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.