RNA binding to human METTL3-METTL14 restricts N^6-deoxyadenosine methylation of DNA *in vitro*

Shan Qi[1,2,#], Javier Mota[1,#], Siu-Hong Chan[3], Johanna Villarreal[1], Nan Dai[3], Shailee Arya[1], Robert A. Hromas[4], Manjeet K. Rao[1], Ivan R. Corrêa Jr.[3], Yogesh K. Gupta[1,2,5*]

[1] Greehey Children's Cancer Research Institute, University of Texas Health at San Antonio
8403 Floyd Curl Drive, San Antonio, TX 78229, USA

[2] Department of Biochemistry and Structural Biology, University of Texas Health at San Antonio
7703 Floyd Curl Drive, San Antonio, TX 78229, USA

[3] New England Biolabs, 240 County Road, Ipswich, MA, 01938, USA

[4] Division of Hematology and Medical Oncology, Department of Medicine, University of Texas Health at San Antonio
7703 Floyd Curl Drive, San Antonio, TX 78229, USA

[5] Lead contact

[*]Corresponding author:

Y.K.G email: guptay@uthscsa.edu

[#] These authors contributed equally to this work

Abstract

Methyltransferase like-3 (METTL3) and METTL14 complex transfers a methyl group from *S*-adenosyl-L-methionine to N^6 amino group of adenosine bases in RNA (m^6A) and DNA (m^6dA). Emerging evidence highlights a role of METTL3-METTL14 in the chromatin context, especially in processes where DNA and RNA are held in close proximity. However, a mechanistic framework about specificity for substrate RNA/DNA and their interrelationship remain unclear. By systematically studying methylation activity and binding affinity to a number of DNA and RNA oligos with different propensities to form inter- or intra-molecular duplexes or single-stranded molecules *in vitro*, we uncover an inverse relationship for substrate binding and methylation and show that METTL3-METTL14 preferentially catalyzes the formation of m^6dA in single-stranded DNA (ssDNA), despite weaker binding affinity to DNA. In contrast, it binds structured RNAs with high affinity, but methylates the target adenosine in RNA (m^6A) much less efficiently than it does in ssDNA. We also show that METTL3-METTL14-mediated methylation of DNA is largely regulated by structured RNA elements prevalent in long noncoding and other cellular RNAs.

Introduction

N^6-methyladenosine (m^6A) is considered a major covalent modification of the adenosine (A) base in coding and noncoding (nc) RNAs[1-5]. It is linked to diverse physiologic processes, including – but not limited to – RNA turnover[6-8], stem cell differentiation[9, 10], oncogenic translation[11-14], and DNA damage repair[15, 16]. In human cells, most m^6A in cellular RNAs is installed by METTL3 and METTL14 methyltransferases[2, 4, 17], both of which belong to the β-class of *S*-adenosyl methionine (SAM)-dependent methyltransferases (N^6-MTases) [18, 19] that also include N^6-deoxyadenosine DNA methyltransferases (m^6dA MTases), especially those from Type III restriction-modification (R-M) systems in bacteria (e.g. $Mod_{A/B}$ dimer of EcoP15I) [20]. The first structures of METTL3-METTL14 complexes suggested common features in EcoP15I and METTL3-METTL14[21, 22]. We analyzed these MTases and found high structural similarity of the cores of human METTL3, METTL14, and EcoP15I, suggesting their evolutionary origin from a common ancestor (fig. 1 a-e and supplementary fig. 1).

While m^6dA in bacterial DNA is deposited in a strictly sequence-dependent manner in double-stranded DNA (dsDNA) [20, 23], m^6A in RNA by human METTL3-METTL14 shows less stringent sequence dependency for sequences flanking the target **A** within the recognition motif DR**A**CH (D=A/G/U, R=A/G, H=A/C/U)[2, 17, 24]. Such relaxed specificity maybe required to expand the repertoire of target transcripts of METTL3-METTL14. However, this does not correlate with low levels of m^6A in cellular RNA, which remains at ~ 0.1% of total pool of ribonucleotides in cultured mammalian cells[25]. Of note, recent evidence suggests that all m^6A is deposited co-transcriptionally, with most occurring on chromatin-associated RNAs, e.g., nascent pre-mRNA

(CA-RNA), promoter-associated RNA, enhancer RNA, and repeat RNA elements[6-8, 26]. Consistently, the METTL3-METTL14 m^{6}A writers can be recruited to chromatin by modified histone tails (e.g., H3K36me3)[27] or the transcription machinery[12, 26]. Certain long noncoding (lnc) RNAs that are essential for regulated transcription, such as chromatin-associated *XIST*[28] and *NEAT2* (*MALAT1*)[29], also harbor m^{6}A marks. A role of m^{6}A in DNA damage response and repair has also been suggested[16, 30]. In those contexts, METTL3 localizes at sites of dsDNA breaks and installs the m^{6}A on DNA-damage associated RNAs[16]. METTL3 is also emerged as a pro-viral host factor in SARS-CoV-2 pathogenesis. Consistently, genetic depletion or inhibition of METTL3 by small molecules triggers innate immune response and limits viral growth[31, 32].

The increasingly appreciated roles of m^{6}A in nucleic acid metabolism thus seem to converge to a common point: the occurrence of m^{6}A at chromatin, a predominantly DNA- and RNA-rich environment inside the nucleus. A recent study showed that METTL3-METTL14 can convert dA to m^{6}dA within equivalent DR**A**CH (D=A/G/U, R=A/G, H=A/C/U) motif in a single-stranded DNA (ssDNA) with a 2.5-fold preference over the equivalent RNA *in vitro*[33]. Since m^{6}dA levels in mammalian DNA are extremely low (0.006 to 0.01%) and thought to originate from RNA catabolism rather than from SAM-dependent MTase reactions[34], fundamental questions about substrate(s) specificity and mechanisms of action of METTL3-METTL14 have emerged. By using classical biochemical and mass spectrometry assays, we show that METTL3-METTL14 preferentially methylates N^{6}dA on ssDNAs to m^{6}dA, and this activity is largely regulated by structured RNA elements prevalent in long noncoding (lnc) RNAs, that are also found in other cellular RNAs. Thus, our work provides a framework to explore a new regulatory axis of RNA-mediated restriction of *N*^{6}-deoxyadenosine (m^{6}dA) methylation in mammalian genomes.

Results

We compared the MTase cores of (Mod_B) of EcoP15I (aa 90 -132, 169 - 261, 385 – 511, PDB: 4ZCF[20]) and METTL14 (aa 116 – 402, PDB: 5IL0[21]). A secondary structure-based superposition[36] of these two structures revealed common features and similar arrangement of canonical MTase motifs (motifs I, and IV-X), including the motif IV (D/EPPY/W/L) that surrounds the Watson-crick edge of the flipped target adenine base (fig. 1 a-e and supplementary fig. 1). Previous studies identified three loops encompassing the sequence intervening MTase motifs IV and V (loop 1), VIII and VIII' (loop 2), and motifs IX and X (loop 3) contribute to substrate binding[21, 22, 35]. Due to lack of experimental structures of METTL3 or Mod_A (in complex with a flipped adenine base), we chose Mod_B and METTL14 for this comparison. While the MTase domain of METTL14 is enzymatically inactive, it harbors features that contribute to substrate binding[21, 22, 35]. These observations prompted us to address the prevailing question about substrate specificity of the human full-length METTL3-METTL14 enzyme complex. We co-purified the full-length human METTL3-METTL14 complex and its truncated form devoid of the RGG motif of METTL14 (METL3-METL14$_{-RGG}$) from insect cells (see details in methods section) (fig. 1f). The RGG motifs are clustered sequences of arginines and glycines. These motifs are commonly present in diverse set of RNA binding proteins that play important roles in physiological processes, such as RNA synthesis and processing[37]. In human METTL14, a total of seven RGG triplets and two RG motifs are present at its c-terminus tail (aa 408-457). The region also harbors a number of aromatic amino acids, which can further stabilize the bound nucleic acids via hydrophobic or stacking interactions. Consistently, this region contributes to substrate binding and m^6A activity of the

METTL3-METTL14 enzyme complex[38]. The sequence encompassing the RGG motifs in METTL14 is well conserved in higher vertebrates[17].

Next, we designed RNA substrates of varied lengths (5 to 30 nucleotides) with at least one canonical m^{6}A signature motif (RR**A**CH), including an RNA oligo (rPal-Top+bat) in which two RR**A**CH sites were arranged in a palindromic fashion (fig. 1g). We included a 14-mer DNA oligo (d6T) according to Woodcock et al[33] that reported that METTL3-METTL14 exhibited highest methyltransferase (MTase) activity on d6T. We also included two structured RNAs with propensity to form a perfect stem-loop (23-mer rTCE23 or also known as SRE[39]) without a DRACH motif, and a 30-mer rNEAT2 (or MALAT1) encompassing one DRACH motif and predicted to form a stem-loop with a bulged stem (fig. 1 h-i).

The m^{6}A mark in mammalian RNAs is widespread and expands beyond internal adenines in mRNAs to lncRNAs (*XIST*[28], *NEAT2* or *MALAT1*[29]), pri-miRNAs[40] and snoRNAs[24] – all of which are crucial to gene regulation and maintenance of cellular homeostasis. Of note, *NEAT2* lncRNA can relocate a large subset of transcription units across cellular compartments from polycomb bodies (PcGs) to interchromatin granules (ICGs)[41]. According to this model, a 49-nt long fragment of *NEAT2* (nt 4917-4966) facilitates the recruitment of a mega transcription complex[41]. We observed that this *NEAT2* fragment harbors one RR**A**CH motif (AA**A**CA) in its loop region and fulfils the criteria of m^{6}A peak signature (RR**A**CH[2, 17] or DR**A**CH[24] or **A**CA[42]) (fig. 1g). Thus, we decided to test the activity of METTL3-METTL14 on this regulatory module of the lncRNA. For simplicity of synthesis and purification, we focused on a slightly shorter version (30-nt; *rNEAT2-30*) of *NEAT2-49* in which both the m^{6}A signature motif and bulged stem regions were well

preserved (fig. 1g-h). The rTCE23 could serve as a control - it forms a perfect stem-loop structure in solution - confirmed by solution NMR previously[39] - but lacks a m^6A signature motif (fig. 1g, i).

We observed higher enzymatic activity of METTL3-METTL14 on the d6T DNA oligo compared with its RNA counterpart (r6T) (fig. 2a, c). By closely examining the sequence of the 14-mer d6T/r6T, we found a propensity of these oligos to form duplex DNA (for d6T) or RNA (for r6T): 6 of the 14 nucleotides at 5'-end will pair with a complementary sequence to create a duplex with an 8-nt overhang at the 3'-end of each strand (fig. 1g and 2d). As a result of this self-annealing, two adenine bases are available for methylation (m^6dA) at the end of the short palindromic sequence (5'-CCG•CGG-3'). We thus hypothesized that d6T (or 6T, as referred to by Woodcock et al[33]) may not truly represent a ssDNA substrate. Thus, we designed a new DNA oligo (d6T*) in which two nucleotides, a deoxycytidine and a deoxyguanosine at -3 and -2 positions (relative to the signature motif RR**A**CH), respectively, were replaced by two thymidines to disrupt base pairing and palindrome formation without affecting the core RR**A**CH motif required for m^6dA installation (fig. 1g). We reasoned that if METTL3-METTL14 indeed prefers a ssDNA for methylation, it should show higher activity for d6T*, a true ssDNA substrate, over d6T. Indeed, we found that the methyltransferase activity of METTL3-METTL14 towards d6T* increased by 2-fold compared with d6T (and >12-fold compared with r6T) (fig. 2a). Consistent with previous observations[33], METTL3-METTL14 showed no activity in the presence of a perfect duplex version of d6T DNA (d6T-ds). These results confirm a robust activity of METTL3-METTL14 on ssDNA substrates.

We observed no methyltransferase activity on rTCE23 RNA, as expected; but surprisingly, some low (but consistently detectable) methyltransferase activity on *rNEAT2* (fig. 2a). We then asked what attributes (sequence, length, and shape) in an RNA substrate are critical for METTL3-METTL14 to yield a high methyltransferase activity. To answer this question, we designed eight RNA oligos with varied length (5- to 22-nt) and sequences (fig. 1g). We also varied the position of the RR**A**CH motif and the flanking sequences in these oligos. All oligos comprised one RR**A**CH motif, except the RNA duplex Pal-Top+bot, where two RR**A**CH sites were arranged in a palindromic fashion within each 10-mer strand. We designed another variant of probe rC2-flip (22-mer, RRACH in the center), rC-flip wherein the RR**A**CH motif resided at the 3'-end. The other three RNA oligos, FoxM1-p1 (13-nt, RR**A**CH at 3'-end), c-Myc-p5 (8-nt), and c-Myc-p6 (11-nt, RR**A**CH at 5'-end) were derived from the *FOXM1* and *MYC* genes because of their involvement in m^{6}A-mediated processes in glioblastoma and acute myeloid leukemia[13, 43]. We covalently attached a fluorescein moiety at the 5'-end of all these oligos to facilitate the quantitative measurement of their binding affinities (equilibrium dissociation constant or Kd) to METTL3-METTL14 (see methods section for details).

We used a radiometric assay and fluorescence polarization (FP) to determine methylation activity and RNA binding, respectively. As shown in fig. 2a,c, RNA oligos with high propensity toward forming secondary structures (rNEAT2, rTCE23, Pal-Top+bot) showed low to negligible levels of methylation by METTL3-METTL14. The other RNA oligos showed some levels of methylation, though they could not be methylated with as much efficiency as the ssDNA substrate, d6T*. The results of binding experiments are shown in fig. 2b and table 1. The full-length METTL3-METTL14 binds to rNEAT2 with the highest affinity (Kd = 2 nM), followed by the hairpin

rTCE23 (Kd = 25 nM), the 3'-overhang r6T (Kd = 114 nM), the palindromic Pal-Top+bot (Kd = 127 nM), and the ssRNA FoxM1-p1 (Kd = 294 nM). Surprisingly, the DNA substrates (d6T and d6T*) that showed highest methyltransferase activity were the poorest binders, with Kd values of 370 and 509 nM, respectively.

If METTL3-METTL14 were to methylate ssDNA *in vivo*, and the evidence to date does not support its direct enzymatic origin[34], then what could be the significance of high affinity RNA binding to structured RNA oligos (rNEAT2 and rTCE23)? One explanation is that structural motifs in ncRNA or mRNAs can serve as a recruitment platform for METTL3-METTL14 in a specific context. But how would the RNA binding activity then affect DNA methylation activity? To answer this question, we measured the methyltransferase activity of METTL3-METTL14 in the presence of roughly equimolar mixtures of d6T* DNA with each RNA oligonucleotide. The methyltransferase activity of METTL3-METTL14 was significantly reduced in the presence of RNA (fig. 2c). Of note, RNAs with propensity to form secondary structures and highest binding affinity to METTL3-METTL14 (rNEAT2, rTCE23) attenuated methyltransferase activity to the most negligible levels (fig. 2c). To further explore the relationship between methyltransferase activity and substrate binding affinity, we synthesized a variant of the stem-loop oligo (rNEAT2*) with the DRACH motif (5'- AAA**C**A - 3') changed to an ideal m^{6}A target sequence i.e. 5'- GG**A**CU - 3' (fig. 1g). Although, higher methylation levels were observed for rNEAT2* compared to rNEAT2, most likely attributed to a perfect GGACU motif in rNEAT2*, yet the methylation was significantly lower than that of d6T* DNA (last panel of fig. 2c). Moreover, rNEAT2* showed ~6-fold reduction in affinity compared to rNEAT2 (fig. 2b). Importantly, rNEAT2* also attenuated methylation of the single stranded d6T* DNA by METTL3-METL14 but to a lesser extent than

the rNEAT2 (fig. 2c). These results suggests that: a) the sequence and shape of the target RNAs dictate the m^6A activity of METTL3-METTL14, and b) that higher affinity substrates tend to show lower methylation levels, possibly due to slow off rates of m^6A RNA.

To further investigate RNA-mediated inhibition, we showed that the two structured RNAs (rTCE23 and rNEAT2) inhibited the DNA methylation activity of METTL3-METTL14 in a dose dependent manner, with rNEAT2 (IC_{50} = 216 nM) showing 4-fold stronger inhibition than rTCE23 (IC_{50} = 887 nM) (fig. 2e). Next, we employed a LC/MS-based oligonucleotide intact mass analysis to further validate the results of the radiometric assay. As shown in fig. 2f, an equimolar (with respect to d6T* DNA) addition of rNEAT2 or rTCE23 significantly reduced the methyltransferase activity from 80.5% to 14.5 and 7.3%, respectively. We then performed nucleoside analysis of the METTL3-METTL14 reactions to unambiguously determine the identity of methylated base and efficiency of methylation (fig. 2g). We confirmed that deoxyadenosine on d6T* and adenosine on rNEAT2 were the only modified bases. In the absence of rNEAT2 or rTCE23, 86.3% of dAs on d6T* were modified. When rNEAT2 or rTCE23 were present, only 14.4 and 7.4%, respectively, of dAs on d6T* were modified. Additionally, 5.8% of the rAs on rNEAT2 were modified. As expected, no methylation on rTCE23 was detected (no RR**A**CH motif). These results are consistent with those obtained by oligonucleotide intact mass analysis. Altogether, the methyltransferase and binding assays confirm that the binding to structured RNAs, especially those lacking the GG**A**CU sequence, almost completely abolishes the methyltransferase activity METTL3-METTL14.

The c-terminal RGG repeat motif of METTL14 contributes to RNA binding and activity of METTL3-METTL14[38] (fig. 3a). Thus, we hypothesized that the absence of RGG motif in

METTL3-METTL14 should diminish its ability to bind RNA and consequently attenuate the effect of RNA on DNA methylation. In fact, we observed a 2- to 10-fold decrease in binding affinity by METTL3-METTL14$_{-RGG}$ (fig. 3b, and table 1). This suggests that the RGG motifs play a major role in RNA and DNA substrates, while other parts of the enzyme (e.g., CCCH-type zinc finger domains in METTL3[44] and MTase core of METTL14[35]) may also contribute to the overall binding (fig. 3b). As expected, the deletion of RGG caused 60% reduction in DNA methylation activity. In the presence of rNEAT2, the activity of METTL3-METTL14$_{-RGG}$ on d6T* DNA was reduced to about 75% of its activity in absence of the RNA; comparatively, a >90% activity reduction was observed for the full-length METTL3-METTL14 in presence of the same RNA (fig. 3c). These results suggest that the RGG motif, as a general RNA binding domain, dictates the shape-specific RNA recognition and promotes an RNA-mediated restriction of the DNA methylation activity of METTL3-METTL14 (table 1).

Discussion

Our results are noteworthy given the relevance of METTL3-METTL14 in regions where RNA and DNA are held in closed proximity, as supported by recent studies suggesting: a) localization of METTL3 to sites of double-stranded DNA breaks, and its role in DNA/RNA hybrid accumulation[16], b) recruitment of METTL3 to chromatin by promoter-bound transcription factors[12], c) UV-induced DNA damage signaling[30], and d) metabolism of R-loops[45]. Earlier studies suggest the existence of a large complex of multiple subunits of ~1 MDa size, possessing RNA N^6-adenosine methylation activity distributed over two sub-complexes of approximate molecular

mass of ~200 and ~800 kDa[2, 46]. One such subunit is WTAP, which alters both the pattern and location of the m^{6}A transcriptome[47]. Future studies could examine the role of WTAP and other partners of METTL3-METTL14 in substrate recognition and specificity, especially within the context of the regulatory role of RNA in DNA methylation.

In summary, we unveiled a new *N*^{6}-deoxyadenosine methylation activity on ssDNA, which is regulated by binding of METTL3-METTL14 to shape-dependent structured regions of ncRNAs and chromatin-associated nascent pre-mRNAs (CA-RNA) (fig. 3d, e). The structured elements in RNA could facilitate the recruitment of METTL3-METTL14 to chromatin and/or keep the writer complex engaged during co-transcriptional occurrence of m^{6}A. Such a topologic arrangement of METTL3-METTL14/RNA interaction could also restrict m^{6}dA deposition, especially to the single-stranded regions of DNA exposed during transcription[26], DNA recombination[16], damage and repair[15], and R-loop metabolism[45].

A major difference in prokaryotic and eukaryotic genomes is the contrasting occurrence of the base methylation. The m^{6}dA is a predominant modification mark in bacterial DNA, whereas the m^{5}dC is more prevalent in eukaryotic genomes. How is the eukaryotic genome then protected from m^{6}dA deposition despite having an efficient enzyme machinery (e.g., METTL3-METTL14) to accomplish such task? A comparative sequence/structural analysis of the bacterial EcoP15I and human METTL3 and METTL14 MTases provides some hints (fig. 1a-c, fig. 3a). Of note, there are two major differences that exist in the sequences of METTL3 and METTL14: a) loss of the target recognition domain (TRD) and b) acquisition of the RGG motif in METTL14. The TRD motif (in EcoP15I MTase) is a major contributor to binding and specificity for a dsDNA[20], whereas

271 the RGG motif is known for RNA binding[37]. It is likely that gene fusion/genome arrangement
272 event(s) could have occurred during evolution[18, 19] to ensure very low/no occurrence[48] of m^6dA in
273 mammalian genome. The high affinity to structured RNAs as demonstrated here could also ensure
274 efficient recruitment of METTL3-METTL14 to specific nuclear compartment(s) in human cells.
275
276 Interestingly, a recent study showed that a comprehensive network of RNA-RNA and RNA-DNA
277 interactions mediated by ncRNAs (including MALAT1) creates distinct nuclear compartments and
278 facilitate recruitment of proteins for RNA processing, heterochromatin assembly, and gene
279 regulation[49]. Thus, the secondary structure and 3D shape of the RNAs appears to be crucial for
280 these important biological processes. In line with these observations, our results suggest that
281 structured elements in RNAs may play an important role in regulating the ssDNA and ssRNA
282 methylation activity of METTL3-METTL14.
283

METHODS

Protein expression, purification, and oligonucleotide preparation

The coding sequence of full-length human METTL3 (NCBI reference sequence GI: 33301371) was used in this study. The gene was cloned into a plasmid suitable for expression of proteins in insect cells with an N-terminal poly histidine tag followed by a tobacco etch virus (TEV) protease site. This plasmid (5TEY (METTL3) was a kind gift from Dr. Cheryl Arrowsmith (Addgene plasmid # 101892; http://n2t.net/addgene:101892; RRID:Addgene_101892). The coding sequence of full-length human METTL14 (NCBI reference number GI: 172045930) was cloned into the pFastBac1 vector between BamHI and NotI restriction enzyme sites. For expression of recombinant proteins, we used ExpiSf Expression System (Thermo Fisher). First, the viral DNA bacmids for each gene were prepared from individual plasmids transformed into MAX Efficiency™ DH10Bac™ competent cells (Thermo Fisher). We confirmed the identity of the inserted genes by PCR amplification and DNA sequencing. The recombinant P0 virus for each gene were generated in ExpiSf9 insect cells following the manufacturer's recommendations (Thermo Fisher). The amount of each virus needed for infection has been optimized for maximal production of the two proteins. Infected cells were grown in ExpiSf CD medium (Thermo Fisher) at 27 ºC on an orbital shaker (speed 125 rpm) in a controlled (non-humidified, air regulated) environment. Cells were harvested at 72 hours post-infection by spinning at 300 X g for 5 minutes, washed with phosphate buffered saline (PBS) twice, and resuspended in a buffer A (0.025 M Tris-HCl pH 8.0, 0.5 M NaCl, 0.005 M Imidazole, 0.1 mM TCEP, 10% [(v/v)] glycerol) supplemented with 0.5% (v/v) Igepal, 2 mini-protease inhibitor cocktail tablets (Roche), DNase I, and lysozyme. Cells were lysed using a microfluidizer (Analytik, UK) and the soluble fraction was separated by

centrifuging the lysate at 40,000 rpm for 40 minutes. The supernatant was passed through a 0.22 µm filter.

The clarified soluble fraction was loaded on to a Nuvia IMAC column (Bio-Rad) pre-equilibrated in buffer A. The proteins were eluted by increasing the concentration of imidazole. The 6XHis tag was then proteolytically removed using TEV protease. Any uncleaved fractions and remaining protein impurities were removed by second passage through an IMAC column. The METTL3-METTL14 complex was further purified by successive passage through HiTrap heparin and HiLoad Superdex 200 columns (GE Healthcare). The protein complex was eluted as a single homogenous species in a final buffer containing 0.02 M Tris-HCl pH 8.0, 0.15 M NaCl, 0.1 mM TCEP. The purified protein complex was concentrated to 3-5 mg/mL and used immediately, or flash frozen in liquid nitrogen and stored at -80 ºC. The METTL3-METT14$_{\text{-RGG}}$ complex lacks the c-terminus RGG repeats motif (amino acids 400 - 457) of METTL14. It was expressed and purified using same method as the full-length METTL3-METTL14 complex.

All RNA and DNA oligonucleotides oligonucleotides used in this study were synthesized, purified by HPLC, and received in deprotected (for RNAs) and desalted form from Dharmacon and Integrated DNA Technologies, respectively. Oligos were dissolved in 1X buffer containing 0.01 M Tris-HCl pH 8.0, 0.05 M NaCl. To prepare double-stranded oligos, the two strands were mixed in roughly equimolar amounts. Annealing was carried out by heating the mixture to 95 ºC for 2 minutes followed by gradual cooling to 25 ºC over 45 minutes in a thermocycler (Eppendorf).

Methyltransferase assays

Radiometric assay

Each reaction was carried out in a 5 µL mixture containing 50 mM HEPES pH 7.5, 5 mM NaCl, 1 mM dithiothreitol (DTT), 5 µM [methyl-^{3}H] *S*-adenosyl methionine (PerkinElmer), 10 µM substrate RNA/DNA probe, and 1 µM co-purified full-length METTL3-METTL14 enzyme complex. The reactions were incubated at 37 ºC for 1 hour and 3 µL of each reaction was quenched by blotting on Hybond-N+ membrane (Amersham). RNA/DNA probes were crosslinked to membrane by exposure to ultraviolet light (254 nm) for 2 minutes. The membranes were successively washed three times by 1X PBS followed by three ethanol washes for 5 minutes each. The membranes were air dried in the hood for 15 minutes and the counts per minute (c.p.m.) of the RNA/DNA probes were measured using a scintillation counter (Beckman LS6500). We also performed this assay in a reaction buffer containing higher salt (50.0 mM NaCl) and observed consistent results (dose dependent inhibition of m^{6}dA activity of METTL3-METTL14 by small, structured RNA oligos) as presented in Figure 2e. Results presented here are an average of three independent experiments ($n = 3$) with one standard deviation (s.d.) for each oligonucleotide shown as error bars. Source data are provided as a Source Data file. The results of two groups in fig. 3c were derived from three independent experiments and were analyzed using two-tailed Student's T-test (unpaired).

Quantitation of methylation and validation of identity of methylated base by mass spectrometry

Methylation reactions were carried out in a 10 µL mixture containing 50 mM Tris-HCl, pH 7.5, 5 mM NaCl, 1 mM DTT, 200 µM SAM, 10 µM substrate RNA/DNA probe, and 1 µM METTL3-METTL14 enzyme complex. The reactions were incubated at 37 °C for 1 hour, followed by incubation with 0.8 units of proteinase K (New England Biolabs) at 37 °C for 10 minutes. Two

353 microliters of each reaction were subjected to oligonucleotide intact mass analysis by Liquid
354 Chromatography-Mass Spectrometry (LC-MS) as described previously[50]. The raw data was
355 deconvoluted using ProMass HR (Novatia, LCC). The deconvoluted-mass peak area ratio between
356 reactants and expected products was used to estimate the percentage of methylation. Results from
357 three independent experiments (n = 3) are shown. To verify the identity of the modified
358 nucleotides, METTL3-METTL14 reactions were also subjected to nucleoside analysis. Briefly, 8
359 µL of each of reaction were purified using Oligo Clean-Up and Concentration Kit (Norgen Biotek)
360 according to manufacturer's instructions. The nucleic acids were eluted in 20 µL nuclease-free
361 water. To seventeen microliters of the eluates were added 2 µL of 10x Nucleoside Digestion Mix
362 reaction buffer and 1 µL of Nucleoside Digestion Mix (New England Biolabs). Reactions were
363 incubated at 37 °C for 1 hour. The entirety of the nucleoside digestion reaction was subjected to
364 Ultra-High-Performance Liquid Chromatography (UHPLC) or UHPLC-MS analysis without
365 purification. UHPLC analysis was performed on an Agilent 1290 Infinity II system equipped with
366 a G7117A diode array detector (DAD). UHPLC-MS analysis was performed an Agilent 1290
367 Infinity II system equipped with a G7117A DAD detector and a 6135 XT mass detector. LC
368 separations were carried out using a Waters XSelect HSS T3 XP column (2.1 × 100 mm, 2.5 µm)
369 with the gradient mobile phase consisting of methanol and 10 mM ammonium acetate buffer (pH
370 4.5). The identity of each nucleoside was confirmed by analysis of the LC retention times relative
371 to authentic standards and by mass spectrometric analysis. The relative abundance of each
372 nucleoside was determined by UV absorbance.
373
374

Determining the nucleic acid-binding affinities of METTL3-METTL14 and METTL3-METTL14-RGG

Fluorescence polarization (FP)-based assay was used to measure equilibrium dissociation binding constants (K_d). The reactions were carried out in a buffer 0.01 M HEPES pH 7.5 and 0.05 M KCl. All oligonucleotides contain a fluorescein moiety covalently attached to the 5'-end. A constant concentration of the fluorescein labeled oligo (1 nM) was used with increasing concentrations of full-length METTL3-METTL14 protein or its truncated version (METTL3-METTL14$_{-RGG}$ (0 - 2000 nM) in a 384-well plate. Significant changes in fluorescence polarization upon increasing concentrations of protein were indicative of direct binding. The fluorescence polarization (emission wavelength = 530 nm, excitation wavelength = 485 nm) value for each dilution was measured using PHERAstar FS (BMG Labtech). The buffer corrected values were used to calculate the equilibrium dissociation constant (K_d) for protein-DNA/RNA binding using a simple 1:1 specific binding model (Y=Bmax*X/(Kd + X), where X denotes the concentration of protein ligand, Y = specific binding, Bmax= maximum binding, Kd = equilibrium dissociation constant in same units as X). Data were fitted by a single-site binding model in GraphPad Prism (GraphPad Software, San Diego, CA). Results presented here are an average of three independent experiments ($n = 3$) with one standard deviation (s.d.) for each oligonucleotide shown as error bars. Source data are provided as a Source Data file.

DATA AVAILABILITY

The information about coding sequences of human METTL3 (NCBI reference sequence GI: 33301371) and METTL14 (NCBI reference sequence GI: 172045930) used in this study is

available at NCBI. Source data are provided as a separate Source Data file. Correspondence and requests for material should be addressed to Y.K.G. (guptay@uthscsa.edu).

401 **REFERENCES**
402
403 [1] Desrosiers, R., Friderici, K., and Rottman, F. (1974) Identification of methylated nucleosides
404 in messenger RNA from Novikoff hepatoma cells, *Proc Natl Acad Sci U S A 71*, 3971-
405 3975.
406 [2] Bokar, J. A., Rath-Shambaugh, M. E., Ludwiczak, R., Narayan, P., and Rottman, F. (1994)
407 Characterization and partial purification of mRNA N6-adenosine methyltransferase from
408 HeLa cell nuclei. Internal mRNA methylation requires a multisubunit complex, *J Biol*
409 *Chem 269*, 17697-17704.
410 [3] Rottman, F. M., Bokar, J. A., Narayan, P., Shambaugh, M. E., and Ludwiczak, R. (1994) N6-
411 adenosine methylation in mRNA: substrate specificity and enzyme complexity,
412 *Biochimie 76*, 1109-1114.
413 [4] Dominissini, D., Moshitch-Moshkovitz, S., Schwartz, S., Salmon-Divon, M., Ungar, L.,
414 Osenberg, S., Cesarkas, K., Jacob-Hirsch, J., Amariglio, N., Kupiec, M., Sorek, R., and
415 Rechavi, G. (2012) Topology of the human and mouse m6A RNA methylomes revealed
416 by m6A-seq, *Nature 485*, 201-206.
417 [5] Meyer, K. D., Saletore, Y., Zumbo, P., Elemento, O., Mason, C. E., and Jaffrey, S. R. (2012)
418 Comprehensive analysis of mRNA methylation reveals enrichment in 3' UTRs and near
419 stop codons, *Cell 149*, 1635-1646.
420 [6] Ke, S., Pandya-Jones, A., Saito, Y., Fak, J. J., Vagbo, C. B., Geula, S., Hanna, J. H., Black,
421 D. L., Darnell, J. E., Jr., and Darnell, R. B. (2017) m(6)A mRNA modifications are
422 deposited in nascent pre-mRNA and are not required for splicing but do specify
423 cytoplasmic turnover, *Genes Dev 31*, 990-1006.
424 [7] Knuckles, P., Carl, S. H., Musheev, M., Niehrs, C., Wenger, A., and Buhler, M. (2017) RNA
425 fate determination through cotranscriptional adenosine methylation and microprocessor
426 binding, *Nat Struct Mol Biol 24*, 561-569.
427 [8] Liu, J., Dou, X., Chen, C., Chen, C., Liu, C., Xu, M. M., Zhao, S., Shen, B., Gao, Y., Han,
428 D., and He, C. (2020) N (6)-methyladenosine of chromosome-associated regulatory RNA
429 regulates chromatin state and transcription, *Science 367*, 580-586.
430 [9] Batista, P. J., Molinie, B., Wang, J., Qu, K., Zhang, J., Li, L., Bouley, D. M., Lujan, E.,
431 Haddad, B., Daneshvar, K., Carter, A. C., Flynn, R. A., Zhou, C., Lim, K. S., Dedon, P.,
432 Wernig, M., Mullen, A. C., Xing, Y., Giallourakis, C. C., and Chang, H. Y. (2014) m(6)A
433 RNA modification controls cell fate transition in mammalian embryonic stem cells, *Cell*
434 *Stem Cell 15*, 707-719.
435 [10] Geula, S., Moshitch-Moshkovitz, S., Dominissini, D., Mansour, A. A., Kol, N., Salmon-
436 Divon, M., Hershkovitz, V., Peer, E., Mor, N., Manor, Y. S., Ben-Haim, M. S., Eyal, E.,
437 Yunger, S., Pinto, Y., Jaitin, D. A., Viukov, S., Rais, Y., Krupalnik, V., Chomsky, E.,
438 Zerbib, M., Maza, I., Rechavi, Y., Massarwa, R., Hanna, S., Amit, I., Levanon, E. Y.,
439 Amariglio, N., Stern-Ginossar, N., Novershtern, N., Rechavi, G., and Hanna, J. H. (2015)
440 Stem cells. m6A mRNA methylation facilitates resolution of naive pluripotency toward
441 differentiation, *Science 347*, 1002-1006.
442 [11] Lin, S., Choe, J., Du, P., Triboulet, R., and Gregory, R. I. (2016) The m(6)A
443 Methyltransferase METTL3 Promotes Translation in Human Cancer Cells, *Mol Cell 62*,
444 335-345.
445 [12] Barbieri, I., Tzelepis, K., Pandolfini, L., Shi, J., Millan-Zambrano, G., Robson, S. C.,
446 Aspris, D., Migliori, V., Bannister, A. J., Han, N., De Braekeleer, E., Ponstingl, H.,

Hendrick, A., Vakoc, C. R., Vassiliou, G. S., and Kouzarides, T. (2017) Promoter-bound METTL3 maintains myeloid leukaemia by m(6)A-dependent translation control, *Nature 552*, 126-131.

[13] Vu, L. P., Pickering, B. F., Cheng, Y., Zaccara, S., Nguyen, D., Minuesa, G., Chou, T., Chow, A., Saletore, Y., MacKay, M., Schulman, J., Famulare, C., Patel, M., Klimek, V. M., Garrett-Bakelman, F. E., Melnick, A., Carroll, M., Mason, C. E., Jaffrey, S. R., and Kharas, M. G. (2017) The N6-methyladenosine (m6A)-forming enzyme METTL3 controls myeloid differentiation of normal hematopoietic and leukemia cells, *Nat Med*.

[14] Choe, J., Lin, S., Zhang, W., Liu, Q., Wang, L., Ramirez-Moya, J., Du, P., Kim, W., Tang, S., Sliz, P., Santisteban, P., George, R. E., Richards, W. G., Wong, K. K., Locker, N., Slack, F. J., and Gregory, R. I. (2018) mRNA circularization by METTL3-eIF3h enhances translation and promotes oncogenesis, *Nature 561*, 556-560.

[15] Xiang, Y., Laurent, B., Hsu, C. H., Nachtergaele, S., Lu, Z., Sheng, W., Xu, C., Chen, H., Ouyang, J., Wang, S., Ling, D., Hsu, P. H., Zou, L., Jambhekar, A., He, C., and Shi, Y. (2017) RNA m6A methylation regulates the ultraviolet-induced DNA damage response, *Nature 543*, 573-576.

[16] Zhang, C., Chen, L., Peng, D., Jiang, A., He, Y., Zeng, Y., Xie, C., Zhou, H., Luo, X., Liu, H., Chen, L., Ren, J., Wang, W., and Zhao, Y. (2020) METTL3 and N6-Methyladenosine Promote Homologous Recombination-Mediated Repair of DSBs by Modulating DNA-RNA Hybrid Accumulation, *Mol Cell 79*, 425-442 e427.

[17] Liu, J., Yue, Y., Han, D., Wang, X., Fu, Y., Zhang, L., Jia, G., Yu, M., Lu, Z., Deng, X., Dai, Q., Chen, W., and He, C. (2014) A METTL3-METTL14 complex mediates mammalian nuclear RNA N6-adenosine methylation, *Nat Chem Biol 10*, 93-95.

[18] 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.

[19] Woodcock, C. B., Horton, J. R., Zhang, X., Blumenthal, R. M., and Cheng, X. (2020) Beta class amino methyltransferases from bacteria to humans: evolution and structural consequences, *Nucleic Acids Res 48*, 10034-10044.

[20] Gupta, Y. K., Chan, S. H., Xu, S. Y., and Aggarwal, A. K. (2015) Structural basis of asymmetric DNA methylation and ATP-triggered long-range diffusion by EcoP15I, *Nat Commun 6*, 7363.

[21] Wang, X., Feng, J., Xue, Y., Guan, Z., Zhang, D., Liu, Z., Gong, Z., Wang, Q., Huang, J., Tang, C., Zou, T., and Yin, P. (2016) Structural basis of N(6)-adenosine methylation by the METTL3-METTL14 complex, *Nature 534*, 575-578.

[22] Sledz, P., and Jinek, M. (2016) Structural insights into the molecular mechanism of the m(6)A writer complex, *Elife 5*.

[23] Malone, T., Blumenthal, R. M., and Cheng, X. (1995) Structure-guided analysis reveals nine sequence motifs conserved among DNA amino-methyltransferases, and suggests a catalytic mechanism for these enzymes, *J Mol Biol 253*, 618-632.

[24] Linder, B., Grozhik, A. V., Olarerin-George, A. O., Meydan, C., Mason, C. E., and Jaffrey, S. R. (2015) Single-nucleotide-resolution mapping of m6A and m6Am throughout the transcriptome, *Nat Methods 12*, 767-772.

[25] Dubin, D. T., and Taylor, R. H. (1975) The methylation state of poly A-containing messenger RNA from cultured hamster cells, *Nucleic Acids Res 2*, 1653-1668.

[26] Slobodin, B., Han, R., Calderone, V., Vrielink, J., Loayza-Puch, F., Elkon, R., and Agami, R. (2017) Transcription Impacts the Efficiency of mRNA Translation via Co-transcriptional N6-adenosine Methylation, *Cell 169*, 326-337 e312.

[27] Huang, H., Weng, H., Zhou, K., Wu, T., Zhao, B. S., Sun, M., Chen, Z., Deng, X., Xiao, G., Auer, F., Klemm, L., Wu, H., Zuo, Z., Qin, X., Dong, Y., Zhou, Y., Qin, H., Tao, S., Du, J., Liu, J., Lu, Z., Yin, H., Mesquita, A., Yuan, C. L., Hu, Y. C., Sun, W., Su, R., Dong, L., Shen, C., Li, C., Qing, Y., Jiang, X., Wu, X., Sun, M., Guan, J. L., Qu, L., Wei, M., Muschen, M., Huang, G., He, C., Yang, J., and Chen, J. (2019) Histone H3 trimethylation at lysine 36 guides m(6)A RNA modification co-transcriptionally, *Nature 567*, 414-419.

[28] Patil, D. P., Chen, C. K., Pickering, B. F., Chow, A., Jackson, C., Guttman, M., and Jaffrey, S. R. (2016) m(6)A RNA methylation promotes XIST-mediated transcriptional repression, *Nature 537*, 369-373.

[29] Liu, N., Dai, Q., Zheng, G., He, C., Parisien, M., and Pan, T. (2015) N(6)-methyladenosine-dependent RNA structural switches regulate RNA-protein interactions, *Nature 518*, 560-564.

[30] Xiang, Y., Laurent, B., Hsu, C. H., Nachtergaele, S., Lu, Z., Sheng, W., Xu, C., Chen, H., Ouyang, J., Wang, S., Ling, D., Hsu, P. H., Zou, L., Jambhekar, A., He, C., and Shi, Y. (2017) RNA m(6)A methylation regulates the ultraviolet-induced DNA damage response, *Nature 543*, 573-576.

[31] Burgess, H. M., Depledge, D. P., Thompson, L., Srinivas, K. P., Grande, R. C., Vink, E. I., Abebe, J. S., Blackaby, W. P., Hendrick, A., Albertella, M. R., Kouzarides, T., Stapleford, K. A., Wilson, A. C., and Mohr, I. (2021) Targeting the m(6)A RNA modification pathway blocks SARS-CoV-2 and HCoV-OC43 replication, *Genes Dev 35*, 1005-1019.

[32] Li, N., Hui, H., Bray, B., Gonzalez, G. M., Zeller, M., Anderson, K. G., Knight, R., Smith, D., Wang, Y., Carlin, A. F., and Rana, T. M. (2021) METTL3 regulates viral m6A RNA modification and host cell innate immune responses during SARS-CoV-2 infection, *Cell Rep 35*, 109091.

[33] Woodcock, C. B., Yu, D., Hajian, T., Li, J., Huang, Y., Dai, N., Correa, I. R., Jr., Wu, T., Vedadi, M., Zhang, X., and Cheng, X. (2019) Human MettL3-MettL14 complex is a sequence-specific DNA adenine methyltransferase active on single-strand and unpaired DNA in vitro, *Cell Discov 5*, 63.

[34] Musheev, M. U., Baumgartner, A., Krebs, L., and Niehrs, C. (2020) The origin of genomic N(6)-methyl-deoxyadenosine in mammalian cells, *Nat Chem Biol 16*, 630-634.

[35] Wang, P., Doxtader, K. A., and Nam, Y. (2016) Structural Basis for Cooperative Function of Mettl3 and Mettl14 Methyltransferases, *Mol Cell 63*, 306-317.

[36] Krissinel, E., and Henrick, K. (2004) Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions, *Acta Crystallogr D Biol Crystallogr 60*, 2256-2268.

[37] Thandapani, P., O'Connor, T. R., Bailey, T. L., and Richard, S. (2013) Defining the RGG/RG motif, *Mol Cell 50*, 613-623.

[38] Scholler, E., Weichmann, F., Treiber, T., Ringle, S., Treiber, N., Flatley, A., Feederle, R., Bruckmann, A., and Meister, G. (2018) Interactions, localization, and phosphorylation of the m(6)A generating METTL3-METTL14-WTAP complex, *RNA 24*, 499-512.

[39] Oberstrass, F. C., Lee, A., Stefl, R., Janis, M., Chanfreau, G., and Allain, F. H. (2006) Shape-specific recognition in the structure of the Vts1p SAM domain with RNA, *Nat Struct Mol Biol 13*, 160-167.

[40] Alarcon, C. R., Lee, H., Goodarzi, H., Halberg, N., and Tavazoie, S. F. (2015) N6-methyladenosine marks primary microRNAs for processing, *Nature 519*, 482-485.

[41] Yang, L., Lin, C., Liu, W., Zhang, J., Ohgi, K. A., Grinstein, J. D., Dorrestein, P. C., and Rosenfeld, M. G. (2011) ncRNA- and Pc2 methylation-dependent gene relocation between nuclear structures mediates gene activation programs, *Cell 147*, 773-788.

[42] Garcia-Campos, M. A., Edelheit, S., Toth, U., Safra, M., Shachar, R., Viukov, S., Winkler, R., Nir, R., Lasman, L., Brandis, A., Hanna, J. H., Rossmanith, W., and Schwartz, S. (2019) Deciphering the "m(6)A Code" via Antibody-Independent Quantitative Profiling, *Cell 178*, 731-747 e716.

[43] Zhang, S., Zhao, B. S., Zhou, A., Lin, K., Zheng, S., Lu, Z., Chen, Y., Sulman, E. P., Xie, K., Bogler, O., Majumder, S., He, C., and Huang, S. (2017) m6A Demethylase ALKBH5 Maintains Tumorigenicity of Glioblastoma Stem-like Cells by Sustaining FOXM1 Expression and Cell Proliferation Program, *Cancer Cell 31*, 591-606 e596.

[44] Huang, J., Dong, X., Gong, Z., Qin, L. Y., Yang, S., Zhu, Y. L., Wang, X., Zhang, D., Zou, T., Yin, P., and Tang, C. (2018) Solution structure of the RNA recognition domain of METTL3-METTL14 N(6)-methyladenosine methyltransferase, *Protein Cell*.

[45] Abakir, A., Giles, T. C., Cristini, A., Foster, J. M., Dai, N., Starczak, M., Rubio-Roldan, A., Li, M., Eleftheriou, M., Crutchley, J., Flatt, L., Young, L., Gaffney, D. J., Denning, C., Dalhus, B., Emes, R. D., Gackowski, D., Correa, I. R., Jr., Garcia-Perez, J. L., Klungland, A., Gromak, N., and Ruzov, A. (2020) N(6)-methyladenosine regulates the stability of RNA:DNA hybrids in human cells, *Nat Genet 52*, 48-55.

[46] 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.

[47] Schwartz, S., Mumbach, M. R., Jovanovic, M., Wang, T., Maciag, K., Bushkin, G. G., Mertins, P., Ter-Ovanesyan, D., Habib, N., Cacchiarelli, D., Sanjana, N. E., Freinkman, E., Pacold, M. E., Satija, R., Mikkelsen, T. S., Hacohen, N., Zhang, F., Carr, S. A., Lander, E. S., and Regev, A. (2014) Perturbation of m6A writers reveals two distinct classes of mRNA methylation at internal and 5' sites, *Cell Rep 8*, 284-296.

[48] Douvlataniotis, K., Bensberg, M., Lentini, A., Gylemo, B., and Nestor, C. E. (2020) No evidence for DNA N (6)-methyladenine in mammals, *Sci Adv 6*, eaay3335.

[49] Quinodoz, S. A., Bhat, P., Ollikainen, N., Jachowicz, J. W., Banerjee, A. K., Chovanec, P., Blanco, M. R., Chow, A., Markaki, Y., Plath, K., and Guttman, M. (2020) RNA promotes the formation of spatial compartments in the nucleus, *bioRxiv*, 2020.2008.2025.267435.

[50] Viswanathan, T., Arya, S., Chan, S. H., Qi, S., Dai, N., Misra, A., Park, J. G., Oladunni, F., Kovalskyy, D., Hromas, R. A., Martinez-Sobrido, L., and Gupta, Y. K. (2020) Structural basis of RNA cap modification by SARS-CoV-2, *Nat Commun 11*, 3718.

[51] Parisien, M., and Major, F. (2008) The MC-Fold and MC-Sym pipeline infers RNA structure from sequence data, *Nature 452*, 51-55.

ACKNOWLEDGMENTS

This work was partially supported by funding from the Max and Minnie Tomerlin Voelcker Foundation, San Antonio Partnership for Precision Therapeutics, IIMS/CTSA pilot award, and laboratory startup funds from the Greehey Children's Cancer Research Institute (GCCRI) of UT Health San Antonio to Y.K.G. S.Q. is supported by the GCCRI. Y.K.G. is a recipient of a high impact/high risk award (RP190534) from the Cancer Prevention and Research Institute of Texas (CPRIT), and a Rising STARs award from the University of Texas System. Y.K.G and M.K.R are supported by CPRIT grant (RP200110). Y.K.G is grateful for the support from NIH grant 1R01AI161363. S-H.C., N.D. and I.R.C. are grateful to Donald Comb, Jim Ellard and Rich Roberts for their support of research at New England Biolabs.

AUTHOR CONTRIBUTIONS

Y.K.G. conceived, designed, and supervised the overall study, and analyzed data; S.Q., J.M. expressed, and purified proteins, performed biochemical and enzymatic assays, and analyzed data; S-H.C., N.D., I.R.C. designed and performed LC/MS-based assays, and analyzed data; J.V., and S.A. optimized conditions of protein expression, and purification, and prepared oligonucleotides for biochemical assays. R.A.H., M.K.R. provided resources/reagents. Y.K.G. wrote the manuscript and all authors read and approved this version.

COMPETING INTERESTS

Y.K.G. is founder of Atomic Therapeutics. S-H. C., N.D., and I.R.C. are employees of New England Biolabs, a manufacturer and vendor of molecular biology reagents. None of these

603 affiliations affect the authors' impartiality, adherence to journal standards and policies, or
604 availability of data.

605

606 **MATERIALS AND CORRESPONDENCE**

607 Correspondence and requests for material should be addressed to Y.K.G. (Email:
608 guptay@uthscsa.edu).

609

FIGURE LEGENDS

Figure 1. Structural similarity, purification of methyltransferases, and substrate designing. ***a-b.*** Domain architecture of Mod subunit of EcoP15I, human METTL3, and human METTL14 methyltransferases (MTases). All three members belong to the β-class of SAM-dependent MTases and exhibit a sequential arrangement of motifs (IV-X followed by motif I)[18, 19, 23]. Motif IV (D/EPPY/W/L) and I are associated with the recognition of target adenine base[20] and co-factor (SAM) binding[21, 22, 35], respectively. ***c.*** Crystal structure of EcoP15I-DNA complex (PDB ID: 4ZCF)[20]: Mod_A MTase (cyan), Mod_B MTase (blue), non-methylating DNA strand (grey), target (methylating) DNA strand (orange), flipped adenine base (red stick). The methyltransferase cores of Mod_B and Mod_A are shown in dark blue and light pink, respectively. The Res subunit of EcoP15I was omitted for clarity. ***d.*** MTase domains of two Mod subunits (ModA/B) of EcoP15I with target DNA strand. The regions flanking the methyltransferase core (e.g., CTD and TRD) are omitted for clarity. Only the region encompassing the MTase core (aa 90-132, 169-261, and 385-511) of EcoP15I Mod was selected for the alignment with the methyltransferase core of METTL3 (aa 358-580) and METTL14 (aa 165-378). ***e.*** An overlay of MTase domains of EcoP15I and METTL14 (PDB ID: 5IL0) shows structural similarity within MTase folds (rmsd = 2.55Å over 278 Cα atoms). RNA strand here was modeled based on the respective methylating strand in the EcoP15I structure. ***f.*** Chromatogram of final size exclusion chromatography (SEC) step of purification showing METTL3-METTL14 complexes (left, full-length; right, METL3-$METTL14_{[-RGG]}$) co-eluted as single homogenous species. Blue, absorbance at 280 nm; green, absorbance at 260 nm (A260). Coomassie stained gels (lower panels) confirm high purity of METTL3-METTL14 proteins in the SEC peak fractions. ***g.*** Sequence of DNA and RNA oligonucleotides used in this study. All oligos have a covalently attached 5'-fluorescein (not shown). ***h.*** Secondary structure of rNEAT2 and its

633 3-D model as predicted by MC-SYM[51]. rNEAT2 harbors a potential DRACH motif (yellow line).
634 ***i.*** Secondary structure of rTCE23 RNA and its solution NMR structure (PDB ID: 2ES5)[39].
635
636 **Figure 2. RNA-mediated restriction of METTL3-METTL14 activity**
637 ***a.*** Methyltransferase activity of METTL3-METTL14 in the presence of various DNA or RNA
638 substrates measured by radiometric assay. CPM, counts per minute. The highest activity was
639 measured with the d6T* oligo. ***b.*** FP-based binding assay for DNA and RNA oligos showing
640 highest affinity for rNEAT2 RNA (green), and lowest affinity for d6T* oligo (red). Equilibrium
641 dissociation constants (Kd) for each oligo are shown on the right side of the isotherms. The data
642 were fit into one site specific binding model (Y=Bmax*X/(Kd + X). See methods section and
643 source data for details. ***c.*** Methyltransferase activity of METTL3-METTL14 on the respective
644 RNA oligos (red), d6T* DNA alone (black) and equimolar mixture of the two (blue), measured by
645 radiometric assay. ***d.*** predicted secondary structures of each oligonucleotide. yellow, RNA strand;
646 black, DNA strand. The values of the equilibrium dissociation constants (Kd) shown for each
647 oligonucleotide indicate an inverse relationship between binding affinity and methyltransferase
648 activity. ***e.*** dose-dependent inhibition of METTL3-METTL14 activity by RNA oligos rNEAT2 or
649 rTCE23, as measured by radiometric assay in a reaction buffer containing 5.0 mM NaCl (upper
650 panel) and 50.0 mM NaCl (lower panel). IC_{50}, concentration of RNA required to achieve 50%
651 inhibition of the METTL3-METTL14 activity. ***f.*** Attenuation of the methyltransferase activity in
652 presence of rNEAT2 or rTCE23, as measured by oligonucleotide intact mass analysis. Quantitation
653 of modified dA (black circle) or rA (red square) is shown in absence or presence of equivalent
654 amounts of rNEAT2 or rTCE23. ***g.*** Nucleoside composition analysis of the METTL3-METTL14
655 reactions. UHPLC chromatograms showing the reaction in absence (blue trace) or in the presence

of either rNEAT2 (red trace) or rTCE23 (green trace). The quantitation of the fraction of modified bases in the nucleoside pool was consistent with the results from the oligonucleotide intact mass analysis shown in ***f***. The insert shows the full chromatographic trace with all detected nucleosides. Results presented in panels ***a-c*** and ***e-f*** are the average of three independent experiments ($n = 3$) with one standard deviation (s.d.) for each oligonucleotide (shown as error bars). Source data are provided as a Source Data file.

Figure 3. Role of RGG motifs and model of RNA-mediated regulation of methyltransferase activity

a. Domain architecture of METTL3 and METTL14. LH, leader helix; NLS, nuclear localization signal; ZnF1/2, zinc-finger domain 1/2; RGG, arginine-glycine rich repeats motif. ***b.*** FP-based binding assay for DNA and RNA oligos showing the highest affinity of METTL3-METTL14$_{(-RGG)}$ for rNEAT2 RNA (green) and lowest affinity for d6T* (red). Equilibrium dissociation constants (Kd) for each oligo are shown. The data were fit into one site specific binding model (Y=Bmax*X/(Kd + X). See methods section and source data for details. ***c.*** Relative methyltransferase activity of full-length METTL3-METTL14 and the truncated enzyme devoid of the RGG motif in METTL14 ([METTL3-METTL14$_{(-RGG)}$]) in the presence of d6T*, rNEAT2, or an equimolar mixture of these two oligos, as measured by radiometric assay. Results presented are the average of three independent experiments ($n = 3$) with one standard deviation (s.d.) for each oligonucleotide (shown as error bars). The results of two groups were analyzed and compared using two-tailed Student's unpaired T-test (P value <0.0001). Details about Student's T-test are provided in the source data file. ***d, e.*** Proposed models showing that the METTL3-METTL14 complex can methylate a target adenine (yellow) in a single-stranded DNA region (black).

679 Structured motifs present in ncRNA/mRNAs (orange) can block the methyltransferase activity by

680 a shape-dependent binding of these RNAs to METTL3-METTL14.

681

682

683 **table 1**

684

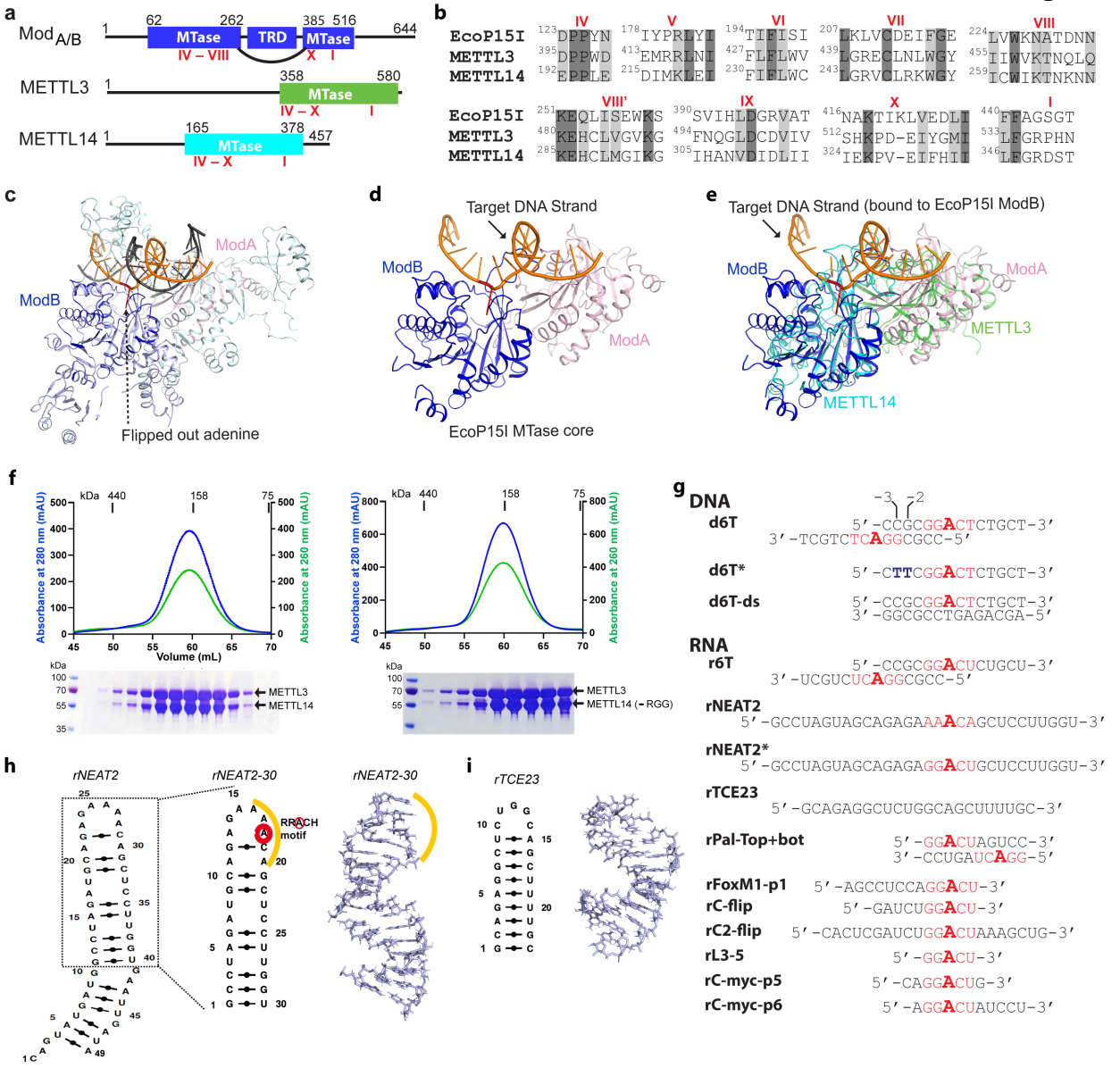
Nucleic acid substrate	Kd in nM (range of Kd in nM)	
	full-length METTL3-METTL14	METTL3-METTL14$_{-RGG}$
RNA		
rNEAT2	2.1 (1.8 – 2.3)	21.2 (19.2 –23.5)
rNEAT2*	13.0 (12.3 – 13.8)	274.1 (150.7 –497.2)
rTCE23	25.6 (23.2 – 28.5)	413 (341 – 502)
rPal-top+bot	127 (97.2 – 167)	257 (132 – 491)
rFoxM1-p1	294 (245 – 354)	920 (480 – 1940)
r6T	114 (103 – 127)	704 (323 – 1668)
DNA		
d6T	370 (343 – 401)	2097 (1494 – >2000)
d6T*	509 (470 – 553)	>2000

685

686

687

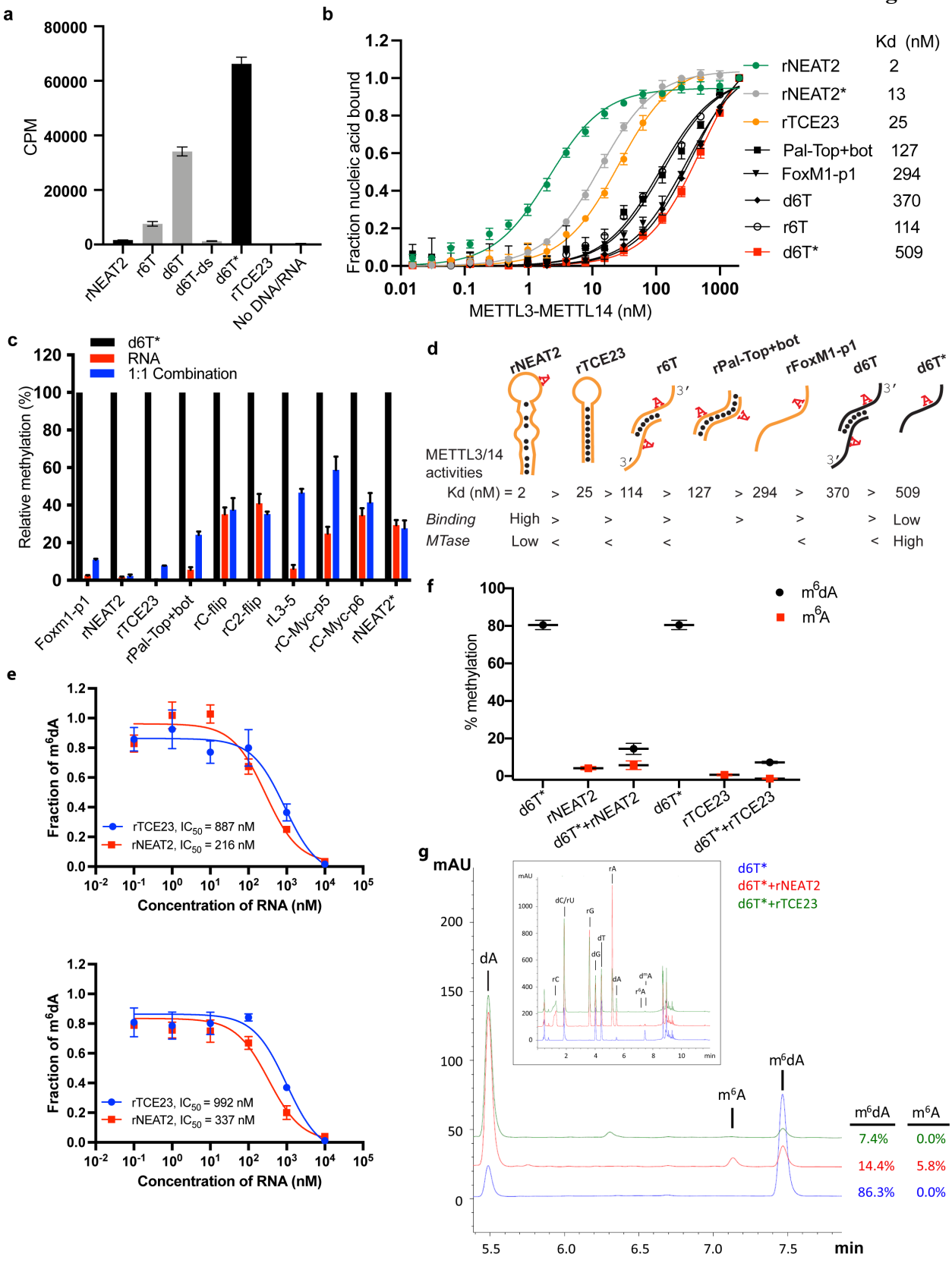
figure 1



688
689

690

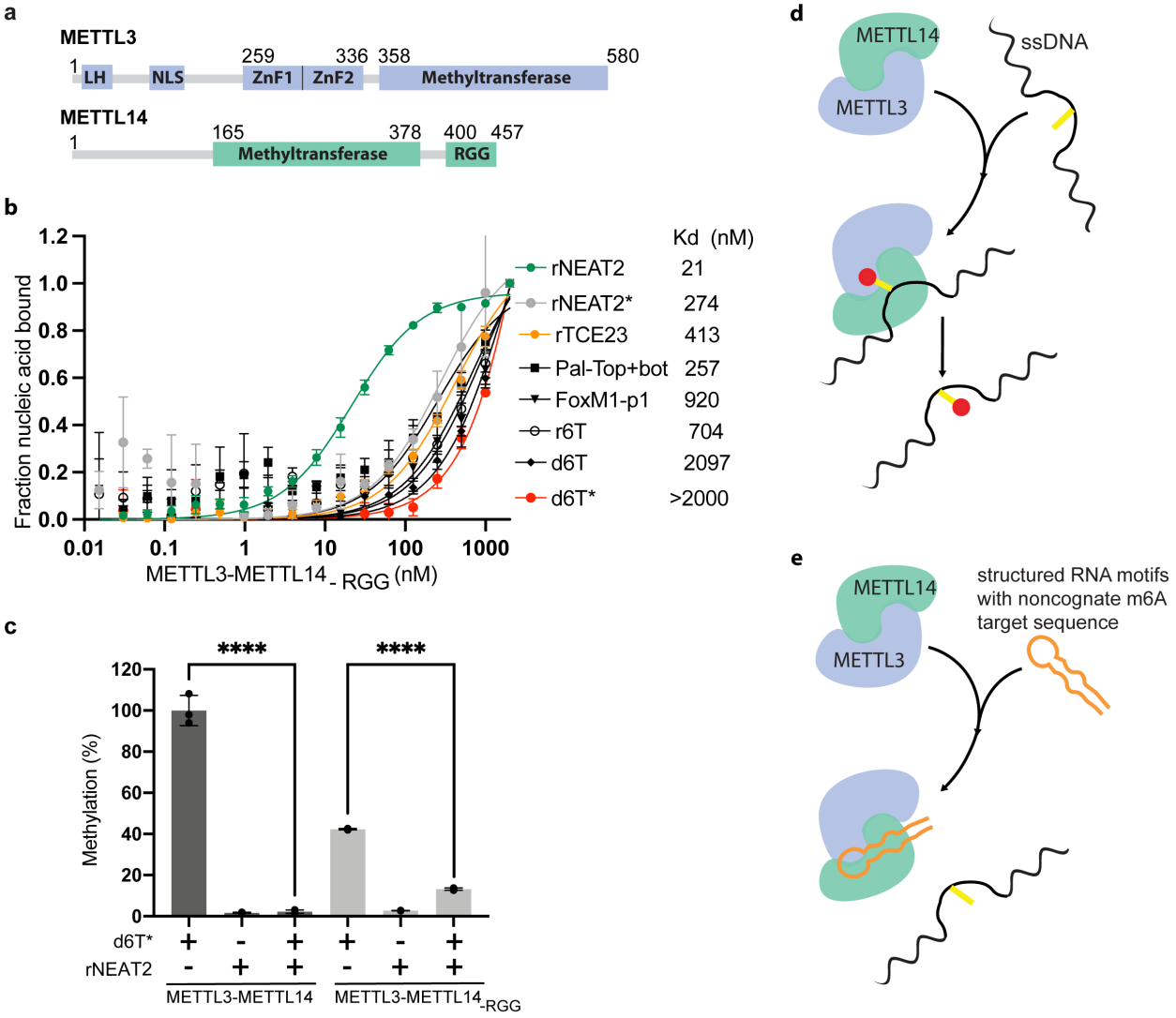
figure 2



691
692

693

figure 3



694

695 **Supplementary information**

696

697 **RNA binding to human METTL3-METTL14 restricts N^6-deoxyadenosine methylation of**

698 **DNA *in vitro***

699

700 Shan Qi[1,2,#], Javier Mota[1,#], Siu-Hong Chan[3], Johanna Villarreal[1], Nan Dai[3], Shailee Arya[1],

701 Robert A. Hromas[4], Manjeet K. Rao[1], Ivan R. Corrêa Jr.[3], Yogesh K. Gupta[1,2,5*]

702

703 [1] Greehey Children's Cancer Research Institute, University of Texas Health at San Antonio

704 8403 Floyd Curl Drive, San Antonio, TX 78229, USA

705

706 [2] Department of Biochemistry and Structural Biology, University of Texas Health at San Antonio

707 7703 Floyd Curl Drive, San Antonio, TX 78229, USA

708

709 [3] New England Biolabs, 240 County Road, Ipswich, MA, 01938, USA

710

711 [4] Division of Hematology and Medical Oncology, Department of Medicine, University of Texas

712 Health at San Antonio

713 7703 Floyd Curl Drive, San Antonio, TX 78229, USA

714

715 [5] Lead contact

716 [*]Corresponding author:

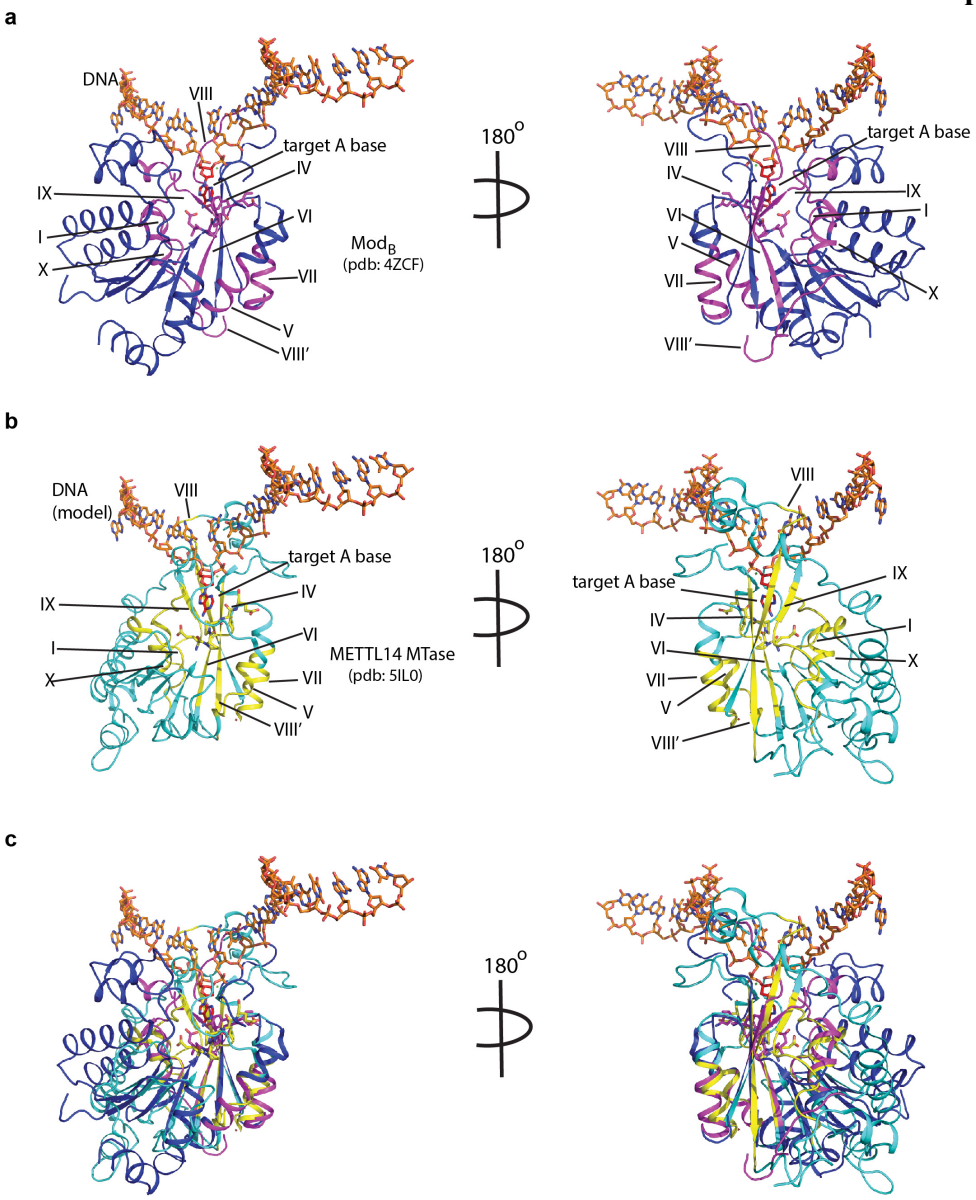
717 Y.K.G email: guptay@uthscsa.edu

718

719 [#] These authors contributed equally to this work

720

supplementary figure 1



S. Figure 1. Mode of DNA recognition. ***a.*** *Blue ribbons,* the MTase core (Mod$_B$) of EcoP15I (aa 90-132, 169-261, 385-511 from PDB: 4ZCF) in complex with the target DNA strand (orange sticks). The conserved motifs in class β MTases are shown in magenta. The motif IV (D/EPPY/W/L) that surrounds the watson-edge edge of the flipped target adenine base (red) is shown in stick mode. ***b.*** *Cyan ribbons,* MTase core of human METTL14 (PDB: 5IL0) with a modeled ssDNA as in Mod$_B$. The conserved motifs are colored in yellow. ***c.*** An overlay of MTases of Mod$_B$ and METTL14 was performed using the SSM method (secondary structure-based) that yielded an rmsd of 2.55 over 282aa of METTL14 aligned with 278aa of Mod$_B$. The canonical MTase motifs, including motif IV, in two MTases overlay well. Due to lack of structures of METTL3 or Mod$_A$ (in complex with a flipped adenine base), we chose Mod$_B$ and METTL14 for this superposition. We could not model RGG motifs due to the lack of predicted structure for this region.