SUPPLEMENTARY INFORMATION

YTHDF2 recognition of N^1-methyladenosine (m^1A)-modified RNA is associated with transcript destabilization

Kyung W. Seo and Ralph E. Kleiner

Department of Chemistry, Princeton University, Princeton, NJ 08540, USA

Table of Contents

Table S1. RNA probes used in this work.

Probe	Sequence	Full mass	Charge (z)	Calculated mass	Observed mass
1	CGAGG-m^{1}A-GG-5DzU-GUAC-biotin	4966.94992	4	1240.47838	1240.48073
2	CGAGGAGG-5DzU-GUAC-biotin	4951.92699	4	1236.97447	1236.97767
3	CGAGG-m^{1}A-GGUGUAC-fluorescein	4799.77193	4	1198.68389	1198.68947
4	CGAGGAGGUGUAC-fluorescein	4784.74901	4	1195.17998	1195.17973
5	CUUUU-m^{1}A-AAGUAC-fluorescein	4321.66319	3	1439.21136	1439.21518
6	CUUUUAAAGUAC-fluorescein	4306.64026	3	1434.53948	1434.53705
7	CGAGG-m^{6}A-GGUGUAC-fluorescein	4798.76466	4	1198.68389	1198.68603
8	CUUUU-m^{6}A-AAGUAC-fluorescein	4320.65591	4	1079.15670	1079.15363

Table S2. RT-PCR primer pairs

Target transcript	Forward primer (5'-3')	Reverse primer (5'-3')
GAPDH	CAATGACCCCTTCATTGACC	TTGATTTTGGAGGGATCTCG
CHPF2	GGCAGCAGTATCGCTCATTT	TCTGAGCCTGCAGTTGTTCT
ARHGAP35	AGGCCCACTTAGAGAAGCTG	CCTCCAGCCACTGGTAGAAA
DCAF8	TTGAGGGCCATACTGGTTGT	GAAACTTGGCCTGGAACACA

Table S3. ΔCt values for qPCR

CHPF2		ARHGAP35		DCAF8	
siControl	siYTHDF2	siControl	siYTHDF2	siControl	siYTHDF2
6.37	4.62	6.74	6.31	6.03	4.91
6.99	5.80	6.48	6.00	6.38	5.91
7.27	6.50			6.17	5.78

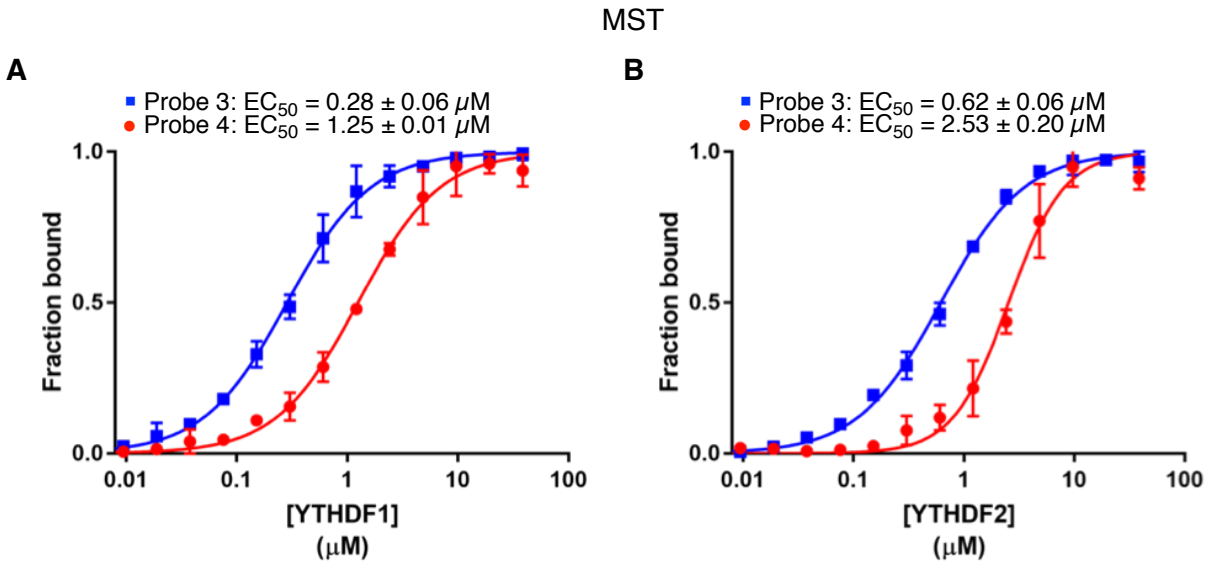

Figure S1. Characterization of YTHDF1/2-m^{1}A interactions via MST. a) Average binding curves of YTHDF1-probe **3**/**4** interactions. b) Average binding curves of YTHDF2-probe **3**/**4** interactions. [probe] = 40 nM for all experiments. 3 replicates were used to characterize all interactions.

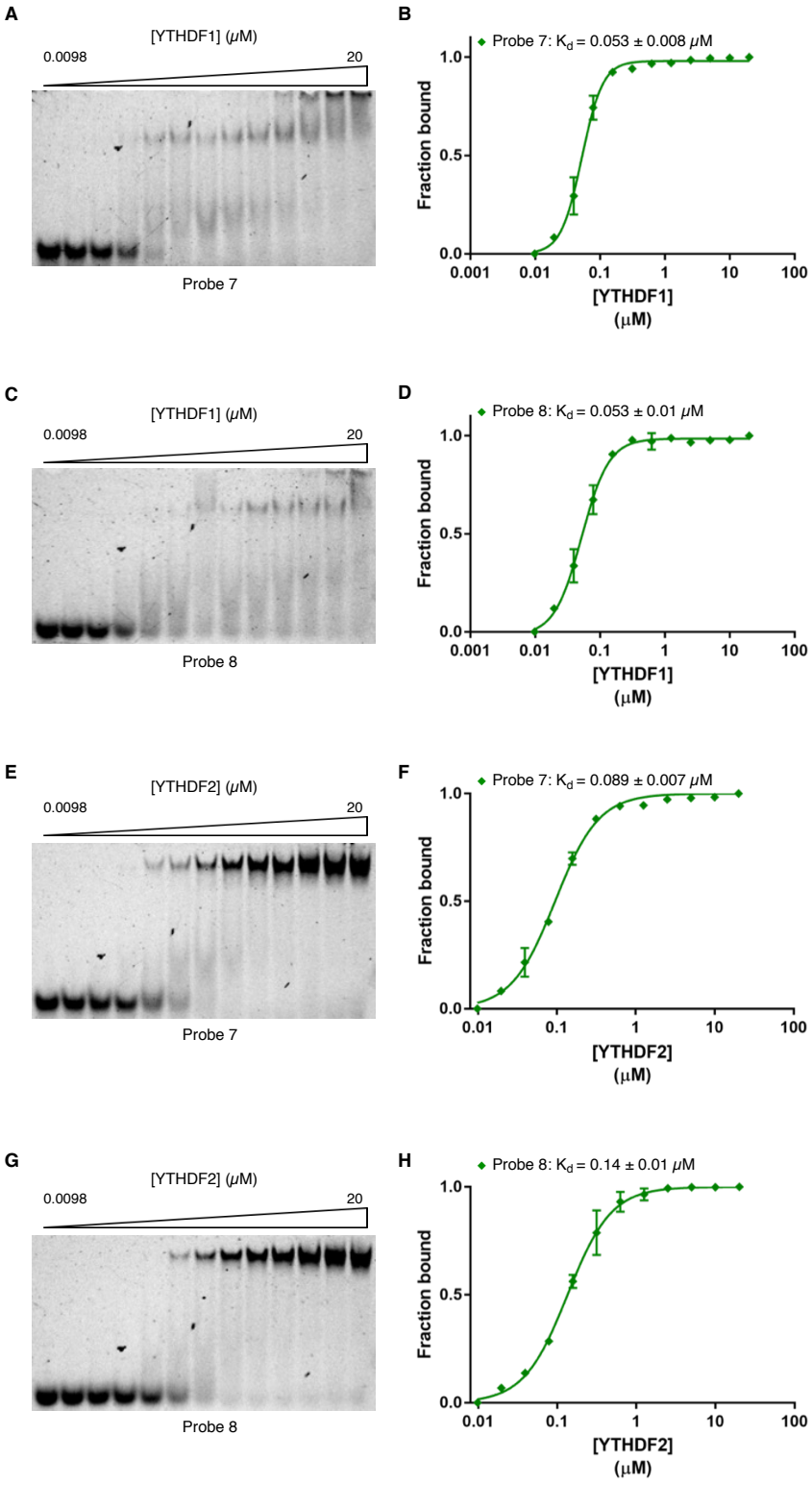

Figure S2. Characterization of YTHDF1/2-m^{6}A interactions. a) Representative EMSA images for YTHDF1 and probe **7**. b) Quantification of YTHDF1-probe **7** interactions. Values represent

Figure S2. (con't) the average and standard deviation of 2 independent measurements. c) Representative EMSA images for YTHDF1 and probe **8**. d) Quantification of YTHDF1-probe **8** interactions. Values represent the average and standard deviation of 2 independent measurements. e) Representative EMSA images for YTHDF2 and probe **7**. f) Quantification of YTHDF2-probe **7** interactions. Values represent the average and standard deviation of 2 independent measurements. g) Representative EMSA images for YTHDF2 and probe **8**. h) Quantification of YTHDF2-probe **8** interactions. Values represent the average and standard deviation of 2 independent measurements.

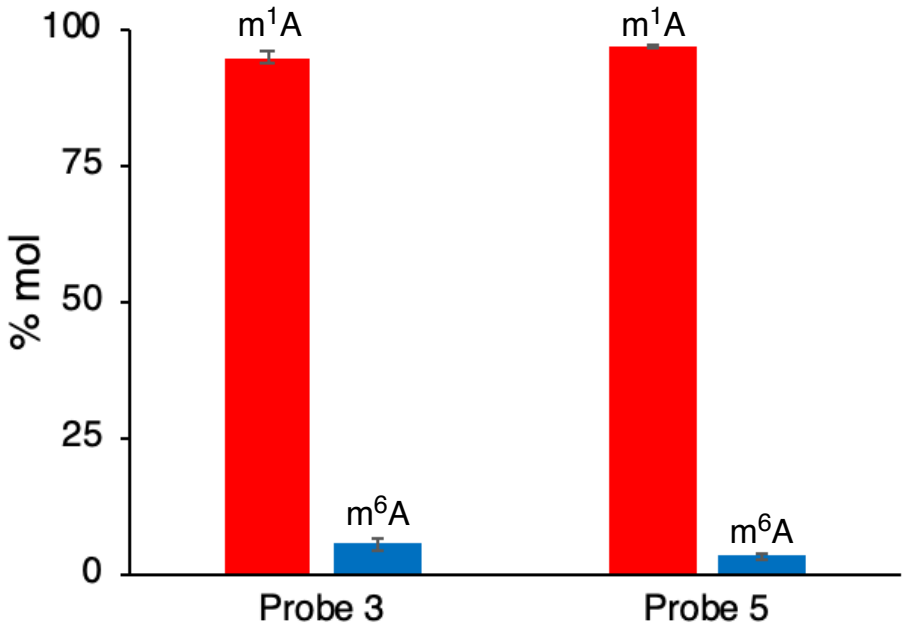

Figure S3. Measurements of Dimroth rearrangement of m^1A probes used in this study with LC-MS/MS.

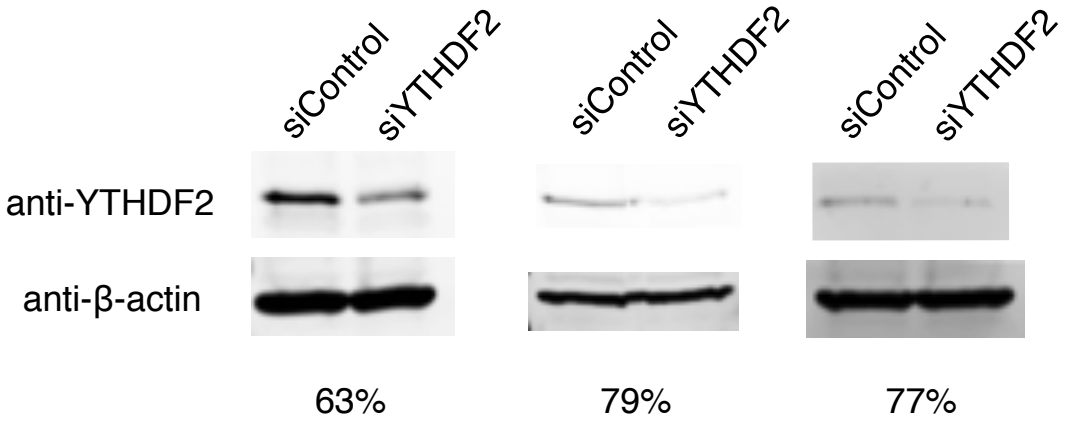

Figure S4. Western blot showing siRNA knockdown of YTHDF2

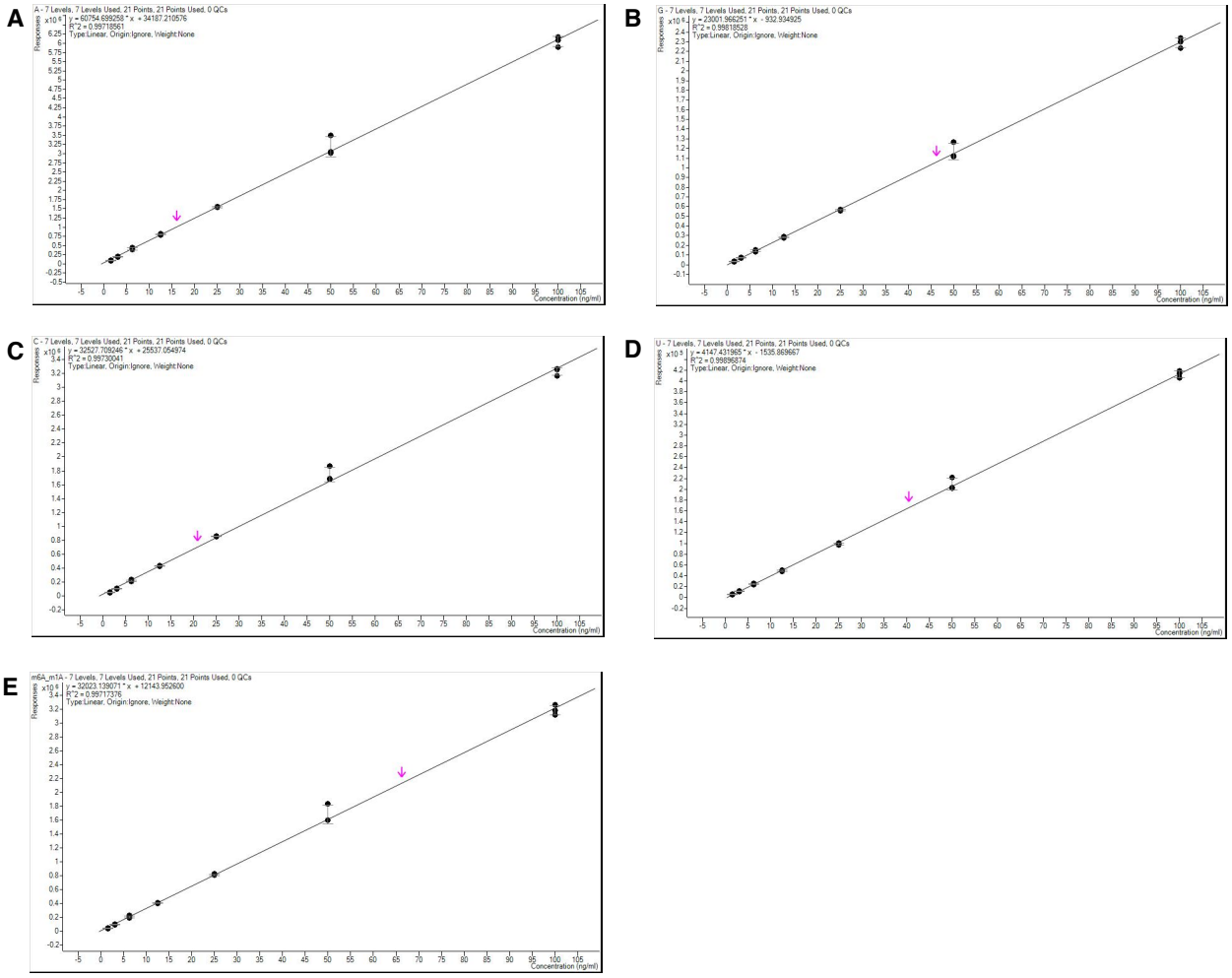

Figure S5. Calibration curves for RIP + LC-MS/MS analysis. a) The calibration curve for adenosine. b) The calibration curve for guanosine. c) The calibration curve for cytidine. d) The calibration curve for uridine. e) The calibration curve for N^1-methyladenosine.

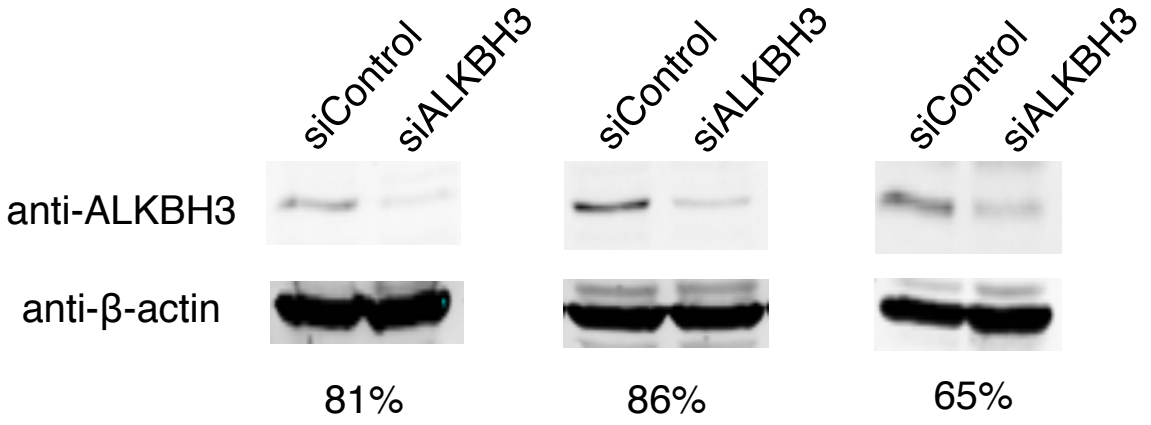

Figure S6. Western blot showing siRNA knockdown of ALKBH3

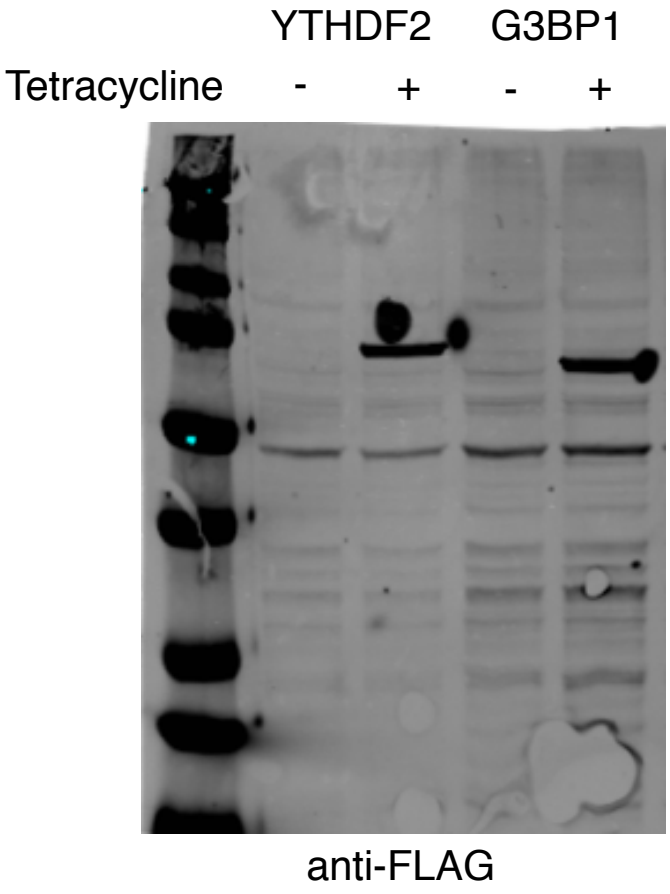

Figure S7. Western blot analysis of 293 Flp-In cell lines containing tetracycline inducible 3xFLAG-YTHDF2 and 3xFLAG-G3BP1 constructs.

Transcript	mRNA lifetime log2(siYTHDF2/siControl)
ACTB	0.884
ARHGDIA	0.523
BRD2*	0.221
HNRNPU	-0.105
ILF3	0.876
MAVS	0.484
PRPF8	1.252
PRUNE*	0.405

Figure S8. Table showing YTHDF2 target transcripts (Wang *et al*.[1]) overlapping with annotated m^1A-containing transcripts (Li *et al*.[2]; Zhou *et al*.[3]). * indicates those annotated in Safra *et al.*[4]

	Global	Li et al	Zhou et al	Safra et al
Log2(siALKBH3/siControl)	-0.001854	-0.02828	-0.05138	0.05903
Number of transcripts	20745	345	189	8

Figure S9. Table showing average ratios of transcript abundance in siALKBH3 RNA-Seq data, overlapping with annotated m^1A-containing transcripts (Li *et al*.[2]; Zhou *et al*.[3]; Safra *et al.*[4]).

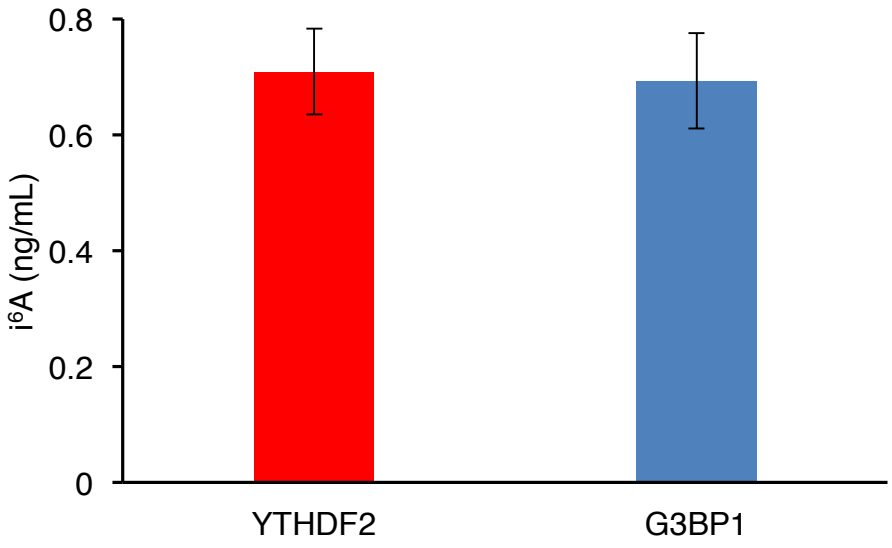

Figure S10. Quantification of N^6-isopentenyladenosine (i^6A) concentration detected in RIP with YTHDF2 and G3BP1.

METHODS

Plasmid construction

Plasmids encoding cDNA were obtained from Addgene: YTHDF2 (# 52300)[1], YTHDF1 (#70087) [1], and YTHDC1 (#85167) [5] . Sequences were cloned into pGEX-6P-1 or a modified pET28a vector for protein expression in *Escherichia coli*. For construction of Flp-In cell lines, YTHDF2 and G3BP1 genes were cloned into a modified pCDNA5/FRT-TO (Life Technologies) vector containing an N-terminal 3x-FLAG affinity tag.

Protein expression

Proteins were expressed in Escherichia coli strain BL21(Rosetta). YTHDF2-YTH was expressed and purified according to literature precedent [6]. YTHDF1-YTH was expressed and purified with the same protocol as YTHDF2-YTH. YTHDC1-YTH was expressed and purified according to literature precedent [7].

Probe synthesis

RNA oligonucleotide probes were chemically synthesized using the UltraMild synthesis protocol (Glen Research). For diazirine-containing RNA probes, oligonucleotides were synthesized with 5-aminoallyl-uridine phosphoramidite (Barry and Associates) and then acylated with NHS-diazirine according to literature precedent[8]. In brief, deprotected and desalted RNA oligos were resuspended in 1X siRNA buffer (Dharmacon) to 200 μM and combined with 0.1 M Na-HEPES pH 7.5 (from a 1M stock) and 10 mM diazirine NHS ester (from a 10X DMSO stock); the final RNA concentration in the reaction was 160 μM. The reaction was vortexed and incubated at ambient temperature for 2-4 hours in the dark. The mixture was purified by reverse-phase HPLC with an Infinity Poroshell 120 EC-C18 analytical column or a Zorbax Eclipse XDB-C18 semiprep column on an Agilent 1260 Infinity instrument. A gradient of 5 to 40% acetonitrile in 0.1

M triethylammonium acetate over 50 minutes was used. Fractions of interest were lyophilized, resuspended in a small amount of water, and characterized by high-resolution mass spectrometry (HRMS) on an Agilent 6220 Accurate-Mass Time-of-Flight LC/MS (ESI-TOF) in negative mode.

Photo-cross-linking reactions with cellular lysate

Samples for proteomic analysis were prepared according to literature precedent [8]. HeLa cells were harvested and lysed by cryomilling. The resulting cell powder (1.5 g) was first extracted with 2.5 mL of low-salt extraction buffer (20 mM Tris HCl pH 7.5, 10 mM NaCl, 2 mM $MgCl_2$, 0.5% Triton X-100, 10% glycerol, protease and phosphatase inhibitor tablet), and then 1.67 mL of high-salt extraction buffer (50 mM Tris HCl pH 7.5, 420 mM NaCl, 2 mM $MgCl_2$, 0.5% Triton X-100, 10% glycerol, protease and phosphatase inhibitor tablet). High-salt and low-salt extracts were pooled and diluted to 3 mg/mL if needed before proceeding to photo-cross-linking.

Photo-cross-linking reactions were performed using 2 mL of extract and 1 μM RNA probe. After addition of the probe, reactions were incubated for 20 minutes on ice, transferred to a pre-chilled 6 cm dish, and cross-linked for 15 min on ice with 365 nm UV (Spectroline ML-3500S). The reaction was incubated with 60 μL of streptavidin agarose 50% bead slurry (Pierce #20357) for 3 hours at 4^{o}C with end-to-end rotation, and the beads were washed with 1% SDS in 1X TBS (3 x 1 mL), 6 M urea in 1X TBS (3 x 1 mL), and 1X TBS (3 x 1 mL). RNA-bound proteins were eluted with 50 μL of RNase elution buffer (10 mM Tris HCl pH 7.5, 40 mM NaCl, 1 mM $MgCl_2$, 25 unit/mL RNase A, and 2000 unit/mL RNase T1) at 37^{o}C for 30 minutes with periodic mixing.

Mass Spectrometry Proteomics Analysis

Eluted proteins were reduced and alkylated, digested with Trypsin Gold (Promega) and desalted using SDB stage-tips [9]. Samples were loaded using an EasynLC 1200 UPLC system directly onto a 45 cm long 75 μm inner diameter nanocapillary column packed with 1.9 μm C18-

AQ (Dr. Maisch, Germany) mated to metal emitter inline with an Orbitrap Fusion Lumos (Thermo Scientific, USA). The mass spectrometer was operated in data dependent mode with the 120,000 resolution MS^1 scan (AGC 4e5, Max IT 50ms, 400-1500 m/z) in the Orbitrap followed by up to 20 MS/MS scans with CID fragmentation in the ion trap. Dynamic exclusion list was invoked to exclude previously sequenced peptides for 60 s if sequenced within the last 30 s and maximum cycle time of 3 s was used. Peptides were isolated for fragmentation using the quadrupole (1.6 Da window). Ion-trap was operated in Rapid mode with AGC 2e3, maximum IT of 300 ms and minimum of 5000 ions.

Raw files were searched using Byonic [10] and Sequest HT algorithms [11] within the Proteome Discoverer 2.1 suite (Thermo Scientific, USA). 15 ppm MS^1 and 0.5Da MS^2 mass tolerances were specified. Carbamidomethylation of cysteine was used as a fixed modification. Oxidation of methionine, and deamidation of asparagine were specified as dynamic modifications. Trypsin digestion with maximum of 2 missed cleavages were allowed. Files searched against the uniprot Homo sapiens database downloaded Dec 20th, 2017 and supplemented with common contaminants.

Scaffold (version Scaffold_4.7.5, Proteome Software Inc., Portland, OR) was used to validate MS/MS based peptide and protein identifications. Peptide identifications were accepted if they could be established at greater than 95.0% probability by the Scaffold Local FDR algorithm. Protein identifications were accepted if they could be established at greater than 99.9% probability and contained at least 2 identified peptides. Protein probabilities were assigned by the Protein Prophet algorithm [12]. Proteins that contained similar peptides and could not be differentiated based on MS/MS analysis alone were grouped to satisfy the principles of parsimony.

Quantitation of protein abundance was performed by spectral counting. All identified peptides were used for counting and proteins were only selected for quantification if they were present in at least 2 replicates. Three independent biological replicates were performed.

To calculate enrichment ratios for proteins identified by only one probe, the minimum spectral count value was set to 0.1.

EMSA (electrophoretic mobility shift assay/gel shift assay)

40 nM RNA probe was combined with various concentrations of protein and incubated at 4°C in binding buffer (10mM HEPES, pH 8.0, 50mM KCl, 1mM EDTA, 0.05% Triton-X-100, 5% glycerol, and 1mM DTT) for 1 hour before being loaded on 5% native TBE gels. 7 µL RNA-protein mixture was loaded to the gel and run at 4°C for 10 min at 130V and 30 min at 90V. Relative fluorescence was measured with ImageQuant LAS 4000 (GE Healthcare Life Sciences). ImageJ was used to quantify the fluorescent signals, and K_D values were calculated with a quadratic function for single-site binding:

$$Y = fraction\ bound$$

$$X = [protein]$$

$$Y = \left\{X + [probe] + K_D - \sqrt{(X + [probe] + K_D)^2 - 4(X \times [probe]}\right\} / (2[probe])$$

K_D estimates for binding curves that did not saturate were obtained with a 4-parameter dose-response equation using GraphPad Prism:

Y=Bottom + (Top-Bottom)/(1 + ((X^HillSlope)/(IC_{50}^HillSlope)))

All biochemical measurements represent the mean and standard deviation from 3 independent replicates.

MST (microscale thermophoresis)

40 nM RNA probe was combined with various concentrations of protein at room temperature in MST buffer (20 mM Tris, pH 7.5, 100 mM NaCl, 5 mM $MgCl_2$, and 0.05% Tween-20) for 2 hours. The mixture was loaded into the capillaries and MST measurements were

taken. Data were recorded in the MO.Control software (Nanotemper) and analyzed in the MO.Affinity Analysis software (Nanotemper) using TJump analysis. The normalized MST values were converted to fraction bound, and K_D values were calculated as described above.

Mammalian cell culture

Mammalian cells were cultured at 37°C in a humidified atmosphere with 5% CO_2 in DMEM (Life Technologies) supplemented with 10% FBS (Atlanta), 1× penicillin-streptomycin and 2 mM L-glutamine (Life Technologies)

siRNA knockdown

Transfection of siRNA (Dharmacon) was performed using Lipofectamine RNAiMAX (Invitrogen) according to the manufacturer's instructions.

siYTHDF2: AAGGACGUUCCCAAUAGCCAAUU

siALKBH3: GAGAGAAGCUUCACUGAAAUU

siControl: UGGUUUACAUGUCGACUAA

25,000 cells were seeded in a 24-well plate 24 hours before transfection. 5 pmol of siRNA per well was used for transfection. Cells were harvested 72 hours after transfection for total RNA isolation and RT-PCR as described below.

RT-PCR

Total RNA was isolated using Quick-RNA™ MiniPrep Plus (Zymo Research #R1057). cDNA was generated using SuperScript III Reverse Transcriptase (Invitrogen) as instructed by the manufacturer. PCR was performed with PowerUp SYBR Green Master Mix (ThermoFisher), and the fluorescence was measured with CFX96™ Real-Time System C1000™ Thermal Cycler (Bio-Rad). All "relative mRNA levels" were measured against GAPDH abundance and the

following equations were used for our calculation:

$\Delta Ct = Ct_{Gene\ of\ Interest} - Ct_{GAPDH}$ (within the same sample)

$\Delta\Delta Ct = \Delta Ct_{knockdown} - \Delta Ct_{control}$

Fold-increase ('relative mRNA level') = $2^{-\Delta\Delta Ct}$

RNA-Seq

Total RNA from ALKBH3-knockdown experiments of HeLa cells was isolated using Quick-RNA™ MiniPrep Plus (Zymo Research #R1057) and used for poly-A enrichment with oligo-dT beads. Pulldown was followed by standard fragmentation, adaptor ligation and PCR amplification for sequencing on the Illumina HiSeq 2000 platform. Sequencing reads were mapped to the human genome hg19 using TopHat 2, set to map stranded reads with default parameters. Reads mapping to exons of each gene were counted using HTseq-count in union mode. RNA-seq data was normalized by total library size. RNA-seq data were visualized using the Integrative Genomics Viewer.

Sequences were uploaded to and demultiplexed on the Princeton HTSeq database system and transferred to the Princeton Galaxy instance, where read quality was assessed using FastQC (Galaxy Version 0.72), Read Distribution (Galaxy Version 2.6.4.1), BAM/SAM Mapping Stats (Galaxy Version 2.6.4), IdxStats (Galaxy Version 2.0.1), and Gene Body Coverage (Galaxy Version 2.6.4.3). Quality control stats were viewed using a MultiQC (Galaxy Version 1.6) report. Sequences were then aligned using RNA STAR (Galaxy Version 2.6.0b-1) with the human reference GRCh38 from the NCBI. Reads aligning to genes according to NCBI's GRCh38 gene annotations were counted using featureCounts (Galaxy Version 1.6.2). Samples were then compared using DESeq2 (Galaxy Version 2.11.40.6), which generated QC plots, rLog normalized counts, and differential expression data.

Generation of stable cell line

Flp-In TRex 293 cells were grown in media containing 15 µg/mL blasticidin and 100 µg/mL zeocin. To generate stable cell lines expressing YTHDF2 or G3BP1, the Flp-In 293 cells were seeded at 0.4 × 10^6 cells per well in six-well plates, and co-transfected with pOG44 (2 µg), the plasmid expressing flp recombinase, and pCDNA5/FRT/TO plasmid containing YTHDF2 or G3BP1 gene (0.2 µg). Following selection in 350 µg/mL hygromycin B and 15 µg/mL blasticidin, colonies were expanded. To confirm expression of the YTHDF2 and G3BP1 genes, cells were grown in the presence or absence of 1 µg/mL tetracycline for 24 hours. Cells were harvested and lysed in NP-40 lysis buffer. Protein concentration in the lysate was quantified by BCA assay and lysate containing same quantities of protein was loaded onto gel for SDS-PAGE. For western blot analysis, the membrane was blocked in 5% BSA for 1 hour after transfer. The membrane was incubated in anti-FLAG M2 antibody (Sigma, 1 µg/mL) for 2 hours, washed with TBST three times for 5 minutes, and stained with IRDye-conjugated donkey anti-mouse 680LT antibody (LI-COR Biosciences, 0.05 µg/mL).

RIP

2.2 x 10^7 Flp-In TRex 293 cells containing FLAG-YTHDF2 or FLAG-G3BP1 were plated (10 x 10-cm plates). The cells were induced with 1 µg/mL tetracycline 24 hours after seeding. The induced cells were harvested after 24 hours, washed once with cold PBS, and pelleted by centrifuge for 5 min at 1,000g. The cell pellet was resuspended with 2 volumes of lysis buffer (150 mM KCl, 10 mM HEPES pH 7.6, 2 mM EDTA, 0.5% NP-40, 0.5 mM DTT, 1:100 protease inhibitor cocktail, 0.4 U/µL RNasin) and was constantly agitated for 30 minutes at 4^{o}C. The lysate was centrifuged at 15,000g for 20 min at 4^{o}C to clear the lysate. The lysate was pre-cleared by incubating with 100 µL of Dynabeads^{TM} Protein G (ThermoFisher 10004D) suspension for 1 hour at 4^{o}C with continuous rotation. Anti-FLAG beads were generated by incubating 200 µL of

Dynabeads™ Protein G with 1 mL of lysis buffer and 20 µL of anti-FLAG M2 antibody (Sigma F1804, 1.0 mg/mL) for 1 hour at room temperature. The pre-cleared lysate was incubated with anti-FLAG beads with constant rotation for 4 hours at 4°C. The beads were washed eight times with 1 mL of ice-cold wash buffer (200 mM KCl, 50 mM HEPES pH 7.6, 2 mM EDTA, 0.05% NP-40, 0.5 mM DTT, 0.4 U/µL RNasin). Washed beads were resuspended in 200 µL of PK buffer (100 mM Tris-HCl pH 7.4, 50 mM NaCl, 10 mM EDTA, 0.2% SDS) and incubated with 10 µL of proteinase K (NEB P8107S, 800 U/mL) shaking at 1,000 rpm for 40 min at 50°C. The supernatant was removed from the beads and treated with TRIzol LS (ThermoFisher) to isolate the RNA. RNA was then subjected to nuclease digestion, dephosphorylation, and LC-MS/MS analysis as described below. Reported values represent the mean and standard deviation from 4 independent biological experiments.

LC-MS/MS

The RNA was digested with nuclease P1 (2U) in 30 µL of buffer containing 7 mM NaOAc pH 5.2, 0.4 mM $ZnCl_2$ at 37°C for 2 hours, followed by the addition of 3.5 µL of 10X Antarctic Phosphatase Reaction buffer and 1.5 µL of Antarctic Phosphatase (NEB M0289). After an additional incubation at 37°C for 2 hours, 1 µL of the solution was analyzed by LC-MS/MS. Nucleosides were separated by reverse phase ultra-performance liquid chromatography on a 3 µm Hypersil GOLD™ 150 x 2.1 mm column (ThermoFisher #25003-152130) with on-line mass spectrometry detection using an Agilent 6470 QQQ triple-quadrupole mass spectrometer in positive electrospray ionization mode. The nucleosides were quantified by using the nucleoside to base ion mass transitions of 282 to 150 (m^1A), and 268 to 136 (A). Quantification was performed in comparison with the standard curve obtained from pure nucleoside standards running on the same batch of samples. The ratio of m^1A to A was calculated based on the calibrated concentrations.

Antibodies

Anti-YTHDF2 (Proteintech, rabbit, 24744-1-AP, 1:2000)

Anti-ALKBH3 (Millipore, rabbit, 09-882, 1:5000)

Anti-FLAG (Sigma, mouse, F1804, 1:1000)

Anti-beta-actin (Cell Signaling Technology, mouse, 8H10D10, 1:2000)

REFERENCES

1. Wang, X.; Lu, Z.; Gomez, A.; Hon, G. C.; Yue, Y.; Han, D.; Fu, Y.; Parisien, M.; Dai, Q.; Jia, G.; Ren, B.; Pan, T.; He, C., N6-methyladenosine-dependent regulation of messenger RNA stability. *Nature* **2014,** *505* (7481), 117-20.
2. Li, X.; Xiong, X.; Zhang, M.; Wang, K.; Chen, Y.; Zhou, J.; Mao, Y.; Lv, J.; Yi, D.; Chen, X. W.; Wang, C.; Qian, S. B.; Yi, C., Base-Resolution Mapping Reveals Distinct m(1)A Methylome in Nuclear- and Mitochondrial-Encoded Transcripts. *Molecular cell* **2017,** *68* (5), 993-1005 e9.
3. Zhou, H.; Rauch, S.; Dai, Q.; Cui, X.; Zhang, Z.; Nachtergaele, S.; Sepich, C.; He, C.; Dickinson, B. C., Evolution of a reverse transcriptase to map N(1)-methyladenosine in human messenger RNA. *Nat Methods* **2019**.
4. Safra, M.; Sas-Chen, A.; Nir, R.; Winkler, R.; Nachshon, A.; Bar-Yaacov, D.; Erlacher, M.; Rossmanith, W.; Stern-Ginossar, N.; Schwartz, S., The m1A landscape on cytosolic and mitochondrial mRNA at single-base resolution. *Nature* **2017,** *551* (7679), 251-255.
5. Patil, D. P.; Chen, C. K.; Pickering, B. F.; Chow, A.; Jackson, C.; Guttman, M.; Jaffrey, S. R., m(6)A RNA methylation promotes XIST-mediated transcriptional repression. *Nature* **2016,** *537* (7620), 369-373.
6. Zhu, T.; Roundtree, I. A.; Wang, P.; Wang, X.; Wang, L.; Sun, C.; Tian, Y.; Li, J.; He, C.; Xu, Y., Crystal structure of the YTH domain of YTHDF2 reveals mechanism for recognition of N6-methyladenosine. *Cell Res* **2014,** *24* (12), 1493-6.
7. Xu, C.; Wang, X.; Liu, K.; Roundtree, I. A.; Tempel, W.; Li, Y.; Lu, Z.; He, C.; Min, J., Structural basis for selective binding of m6A RNA by the YTHDC1 YTH domain. *Nat Chem Biol* **2014,** *10* (11), 927-9.
8. Arguello, A. E.; DeLiberto, A. N.; Kleiner, R. E., RNA Chemical Proteomics Reveals the N(6)-Methyladenosine (m(6)A)-Regulated Protein-RNA Interactome. *J Am Chem Soc* **2017,** *139* (48), 17249-17252.
9. Rappsilber, J.; Mann, M.; Ishihama, Y., Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nat Protoc* **2007,** *2* (8), 1896-906.
10. Bern, M.; Kil, Y. J.; Becker, C., Byonic: advanced peptide and protein identification software. *Curr Protoc Bioinformatics* **2012,** *Chapter 13*, Unit13 20.
11. Eng, J. K.; McCormack, A. L.; Yates, J. R., An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J Am Soc Mass Spectrom* **1994,** *5* (11), 976-89.
12. Nesvizhskii, A. I.; Keller, A.; Kolker, E.; Aebersold, R., A statistical model for identifying proteins by tandem mass spectrometry. *Anal Chem* **2003,** *75* (17), 4646-58.