Supporting Information for

"Targeted Quantitative Profiling of Epitranscriptomic Reader, Writer and Eraser Proteins Using Stable Isotope-Labeled Peptides"

Tianyu F. Qi[1], Xiaochuan Liu[2], Feng Tang[2], Jiekai Yin[1], Kailin Yu[2], and Yinsheng Wang[1,2,*]

[1]*Environmental Toxicology Graduate Program, University of California Riverside, Riverside, CA 92521-0403.*

[2]*Department of Chemistry, University of California Riverside, Riverside, CA 92521-0403.*

[*]Correspondence author: Yinsheng Wang. Telephone: (951)827-2700; E-mail: yinsheng.wang@ucr.edu

Table of Contents

Materials and Methods

Cell culture

HEK293T cells were cultured in Dulbecco's modified Eagle's medium (DMEM) complemented with 10% fetal bovine serum (FBS) and 1% penicillin-streptomycin. HEK293T cells with *ALKBH5*, *FTO*, and *METTL3* genes being individually ablated with CRISPR/Cas9[1] were maintained at 37°C in a humidified chamber supplemented with 5% CO_2.

Crude SIL peptides

A total of 48 crude SIL peptides ([$^{13}C_6$,$^{15}N_2$]-Lys and [$^{13}C_6$,$^{15}N_4$]-Arg) representing 45 RWE proteins were synthesized and purified by Vivitide (Gardner, MA). The peptide purity was approximately 75% and isotopic purity was around 99%. The complete list of the SIL peptides is shown in Table S1a. Each SIL peptide was reconstituted in 15% acetonitrile in 0.1% formic acid. All SIL peptides were mixed to give a stock solution for spiking into proteomic samples.

Proteomic sample preparation

ALKBH5$^{-/-}$, *FTO*$^{-/-}$, and *METTL3*$^{-/-}$ and the isogenic parental HEK293T cells were harvested, and proteomic samples were prepared using a filter-aided sample preparation (FASP) method with minor modifications as described elsewhere.[2] After desalting using Pierce C18 Tips (Thermo Fisher), the resulting tryptic peptides were spiked-in with a mixture of SIL peptides at a final concentration at 2 fmol/μL. Peptides (500 ng – 1 μg) and SIL peptide mixture (4 fmol) were subjected to LC-PRM analysis. Two LC-PRM runs were carried out for profiling epitranscriptomic RWE proteins in the PRM library. Three replicate samples were initially prepared; however, one replicate of *METTL3*$^{-/-}$ was removed from the analysis due to a contamination concern observed from the abnormal total-ion chromatogram. Therefore, only two replicates of *METTL3*$^{-/-}$ cell samples were analyzed.

LC-PRM data acquisition

The setting of the isotope modifications of the PRM library provided in ProteomeXchange Consortium with the dataset identifier PXD030387 was adjusted in Skyline[3] to reflect [$^{13}C_6$,$^{15}N_2$]-Lys- and [$^{13}C_6$,$^{15}N_4$]-Arg-labeled SIL peptides, which differed from the isotope modifications in stable isotope labelling by amino acids in cell culture (SILAC). Procedures of LC-PRM data acquisition were the same as those described elsewhere.[4]

LC-PRM data analysis

A PRM spectral library of RWE proteins was established based on previously acquired shotgun proteomic data deposited in the folder named "library" in PeptideAtlas with the identifier number of PASS01177.[5] After raw data were imported to Skyline, PRM traces were manually curated to remove potential interfering fragment ions, which were not overlaid with other fragment ions and had poor mass accuracy (>20 ppm). A dotp value[6] of > 0.7 and 4-6 fragment ions with the same retention time were employed as the criteria for positive peptide identification. Quantification results, including protein name, peptide name, replicate number, isotope, total peak area, retention time, and library dotp, were exported from Skyline to Excel. Additional data processing was conducted in Excel, shown in Table S1d-g. In summary, the ratio of each peptide representing a specific RWE protein was calculated using a two-step normalization procedure: (1) the peak area of an endogenous peptide is normalized to that of its corresponding SIL peptide or a surrogate standard; (2) the ratio from the first-step normalization is further normalized against the ratio of

the sum of peak areas for all light peptides over that for all heavy peptides in each LC-PRM run. The peptide ratio in each sample, averaged from the quantification results of 2-3 biological replicates, was represented by mean ± S.D. The relative ratio of the peptide in knockout cells vs. HEK293T cells was further represented by ratio ± propagation error. The ratio of a specific RWE protein in knockout cells relative to HEK293T cells was represented by the mean ratio from relative peptide ratios ± the new propagation error. It is worth noting that if multiple peptides were detected from one RWE protein, only the relative peptide ratio with the propagation error was used to calculate the mean protein ratio and the new propagation error.

The LC-PRM mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE[7] partner repository with the dataset identifier PXD036032.

Western blots

ALKBH5$^{-/-}$, *FTO*$^{-/-}$ and *METTL3*$^{-/-}$ cells and the isogenic parental HEK293T cells were harvested, and lysed with CelLytic M supplemented with 1-2% protease inhibitor cocktail. After centrifugation at 16,100 g for 25 min, Bradford assay was conducted for total protein quantification. The same amount of total proteins from each sample was incubated Laemmli loading buffer at 95 °C for 10 min, resolved by SDS-PAGE, and proteins were transferred to a nitrocellulose membrane at 90 V for 1 h at 4 °C. After blocking the membrane with 5% non-fat dry milk in PBS-T (PBS with 0.1% Tween 20) for 40 min, the membrane was cut into pieces based on the apparent molecular weight of each protein of interest according to the product information provided on https://www.ptglab.com/. Each membrane was incubated at 4 °C overnight with the following antibodies: NOP2 (Proteintech, 10448-1-AP, 1:2000), PUS3 (Proteintech, 17248-1-AP, 1:1000), NSUN6 (Proteintech, 17240-1-AP, 1:1000), and GAPDH (Santa Cruz, sc-32233, 1:10,000). The membranes were thoroughly washed with PBS-T for five times followed by incubation with donkey anti-rabbit secondary antibody (Sigma, A0545, 1:5,000) for NOP2, PUS3, and PUS1, and anti-mouse secondary antibody (Santa Cruz, m-IgGκ BP-HRP, 1:5,000) for GAPDH. After thorough washing for five times, the membranes were visualized using Amersham ECL™ Western Blot Detecting Reagent. Quantification of Western blot was carried out using Image Studio Lite Ver 5.2.

Bioinformatic analysis of m^{6}A mapping data in HEK293 cells

Four custom tracks from GSE63753 were imported into UCSC genome browser, including GSE63753_hek293.abcam.CIMS.C2T.bedgraph.gz (CIMS C2T Profile); GSE63753_hek293.abcam.CIMS.tag.uniq.bedgraph.gz (CIMS Unique Tag Profile); GSE63753_hek293.sysy.CITS.m6A.12051.bed.gz (CITS m6A); GSE63753_hek293.sysy.CITS.tag.uniq.bedgraph.gz (CITS Unique Tag Profile). From the location of CIMS C2T and CITS m^{6}A, we were able to identify m^{6}A at single-nucleotide resolution. From CIMS and CITS unique tag profiles, we were able to locate the m^{6}A-enriched regions.

References

(1) Yang, Y.-Y.; Yu, K.; Li, L.; Huang, M.; Wang, Y. Proteome-wide interrogation of small GTPases regulated by N^6-methyladenosine modulators. *Anal. Chem.* **2020**, *92*, 10145-10152.
(2) Wisniewski, J. R.; Zougman, A.; Nagaraj, N.; Mann, M. Universal sample preparation method for proteome analysis. *Nat. Methods.* **2009**, *6*, 359-362.

(3) MacLean, B.; Tomazela, D. M.; Shulman, N.; Chambers, M.; Finney, G. L.; Frewen, B.; Kern, R.; Tabb, D. L.; Liebler, D. C.; MacCoss, M. J. Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics.* **2010**, *26*, 966-968.
(4) Qi, T. F.; Miao, W.; Wang, Y. Targeted profiling of epitranscriptomic reader, writer, and eraser proteins accompanied with radioresistance in breast cancer cells. *Anal. Chem.* **2022**, *94*, 1525-1530.
(5) Miao, W.; Li, L.; Wang, Y. A targeted proteomic approach for heat shock proteins reveals DNAJB4 as a suppressor for melanoma metastasis. *Anal. Chem.* **2018**, *90*, 6835-6842.
(6) Kawahara, R.; Bollinger, J. G.; Rivera, C.; Ribeiro, A. C. P.; Brandão, T. B.; Paes Leme, A. F.; MacCoss, M. J. A targeted proteomic strategy for the measurement of oral cancer candidate biomarkers in human saliva. *Proteomics.* **2016**, *16*, 159-173.
(7) Perez-Riverol, Y.; Bai, J.; Bandla, C.; Garcia-Seisdedos, D.; Hewapathirana, S.; Kamatchinathan, S.; Kundu, D. J.; Prakash, A.; Frericks-Zipper, A.; Eisenacher, M.; Walzer, M.; Wang, S.; Brazma, A.; Vizcaino, J. A. The PRIDE database resources in 2022: a hub for mass spectrometry-based proteomics evidences. *Nucleic Acids Res.* **2022**, *50*, D543-D552.

Figure S1. PRM traces showing the evidence of successful knockout of *ALKBH5*, *METTL3*, and *FTO* in HEK293T. Dotp values are 0.3, 0.33, and 0.35 of VSEPVLSLPVR from ALKBH5 in *ALKBH5*$^{-/-}$, NPEAALSPTFR from METTL3 in *METTL3*$^{-/-}$, and ILIGNPGCTYK from FTO in *FTO*$^{-/-}$ respectively, compared with 0.96, 0.99, 0.9 in their isogenic HEK293T cells. The SIL peptide of ILIGNPGCTYK was not detected.

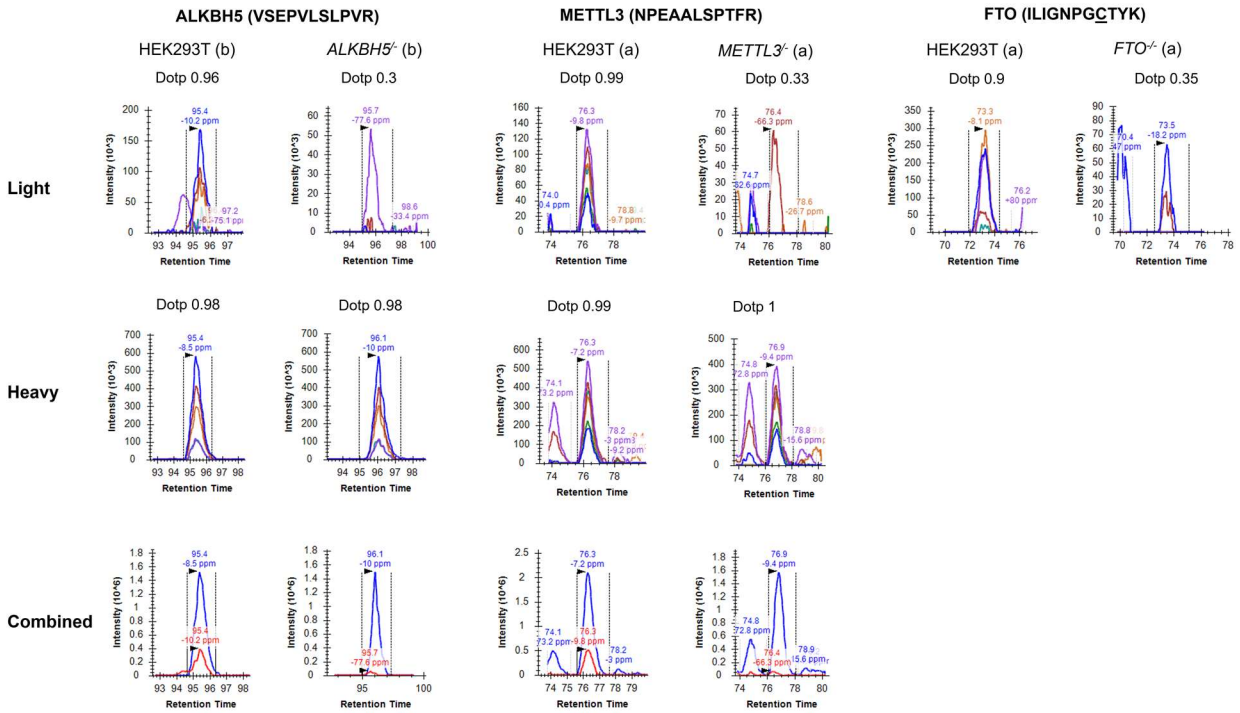

Figure S2. Mapping results for m^{6}A in *NOP2* (a), *PUS3* (b), *TGS1* (c), and *RBMX* (d) mRNAs in GSE63753 dataset. The lower panel is the full view of the gene. The upper panel shows the zoom-in view of the marked signal to check genomic sequences. The potential m^{6}A sequence motif is highlighted in red boxes. For those genes located on the reverse strand, the converted complementary sequence was displayed on the top, together with the identification of m^{6}A site (labeled as a red circle below) determined based on the location of CIMS C→T mutational signature and/or CITS deletion signature.

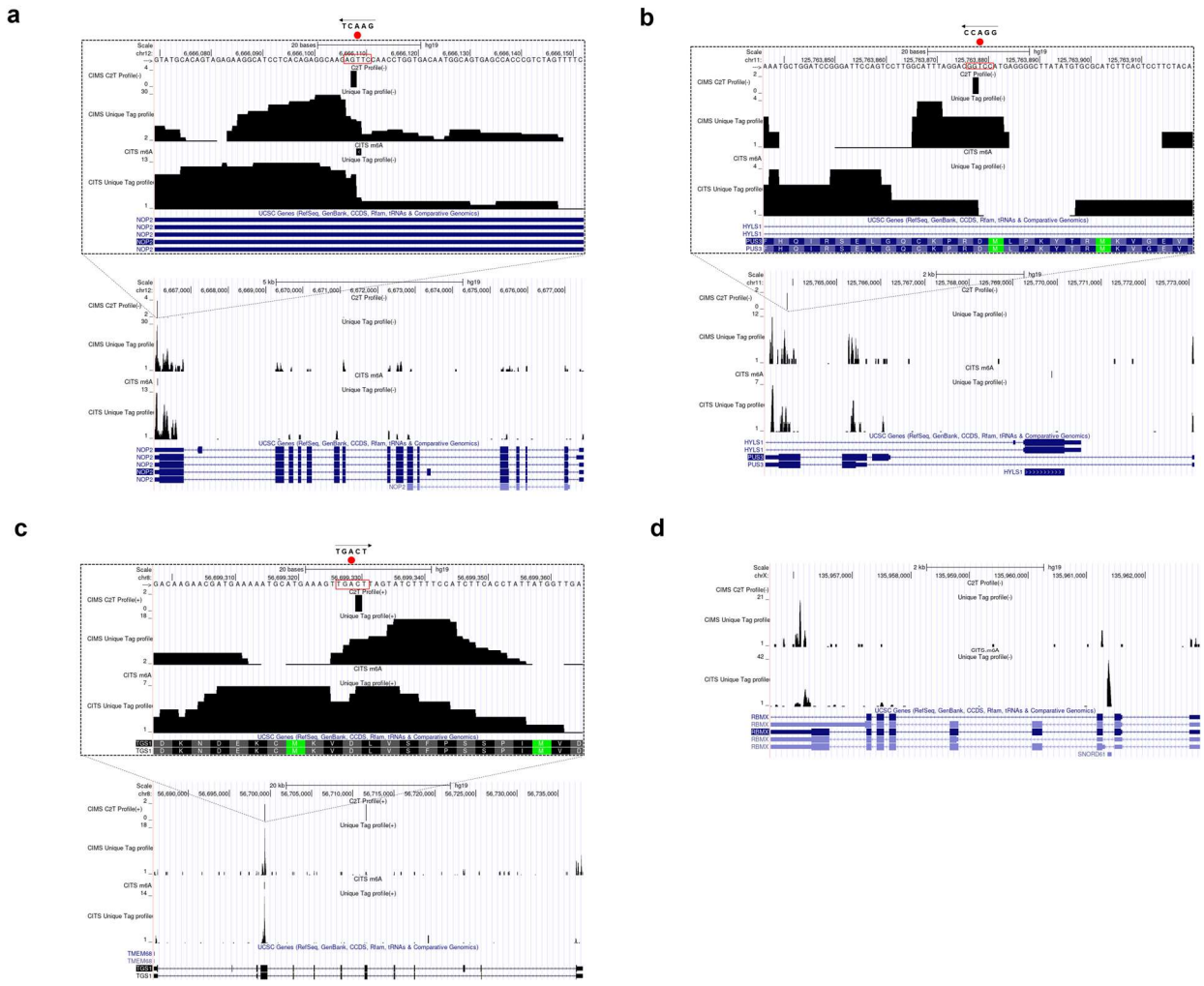

Figure S3. (a) LC-PRM quantification results of PUS1, PUS3, TRUB1, PUS7, and PUS7L in *ALKBH5*$^{-/-}$, *FTO*$^{-/-}$ and *METTL3*$^{-/-}$ cells, compared with the isogenic parental HEK293T cells. (b) m^{6}A mapping results for the transcripts of *PUS1*, *TRUB1*, *PUS7*, and *PUS7L* genes in GSE63753. Additional description is provided in the legend of Figure S2.