Structural basis for selective binding of m6A RNA by the YTHDC1 YTH domain

Chao Xu[1#,*], Xiao Wang[2,3#], Ke Liu[1#], Ian A. Roundtree[3,4], Wolfram Tempel[1], Yanjun Li[1], Zhike Lu[2,3], Chuan He[2,3*], Jinrong Min[1,5*]

[1]Structural Genomics Consortium, University of Toronto, 101 College Street, Toronto, ON, M5G 1L7, Canada.

[2]Department of Chemistry and Institute for Biophysical Dynamics, The University of Chicago, 929 East 57th Street, Chicago, Illinois 60637, USA.

[3]Howard Hughes Medical Institute, The University of Chicago, 929 East 57th Street, Chicago, Illinois 60637, USA.

[4]Medical Scientist Training Program and Department of Biochemistry and Molecular Biology, The University of Chicago, 929 East 57th Street, Chicago, Illinois 60637, USA.

[5]Department of Physiology, University of Toronto, Toronto, ON, Canada M5S 1A8.

#: These authors contributed equally to the work.

*: Corresponding author: jr.min@utoronto.ca, chuanhe@uchicago.edu, chao.xu@utoronto.ca

Abstract

N^6-methyladenosine (m^6A) is the most abundant internal modification of nearly all eukaryotic mRNAs, and has recently been reported to be recognized by the YTH domain containing proteins. Here we present the crystal structures of the YTH domain of YTHDC1 and its complex with an m^6A-containing RNA. Our structural studies, together with PAR-CLIP and biochemical experiments reveal the specific mode of m^6A-YTH binding, but also explain the preferential recognition of the GG(m^6A)C sequences by YTHDC1.

Methylation of internal adenosines to form N^6-methyladenosine (m^6A) is a key processing event during maturation of eukaryotic messenger RNAs (mRNAs), in concert with 5' capping, 3' polyadenylation and splicing[1]. m^6A is highly conserved across all eukaryotes from yeast[2] to human[3] within a G(m^6A)C (70%) or A(m^6A)C (30%) motif, and is physiologically essential to metazoans[4-6]. Transcriptome-wide studies have revealed that m^6A is a widespread modification present in over 7,000 human genes[7,8]. Analogous to other chemical codes that overlay primary sequence, such as DNA methylation and histone marks, the m^6A modification on mRNA and certain non-coding RNA is reversible and carries regulatory functions[9-11]. The biology of m^6A is achieved by three types of proteins[9]: m^6A methyltransferase complex (such as METTL3-METTL14) that installs the methyl group[12,13]; demethylases (FTO[10] and ALKBH5[11]) that reverse methylation; and m^6A-specific RNA-binding proteins that recognize m^6A ("readers") [7,9,14]. Identification of m^6A readers is especially important in determining the cellular function of m^6A. YTHDF2 has recently been characterized as the first m^6A reader that regulates the cytoplasmic stability of methylated RNA[14], implicating reversible RNA modification as a new layer of gene regulation at the post-transcriptional level [9,14].

The YTH RNA-binding domain is conserved in eukaryotes[15]. In mammalian cells, five proteins (YTHDF1-3 and YTHDC1-2) contain the conserved YTH domain (**Supplementary Results**, **Supplementary Fig. 1**). We have previously shown that three of these proteins, YTHDF1, YTHDF2, and YTHDF3, are cytoplasmic m^6A-specific binders[14]; however, the molecular basis for selective m^6A recognition by the YTH domain is unclear. YTHDC1 (YTH domain-containing 1), another member of the YTH domain family, resides in nucleus and is particularly interesting. YTHDC1 forms a novel YT body at transcriptionally active sites and adjacent to RNA processing speckles[16]. It interacts with other splicing factors and has also been reported to be a potential tumor repressor in endometrial cancer[17,18].

Studies of m^6A RNA methylomes in different organisms, including human, mouse and yeast, reveals that m^6A often exists in a RRACN (R is G or A) consensus motif [2,7,8], and the YTH domain of YTHDF2 preferentially binds to a conserved motif of G(m^6A)C[14]. In order to

corroborate the binding of the YTH domain of YTHDC1 to m^{6}A-containing RNA and determine its sequence preference, we first measured its binding affinity to the previously reported 42mer RNA oligonucleotides by gel shift assay and found that the YTH domain of YTHDC1 preferentially binds the methylated RNA with a dissociation constant (K_d) of around 0.4 μM (**Supplementary Fig. 2**), similar to that of YTHDF proteins[15]. The selective binding of YTHDC1 to methylated RNA was further confirmed by *in vitro* pull-down assay (**Fig. 1a**). We next applied photoactivatable ribonucleoside crosslinking and immunoprecipitation (PAR-CLIP)[19] to identify the RNA binding sites of YTHDC1. Among the 10,245 identified PAR-CLIP peaks, 51% overlap with the previously reported m^{6}A sites in human HeLa cell (**Supplementary Fig. 3a**)[11] and 21% overlap with the YTHDF2 binding sites (**Supplementary Fig. 3b**). RGAC is enriched in the PAR-CLIP peaks of YTHDC1 (**Supplementary Fig. 3c**) and the top binding motif of YTHDC1 is GGAC (**Fig. 1b**, **Supplementary Fig. 3d-3f**), the same as that of YTHDF2 and consistent with the reported high-resolution mapping of m^{6}A sites[8]. The YTHDC1 PAR-CLIP sites mainly distribute on exons and peaks around stop codon, hence resembling the sub-transcript distribution of m^{6}A methylome (**Fig. 1c**), and indicating that YTHDC1 is a *bona fide* m^{6}A reader in the cell. Functional clustering of the YTHDC1 gene targets reveals a significant portion of transcription regulators and implies an important role of YTHDC1 in gene expression (**Supplementary Fig. 4**).

In order to understand the sequence selectivity of YTHDC1 and search for suitable m^{6}A RNA oligonucleotides for crystallography, we measured the binding affinities of the YTH domain of YTHDC1 to a series of RNA oligos of different lengths by the quantitative ITC (isothermal titration calorimetry) binding assay (**Table 1** and **Supplementary Table 1**). We found that: 1) The YTHDC1 YTH domain recognizes RNA in an m^{6}A-dependent manner regardless of the RNA length; 2) The YTH domain of YTHDC1 prefers a G nucleotide at the -1 position relative to the m^{6}A, and an adenosine is least favored at this position; and 3) the YTH domain of YTHDC1 exhibits a slight preference for G and C at the -2 or +1 positions, respectively. Most importantly, our ITC binding results are consistent with our PAR-CLIP results, supporting the assertion that YTHDC1 preferentially binds to the GG(m^{6}A)C sequence

(**Table 1**).

In order to gain insights into the molecular mechanism of specific m^6A recognition by the YTH domain, we determined crystal structures of the YTH domain of human YTHDC1 and its complex with a 5mer m^6A RNA (GG(m^6A)CU) (**Supplementary Table 2**). Consistent with the previously determined solution structure of apo-YTHDC1 (PDB entry 2YUD), the YTH domain of YTHDC1 adopts an open α/β fold. Part of the loop linking the β4 and β5 strands (called L_{45} thereafter) is disordered in the apo structure (**Supplementary Fig. 5**), but resolved in the m^6A RNA complex structure (**Fig. 2a**). In addition, The YTHDC1 YTH domain contains 3_{10} helices (η1-η3), one following L_{45} (η1) and the other two between β5 and β6 strands (η2 and η3) (**Supplementary Fig. 1b**). Like L_{45}, η1 is missing in the apo structure, but visible in the complex structure. Other than that, the apo and m^6A complex structures are almost identical with a root-mean-square deviation (R.M.S.D) of only 0.42Å between these two structures. In the YTHDC1-GG(m^6A)CU complex structure, all five nucleotides are visible (**Fig. 2a**).

The base of N^6-methyladenosine is snugly accommodated in a deep pocket formed by residues from β1, the loop between β1 and α1, α1, β2, and L_{45} (**Fig. 2a-2b**). Notably, the methyl group of m^6A is recognized by an aromatic cage consisting of W377, W428 and L439 (**Fig. 2c**, and **Supplementary Fig. 6**), and the distances between the N6 atom and the Leu439 and the indole planes of W377 and W428 are similar to those between the N^{ε} of the methylated lysines of histone H3 and the rings of the aromatic cage residues of HP1 or JARID1 **Supplementary Fig. 7**)[20,21]. Mutating either W377 or W428 to alanine completely disrupts its binding to m^6A RNA (**Table 1**), underscoring the importance of the cage residues in m^6A recognition. In addition to the cage accommodation, the adenine moiety of m^6A also forms three hydrogen bonds with the main chain carbonyl oxygen of S378 (6-amino group), the side chain NH_2 of N367 (N^1 in purine), and the main chain NH group of N363 (N^3). The C2'-ribosyl hydroxyl oxygen of m^6A forms another hydrogen bond with the side chain of N363 (**Supplementary Fig. 7**).

Our complex structure not only sheds light on why the YTH domain specifically recognizes

m^6A, but also provides structural basis for the preference of a G nucleotide at the -1 position (G-1). The carbonyl oxygen (guanine 6-oxo) of the guanosine G-1 forms a hydrogen bond with the main chain NH group of V382 (**Supplementary Fig. 7** and **Supplementary Fig. 8a**). Replacing G-1 with any other nucleotide would disrupt this hydrogen bond (**Supplementary Fig. 8b-8d**). Furthermore, replacing G-1 with adenosine could introduce steric clashes with V382 between the NH_2 group of adenosine and the main chain NH group of V382 (**Supplementary Fig. 8b**).

The cytidine at the +1 position (C+1) forms a water-mediated hydrogen bond with the side chain of N466 via its base carbonyl oxygen, two hydrogen bonds via its ribosyl hydroxyl oxygen, and two hydrogen bonds respectively with the side chain of R475 and the backbone of D476 via its phosphate group. Most importantly, this cytosine is stacked between the guanidinium group of R475 of YTHDC1 and a uracil at the +2 position, forming cation-π and π-π interactions, respectively (**Supplementary Fig. 6**). The similar stacking mode has been found in other protein-nucleic acids complexes, (**Supplementary Fig. 9**)[22]. The importance of this R475-mediated interaction has also been verified by our mutagenesis studies. Mutating R475 to phenylalanine diminishes the binding affinity by 9 fold, while mutating R475 to alanine decreases the binding affinity over 100 fold (**Table 1**). On the other hand, we found that this stacking interaction is not base-specific, consistent with our ITC data that replacing C+1 with other nucleotides only slightly affects the binding affinity (**Table 1**).

The guanosine at the -2 position (G-2) forms two base specific hydrogen bonds with the main chain carbonyl oxygen of YTHDC1 D476, one formed directly via its NH_2 group (N^2) and the other water-mediated hydrogen bond via its NH group (N^1) (**Supplementary Fig. 6**). Substituting G-2 for other nucleotides will disrupt both hydrogen bonds, but will not introduce steric clashes (**Supplementary Fig. 10**). Accordingly, m^6A binding to YTHDC1 is only slightly reduced when replacing G-2 with other nucleotides (1.9-2.1 fold, **Table 1**). The uridine at the +2 position (U+2) forms five hydrogen bonds with YTHDC1 via phosphate and hydroxyl groups (**Supplementary Fig. 6**), consistent with a lack of selectivity at this position. Although YTHDC1 does not display sequence selectivity at the +2 position, it may also

contribute to binding because a 3mer G(m^{6}A)C RNA displays 14-fold weaker binding affinity than the 5mer GG(m^{6}A)CU RNA (**Table 1**).

Structural and binding studies reported in this study unveil a general function of the YTH domain as a specific m^{6}A RNA reader, and a guanine nucleotide is favored at the position preceding m^{6}A. Previous m^{6}A methylome profiling shows that m^{6}A is often embedded in a RRACU sequence (R=G/A) in mouse and human[7,8], and in a RGAC motif in yeast[2], which is consistent with earlier biochemical studies showing that the m^{6}A methyltransferases specifically modify adenine within the consensus sequence motifs of GAC (70%) and AAC (30%)[3]. In addition to the YTH domain proteins, a RRM domain protein ELAVL1 has also been identified as an m^{6}A reader from an m^{6}A RNA pull-down assay[7]. Without any doubt, more m^{6}A readers would be identified in the future, which could potentially selectively recognize non-GAC m^{6}A RNA.

As the most abundant internal modification in RNA, the m^{6}A modification is reversible and plays regulatory roles similar to reversible DNA and histone modifications[9]. The m^{6}A mark could exhibit biological functions by disturbing the binding of normal protein factors, altering RNA secondary structure, or recruiting specific reader proteins. We show here that the YTH domains possess an exquisite pocket for specific recognition of the methyl group of m^{6}A, which is distinct from that of the 5mC (5-methylcytosine) by the MBD and SRA domains[22] and that of m7G (7-methyl-GDP) by eIF4E[23] (**Supplementary Fig. 9**). Instead, the recognition mode is similar to that of the methyl-lysine/arginine in histones[20,21] (**Supplementary Fig. 7**), which has not been observed for nucleic acid methylations in the past; the evolution of the similar methyl-binding mode of m^{6}A recognition as in histones suggests functional importance of this reading process in mammalian cells. We also provide molecular basis for the preference of the GG(m^{6}A)C sequence by the YTH domain proteins. Therefore, the current work serves as a prototype for selective recognition of the new emerging mode of m^{6}A RNA methylation in biological regulation.

Supplementary Information is available in the online version of the paper.

ACCESSION CODES

The crystal structures of YTHDC1 were deposited in the Protein Data Bank under accession codes 4R3H (apo YTHDC1 YTH domain) and 4R3I (YTHDC1-m6A RNA complex), respectively. The PAR-CLIP sequencing data were deposited in the Gene Expression Omnibus (http://www.ncbi.nlm.nih.gov/geo) under accession number GSE58352.

ACKNOWLEDGEMENT

We would like to thank John Walker for assistance in structure determination. The SGC is a registered charity (number 1097737) that receives funds from AbbVie, Boehringer Ingelheim, the Canada Foundation for Innovation, the Canadian Institutes for Health Research, Genome Canada through the Ontario Genomics Institute [OGI-055], GlaxoSmithKline, Janssen, Lilly Canada, the Novartis Research Foundation, the Ontario Ministry of Economic Development and Innovation, Pfizer, Takeda, and the Wellcome Trust [092809/Z/10/Z, J.M.]. C.H. is supported by Howard Hughes Medical Institute as an investigator. Results shown in this report are derived from work performed at Argonne National Laboratory, Structural Biology Center at the Advanced Photon Source. Argonne is operated by UChicago Argonne, LLC, for the U.S. Department of Energy, Office of Biological and Environmental Research under contract DE-AC02-06CH11357.

Author Contributions

C.X. and J.M. conceived the project; C.X performed the structural and binding experiments with assistance from K.L, W. T. and Y. L.; X. W. and I. A. R. conducted the PAR-CLIP experiment of YTHDC1; Z. L. and X. W. analyzed the PAR-CLIP data. C.X., X. W., C.H. and J.M wrote the manuscript. All authors contributed to editing the manuscript. C. H. and J.M. supervised the project.

Competing Financial Interests

The authors declare no competing financial interests.

Additional information

Supplementary information is available in the online version of the paper. Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to C.X. (chao.xu@utoronto.ca), C.H. (chuanhe@uchicago.edu), or J.M. (jr.min@utoronto.ca).

FIGURE LEGENDS

Figure 1. YTHDC1 is a nuclear m^{6}A reader. (a). LC-MS/MS showed that the recombinant-expressed YTH domain of YTHDC1 was able to enrich m^{6}A-containing RNA from poly(A)-tailed RNAs of human HeLa cell. When the YTHDC1 protein was crosslinked with its associated RNA by ultraviolent light and subjected to partial digestion, the enrichment of m^{6}A at the YTH-bound portion was increased compared to that without crosslinking treatment Blue columns, immunoprecipitation (IP) of the YTH domain of YTHDC1 with its associated RNA; orange columns, protein-RNA complexes are UV-crosslinked and partially RNase T1 digested before IP (CLIP). Error bars, mean ± s.t.d. (standard deviation), n = 2, technical replicates. (b). Binding motif identified by HOMER motif analysis (p = 1.0×10^{-246}, 2778 sites were found under this motif). (c). Pie chart depicting the region distribution of YTHDC1-binding sites identified by PAR-CLIP.

Figure 2. Structural basis of preferential recognition of the GG(m^{6}A)C consensus motif

by the YTH domain of YTHDC1. (a). Overall structure of the YTHDC1 YTH domain in complex with RNA GG(m^6A)CU. The protein is shown in blue cartoon except loop L_{45} and two α helices (α0 and α4), which are colored in orange and red, respectively. The 5mer RNA is shown in yellow sticks. The composite simulated anneal |Fo| − |Fc| omit map of the 5mer RNA is contoured at 2.5σ. (b). The electrostatic surface of the YTHDC1 YTH domain in complex with the 5mer RNA GG(m^6A)CU (yellow stick). The m6A RNA resides in a highly positively charged binding groove formed by the C-terminal ends of strands β1, β2 and β3 and subsequent loops, helix α1, L_{45} and the loop preceding β6. (c). The m^6A binding pocket of the YTHDC1 YTH domain. The protein and m^6A are shown in blue ribbon and yellow stick, respectively. Residues involved in recognizing m^6A are shown in blue sticks. The distances between N^6 of m^6A and the three residues are labeled in black, and the hydrogen bonds are indicated in gray dash.

Online Methods

Cloning, expression and purification of human YTHDC1 YTH domain for crystallography and ITC

The YTH domains of humanYTHDC1 (residues 345-509) was subcloned into pET28a-MHL vector. The recombinant YTHDC1 YTH domain was over-expressed at 18 °C with amino-terminal His-tag in *E. coli* BL21 (DE3) Codon plus RIL (Stratagene) in the presence of 1 mM IPTG. The overnight cell cultures were harvested by centrifuge and dissolved in the lysis buffere containing 20 mM Tris-HCl, pH 7.5, 400 mM NaCl, 0.5 mM PMSF. The cells were lysed by sonication. Supernatant was collected after centrifugation at 16000 g for 40 minutes, and then applied to Ni-NTA resin (Qiagen). The target protein was washed with lysis buffer plus 30 mM imidazole and then eluted in a buffer containing 20 mM Tris-HCl, pH 7.5, 400 mM NaCl and 0.5 M imidazole. Tobacco etch virus (TEV) protease was added to remove the N-terminal tag of the recombinant protein and dialyzed with lysis buffer overnight. The mixture was applied to another Ni-NTA resin to remove the protease and uncleaved proteins. Then the cleaved recombinant proteins were further purified by Superdex 75 gel filtration (GE health care) in a buffer containing 10 mM Tris-HCl, pH 7.5, 150 mM NaCl and 1 mM DTT. The purified protein was concentrated to 20 mg/ml and stored at -80 ^{o}C.

The mutants of YTHDC1 YTH domain were cloned using the Site-Directed mutagenesis kits (Invitrogen), and were expressed and purified in the same way as the wild type (**Supplementary Fig. 11** and **Supplementary Fig. 12**).

Crystallization, data collection and structural determination

For crystallization of the YTHDC1 YTH domain (345-509), 1 µl protein (15mg/ml) was

mixed with 1 µl crystallization buffer using hanging drop vapour diffusion method at 18 °C. The diffracting crystal was crystallized in a buffer containing 0.1 M Bis-Tris, pH 6.5, 0.2 M Ammonium Sulfate and 25% PEG 3350. For crystallization of YTHDC1 (345-509)-GG(m^6A)CU complex, 10 mg/ml protein (final concentration) was mixed with the modified 5-mer RNA GG(m^6A)CU (Thermo Fisher Scientific, Inc.) in a molecular ratio of 1:2. The mixture was incubated on ice for 0.5 h before crystallization. The RNA complex was crystallized in the same condition as the apo YTHDC1.

Diffraction data were collected using a copper rotating anode x-ray source, integrated with XDS[24] and scaled with POINTLESS/AIMLESS[25]. In the case of the YTHDC1-RNA complex, data were also integrated and scaled in HKL-3000 for use in molecular replacement and early refinement steps. Structures were solved by molecular replacement with PHASER[26]. A crystal structure isomorphous to the YTHDC1-apo crystal, below referred to as "crystal form A" was solved with a search model derived from PDB entry 2YUD. Model coordinates of crystal form A served as search model for the YTHDC1-RNA complex structure. RNA geometry restraints were supplemented with JLIGAND[27]. The current models were iteratively rebuilt, refined and validated with COOT, REFMAC and MOLPROBITY[28,29], respectively. A summary of crystallographic statistics is shown in **Supplementary Table 2**.

Isothermal titration calorimetry (ITC) measurements

All Isothermal Titration Calorimetry (ITC) measurements were recorded at 25 °C using a MicroCal ITC200 (GE healthcare). All RNAs used for ITC binding experiments were purchased from Thermo Fisher Scientific except the unmodified GGACU, which was purchased from Integrated DNA Technologies, Inc. The purity of all purchased RNAs is > 90%. All proteins and RNAs are dialyzed or dissolved in the same buffer containing 20 mM Tris, pH 7.5, 150 mM NaCl before use. 10-15 injections were performed by injecting 2 µl 400-700 µM RNAs into a sample well containing 10-60 µM of proteins. The concentration of the proteins and RNAs were estimated with absorbance spectroscopy using the extinction coefficients, OD280 and OD260, respectively. Binding isotherms were plotted, analyzed and

fitted in a one-site binding model by Origin Software (MicroCal Inc.) after subtraction of respective controls. The Kd, entropy, enthalpy and Gibbs free energy, as well as their deviations were also calculated by Origin Software (MicroCal Inc.) during the ITC curve fitting in a one-site model (**Table 1** and **Supplementary Table 1**).

Plasmid construction and protein expression for PAR-CLIP and *in vitro* IP C-terminal flag-tagged YTHDC1 was cloned from commercial cDNA (Open Biosystems, clone ID: 5541053) into vector pcDNA 3.0 (*EcoR*I, *Xho*I; forward primer: CGTCAGAATTCATGGCGGCTGACAGTC; reverse primer: GGCATCTCGAGTTACTTGTCATCGTCGTCCTTGTAATCTCTTCTATATCGACCTCT CTC). Plasmids with high purity for mammalian cell transfection were prepared with a Maxiprep kit (QIAGEN). Recombinant His-tagged YTH domain of YTHDC1 was constructed by sub-cloning into vector pET-28a (*NdeI*, *XhoI*; forward primer: CGTCACATATGCAAACCAGTAAACTCAAATATGTGC; reverse primer: GGCATCTCGAGTCAGTGACGCATTTTATGAATGACCTG). The resulting clones were transfected into the *E. coli* strain BL21 and expression was induced at room temperature with 1 mM IPTG for 20 h. The pellet collected from 2 L bacteria culture was then lysed in 30 mL lysis buffer (20 mM Tris, pH 7.5, 200 mM NaCl, 0.1% (v/v) Triton X-100) and sonicated for 10 min. After removing cell debris by centrifuge at 12000 rpm for 30 min, the supernatant was purified by Ni-NTA cartridge (QIAGEN, 5 mL) following manufacture's instruction. The crude products were further purified by gel-filtration chromatography in GF buffer (10 mM Tris, pH 7.5, 200 mM NaCl, 3 mM DTT). The yield was around 10 mg per liter of bacterial culture.

Gel shift binding assay and *in vitro* RNA IP was conducted with recombinant His-taggged YTH domain of YTHDC1 following previous reported procedure[14]. The procedure of *in vitro* RNA CLIP (crosss-linking and IP) is the same as *in vitro* RNA IP, except for the following modifications: (1) before immunoprecipitation with anti-His maganetic beads, the RNA-protein mixture was transferred to a UV-transparent 96 well plate (placed on ice and without cover) and cross-linked three times with 0.15 J/cm^2 of 254 nm UV light each time in

a Stratalinker 2400 (Stratagene); (2) after UV-cross-linking, the RNA-protein mixture was quenched on ice and subjected to RNase T1 digestion (1 U/μL RNase T1 for 8 minutes at 22^{o}C); (3) after IP and washing, the RNA was detached from magnetic beads by protease K digestion (1 mg/mL final concentration, 50^{o}C, 30 min) and the RNA was further recovered by Zymo RNA clean and concentrator.

Mammalian cell culture and plasmid transfection Human HeLa cells used in this study were purchased from ATCC (CCL-2) and cultured in DMEM (Gibco, 11965) media supplemented with 10% FBS and 1% 100 X Pen Strep (Gibco). Transfection was achieved using Lipofectamine 2000 (Invitrogen) according to the manufacturer's protocol.

PAR-CLIP was performed based on a previously reported procedure[30] with the following modifications. 4 x 15 cm plates of cells were transfected with Lipofectamine 2000 according to manufacturer's protocol. After 6 hours the media was replaced and cells were cultured in fresh media supplemented with 200 μM 4-thiouridine (Sigma) overnight. The first RNase T1 digestion was conducted at 1 U/μL RNase T1 for 8 minutes. For the second digestion, the concentration of RNase T1 was reduced from 100 U/μL to 20 U/μL. Following dephosphorylation, 1/10 of the sample was partitioned for ^{32}P labeling while the remaining volume was treated with 1 U/μL T4 PNK at 37°C for 10 minutes, followed by addition of ATP to 1 μM for 30 minutes at 37 °C. This sample was then washed and digested with proteinase K. RNA was purified using Zymo RNA Clean & Concentrator before library construction by Tru-seq small RNA sample preparation kit (Illumina). The cDNA library was sequenced by illumina Hiseq2000 with single end 50-bp read length. The adapters were trimmed by using FASTX-Toolkit[31]. The deep sequencing data were mapped to Human genome version hg19 by Tophat version 2.07 [32] without any gaps and allowed for at most two mismatches. PAR-CLIP data were analyzed by PARalyzerv1.1 [33] with default settings. The motif of PAR-CLIP peaks was generated by HOMER motif analysis. The motif with lowest *p*-value was shown in Fig. 1b. The distances between GGAC and T to C mutation sites were calculated by setting the A of GGAC as position zero (**Supplementary Fig. 3e**).

REFERENCES

1. Tuck, M.T. The formation of internal 6-methyladenine residues in eucaryotic messenger RNA. *Int J Biochem* **24**, 379-86 (1992).
2. Schwartz, S. et al. High-resolution mapping reveals a conserved, widespread, dynamic mRNA methylation program in yeast meiosis. *Cell* **155**, 1409-21 (2013).
3. Wei, C.M. & Moss, B. Nucleotide sequences at the N6-methyladenosine sites of HeLa cell messenger ribonucleic acid. *Biochemistry* **16**, 1672-6 (1977).
4. Zhong, S. et al. MTA is an Arabidopsis messenger RNA adenosine methylase and interacts with a homolog of a sex-specific splicing factor. *Plant Cell* **20**, 1278-88 (2008).
5. Bokar, J.A. Fine-Tuning of RNA Functions by Modification and Editing. Vol. 12 (ed. Grosjean, H.) 141-177 (Springer-Verlag, Berlin Heidelberg, 2005).
6. Hongay, C.F. & Orr-Weaver, T.L. Drosophila Inducer of MEiosis 4 (IME4) is required for Notch signaling during oogenesis. *Proc Natl Acad Sci U S A* **108**, 14855-60 (2011).
7. Dominissini, D. et al. Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. *Nature* **485**, 201-6 (2012).
8. Meyer, K.D. et al. Comprehensive analysis of mRNA methylation reveals enrichment in 3' UTRs and near stop codons. *Cell* **149**, 1635-46 (2012).
9. Fu, Y., Dominissini, D., Rechavi, G. & He, C. Gene expression regulation mediated through reversible m(6)A RNA methylation. *Nat Rev Genet* **15**, 293-306 (2014).
10. Jia, G. et al. N6-methyladenosine in nuclear RNA is a major substrate of the obesity-associated FTO. *Nat Chem Biol* **7**, 885-7 (2011).
11. Zheng, G. et al. ALKBH5 is a mammalian RNA demethylase that impacts RNA metabolism and mouse fertility. *Mol Cell* **49**, 18-29 (2013).
12. Liu, J. et al. A METTL3-METTL14 complex mediates mammalian nuclear RNA N6-adenosine methylation. *Nat Chem Biol* **10**, 93-5 (2014).
13. Wang, Y. et al. N6-methyladenosine modification destabilizes developmental regulators in embryonic stem cells. *Nat Cell Biol* **16**, 191-8 (2014).
14. Wang, X. et al. N6-methyladenosine-dependent regulation of messenger RNA stability. *Nature* **505**, 117-20 (2014).
15. Zhang, Z. et al. The YTH domain is a novel RNA binding domain. *J Biol Chem* **285**, 14701-10 (2010).
16. Nayler, O., Hartmann, A.M. & Stamm, S. The ER repeat protein YT521-B localizes to a novel subnuclear compartment. *J Cell Biol* **150**, 949-62 (2000).
17. Zhang, B. et al. Alternative splicing-related factor YT521: an independent prognostic factor in endometrial cancer. *Int J Gynecol Cancer* **20**, 492-9 (2010).
18. Hirschfeld, M. et al. Hypoxia-dependent mRNA expression pattern of splicing factor YT521 and its impact on oncological important target gene expression. *Mol Carcinog (2013).* (e-pub ahead of print 13 June 2013; DOI: 10.1002/mc.22045).
19. Hafner, M. et al. Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell* **141**, 129-41 (2010).
20. Jacobs, S.A. & Khorasanizadeh, S. Structure of HP1 chromodomain bound to a lysine 9-methylated histone H3 tail. *Science* **295**, 2080-3 (2002).

21. Wang, G.G. et al. Haematopoietic malignancies caused by dysregulation of a chromatin-binding PHD finger. *Nature* **459**, 847-51 (2009).
22. Zou, X., Ma, W., Solov'yov, I.A., Chipot, C. & Schulten, K. Recognition of methylated DNA through methyl-CpG binding domain proteins. *Nucleic Acids Res* **40**, 2747-58 (2012).
23. Marcotrigiano, J., Gingras, A.C., Sonenberg, N. & Burley, S.K. Cocrystal structure of the messenger RNA 5' cap-binding protein (eIF4E) bound to 7-methyl-GDP. *Cell* **89**, 951-61 (1997).
24. Kabsch, W. Xds. *Acta Crystallogr D Biol Crystallogr* **66**, 125-32 (2010).
25. Evans, P.R. & Murshudov, G.N. How good are my data and what is the resolution? *Acta Crystallogr D Biol Crystallogr* **69**, 1204-14 (2013).
26. McCoy, A.J. et al. Phaser crystallographic software. *J Appl Crystallogr* **40**, 658-674 (2007).
27. Lebedev, A.A. et al. JLigand: a graphical tool for the CCP4 template-restraint library. *Acta Crystallogr D Biol Crystallogr* **68**, 431-40 (2012).
28. Murshudov, G.N. et al. REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr D Biol Crystallogr* **67**, 355-67 (2011).
29. Chen, V.B. et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* **66**, 12-21 (2010).
30. Hafner, M. et al. PAR-CliP--a method to identify transcriptome-wide the binding sites of RNA binding proteins. *J Vis Exp*, 2 Jul 2010 (DOI:10.3791/2034).
31. Pearson, W.R., Wood, T., Zhang, Z. & Miller, W. Comparison of DNA sequences with protein sequences. *Genomics* **46**, 24-36 (1997).
32. Trapnell, C., Pachter, L. & Salzberg, S.L. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25**, 1105-11 (2009).
33. Corcoran, D.L. et al. PARalyzer: definition of RNA binding sites from PAR-CLIP short-read sequence data. *Genome Biol* **12**, R79 (2011).

Table 1. Binding affinities of RNA constructs to the YTH domain of wild type and mutant YTHDC1 measured by ITC.

	RNAs	Kds (μM)
1	GAACCGG(m^6A)CUGUCUUA	0.30 ± 0.06
2	GAACCGA(m^6A)CUGUCUUA	2.0 ± 0.4
3	GAACCGU(m^6A)CUGUCUUA	0.50 ± 0.12
4	GAACCGC(m^6A)CUGUCUUA	0.40 ± 0.07
5	GG(m^6A)CU	2.0 ± 0.1
6	GA(m^6A)CU	15 ± 2
7	GU(m^6A)CU	5.7 ± 0.7
8	GC(m^6A)CU	5.6 ± 0.9
9	GG(m^6A)AU	6.2 ± 0.4
10	GG(m^6A)UU	3.3 ± 0.2
11	GG(m^6A)GU	2.8 ± 0.3
12	AG(m^6A)CU	3.8 ± 0.4
13	UG(m^6A)CU	4.3 ± 0.5
14	G(m^6A)C	28 ± 5
15	GAACCGGACUGUCUUA	*NB
16	GGACU	NB
17	GG(m^6A)CU	NB (W377A)
18	GG(m^6A)CU	NB (W428A)
19	GG(m^6A)CU	18 ± 2 (R475F)
20	GG(m^6A)CU	210 ± 20 (R475A)

*NB: No detectable binding. Note: The data represent mean value ± s.t.d (standard deviation), and standard deviations are calculated from the ITC curve fitting by Microcal Origin software.

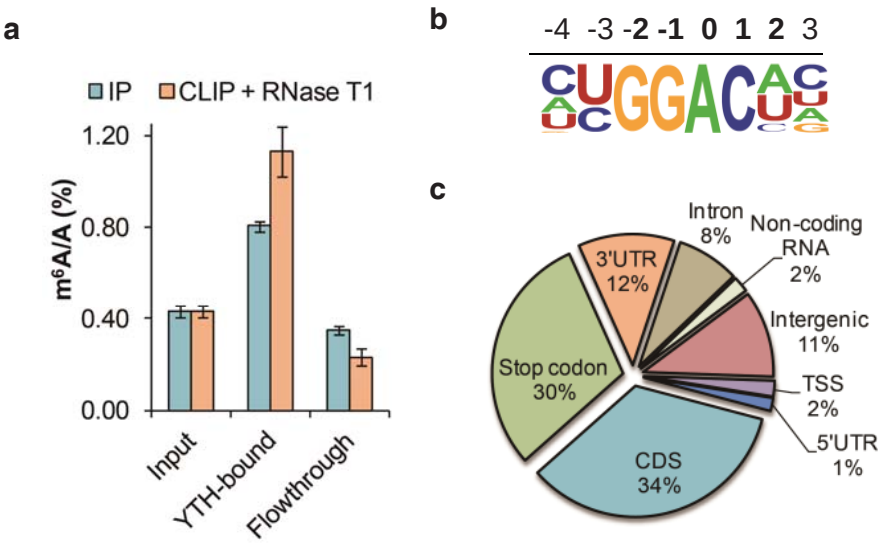

a
IP
CLIP + RNase T1
m6A/A (%)
1.20
0.80
0.40
0.00
Input
YTH-bound
Flowthrough
b
-4 -3 -2 -1 0 1 2 3
c
Intron 8%
Non-coding RNA 2%
3'UTR 12%
Intergenic 11%
Stop codon 30%
TSS 2%
5'UTR 1%
CDS 34%

Figure 1

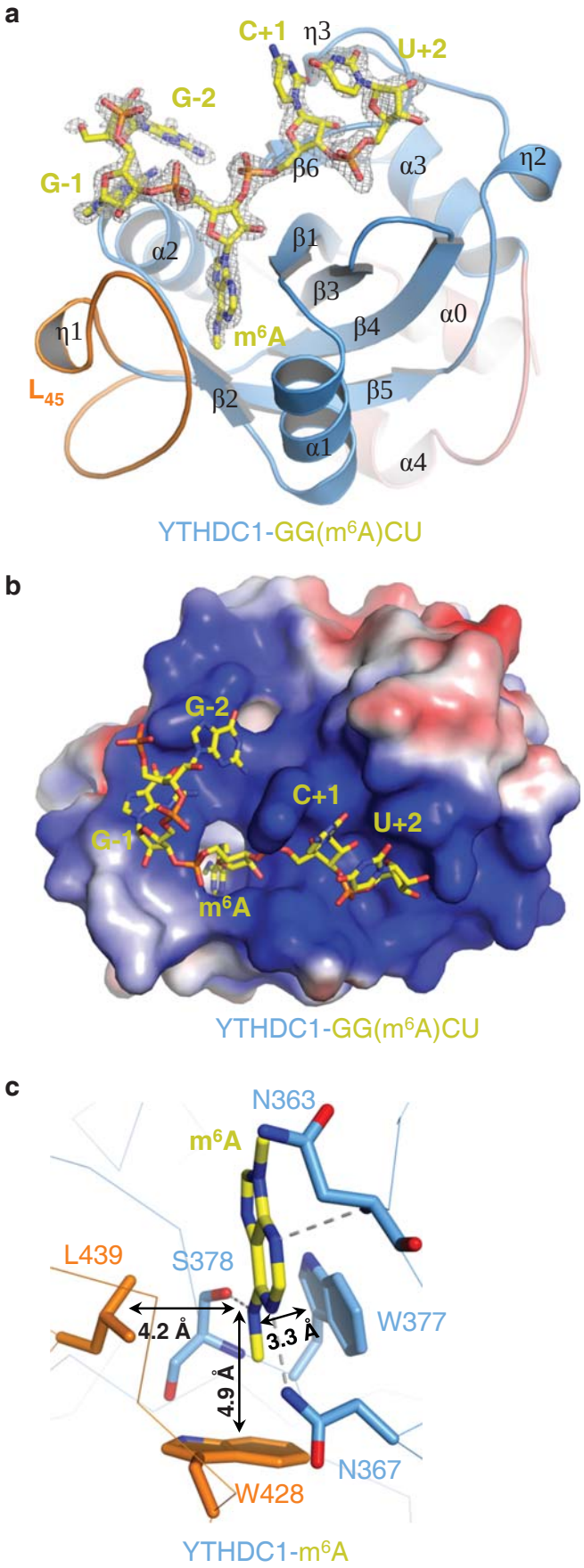
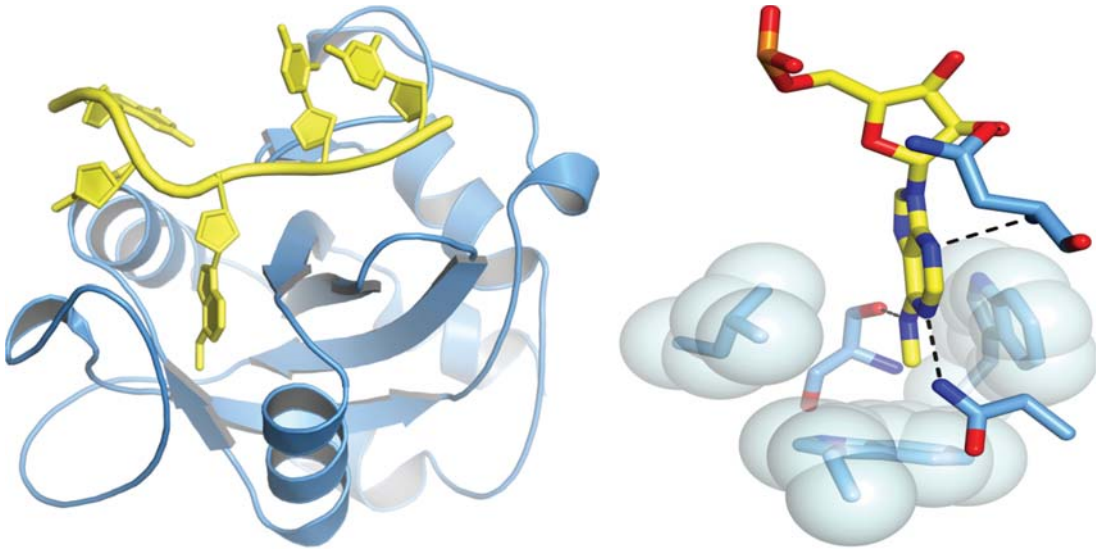

a
C+1
η3
U+2
G-2
G-1
β6
α3
η2
α2
β1
β3
α0
η1
m6A
β4
L45
β2
β5
α1
α4
YTHDC1-GG(m6A)CU
b
G-2
C+1
U+2
G-1
m6A
YTHDC1-GG(m6A)CU
c
N363
m6A
L439
S378
4.2 Å
3.3 Å
W377
4.9 Å
N367
W428
YTHDC1-m6A

Figure 2



Graphic Abstract