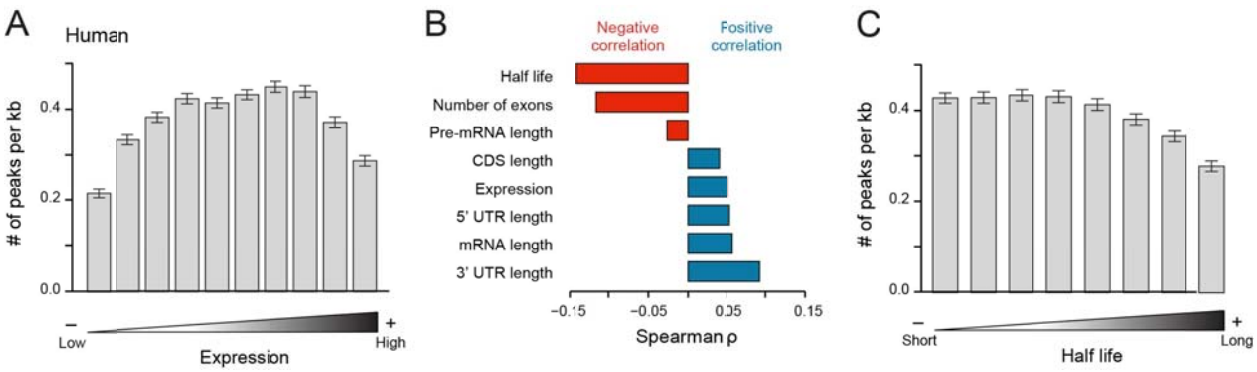
Kmer	Fold change	Bonferonni *p* value
AGAAGA	4.73	3.25×10^{-134}
AAGAAG	4.71	4.42×10^{-118}
GAAGAA	4.65	2.56×10^{-119}
GAAGAG	4.04	2.70×10^{-87}
GCGGAA	3.84	4.69×10^{-11}
AAGAGG	3.83	7.99×10^{-55}
GAAGA	3.68	2.52×10^{-242}
GGAAGA	3.63	1.55×10^{-73}
GAACGA	3.62	9.46×10^{-8}
AACGAG	3.62	4.75×10^{-9}

Supplemental Figure S1 related to Figure 2. (A-D) Knockdown efficiency. (A-C) Shown are sample transcript knockdown levels (Y axis, expressed as proportion of average gene expression among non-perturbed samples), (upon perturbation of (**A**) WTAP or METTL3 (X axis) using shRNAs in mouse embryonic fibroblasts; (**B**) WTAP, METTL3, METTL14 using shRNAs in A549 cells, or (**C**) of WTAP, METTL3, METTL14 and KIAA1429 using siRNAs in human A549 cells. Transcript levels were determined based on TMM-normalized values from the (input) RNA-seq experiments. Boxplots present the distribution of the mean-normalized expression levels across all non-perturbed samples, with whiskers extending until the minimum and maximum of the distribution. Red dots indicate levels in the perturbed samples. Some red dots are additionally labelled with information on the experiment or hairpin used **(D)** Protein knockdown levels (by Western blots) following siRNA mediated depletion of METTL3, METTL14 and KIAA1429 from A549 cells. Cox4 was used as a loading control. **(E,F)** Identification of optimal set of consensus sites. For each of 64 pentamers harboring an AC at position 3-4 (NNACA), presented is a fold change quantifying enrichment in the set of 10,000 top motifs in human **(E)** and mouse **(F)** over an equally sized shuffled dataset (Y-axis), along with a centrality score (X-axis) denoting the extent to which a motif is biased towards appearing at the very center of the detected peak (**Methods**). The motifs in red, scoring highest in both metrics and identical between human and mouse, were selected as the set of consensus motifs used throughout the manuscript. **(G,H)** Sequence features associated with false positive methylation sites. (**G**) Shown are the top scoring pentamers (first column) enriched in sites from the 'false-positive' cluster, along with their fold enrichment (second column) and a Bonferroni-corrected P-value for a chi-squared test assessing their enrichment when comparing them to a set of WTAP-dependent sequences. (**H**) Sequence logo depicting a motif obtained from clustering all enriched pentamers with fold change >2 and corrected P value < 0.05. **(I-K)** Hallmarks of methylation in human A549 cells. Distributions of distances to consensus **(I)**, lengths of internal exons harboring methylated sites **(J)**, and sites across defined gene segments **(K)**, for the sites in the three clusters (WTAP-dependent true positives: TP; False positives: FP; and TSS related: TSS) in Fig. 2F**.** Distributions and gene segments defined as in **Fig. 2B-D**.

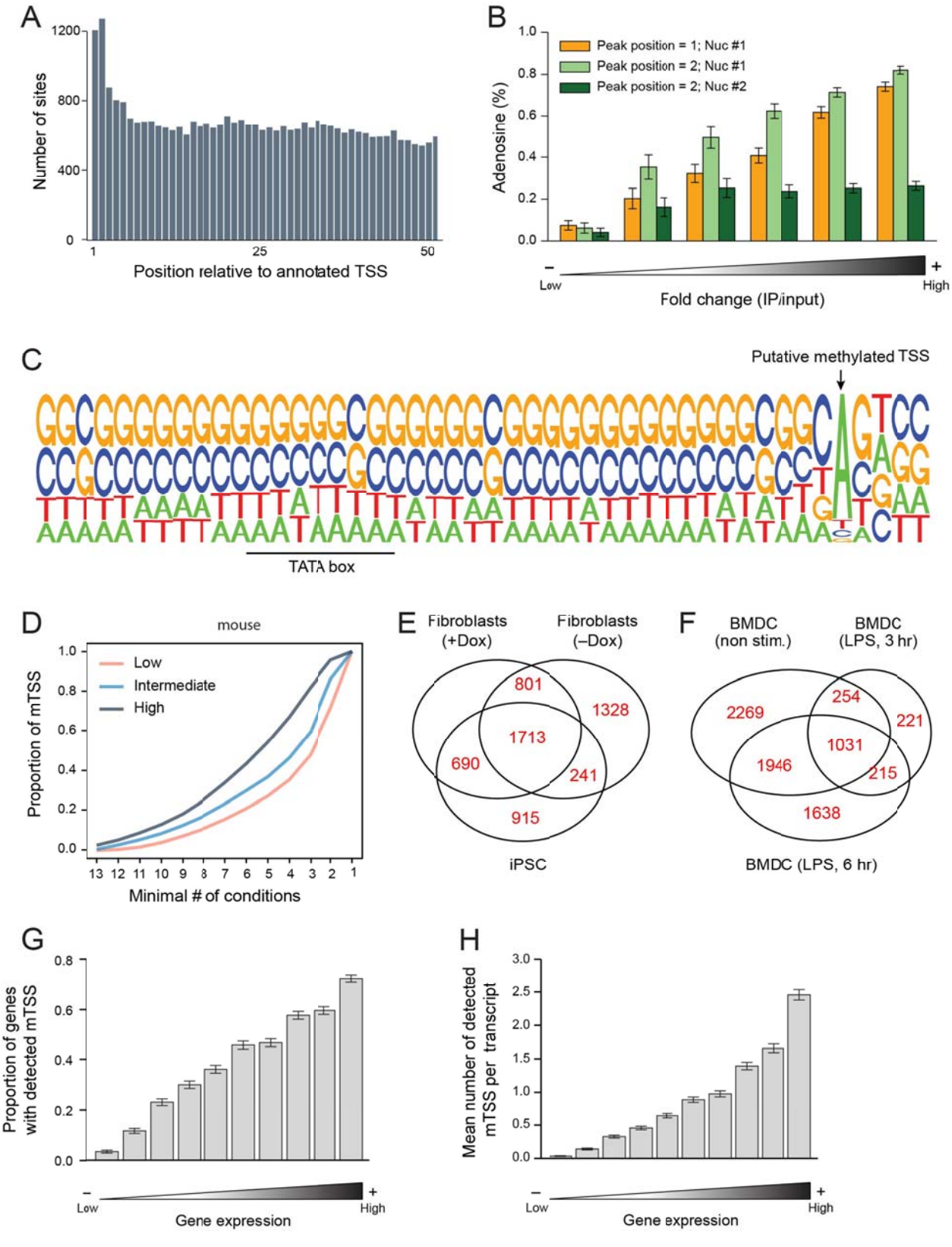


Supplemental Figure S2 related to Figure 3. Characterization of methylation sites. (A) Absolute distance from consensus for sets of sites with 'high', 'intermediate' and 'low' confidence, in human datasets, defined as described in the text. **(B)** Cumulative distributions of exon lengths in the 'high', 'intermediate' and 'low' confidence sets. The X-axis is truncated at 1000 nt. The grey dashed line indicates the median exon length across the three sets. **(C)** Proportion of sites comprised in the stop codon segments for each of the three sets. **(D,E)** Distribution of expression levels in the two brain samples and three dendritic cell samples for

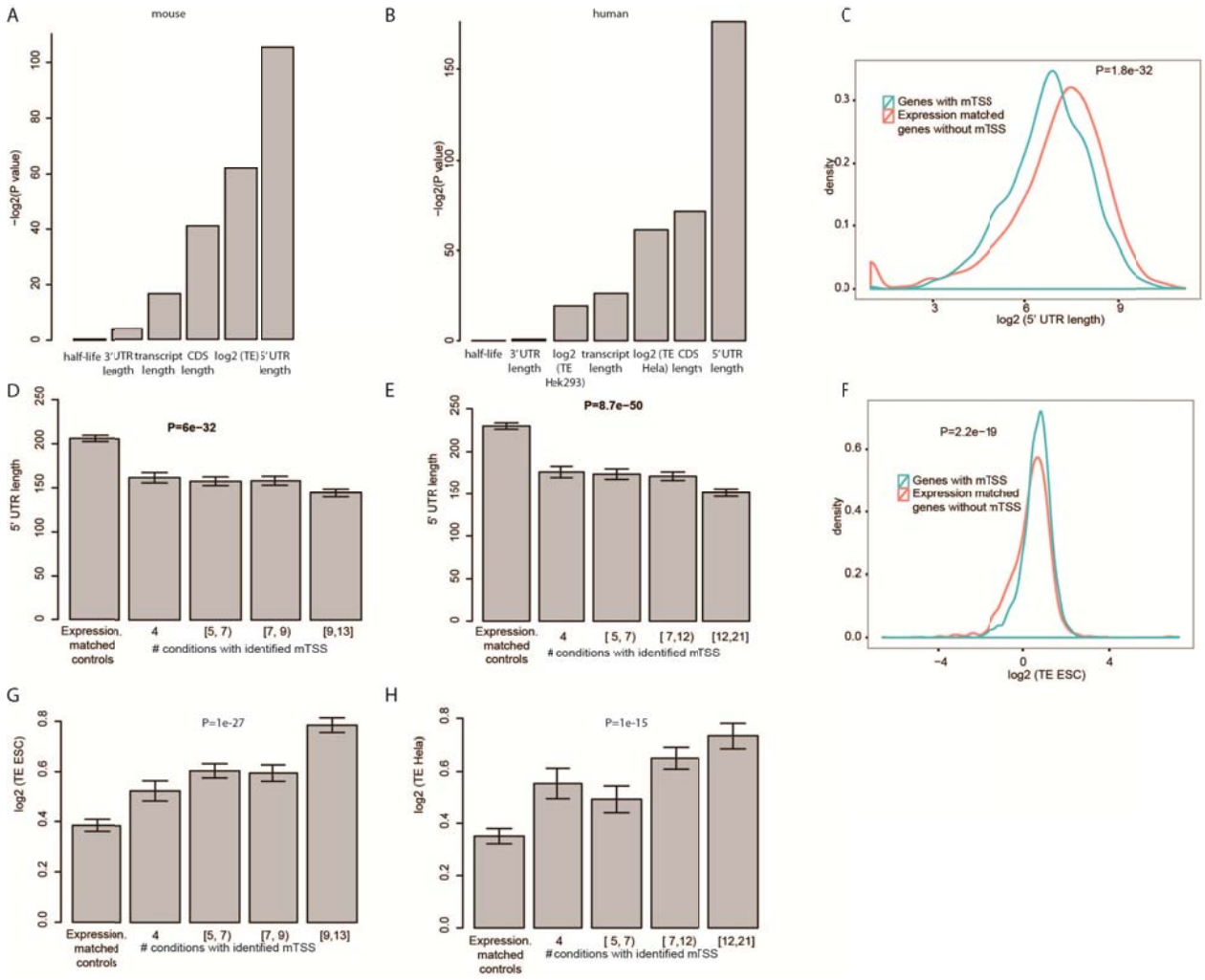
genes harboring sites present only in brain (D) or only in DCs (E). **(F,G)** GO terms enriched in genes from (D) and (E), respectively. **(H)** Variability of POI scores across the two dynamic datasets in human (reprogramming into iPSCs and differentiation into NPCs) compared to variability of these scores across different experiments performed in A549 cells. For this analysis, we designated a control cohort of perturbation experiments in A549 cells in which we failed to observe a substantial impact on methylation (replicates of the shMETTL3, shMETTL14, and shGFP replicates). For each site in each experiment in this cohort, a distribution of Z scores was calculated based on all remaining experiments in the cohort. Similarly, distributions of Z scores with respect to the control cohort were calculated for all other conditions, including the dynamic experiments and WTAP perturbation. Cumulative distributions of Z scores for a subset of these conditions are displayed. These analyses indicated the perturbations of WTAP as clear outliers, but the distributions of the remaining **(I)** Analysis as in (H) but for mouse. As cohort controls in this experiment, we made use of shMETTL(3), shHNRPDL, shRFP, shGFP, as well as the dendritic cells and the brain samples.



Supplemental Figure S3 related to Figure 4. Features associated with methylation site density in human genes. (A) Association of gene expression levels (X axis) with methylation peak density (Y axis). **(B)** Spearman's ρ (X axis) between methylation densities and different gene level features (Y axis), performed as in **Fig. 4C**. **(C)** Association of mRNA half life with methylation density. Analyses in this figure were done as presented in Figure 4, but for the human data.


A
Number of sites
1200
800
400
0
1
25
50
Position relative to annotated TSS
B
Peak position = 1; Nuc #1
Peak position = 2; Nuc #1
Peak position = 2; Nuc #2
Adenosine (%)
1.0
0.8
0.6
0.4
0.2
0.0
–
+
Low
High
Fold change (IP/input)
C
Putative methylated TSS
TATA box
D
mouse
Low
Intermediate
High
Proportion of mTSS
1.0
0.8
0.6
0.4
0.2
0.0
13 12 11 10 9 8 7 6 5 4 3 2 1
Minimal # of conditions
E
Fibroblasts (+Dox)
Fibroblasts (–Dox)
801
1328
1713
690
241
915
iPSC
F
BMDC (non stim.)
BMDC (LPS, 3 hr)
2269
254
221
1031
1946
215
1638
BMDC (LPS, 6 hr)
G
Proportion of genes with detected mTSS
0.8
0.6
0.4
0.2
0.0
–
+
Low
High
Gene expression
H
Mean number of detected mTSS per transcript
3.0
2.5
2.0
1.5
1.0
0.5
0.0
–
+
Low
High
Gene expression

Supplemental Figure 4 related to Figure 5. Anal ysis of methylated TSSs in human and mouse. (A) Distribution of relative positions of all mouse methylated sites harboring >10 reads (across all IP samples) with respect to annotated transcription start sites. **(B)** Proportion of positions containing an adenosine (Y axis), as a function of fold change between number of reads in the IP sample compared to the input sample (X axis), for 1,247 sites occurring at the very first annotated position (orange bars), and for 1,292 sites occurring at the second annotated position, showing separately for the latter the first position (light green bars) and the second (dark green bars). Error bars indicate the standard error of the proportion. **(C)** Sequence logo surrounding the human internal mTSSs (position >4) derived from 3,552 sites with log2 fold changes >4.5, as shown for mouse in **Fig. 5b. (D)** Proportion of methylated TSS sites (Y axis) in each of three equally sized fold-change bins (low, intermediate and high) detected in at least a given number of experiments among 13 experiments in mouse (X axis), including both the perturbation experiments (displayed in **Fig. 2**) and the experiments examining the dynamics of the methylations (displayed in **Fig. 3**). **(E,F)** Venn-diagrams showing the extent of overlap in detection of mTSSs across pluripotency reprogramming (human, **E**), and dendritic cell stimulation (mouse, **F**) and. Only sites with fold- changes >4 were considered as present in this analysis. Gene expression levels were not controlled for in this analysis. **(G)** Evaluation of saturation in mTSS detection in MEFs. Genes expressed above the 60^{th} percentile were binned into 10 equally sized bins based on their expression (X axis). The proportion of genes for which an mTSS was identified is plotted for each bin (Y axis). Error bars – standard error of proportion. **(H)** As in (G), but quantifying the average number of unique mTSSs identified per transcript, as a function of expression. Error bars – SEM.



Supplemental Figure 5 related to Figure 5. Features correlating with mTSS presence. **(A-B)** Log-transformed P values (Y axis), obtained for a non-parametric Mann-Whitney test comparing distributions of each indicated variable (X axis) in genes harboring an mTSSs and in expression matched controls in mouse (A) and human (B) datasets. **(C)** Distributions of 5' UTR lengths within genes harboring an mTSS (blue) and matched controls (red) in mouse. **(D-E)** Quantitative relationship between binned number of samples in which an mTSS was identified (X axis) and mean 5' UTR length (Y axis) in mouse (D) and human (E). Lengths are also shown for expression matched controls (left most bar). Error bars – standard error of the mean (SEM). P values derived using the Kruskal-Wallis test are indicated; Post-hoc paired wilcox-test were significant for all pairwise comparisons involving the two extreme bins. (F) Distribution of translational efficiencies within genes harboring an mTSS (blue) and matched controls (red), in mouse, measured in mES cells. **(G-H)** Quantitative relationship between binned number of samples in which an mTSS was identified (X axis) and mean translational efficiencies (Y axis) in mouse (G) and human (H). Efficiencies are also shown for expression matched controls (leftmost bar). Error bars – standard error of the mean (SEM). P values derived using the Kruskal-Wallis test are indicated; In human, post-hoc paired Mann-Whitney tests were significant across all

pairwise comparison in human with the exception of the second and third bins, whereas in mouse only the pairwise interactions involving either of the extreme bins were significant.

Supplemental Table legends

Supplemental Table 1, related to Figure 1: All protein-protein / RNA-protein putative interactions (fold changes >1.5) identified in this study. For each association, the bait and target are listed, the fold-change of the association, the number of quantified peptides and spectra, and in indication of abundance in the CRAPOME database, which was used for filtering purposes.

Supplemental Table 2, related to Figures 2 and 3: Set of m6A peaks identified in this study. Genomic coordinates are provided for each peak, along with a UCSC id of the gene harboring it, transcription start and end sites of the gene, distance from the peak to the nearest consensus site, a classification score (based on a linear combination of dependence on WTAP and other features, see text) and a peak confidence category based on it, and a set of POI scores for the peak strength across the different experiments.

Supplemental Table 3, related to Figure 5: Set of putative mTSS sites identified in human and mouse. Genomic coordinates and gene annotations are provided, along with number of experiments in which the given site obtained a fold-change exceeding 4, the overall fold change obtained for a site when summarizing IP and input reads across all experiments, and an indicator, for each condition, of whether or not a site was detected across each individual sample.

Supplemental Table 4, related to Methods: Information pertaining to the proteomic experiments. Promoters and 3' tags of constructs are provided, the nature of the control experiment against which the pulldown was quantified, and the number of replicates performed for the experiment.

Supplemental Table 5, related to Methods: Information pertaining to the reagents used for the RNAi experiments. Product ids for the siRNAs and shRNAs are indicated.

Supplemental Experimental Procedures

Mammalian cell culture

Human 293T, A549 and p53-/- mouse embryonic fibroblast (MEF) cell lines were cultured in DMEM media supplemented with 10% FBS and 1% penicillin/streptomycin.

Protein pull-down and mass-spectrometry

Human 293T cells were grown in 10 cm plates to 70% confluency, and transfected with plasmids encoding 3' tagged proteins (see **Table S5** for details on epitopes) using Lipofectamine 2000 (Life Technologies), following the manufacturer's protocol. Cells were harvested 24 hours after transfection, pelleted by centrifugation and washed in PBS. The cell pellet was then resuspended in 600μL of lysis buffer (150mM NaCl, 1.5mM MgCl2, 0.5% NP-40, 50mM Tris-HCl) with protease inhibitor (Roche). Cells were lysed on ice for 30 minutes and then centrifuged at 4 degrees for 15 minutes at 10000 rpm. The supernatant was collected and BCA was used to estimate protein amounts. 50μL of Dynabeads Protein G beads were washed in 250μL of lysis buffer three times and then resuspended in 500μL. They were then conjugated to Anti-HIS (GenScript) or Anti-V5 (Life Technologies) antibodies by rotating at 4 degrees for 2 hours. The beads were then washed three times in 250μL of lysis buffer and then the lysate was added followed by a 2 hour rotation at 4 degrees. The beads were again washed six times in 500μL of lysis buffer followed by an elution into elution buffer (50mM DTT, 25% LDS) at 80 degrees for 10 minutes.

Proteins were precipitated by adding -20°C cold acetone to the eluate (acetone to eluate ratio 7:1) and overnight incubation at -20°C. The proteins were pelleted by centrifugation at 20000xg for 15min at 4°C. The supernatant was discarded and the pellet was left to dry by evaporation. The

protein pellet was reconstituted in 50 µl lysis buffer (8M Urea, 75mM NaCl, 50mM Tris/HCl pH 8.0, 1mM EDTA) and protein concentrations were determined by BCA assay (Pierce, Rockford, IL). 5 to 10µg total protein per affinity enrichment were obtained condition. Disulfide bonds were reduced with 5mM dithiothreitol and cysteines were subsequently alkylated with 10mM iodoacetamide. Samples were diluted 1:4 with 50mM Tris/HCl (pH 8.0) and sequencing grade modified trypsin (Promega, Madison, WI; V5113) was added in an enzyme-to-substrate ratio of 1:50. After 16h of digestion, samples were acidified with 1% formic acid (final concentration). Tryptic peptides were desalted on C18 StageTips according to (Rappsilber et al., 2007) and evaporated to dryness in a vacuum concentrator.

Desalted peptides were labeled with the iTRAQ reagent according to the manufacturer's instructions (AB Sciex, Foster City, CA) and to (Mertins et al., 2012). Briefly, for 10µg of peptide 0.33 units of iTRAQ reagent was used. Peptides were dissolved in 10µl of 0.5M TEAB pH 8.5 solution and the iTRAQ reagent was added in 25µl of ethanol. After 1h incubation the reaction was stopped with 50mM Tris/HCl (pH 8.0). Differentially labeled peptides were mixed and subsequently desalted on C18 StageTips (Rappsilber et al., 2007) and evaporated to dryness in a vacuum concentrator. Peptides were reconstituted in 20µl 3% MeCN/0.1% formic acid. LC-MS/MS analysis was performed as described in (Mertins et al., 2013).

Gene silencing using siRNAs

Human A549 cells were plated in 6-well plates at 20% confluency. siRNAs targeting the relevant genes (**Table S8**) were transfected using Lipofectamine RNAiMAX (Life Technologies) following the manufacturer's protocols, with two additional siRNA boosts delivered at 48 and 96 hours following transfection. Cells were harvested at 144 hours.

Gene silencing using shRNA

Human A549 cells and $p53^{-/-}$ mouse fibroblasts were plated in 6-well plates at 20% confluency. Cells were infected with shRNAs targeting the relevant genes in human and mouse (**Table S5**), in media supplemented with 8μg/mL polybrene followed by centrifugation at 2200 rpm for 30 minutes. 18 hours later media was removed and replaced with media supplemented with 2μg/mL puromycin. Cells were harvested 72-96 hours later. For identifying effective shRNAs, we screened – for each gene – 4-6 shRNAs available through the Broad RNAi Consortium (TRC), and quantified decreases in expression level using qPCR; We selected 1-2 shRNAs for which we observed maximal decreases in expression level, and for which these levels were below 30% (typically even less) of WT expression levels.

HPLC-MS of nucleosides

50 ul oligo-dT selected mRNA fractions and flowthrough RNA fractions were digested with 2 units of P1 nuclease (US biological) at 50oC in 50 mM ammonium acetate buffer pH 5.3 with 5 mM zinc chloride for 2 hours. Nucleotides were dephosphorylated by addition of 5 units of CIP (New England Biolabs) for another 2 hours at 37oC, and then diluted 1:10 in acetonitrile. Metabolome data were acquired on a Dionex UltiMate 3000 UPLC coupled to a Q Exactive orbitrap mass spectrometer (Thermo Fisher Scientific). Separation was performed on a Luna 3 μm Amino Column (150 x 2 mm; Phenomenex) using a gradient of 5 mM ammonium acetate + 0.2% v/v ammonium hydroxide, pH 10, and acetonitrile. The mass spectrometer was operated in the targeted SIM mode with DZA as the internal standard. Metabolites were identified by accurate mass (+/- 5 ppm) and retention time. Peak areas were calculated with QuanBrowser (Thermo Fisher Scientific).

M6A-Seq

Isolation of total RNA, preparation of poly(A) RNA, the m6A pull-down procedure, and library preparation were performed as detailed in (Schwartz et al., 2013). Anti-m6A antibodies were either obtained from Synaptic Systems (catalog number 202 003) or generously provided by New England Biolabs (Meyer et al., 2012). Samples were sequenced on Illumina HiSeq 2000 and HiSeq 2500 platforms in paired end mode, with 25-30 nucleotides sequenced from each of the two ends.

Read alignment

Reads were initially mapped against a set of human or mouse ribosomal RNA (rRNA) sequences using Bowtie (version 0.12.7), and all reads aligning to the rRNA were discarded. All remaining reads were aligned against the human (hg19) or mouse (mm9) genome using Tophat (version 1.4.1). Parameters used were '--max-multihits 1 –prefilter-multihits' and '–transcriptome-index', for which we assigned a pre-indexed version of the relevant transcriptomes, based on the UCSC Known Genes set of annotations. An in-house script was then used to project all reads aligning to the genome upon the human and mouse transcriptomes. Only reads fully matching a transcript structure, as defined by the 'UCSC Known Genes' transcriptome annotation, were retained. Such reads were computationally extended in transcriptome space from the beginning of the first read to the end of its mate, and coverage in transcriptome-space was calculated for each nucleotide across all transcripts.

Detection of putative m6A sites

Putative m6A sites were identified using a 3 step-approach, consisting of: (**1**) Examination of the IP sample, to identify regions within genes in the IP samples that were enriched in comparison to background gene levels; (**2**) Comparison of IP sample with input sample, to ensure that these

regions were not enriched in corresponding input samples; and (**3**) Comparison across multiple replicates of IP and input samples. Below we provide a detailed description of these steps, which is very similar to the approach used in (Schwartz et al., 2013), with a few extensions and modifications.

(1) Peak detection within genes. To search for enriched peaks in the m6A IP samples, we scanned each gene using sliding windows of 100 nucleotides with 50 nucleotides overlap. Each window was assigned a Peak Over Median (POM) score, defined as mean coverage in the window / median coverage across the gene. Windows with POM scores greater than 4 (*i.e.*, greater than 4-fold enrichment) and with a mean coverage >10 reads were retained. Overlapping windows were merged together, and for each disjoint set of windows in transcriptome space we recorded its start, end, and peak position, corresponding to the position with the maximal coverage across the window.

(2) Ensuring that peaks were absent in input. We repeated the procedure in step (1) for the input sample. We eliminated from all subsequent analysis all windows that were detected in both step (1) and (2).

(3) Comparison of multiple samples and criteria for WTAP-dependence. To compare between different perturbations and/or conditions, we applied the following strategy. We first merged the coordinates of all windows from all samples passing step (1) and (2), to define a set of disjoint windows passing these filters in at least one of the samples. For each such window, we recalculated the peak start, end, peak position, and POM score (as defined above) across each of the samples using the approach in step (1). In addition, for each window we calculated a Peak Over Input (POI) score, corresponding to the fold-change of coverage across the window in the IP sample over the coverage in the input sample. To account for differences in sample depth, we

estimated the mean difference between IP and input samples across the 500 most highly expressed genes, which we used as an estimate for background. We subtracted this background from the POI score. Of note, the POM and POI scores generally correlated well with each other; nevertheless, we empirically found that in some cases it was more informative to filter based on the one than on the other.

For each sample, we then calculated the maximal POI across all samples, as well as the number of samples in which the POI scores were ≥3. As a measure of WTAP dependence, we calculated for each site the difference between the mean POI scores across non-perturbed samples (*e.g.*, shGFP, shRFP, siControl) and the WTAP perturbed counterparts (*e.g.*, shWTAP, siWTAP), and required that these reflect a 50% or 100% increase compared to the non-perturbed samples (precise thresholds are indicated in the figure legends of the respective experiments).

Sites were grouped assigned into one of five gene segments: (1) TSS segment, comprising the first 200nt of each gene, (2) 5' UTR segment, comprising all non-TSS segment 5' UTR positions, (3) stop codon segment, comprising all positions within 200nt of the annotated stop codon, (4) CDS segment, comprising all non-stop segment positions in the CDS, and (5) 3'UTR segment, comprising all non-stop segment 3' UTR positions.

HUES9 human pluripotent stem cell and neural progenitor culture

For HUES9 (Harvard Stem Cell Institute iPS Core Facility) human pluripotent stem cells, standard stem cell media consists of mTeSR1 (STEMCELL Technologies) supplemented with 1X Normocin (InvivoGen). Cells were maintained using the Enhanced Culture Protocol described elsewhere (Schinzel et al., 2011). For HUES9 human neural progenitor cells (NPCs), standard media consists of 50% N2 media and 50% B27 media (all from Life Technologies

unless indicated otherwise). N2 media: DMEM-F12 supplemented with 1X N2 supplement, 5μg/ml insulin, 100μM β-mercaptoethanol, 1X Glutamax, 1X non-essential amino acids, 1X Normocin (InvivoGen). B27 media: Neurobasal supplemented with 1X B27 supplement, 1X sodium pyruvate, 1X Glutamax, 5ml non-essential amino acids, 1X Normocin (InvivoGen). All culture dishes were coated with Geltrex (1:100 in DMEM) for at least 1 hour at 37°C.

NPCs were differentiated from HUES9 cells using standard dual SMAD differentiation (Chambers et al., 2009; Shi et al., 2012). Briefly, stem cells were passaged the day before differentiation at sufficient density to achieve 100% confluence within 24 hours. Once at 100% confluence, media was switched to standard neural progenitor media supplemented with 200ng/mL rhNoggin (R&D Systems) and 10μM SB431542 (Stemgent) and changed daily for the first 10 days of neural induction. After 10 days, cells were passaged and replated at 200,000 cells per cm^2 in standard neural progenitor media supplemented with 10μM Y-27632. Twenty-four hours after passaging, media was replaced with standard neural progenitor media alone. Between 12 and 17 days after neural induction, cultures were monitored closely to verify formation of neural rosettes. After this stage, NPCs and neurons were split as necessary (with 10μM Y-27632) until day 25 post-induction, when cells were rinsed once in PBS and then lysed in RLT Plus with β-mercaptoethanol, snap-frozen on dry ice, and stored at -80°C until RNA extraction.

Mice brain samples

Male mice were sacrificed at the age of 3 months, and their hippocampus was micro-dissected. Mice embryos were sacrificed at the age of E16, and their brain tissue was extracted. Brain samples from siblings were pooled together (6-8 animals per litter). The brain tissue was inserted after dissection immediately into RNA-later stabilization reagent (Qiagen), kept in 4°C for 24 hours, and then transferred to -80°C for storage. All experiments were performed in compliance

with the institutional guidelines and were reviewed and approved by the Institutional Animal Use and Care Committee (IAUCC) of the Massachusetts Institute of Technology (Protocol number: 0411-040-14).

Bone marrow derived dendritic cells

Cultures of bone marrow derived dendritic cells (BMDCs from 6-8 week old female C57BL/6J mice) were prepared as previously described (Amit et al., 2009). At 9 days of *in vitro* culture, cells were stimulated with lipopolysaccharide (LPS, Invivogen) as previously described (Amit et al., 2009). Cells were harvested either at time of stimulation, or 3 and 6 hours later. All experiments were performed in compliance with the institutional guidelines and were reviewed and approved by the Institutional Animal Use and Care Committee (IAUCC) of the Massachusetts Institute of Technology (Protocol number: 0612-058-15).

Induction of pluripotency

hTERT immortalized fibroblasts differentiated from iPSCs harboring a Doxycycline-inducible polycistronic OCT4-KLF4-MYC-SOX2 (Cacchiarelli et al., manuscript in preparation) were collected in growth conditions in absence of Doxycycline, upon 5 days of Doxycycline treatment, or in absence of Doxycycline upon derivation of the iPSC line. Reprogramming conditions are as described in (Chan et al., 2009).

Consensus sequence

We defined the m6A consensus sequence as a motif matching any of the following sequences: GGACA, GGACU, GGACC, GAACU, AGACU, AGACA and UGACU. This set of sequences was obtained by taking a 60nt region surround the 10,000 top scoring peaks in human and mouse ('test' sequences), and mapping all NNACN pentamers within this region. As random controls,

we scrambled this set of sequences. For each pentamer, we quantified the fold-change in abundance between the test and control sequences. In parallel, as methylated motifs are biased to be in the very center of the 60nt region, we scored each pentamer based on its centrality. For each pentamer, we then generated 12-bin histograms of the distributions of distances from the center of the sequence, and divided the density at the center of the histogram with the mean densities in the first and last bin ('centrality score'). Plotting each pentamer as a function of these two metrics yielded the above defined set enriched in both metrics (and which was identical in human and mouse) (**Fig. S1E-F**).

Overlap of mTSS and CAGE data

For examining the overlap between our set of mTSSs and CAGE data, we used data from a CAGE experiment performed on long, cellular, poly(A) mRNA in the ENCODE project (Consortium et al., 2012). Processed data summarizing the number of reads beginning at each genomic position was downloaded from http://ccg.vital-it.ch/mga/hg19/encode/GSE34448/GSE34448.html (A549_cell_longPolyA_rep1.sga). This data was merged with the set of annotated mTSSs based on common, genomic coordinates.

mRNA stability

Half-lives for mouse mRNAs in NIH3T3 fibroblasts were obtained from (Schwanhausser et al., 2011). Half-lives for human mRNAs were obtained based on (Duan et al., 2013), which measured half-lives across 7 human HapMap lymphoblastoid cell lines. We used the median value across these cell lines as an estimate of half-life.

Statistical analysis

All statistical analyses and visualizations were performed in R: Sequence logos were prepared using the SeqLogo package (Bembom, 2011), heatmaps were generated using the gplots package (Warnes, 2012); many other plots were generated using the ggplot2 package (Wickham, 2009).

References

Amit, I., Garber, M., Chevrier, N., Leite, A.P., Donner, Y., Eisenhaure, T., Guttman, M., Grenier, J.K., Li, W., Zuk, O.*, et al.* (2009). Unbiased reconstruction of a mammalian transcriptional network mediating pathogen responses. Science (New York, NY *326*, 257-263.

Bembom, O. (2011). seqLogo: Sequence logos for DNA sequence alignments.

Chambers, S.M., Fasano, C.A., Papapetrou, E.P., Tomishima, M., Sadelain, M., and Studer, L. (2009). Highly efficient neural conversion of human ES and iPS cells by dual inhibition of SMAD signaling. Nature biotechnology *27*, 275-280.

Chan, E.M., Ratanasirintrawoot, S., Park, I.H., Manos, P.D., Loh, Y.H., Huo, H., Miller, J.D., Hartung, O., Rho, J., Ince, T.A.*, et al.* (2009). Live cell imaging distinguishes bona fide human iPS cells from partially reprogrammed cells. Nature biotechnology *27*, 1033-1037.

Consortium, E.P., Bernstein, B.E., Birney, E., Dunham, I., Green, E.D., Gunter, C., and Snyder, M. (2012). An integrated encyclopedia of DNA elements in the human genome. Nature *489*, 57-74.

Duan, J., Shi, J., Ge, X., Dolken, L., Moy, W., He, D., Shi, S., Sanders, A.R., Ross, J., and Gejman, P.V. (2013). Genome-wide survey of interindividual differences of RNA stability in human lymphoblastoid cell lines. Scientific reports *3*, 1318.

Ingolia, N.T., Lareau, L.F., and Weissman, J.S. (2011). Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. Cell *147*, 789-802.

Mertins, P., Qiao, J.W., Patel, J., Udeshi, N.D., Clauser, K.R., Mani, D.R., Burgess, M.W., Gillette, M.A., Jaffe, J.D., and Carr, S.A. (2013). Integrated proteomic analysis of post-translational modifications by serial enrichment. Nat Methods *10*, 634-637.

Mertins, P., Udeshi, N.D., Clauser, K.R., Mani, D.R., Patel, J., Ong, S.E., Jaffe, J.D., and Carr, S.A. (2012). iTRAQ labeling is superior to mTRAQ for quantitative global proteomics and phosphoproteomics. Molecular & cellular proteomics : MCP *11*, M111 014423.

Meyer, K., Saletore, Y., Zumbo, P., Elemento, O., Mason, C., and Jaffrey, S. (2012). Comprehensive Analysis of mRNA Methylation Reveals Enrichment in 3' UTRs and near Stop Codons. Cell *149*, 1635-1646.

Rappsilber, J., Mann, M., and Ishihama, Y. (2007). Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. Nature protocols *2*, 1896-1906.

Schinzel, R.T., Ahfeldt, T., Lau, F.H., Lee, Y.K., Cowley, A., Shen, T., Peters, D., Lum, D.H., and Cowan, C.A. (2011). Efficient culturing and genetic manipulation of human pluripotent stem cells. PloS one *6*, e27495.

Schwanhausser, B., Busse, D., Li, N., Dittmar, G., Schuchhardt, J., Wolf, J., Chen, W., and Selbach, M. (2011). Global quantification of mammalian gene expression control. Nature *473*, 337-342.

Schwartz, S., Agarwala, S.D., Mumbach, M.R., Jovanovic, M., Mertins, P., Shishkin, A., Tabach, Y., Mikkelsen, T.S., Satija, R., Ruvkun, G.*, et al.* (2013). High-Resolution Mapping Reveals a Conserved, Widespread, Dynamic mRNA Methylation Program in Yeast Meiosis. Cell.

Shi, Y., Kirwan, P., and Livesey, F.J. (2012). Directed differentiation of human pluripotent stem cells to cerebral cortex neurons and neural networks. Nature protocols *7*, 1836-1846.

Subtelny, A.O., Eichhorn, S.W., Chen, G.R., Sive, H., and Bartel, D.P. (2014). Poly(A)-tail profiling reveals an embryonic switch in translational control. Nature *508*, 66-71.

Warnes, G.R. (2012). gplots: Various R programming tools for plotting data.

Wickham, H. (2009). ggplot2: elegant graphics for data analysis. Springer New York.