

# Structural insights into molecular mechanism for *N*<sup>6</sup>-adenosine methylation by MT-A70 family methyltransferase METTL4

Received: 1 September 2021

Accepted: 12 September 2022

Published online: 26 September 2022

Check for updates

Qiang Luo<sup>1,2,8</sup>, Jiezhen Mo<sup>3,8</sup>, Hao Chen<sup>4,8</sup>, Zetao Hu<sup>1</sup>, Baihui Wang<sup>2</sup>, Jiabing Wu<sup>2</sup>, Ziyu Liang<sup>3</sup>, Wenhao Xie<sup>2</sup>, Kangxi Du<sup>2</sup>, Maolin Peng<sup>2</sup>, Yingping Li<sup>2</sup>, Tianyang Li<sup>2</sup>, Yangyi Zhang<sup>4</sup>, Xiaoyan Shi<sup>4</sup>, Wen-Hui Shen<sup>2,5</sup>, Yang Shi<sup>6</sup>✉, Aiwu Dong<sup>2</sup>✉, Hailin Wang<sup>3,7</sup>✉ & Jinbiao Ma<sup>1</sup>✉

METTL4 belongs to a subclade of MT-A70 family members of methyltransferase (MTase) proteins shown to mediate *N*<sup>6</sup>-adenosine methylation for both RNA and DNA in diverse eukaryotes. Here, we report that Arabidopsis METTL4 functions as U2 snRNA MTase for *N*<sup>6</sup>-2'-O-dimethyladenosine (m<sup>6</sup>Am) in vivo that regulates flowering time, and specifically catalyzes *N*<sup>6</sup>-methylation of 2'-O-methyladenosine (Am) within a single-stranded RNA in vitro. The apo structures of full-length Arabidopsis METTL4 bound to S-adenosyl-L-methionine (SAM) and the complex structure with an Am-containing RNA substrate, combined with mutagenesis and in vitro enzymatic assays, uncover a preformed L-shaped, positively-charged cavity surrounded by four loops for substrate binding and a catalytic center composed of conserved residues for specific Am nucleotide recognition and *N*<sup>6</sup>-methylation activity. Structural comparison of METTL4 with the mRNA m<sup>6</sup>A enzyme METTL3/METTL14 heterodimer and modeling analysis suggest a catalytic mechanism for *N*<sup>6</sup>-adenosine methylation by METTL4, which may be shared among MT-A70 family members.

Enzymatic RNA covalent modifications represent an essential epigenetic mechanism that exquisitely and plastically regulates multiple cellular activities in eukaryotes<sup>1</sup>. One of the most abundant modifications is the methylation decorated on the *N*<sup>6</sup> of adenosine (m<sup>6</sup>A), which is widespread in mRNA<sup>2,3</sup>, small nuclear RNA (snRNA)<sup>4</sup>, long noncoding RNA (lncRNA)<sup>5,6</sup>, and ribosomal RNA (rRNA)<sup>7</sup>. The m<sup>6</sup>A decoration of these RNA molecules regulates RNA structure or modulates protein-RNA interaction<sup>8,9</sup>, impacting RNA metabolism and diverse signaling

pathways essential for cell survival and differentiation<sup>10,11</sup>. Abnormal m<sup>6</sup>A-based RNA metabolism has been suggested to cause human diseases including obesity<sup>12</sup> and cancers<sup>13,14</sup>. Recent studies have identified the enzymes that mediate m<sup>6</sup>A modification on diverse RNAs, including the METTL3/METTL14 heterodimer<sup>15</sup>, METTL16<sup>16</sup>, ZCCHC4<sup>17</sup>, and METTL5<sup>18</sup> with exquisite sequence specificity. METTL3/METTL14 primarily catalyzes the modification of mRNA and lncRNA nearby the stop code and in the 3'UTR by recognizing a DRACH (D: A, G, U; R: G, A; H: A,

<sup>1</sup>State Key Laboratory of Genetic Engineering, Collaborative Innovation Center of Genetics and Development, Department of Biochemistry and Biophysics, Institute of Plant Biology, School of Life Sciences, Fudan University, 200438 Shanghai, China. <sup>2</sup>State Key Laboratory of Genetic Engineering, Collaborative Innovation Center of Genetics and Development, International Associated Laboratory of CNRS-Fudan-HUNAU on Plant Epigenome Research, Department of Biochemistry and Biophysics, Institute of Plant Biology, School of Life Sciences, Fudan University, 200438 Shanghai, China. <sup>3</sup>State Key Laboratory of Environmental Chemistry and Ecotoxicology, Research Center for Eco-Environmental Sciences, Chinese Academy of Sciences, University of Chinese Academy of Sciences, 100085 Beijing, China. <sup>4</sup>Department of Human Cell Biology and Genetics, School of Medicine, Southern University of Science and Technology, 518055 Shenzhen, Guangdong, China. <sup>5</sup>Institut de Biologie Moléculaire des Plantes, CNRS, Université de Strasbourg, 12 rue du Général Zimmer, 67084 Strasbourg Cédex, France. <sup>6</sup>Ludwig Institute for Cancer Research, Oxford University, Old Road Campus Research Building, Roosevelt Dr, Headington, Oxford OX3 7DQ, UK. <sup>7</sup>Institute of Environment and Health, Jiangnan University, 430056 Wuhan, Hubei, China. <sup>8</sup>These authors contributed equally: Qiang Luo, Jiezhen Mo, Hao Chen. ✉e-mail: [yang.shi@ludwig.ox.ac.uk](mailto:yang.shi@ludwig.ox.ac.uk); [aiwudong@fudan.edu.cn](mailto:aiwudong@fudan.edu.cn); [hlwang@rcees.ac.cn](mailto:hlwang@rcees.ac.cn); [majb@fudan.edu.cn](mailto:majb@fudan.edu.cn)

C, U) consensus sequence<sup>19–21</sup>. Some mRNAs and snRNAs with an UACAGAGAA motif in the stem-loop region are decorated by METTL16<sup>16</sup>.

METTL4, which belongs to a subclade of the MT-A70 family of proteins but being separated from the METTL3 and METTL14 subclades<sup>22</sup>, has been reported to be a U2 snRNA *N*<sup>6</sup>-adenosine methyltransferase in human and *Drosophila*<sup>23–25</sup>. Of note, while human METTL4 only methylates U2 snRNA on A30 when the 2'-OH on the ribose ring is methylated<sup>23,24</sup>, *Drosophila* METTL4 does not have this requisite and can methylate the same adenosine in U2 snRNA without 2'-O methylation<sup>25</sup>. Interestingly, METTL4 or its homologs have also been shown to deposit adenine *N*<sup>6</sup> methylation in DNA (*m*<sup>6</sup>A) in the genome of mammalian tissues<sup>26–28</sup> or mediate DNA *m*<sup>6</sup>A methylation in mitochondria<sup>29</sup>. However, the exact substrate of METTL4 is still under debate in part because of the lack of structural evidence for any proposed METTL4 enzymatic activities discussed above. In fact, only apo structures of MT-A70 family homolog proteins, METTL3/METTL14 heterodimer in complex with cofactors were reported previously<sup>15,30,31</sup>.

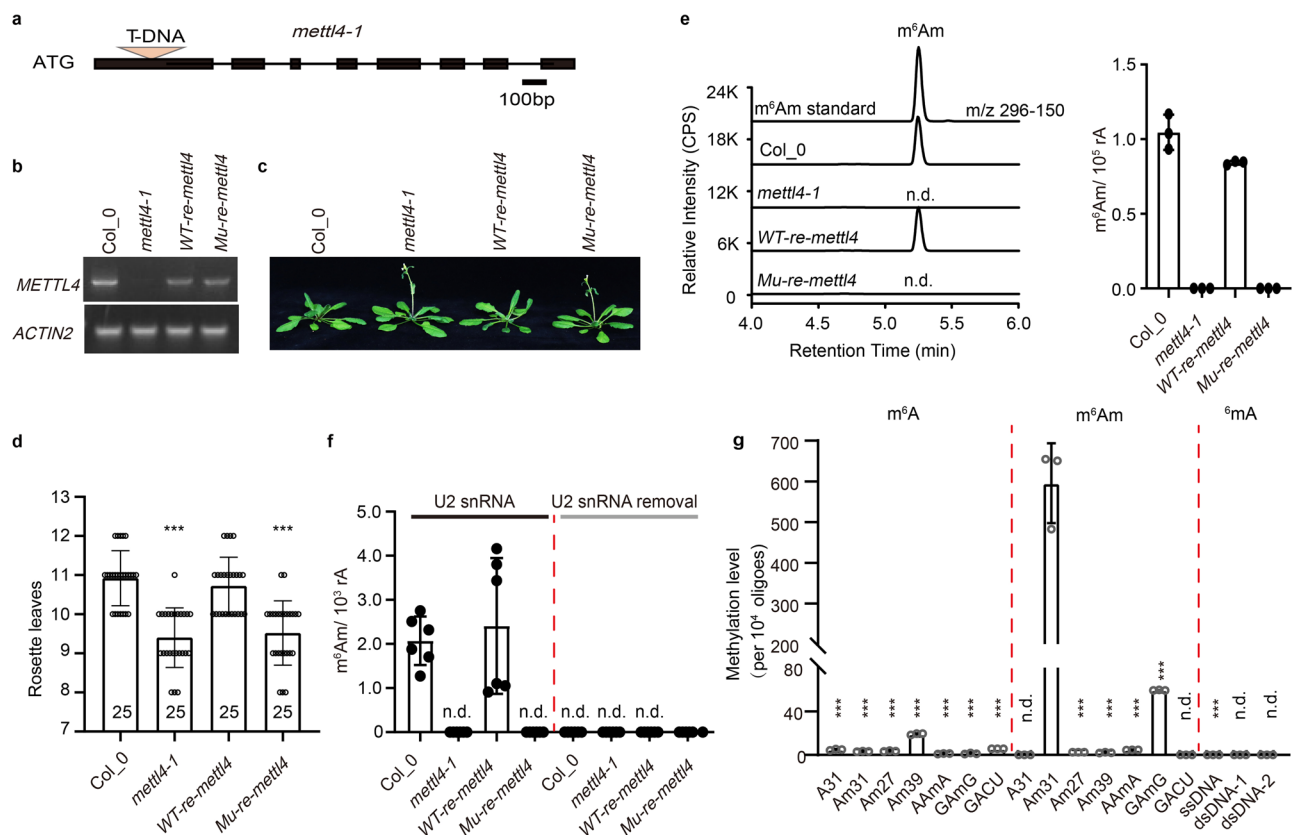
Here we show that *Arabidopsis* METTL4 functions as U2 snRNA MTase for *N*<sup>6</sup>-2'-O-dimethyladenosine (*m*<sup>6</sup>Am) in vivo, and specifically catalyzes *N*<sup>6</sup>-methylation of 2'-O-methyladenosine (Am) within a single-stranded RNA in vitro. We determined crystal structures of full-length METTL4 protein and its complexes with S-adenosyl-L-methionine (SAM) or its analogs, and co-crystal structure of METTL4 bound an Am-containing RNA substrate in the presence of S-adenosyl-L-

homocysteine (SAH). In combination with biochemical and modeling analysis, our findings shed light on the specificity of Am recognition by METTL4 and a novel molecular mechanism for *N*<sup>6</sup>-adenosine methylation, which is likely to be shared by MT-A70 family MTases, such as METTL3/METTL14.

## Results

### In vivo functions of METTL4 in *Arabidopsis*

Based on phylogenetic analysis<sup>32</sup>, we identified At1G19340 in *Arabidopsis thaliana* as the mammalian METTL4 ortholog and systematically investigated its potential substrates in vitro and in vivo. We firstly confirmed a T-DNA insertion mutant line, *mettl4-1*, in which the T-DNA was inserted into the first exon of the *METTL4* gene thus resulting in the disruption of the full-length *METTL4* transcript (Fig. 1a, b). This *mettl4-1* mutant displayed an early flowering phenotype under long-day (LD) conditions compared with the wild-type Col-0 (Fig. 1c). The early flowering phenotype of *mettl4-1* mutant was rescued by re-introduction of a wild type copy of the *METTL4* gene but not a catalytic inactive mutant, in which the conserved DPPW motif critical for catalysis was mutated to APPA (Fig. 1c, d), indicating that the developing defect of *mettl4-1* early flowering phenotype is due to the loss of the catalytic activity of METTL4. Next, we investigated the potential substrates of *Arabidopsis* METTL4 using ultra-high-performance liquid chromatography-triple quadrupole mass spectrometry, coupled with multiple-reaction monitoring (UHPLC-MRM-MS/MS). We measured four RNA modifications, including *N*<sup>6</sup>-methyladenosine (*m*<sup>6</sup>A), 2'-O-methyladenosine (Am), *N*<sup>6</sup>-



**Fig. 1 | METTL4 mediates solely *m*<sup>6</sup>Am methylation of U2 snRNA in *Arabidopsis thaliana*.** **a** The diagrammatic sketch of the T-DNA insertion site in the *METTL4* gene. **b** RT-PCR validation of the full-length transcript of the *METTL4* gene, the *ACTIN2* transcript acts as a control. **c** The phenotype of the *Arabidopsis* plants grown in long-day (LD) condition (22 °C and 16-h-light/8-h-dark). **d** Statistics of the number of rosette leaves when flower buds first appeared, which correlates with flowering time. And the number of the calculated plants were shown in the bar column. The APPA mutant was generated from conserved DPPW motif. Wild-type

*METTL4* and APPA mutant gene rescued *mettl4-1* plants are shown as WT-re-mettl4 and Mu-re-mettl4, respectively. \*\*\*Significant differences between WT and mutant (Student's *t* test: \*\*\**P* < 0.001). **e** UPLC-MS/MS chromatograms (left) and quantification of *m*<sup>6</sup>Am abundance in total RNA from *Arabidopsis thaliana* (right). The error bars represent standard deviations. **f** The quantification of *m*<sup>6</sup>Am abundance in U2 snRNA and residual RNAs. **g** Enzymatic assay of METTL4 using synthetic U2 snRNA fragments and various RNA and DNA oligoes as substrates in vitro. The error bars represent standard deviations. n.d., not detectable; *p* < 0.001.

methyladenosine (m<sup>6</sup>A), and N<sup>6</sup>-2'-O-dimethyladenosine (m<sup>6</sup>Am), in total RNA extracted from Col-0, *mettl4-1* mutant and the two rescued plants (Supplementary Fig. 1a–f and Fig. 1e). The quantification of m<sup>1</sup>A, Am and m<sup>6</sup>A showed similar levels in wild-type and *mettl4-1* mutant plants, suggesting that these modifications were not affected by the loss of METTL4. However, the m<sup>6</sup>Am level was significantly compromised in the *mettl4-1* mutant plant but restored by the reintroduction of a wild-type copy of *METTL4*. Importantly, the m<sup>6</sup>Am level of U2 snRNA was barely detectable in the *mettl4-1* mutant or in the catalytic mutant *METTL4 APPA* rescued plant, but the level was restored in the wild-type *METTL4* rescued plant (Fig. 1f). Moreover, when the RNA extracts were deprived of U2 snRNA, no m<sup>6</sup>Am was detected (Fig. 1f), suggesting that METTL4-dependent internal m<sup>6</sup>Am modification is likely only present in U2 snRNA in Arabidopsis. Taken together, these results suggested that METTL4 is the sole m<sup>6</sup>Am RNA methyltransferase in Arabidopsis and is exclusively responsible for U2 snRNA m<sup>6</sup>Am modification in vivo.

To address why depletion of *METTL4* (Supplementary Fig. 2a) resulted in an early flowering phenotype (Fig. 1c), we performed RNA sequencing to investigate the molecular function of *METTL4* in Arabidopsis (Supplementary Fig. 2a). Compared with the wild-type Col-0, there are 127 up-regulated genes and 119 down-regulated genes in the *mettl4* mutant, respectively (Supplementary Fig. 2b). Gene Ontology (GO) analysis showed that a set of genes involved in photosynthesis and response to cold or freezing are among these 127 up-regulated and 119 down-regulated genes, which are related to flowering time regulation (Supplementary Fig. 2c). Interestingly, many genes (497) displayed splicing changes in the *mettl4* mutant compared to the wild-type Col-0 (Supplementary Fig. 2d), which are mainly involved in the basic RNA biosynthetic and metabolic processes (Supplementary Fig. 2e). Although there is no direct association between early flowering phenotype and the genes with obvious splicing alterations, it is plausible that Arabidopsis METTL4 affected pre-messenger RNA (pre-mRNA) splicing of a set of these genes, similar to the function of mammalian METTL4<sup>23,24</sup>.

### Substrate features of the Arabidopsis METTL4 in vitro

To further clarify the substrate specificities of the Arabidopsis METTL4, we performed in vitro enzymatic activity analysis using full-length recombinant METTL4 protein purified from *E. coli*. Unlike the human U2 snRNA, which has only one Am decorated site (Am30), the Arabidopsis U2 snRNA contains two Am sites (Am31 and Am39)<sup>33</sup>. Therefore, we designed three RNA substrates based on the Arabidopsis U2 snRNA sequences: Am31 for 2'-OH and 2'-OMe at site 31, Am39 for 2'-OMe at site 39, and Am27 for 2'-OMe at site 27 as a control. Interestingly, only Am31, but not Am39 and Am27, could be methylated by METTL4 in vitro (Fig. 1g and Supplementary Fig. 3a, b). Changing the sequence around the Am31 site from AAmG to GAmG reduced the in vitro activities of METTL4, and the enzyme activity is almost completely lost when the wild-type sequence was changed to AAmA (Fig. 1g and Supplementary Fig. 3a, b). These results suggest that sequences or secondary structures around the Am nucleotide may be critical for the enzyme activity. Next, we tested whether the 2'-O-methyl modification is necessary for the enzyme activity of METTL4 in vitro. METTL4 showed weak m<sup>6</sup>A enzyme activities towards probes Am39 and GACU with two 5'-GGACU-3' motifs (optimal sequence for METTL3/METTL14 heterodimer<sup>34</sup>), and negligible activities towards most other probes, in comparison to the high m<sup>6</sup>Am enzyme activity to Probe Am31 (Fig. 1g and Supplementary Fig. 3a–c), indicating that METTL4 prefers Am over A in vitro. In addition, the quantification of m<sup>6</sup>A modification on U2 snRNA and total RNA showed similar levels in wild-type and *mettl4-1* mutant plants (Supplementary Fig. 1g), suggesting that METTL4 decorates only on Am31 in U2 snRNA but not on A or Am at other positions of U2 snRNA or other RNAs in vivo. These results strongly suggest that Arabidopsis METTL4 has stringent specificities for the sequences and secondary structures surrounding

Am31 nucleotide in the Arabidopsis U2 snRNA. We also measured Arabidopsis METTL4 activities on ssDNA and dsDNA. Compared to the RNA substrates, only a negligible level of activity was observed when a single strand DNA (ssDNA) probe containing the identical sequence of Probe Am31 was used as a substrate, and no detectable DNA m<sup>6</sup>A signal was observed on double-stranded DNA (dsDNA) substrates (Fig. 1g and Supplementary Fig. 3a, b). Taken together, these results suggest that Arabidopsis METTL4 prefers Am-containing RNA as a substrate in vitro.

### The crystal structure of the full-length METTL4

To elucidate the molecular mechanism of METTL4-mediated N<sup>6</sup>-adenosine methylation on U2 snRNA, we performed structural analysis by crystallization of the recombinant full-length protein. Sequence alignment suggested that Arabidopsis METTL4 contains a conserved MTase domain (residues 203–403) with a strictly conserved DPPW motif shared by the MT-A70 m<sup>6</sup>A methyltransferases, as well as an N-terminal extended region without known structural information (Supplementary Fig. 4). We solved the crystal structures of the full-length METTL4 alone (apo), as well as METTL4 bound to the methyl donor S-adenosyl-L-methionine (SAM), or its product S-adenosyl-L-homocysteine (SAH), or its analog Sinefungin (SFG) (Supplementary Fig. 5a–d and Tables 1 and 2). Due to the high structural similarity (Supplementary Figs. 5 and 6), we chose the structure of METTL4-SAM complex for subsequent analysis. Of note, the central MTase domain of METTL4 possessing a classic sandwich-fold similar to METTL3, is embraced by two domains from the N-terminal extended region, i.e., the N-terminal domain (NTD), and the middle domain (MID; Fig. 2a–c). NTD is mainly composed of five  $\alpha$ -helices ( $\alpha$  1–5) and three  $\beta$ -strands ( $\beta$  1–3). A loop between  $\alpha$ 13 and  $\beta$ 14 in MTase domain, referred to as Interface loop in METTL3, is stabilized by the long  $\alpha$ 4 of NTD via extensive hydrophobic interactions in combination with charged interactions (Fig. 2d), which is likely conserved based on sequence alignment (Supplementary Fig. 4). MID, mainly composed of five  $\beta$ -strands and three short  $\alpha$ -helices with an Ig-like fold, leans to the side of the  $\beta$ 9 strand of the MTase domain primarily via antiparallel  $\beta$ -sheet ( $\beta$ 6 $\downarrow$ - $\beta$ 9 $\uparrow$ ) and hydrophobic interactions (Fig. 2e). In addition, the C-terminal extension (CE, aa:404–414), stretching out from the Mtase domain, forms an antiparallel beta sheet with NTD  $\beta$ 3 strand and hydrophobic interactions with I200 from the MID domain (Fig. 2f), which locks the NTD and MID domains together like a zip. The SAM molecule is bound adjacent to the conserved DPPW motif of Mtase domain in the complex structure (Fig. 2g and Supplementary Fig. 7), and particularly, aspartic acid D233 of DPPW motif forms a hydrogen bond with the amino group of SAM tail. Most of the residues involved in SAM binding in METTL4 are conserved in METTL3 (Supplementary Fig. 8), suggesting that MTase domains of MT-A70 family proteins share a common SAM binding mode. However, unlike METTL3, which requires METTL14 to form a fully functional enzyme, METTL4 alone forms an integrated structure by extensive interactions of the central MTase domain with NTD, MID and CE, to perform enzyme function.

### Complex structure of METTL4 bound to Am

To further investigate the Am specific N<sup>6</sup>-methylation mechanism of METTL4, we solved the crystal structure of METTL4 complexed with 3-nucleotides (3-nt) Am-containing RNA oligo (5'-AAmG-3') (Fig. 3a), a sequence taken from the Arabidopsis U2 snRNA around Am31, as a substrate in the presence of SAH (Table 2). In the complex structure, though only the central Am is visible in the electron-density map (Fig. 3d, e), we found the base of Am inserted into the catalytic pocket located in the turning corner of a L-shaped cavity with highly positive charges on the most conserved surface across the METTL4 homologs (Fig. 3a–c). There are four characteristic loops around the L-shaped cavity (Fig. 3d): Gate Loop1, Gate Loop2 and Interface Loop, named after the METTL3 structure<sup>15</sup>, and Entry Loop. The N<sup>6</sup> atom of Am

**Table 1 | The data collection of SeMETTL4 for structure determination**

SeMETTL4-MAD			
Data collection			
Space group	P21	P21	P21
Cell parameter			
<i>a, b, c</i> (Å)	60.1, 94.2, 95.7	60.3, 94.4, 96.0	60.2, 94.3, 95.5
$\alpha, \beta, \gamma$ (°)	90.0, 101.8, 90.0	90.0, 101.8, 90.0	90.0, 101.8, 90.0
Wavelength(Å)	0.96108	0.97956	0.97943
Resolution (Å)	30.0-2.80	30.0-2.80	30.0-2.80
Last shell (Å)	(2.90-2.80)	(2.90-2.80)	(2.90-2.80)
Completeness (%)	99.3 (99.0)	99.6 (98.0)	99.2 (98.1)
Redundancy	1.9 (1.9)	3.7 (3.5)	1.9 (1.9)
<i>I</i> / $\sigma$ ( <i>I</i> )	9.9 (1.2)	10.5 (0.8)	9.9 (1.0)
<i>R</i> <sub>merge</sub> (%)	6.2 (50.8)	9.2 (108.8)	5.9 (57.3)
No. reflections	50,092 (5006)	26,052 (2544)	50,458 (4970)
CC <sup>a</sup>	0.844 <sup>a</sup>	0.691 <sup>a</sup>	0.811 <sup>a</sup>

<sup>a</sup>The values in parentheses are for highest resolution shell.

<sup>b</sup>The statistics CC<sup>b</sup> is an estimate of the “true” CC data under examination to the (unknown) true intensities.

**Table 2 | Data collection and refinement statistics**

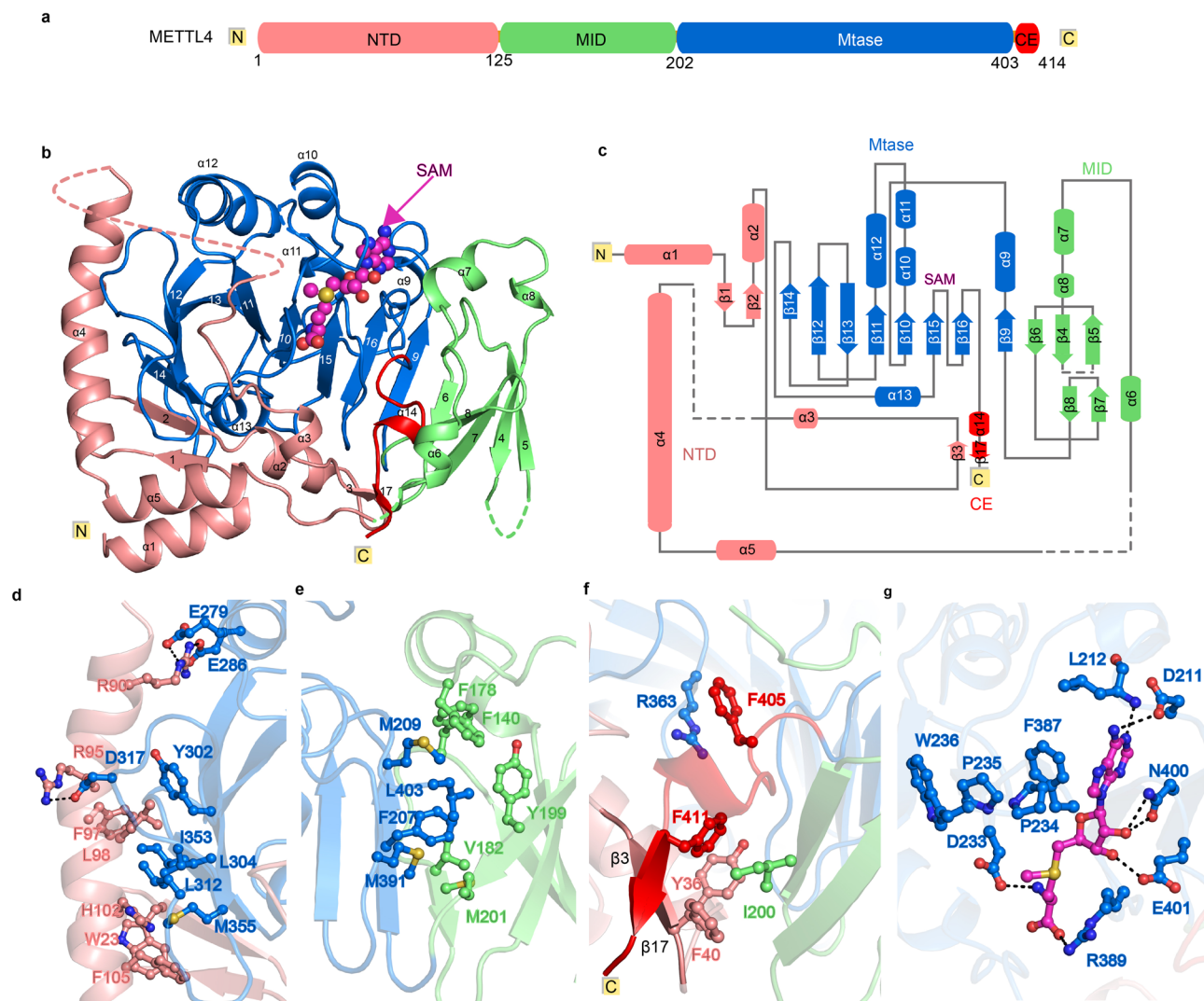
	SeMETTL4	METTL4- SAM	METTL4- SAH	METTL4- SFG	METTL4-Am-SAH
Space group	P21	P21	P21	P212121	I222
<i>a, b, c</i> (Å)	60.8,94.3, 95.7	61.9,93.7, 96.3	63.6,93.5, 96.6	52.2,84.6, 94.1	93.8,130.0, 141.3
$\alpha, \beta, \gamma$ (°)	90.0,101.7,90.0	90.0,104.3,90.0	90.0,104.7,90.0	90.0,90.0, 90.0	90.0,90.0, 90.0
Wavelength(Å)	0.97930	0.97916	0.97776	0.97853	0.97852
Resolution (Å)	30.0-2.50	30.0-2.30	30.0-2.45	30.0-2.35	30.0-3.0
Last shell (Å)	(2.59-2.50) <sup>a</sup>	(2.38-2.30) <sup>a</sup>	(2.54-2.45) <sup>a</sup>	(2.43-2.35) <sup>a</sup>	(3.11-3.00) <sup>a</sup>
Completeness (%)	98.2(93.7)	98.1(91.1)	97.6(88.6)	98.8(91.5)	82.4(54.7)
Redundancy	4.8(3.3)	3.4(3.1)	5.3(4.0)	11.8(8.6)	5.8(2.4)
<i>I</i> / $\sigma$ ( <i>I</i> )	29.9(7.2)	25.6(2.0)	17.5(2.4)	24.2(3.3)	7.8(1.9)
<i>R</i> <sub>merge</sub> (%)	13.2(41.3)	6.3(59.7)	10.4(39.2)	17.4(60.6)	11.2(40.0)
No. reflections	36127 (3421)	46526 (4278)	39010 (3536)	17743 (1596)	14468 (938) <sup>a</sup>
Refinement					
Resolution (Å)	2.50	2.30	2.45	2.35	3.01
<i>R</i> <sub>work</sub> (%) / <i>R</i> <sub>free</sub> (%)	22.6/26.5	22.0/25.1	21.1/25.4	20.1/23.7	24.0/30.0
No. of atoms					
Protein	5326	5515	5639	2741	2877
RNA/DNA	0	0	0	0	48
Ligand/ion	0	54	96	45	26
Water	11	57	22	61	19
<i>B</i> factors	50.0	72.0	55.0	47.0	94.0
Bond length (Å)	0.007	0.006	0.005	0.003	0.019
Bond angle (°)	1.010	1.080	0.974	0.072	1.982
Ramachandran plot (%)					
Most favored	96.31	97.41	98.11	5.38	89.82
Additional allowed	3.69	2.59	1.89	4.62	10.15
Outliers	0	0	0	0	0
PDB number	7CVA	7CV7	7CV9	7CV8	7CV6

Data are from one crystal for each structure.

<sup>a</sup>Values in parentheses are for highest-resolution shell.

forms a tetrahedral like hydrogen network with the side chain of D233 and Y247, as well as the main chain O atom of P234 from Gate Loop1 (Fig. 3f, g). In addition, the *N*<sup>7</sup> atom of Am forms a hydrogen bond with the side chain of S362 from Gate loop2, and the *N*<sup>3</sup> atom of Am directly binds to the side chain of E325 (Fig. 3f, g). These

interactions allow the *N*<sup>6</sup> atom point to the sulfur atom of SAH with 4.7 Å, which is similar to the corresponding distance observed for the *N*<sup>6</sup>-adenine DNA methyltransferase M<sup>1</sup> *TaqI* (PDB code: 1G38)<sup>35</sup>. Notably, the dual conformations of Y247 were captured in the complexes, a free conformation obtained from the Am-absent complex



**Fig. 2 | Crystal structure of full-length METTL4 bound to SAM.** **a** Schematic domain structures of METTL4. **b** Cartoon representation of full-length METTL4 bound to SAM, with the  $\beta$ -strands and  $\alpha$ -helices numbered numerically and the invisible residues in electric map shown as dashed lines. The SAM is shown as magenta sphere. **c** Schematic representation of the full-length structure of METTL4. **d–f** Close-up view of the interaction between Mtase domain and NTD

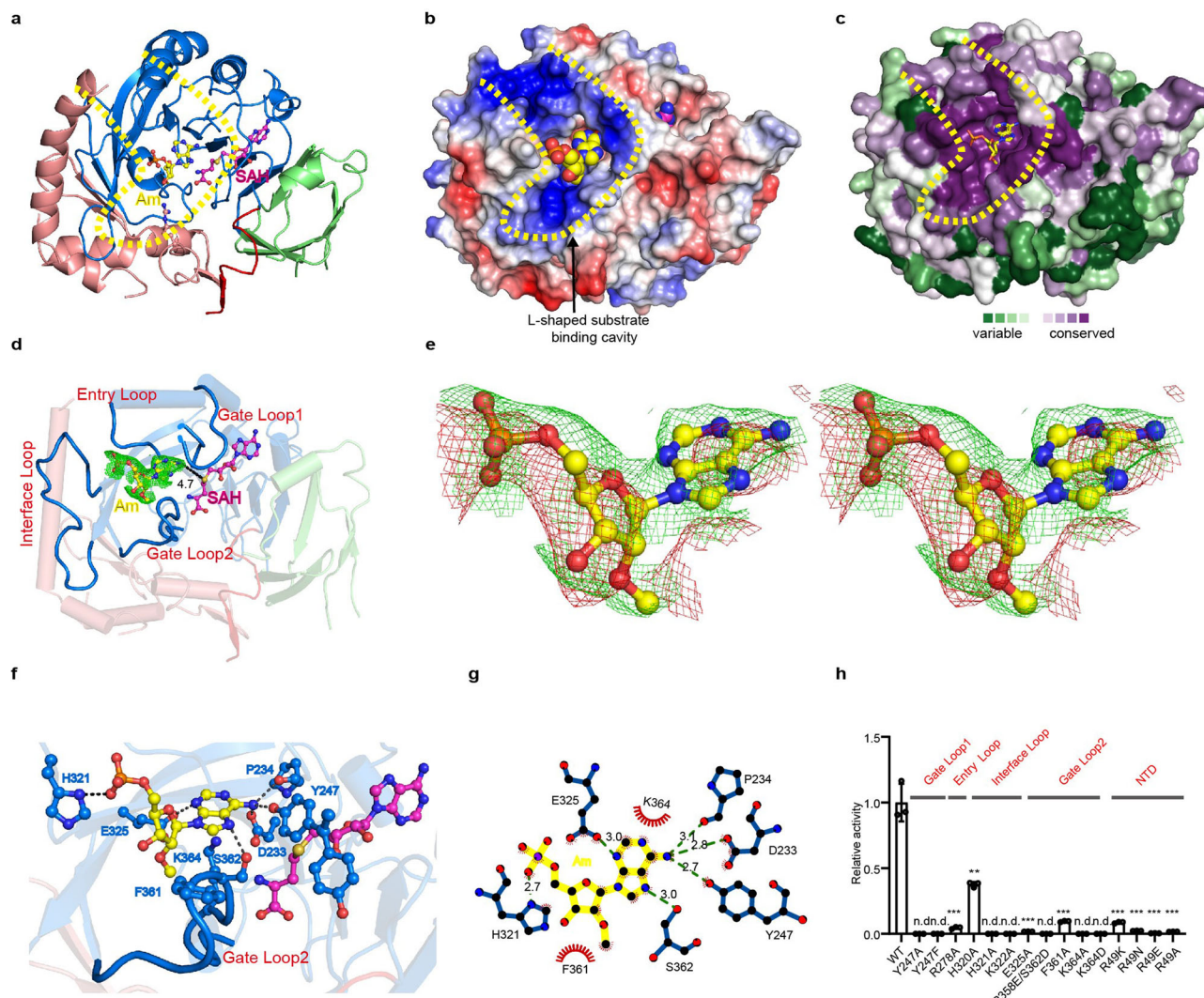
(**d**), MID (**e**), and CE domain (**f**). **g** The interaction network between METTL4 and SAM. The hydrogen bond interactions are shown as green dashed lines and the Vander Waal interactions are shown as red dashed fans. **f** Measurement of the binding affinity between SAM and the METTL4 wild-type and mutated proteins using ITC, n.d., not detected.

and an active but restricted conformation obtained from the Am substrate-containing complex. Compared to the free conformation that flips out, the side chain of Y247 in active conformation turns  $-131^\circ$  towards the active site, resulting in the N6 atom of Am in the active site being attracted by the OH-group of Y247 (Supplementary Fig. 7d). Mutation of Y247 to alanine (Y247A) or phenylalanine (Y247F) and E325 to alanine (E325A) abolished the activity (Fig. 3h), highlighting the important role of these residues in placing Am base into the precise position for  $N^6$ -methylation. The phosphate group of the Am is bound by the H321 from Interface Loop, and mutations of the positively charged residues, such as R278 in the Entry Loop and K322 in the Interface loop, impaired the enzymatic activity, suggesting that these positively charged residues in the four loops are important for substrate recognition.

#### Specific recognition of Am-containing substrate by METTL4

When we superimposed METTL4 with the METTL3-METTL4 complex, there are extensive clashes of the NTD domain with METTL4 (Fig. 4c), and the 2'-O-methyl (2'-OMe) of Am is clearly clashing with the Gate Loop2 even though most of residues involved in the Am interaction are

conserved in METTL3 and METTL4 (Fig. 4d, e and Supplementary Fig. 8). Sequence alignment suggested that an aromatic residue F361 in Gate Loop2, which is relatively conserved in METTL4 proteins (Phe or His), provides a more hydrophobic environment for the methyl group. However, the residue Phe or His in Gate Loop2 of METTL4 is threonine (T) in METTL3, and mutation of F361A significantly reduced the activity (Fig. 3f), suggesting that Gate Loop2 may play a key role in distinguishing 2'-OMe and 2'-OH by METTL4. In addition, unlike the Interface Loop of METTL3, which is bound by METTL4 thus providing an RNA-binding scaffold<sup>15</sup>, the Interface Loop of METTL4 is stabilized by  $\alpha$ 4-helix in NTD (Figs. 2d and 3b), suggesting that NTD of METTL4 plays a similar role of METTL4 for METTL3. We note that the two highly conserved residues in NTD, S48, and R49, interact with D360 and G359 in Gate loop2, respectively (Fig. 4f), which stabilizes the conformation of Gate loop2 for the Am recognition and especially for the 2'-O-methyl group interaction (Supplementary Fig. 9). Consistently, mutations of R49 to alanine (A), asparagine (N) or negatively charged glutamic acid (E), dramatically reduced activity (Fig. 3f). These results suggest a critical role for NTD in the regulation of substrate recognition by METTL4.



**Fig. 3 | Crystal structure of the ternary complex of METTL4-Am-SAH uncovers a L-shaped cavity for the substrate binding and a conserved pocket for Am recognition.** **a** Overall view of the METTL4-Am-SAH complex. **b** Electric surface of METTL4. The positive charged cleft is outlined by yellow dashed line, and the Am is shown as sphere. **c** ConSurf analysis of the conservation of METTL4 proteins. **d** Four conserved loops form substrate-binding cavity, and key residues involved in Am recognition are shown in stick and ball. The 2Fo-Fc map of Am is contoured at the 1.0 sigma level, and the invisible residues in electron density map of Gate Loop 1 shown as dashed lines. **e** Cross-eyed stereo view of the 2Fo-Fc map in green color at 1.0 sigma level generated by CCP4 program and Fo-Fc omit map in red color at

0.6 sigma level generated by Phenix program of the Am molecule in the catalytic pocket of METTL4. The estimated occupancy of the Am molecule in the catalytic pocket is 0.78 according to phenix.refine program. **f** Close-up view of the interactions between Am and MTase domain. The hydrogen bond interactions are depicted as yellow dashed lines. **g** The diagram of interactions involving in Am was analyzed by LigPlot. The hydrogen bonds are indicated by dashed lines and the hydrophobic contacts are shown as an arc with spokes radiating. **h** Enzymatic assay of METTL4 mutants measured by UHPLC-MRM-MS/MS method. The data are mean  $\pm$  SD. Statistical analysis used Student's test for the difference wild-type and mutants, n.s. not significant, n.d. not detectable; \*\* $p < 0.01$ , \*\*\* $p < 0.001$ .

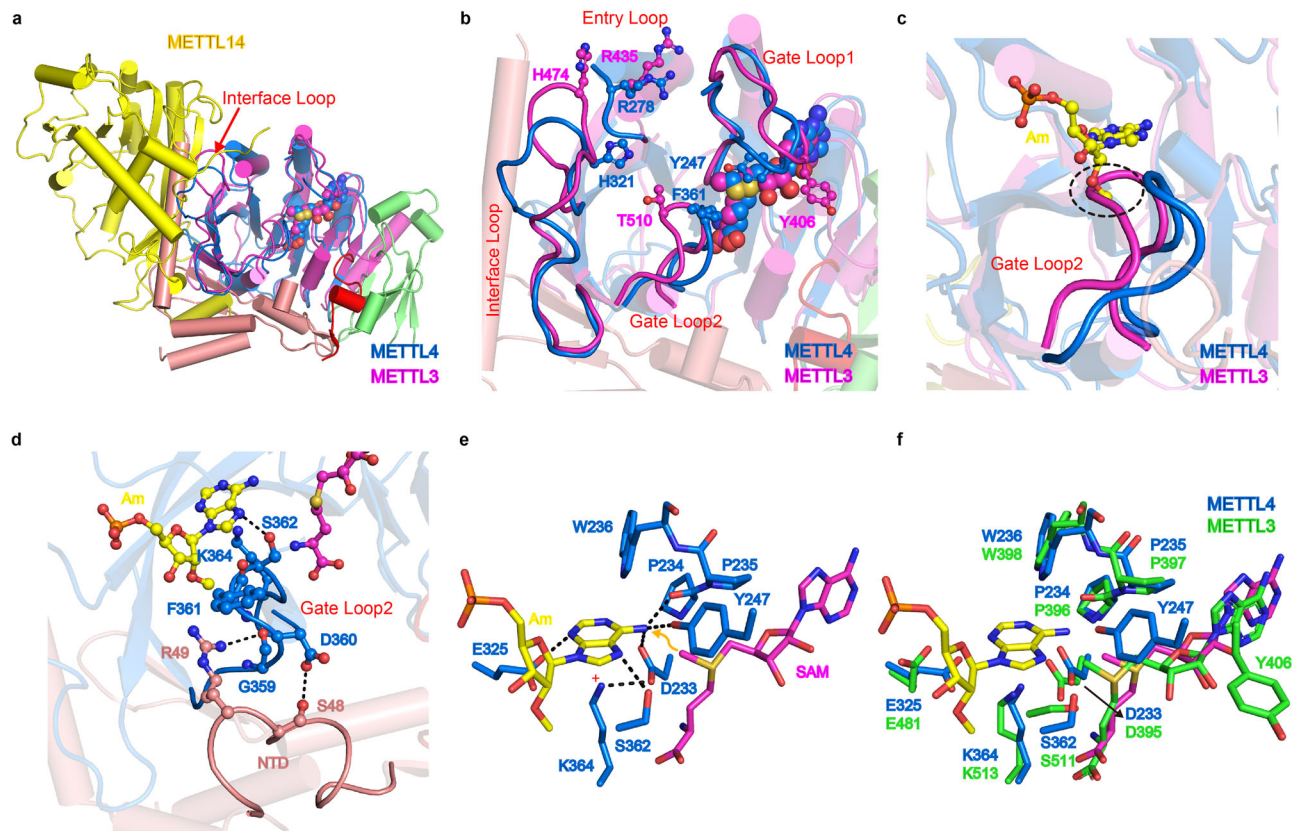
### Conserved residues involved in binding $N^6$ of adenosine

To further reveal the catalytic mechanism, we modeled the catalytic center by replacing SAH with SAM in the structure of the METTL4-Am-SAH complex (Fig. 4e), and noticed that K364, an absolutely conserved residue in Gate loop2 across MT-A70 family proteins (Supplementary Fig. 8), perpendicularly interacts with the Am base. It is plausible that the side chain of K364 forms cation- $\pi$  interaction with the Am base in the catalytic center, when the amino group of K364 side chain forms a salt bridge ionic interaction with D233, which also forms a hydrogen bond with the  $N^6$  atom of the Am base (Fig. 4f). Interestingly, D233 interacts with SAM in the structure of the METTL4-SAM complex. Therefore, with the assistance of D233, K364 may function as a general base to promote the deprotonation of the  $N^6$  amino group of Am and catalyze the methyl group transfer from SAM to the  $N^6$  atom of adenosine. Consistently, mutation of K364 to alanine (K364A) or aspartic acid (K364D) abolished the catalytic activity (Fig. 3f). Superposition of

the conserved residues in the modeled active site of METTL4 with corresponding residues in METTL3 clearly showed that most key residues are in the same conformation (Fig. 4f), suggesting that MT-A70 family proteins may share the novel common catalytic mechanism of  $N^6$  methylation of adenosine. The only exception is the residue Y247 in METTL4 that forms hydrogen bond with  $N^6$  of Am substrate, but the corresponding residue Y406 in METTL3 flips away without binding substrate, indicating that Y247 or Y406 may play a key role in recognition of  $N^6$  of substrate adenosine base.

### A proposed catalytic mechanism of Am by METTL4

Based on structural and modeling studies, we propose a model for the catalytic progresses in the catalytic pocket (Fig. 5). First, the backbone of the target base is recognized by the Interface Loop, thus providing the opportunity to fit the base into the catalytic pocket. Second, the target base is stabilized via the interactions mediated by the 2'-O-



**Fig. 4 | Structural comparison of METTL4 and METTL3/METTL4.**

**a** Superposition of the METTL4 with METTL3-METTL4 heterodimer. **b** Close-up view of the catalytic pocket formed by the four loops in the MTase domains of METTL4 and METTL3, and the key residues were shown as sticks. **c** Structural

superposition of METTL4 and METTL3 shows the Gate Loop2 makes clashes with Am. **d** The NTD makes interaction with Gate Loop2 of MTase domain. **e** Modeled catalytic center of METTL4 by replacing SAH with SAM. **f** Structural superposition of the catalytic center of METTL4 and METTL3.

methyl group with the Gate Loop2, and the hydrogen bond between  $N^6$  of adenosine base and Y247 guiding the  $N^6$  atom to an optimal catalytic position. Third, in the catalytic position, the  $N^6$  atom forms hydrogen bonds with the side chain D233 and the main chain P234 resulting in the deprotonation of  $N^6$  amine group, and the deprotonated  $N^6$  amine group attacks the activated methyl group of SAM. Fourth, the methyl group at the  $N^6$  atom makes extensive clashes with the D233 and P234, which promotes the  $N^6$ -methylated adenosine base to flip out from the catalytic pocket.

## Discussion

This study demonstrated that the Arabidopsis METTL4, similar to human METTL4<sup>20</sup>, is responsible for  $m^6$ Am modification of U2 snRNA in vivo and in vitro, and plays a similar role in the regulation of RNA alternative splicing. The crystal structure of the METTL4-Am-SAH complex in this study provided critical insights into the molecular mechanism for  $N^6$ -adenosine methylation of an Am-containing RNA substrate. Based on the sequence and structural comparisons (Fig. 4c and Supplementary Fig. 8), the catalytic mechanism proposed here (Fig. 5) is probably shared by another MT-A70 family methyltransferase METTL3, which is consistent with the phylogenetic analysis that METTL3 and METTL4 belong to the same MT-A70 clade<sup>22</sup>. The topologies of the MTase domains of METTL4 and METTL3 are similar but different from other RNA  $m^6$ A or  $m^6$ Am methyltransferases, such as METTL5<sup>18</sup>, ZCCHC4<sup>36</sup>, METTL16<sup>37</sup>, and CAPAM<sup>38</sup>. Unlike the aromatic residue F or Y in the conserved [D/N]PP[F/Y] motif, which functions to stabilize the target base in the catalytic pocket, the aromatic residue W236 in the DPPW motif in METTL4 is buried in a hydrophobic pocket adjacent to the Am binding site (Supplementary Fig. 9), implying that the catalytic mechanism of  $N^6$ -methylation mediated by METTL4 is

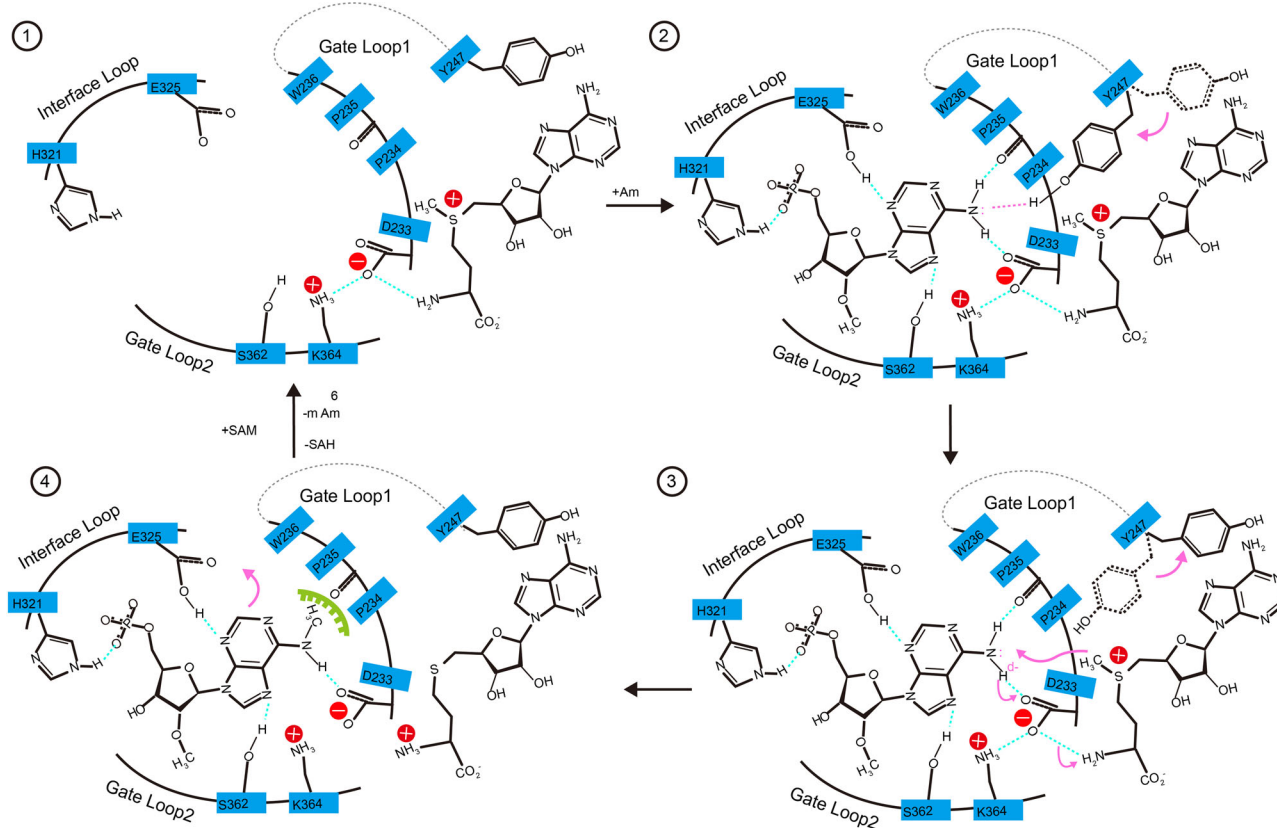
different from the previously proposed  $N^6$ -methyl-ammonium cation mechanism for the  $N^6$ -adenine DNA methyltransferase *MtaTaqI*<sup>35</sup>. Owing to the highly conserved residues in the active site of METTL4 and METTL3 (Fig. 4f), the adenosine base of the substrate is recognized by two MT-A70 family enzymes in a similar way. Therefore, the crystal structure and modeling study of METTL4 in complex with the substrate may help to design more specific inhibitors of METTL3 in developing drugs as a therapeutic strategy against  $m^6$ A-associated human diseases<sup>39</sup>.

The Am-containing RNA substrate is bound by an L-shaped, positively charged cavity formed by four loops in METTL4, which are also conserved in METTL3. However, the Interface Loop in METTL4 is bound by a long helix from NTD, which replaces METTL14 that stabilizes METTL3 and coordinates with METTL3 to form a substrate binding cavity<sup>15</sup>. Therefore, METTL4 functions as a monomer (Supplementary Fig. 6c), but METTL3 forms a functional heterodimer with METTL14<sup>15</sup>. Another difference is that METTL4 prefers Am while METTL3 prefers A instead of Am as the substrate, although *Drosophila* METTL4 appears to act on A in U2 snRNA in vivo<sup>22</sup>, maybe due to the variation in sequences within Gate Loop2 (Supplementary Fig. 4). Lastly, the sequences of NTD are less conserved in METTL4 proteins across diverse species (Supplementary Fig. 4), which may account for their different preferences for 2'-OH or 2'-OME of the adenosine substrates. Further study is needed for the Am preference by METTL4 from the evolutionary perspective.

## Methods

### METTL4 construction, expression, and purification

The open reading frame (ORF) of *METTL4* (AT1G19340) was amplified from an *Arabidopsis thaliana* cDNA library and the ORF segment was



**Fig. 5 | Proposed catalytic mechanism of METTL4.** Diagrammatic presentation of the catalytic mechanism of METTL4. In the initial step (1) the backbone phosphate of Am may be bound by H321 in interface loop; then, Y247 in the Gate Loop1

interacts with N<sup>6</sup> of Am (2), followed by hydrogen bonding from D233 and P234 and methyl group transferred from SAM to N<sup>6</sup> of Am (3). Finally, N<sup>6</sup>-methylated Am is released from the pocket due to the steric clash (4).

cloned into the pET28vector containing a SUMO tag between *Nde* I and *Bam*H I. The protein was expressed in *E.coli* BL21(DE3) and the cells were induced at OD<sub>600</sub> -0.6 by isopropyl β-D-1-thiogalacto-pyranoside (IPTG) with a final concentration of 0.2 mM for 20 h at 14 °C. Cells were harvested by centrifugation at 6000 rpm for 15 min and homogenized in ice-cold buffer1 containing 20 mM Tris-HCl, pH 8.0, 500 mM NaCl and 25 mM imidazole. The cells were crushed by high pressure machine and followed by ultracentrifugation at 18,000 rpm for 1 h at 4 °C. The supernatant was loaded onto a Ni-NTA column (GE healthcare) and the sample was eluted by gradient using buffer2 (20 mM Tris-HCl, pH 8.0, 500 mM NaCl and 500 mM imidazole). The tag was cleaved by ulp1 protease and dialyzed against buffer containing 20 mM Tris-HCl, pH 8.0 and 500 mM NaCl. Then, sample was loaded onto a Ni-NTA column again and collected the flow through, and the sample was further purified by Superdex 200 16/600 column (GE healthcare) equilibrated by buffer containing 20 mM Tris-HCl, pH 8.0, 100 mM NaCl and 5 mM DTT. The peak fraction was concentrated to about 30 mg/ml using Amicon 30-kDa cutoff (Millipore). The mutations were generated by overlap PCR method and the purification of mutated proteins was the same as the wild-type.

### Crystallization

The initial crystallization was performed by Gryphon crystallization robot system using sitting-drop vapor diffusion by mixing 0.2 μl protein sample (6 mg/ml) with 0.2 μl mother solution at 18 °C and crystallization kits from the Hampton research company. All optimization procedures were used the hanging-drop vapor diffusion method at 18 °C. For METTL4 complex, the protein was incubated with 1mM SAM (S-adenosylmethionine purchased from Sigma, CAS number: 86867-01-8), or SFG (sinefungin, purchased from SANTA CRUZ

BIOTECHNOLOGY, CAS number: 58944-73-3) or SAH (S-adenosyl-L-homocysteine purchased from Sigma, CAS number: 53186-57-5) in final concentration. The crystals of Se-METTL4 grew in the condition including 0.1M tacsimate pH5.5 and 8% PEG3350, and the crystals of METTL4-SAM complex grew in the condition including 0.1M sodium malonate pH5.5 and 6% PEG3350. The crystal of METTL4 bound to SFG were grown in the condition of 1M Lithium chloride 0.1M HEPES pH 7.0 10% PEG6000, and the crystal of METTL4 bound to SAH was obtained from the condition of 1600 mM Magnesium sulfate 100 mM MES/ Sodium hydroxide pH6.5. For the complex of METTL4-Am-SAH, the crystal of METTL4-SAH grew in the condition consist of 0.49 M NaH<sub>2</sub>PO<sub>4</sub> and 0.91 M K<sub>2</sub>HPO<sub>4</sub>, and the crystals were soaked with 10 mM RNA oligo (5'-AAMG-3') at final concentration, and the RNA oligo (5'-AAMG-3') was synthesized and purified in Cao's lab at Shanghai Institute of Organic Chemistry, Chinese Academy of Sciences. For the METTL4-DNA-SAH complex, the synthetic DNA sequence 5'-GCCGCGTGATCACGCGGC-3' was annealed and incubated with METTL4 (6 mg/ml) with the molar ratio 1.1:1.0. The crystals were grown in the condition including 0.1M sodium malonate pH5.0 and 12% PEG3350.

### Data collection and structure determination

All crystals were transferred to a cryoprotectant solution containing mother liquor and 20% glycerol. All data were collected at the shanghai synchrotron radiation facility (SSRF) on beamlines BL17U, BL18U, or BL19U using a CCD detector cooled under liquid nitrogen. The data of selenomethionine-labeled METTL4 (Se-METTL4) crystals were collected by multiwavelength anomalous dispersion (MAD), and the data of METTL4-SAM, METTL4-SAH, METTL4-SFG and METTL4-Am-SAH complex were collected by single wavelength anomalous dispersion



(SAD) and processed with HKL2000 or HKL3000 program. The structure of METTL4-SAM, METTL4-SAH, METTL4-SFG, and METTL4-Am-SAH, METTL4-DNA-SAH were solved by molecular replacement of ccp4i program using the Se-METTL4 as a search model. All crystal structures were built using Coot<sup>40</sup> and refined by ccp4<sup>41</sup> and Phenix<sup>42</sup> program. The conservation of the surface of METTL4 was analyzed by ConSurf<sup>43</sup>, and the diagrams of interactions involving in Am and SAM were analyzed by LigPlot<sup>44</sup>. All structure pictures representing in paper were prepared with PyMOL.

### ITC

ITC experiments were performed at 25 °C using a MicroCal iTC200 microcalorimeter. Before the reactions, proteins and SAM were dialyzed against buffer containing 100 mM NaCl and 10 mM HEPES (pH 7.5). In all, 3 mM SAM with an initial injection volume of 0.5 µL (omitted from the analysis) followed by 19 injections (each 2 µL) were titrated into 150 µM proteins. Data were analyzed using Origin 7 software, and the heat of dilution was subtracted from the raw values. Dissociation constant ( $K_d$ ) values were calculated by fitting the isotherm.

### Plant materials

*Arabidopsis mett14-1* mutant was obtained from the Saskatoon collection (<http://aafc-aac.usask.ca/FST/>), and all *Arabidopsis thaliana* strains used in this study were in the Columbia (Col) background. Plants grown under long-day (LD) condition (22 °C and 16-h-light/8-h-dark) were used for flowering phenotype analysis. For RNA extraction, plants were cultured for 14-days on agar-solidified Murashige and Skoog medium M0255 (Duscheffa) supplemented with 0.9% sucrose at 21 °C under a 16-h-light/8-h-dark photoperiod in a Percival AR41L5 growth chamber in 60 to 90 µmol m<sup>-2</sup> s<sup>-1</sup> white light. The *mett14-1* mutant was rescued respectively by introducing *P<sub>METTL4</sub>::HA-METTL4-Flag* and *P<sub>METTL4</sub>::HA-METTL4(D233A/W236A)-Flag* constructs, in which HA- and Flag-tagged wild-type METTL4 and mutated METTL4 (D233A/W236A) were expressed under *METTL4* native promoter.

### U2 snRNA purification

Total RNA was extracted from 12-day seeding *Arabidopsis* by a standard method using Trizol. A synthetic oligo complementary to the *Arabidopsis* U2 snRNA (5'-biotin-GATACTACACTTGATCTTAGCCAAAAGGC CGAGAA, 500 ng) were incubated with total RNA (10 µg), in 50 µL hybridization buffer (150 mM NaCl, 50 mM Tris-HCl, pH 7.9). The hybridization mixture was incubated at 90 °C for 7 min, and slowly cooled to room temperature (25 °C) to allow hybridization to occur. The annealed products were mixed with Dynabeads M-280 Streptavidin (500 µg, Invitrogen, 11205D) at room temperature for 30 min. The annealed probes were mixed with Dynabeads M-280 Streptavidin (500 µg, Invitrogen, 11205D) at room temperature for 30 min. Place the mixture on magnet until the liquid become clear, then remove and collect the flow through (the flow through was called as U2 RNA removal in Fig. 1f). After that, wash the beads with 50 µL hybridization buffer for twice. And the sample were eluted in nuclease-free H<sub>2</sub>O by heating at 70 °C for 5 min. The elution and flow through were used for following digestion and UHPLC-MS/MS analysis.

### RNA digestion

For the in vitro enzyme assay: The probes were dissolved into nuclease-free H<sub>2</sub>O and annealed by heating briefly to 95 °C for 5 min, followed by gradually cooling to 4 °C in PCR machine.

Each probe (90 pmol) was incubated with 4 µg METTL4 protein in 30 µL reaction buffer (0.8 mM d3-SAM, 10% glycerol, 10 mM Tris-HCl pH 7.0) at 28 °C for 1 h. After reaction, all sample were desalted and purified by ice ethanol precipitation. The recovered probes or *Arabidopsis* total RNA were digested into single nucleosides by 0.5 U Nuclease P1 (NP1, Sigma, N8630) and 0.5 U alkaline phosphatase (CIP,

New England Biolabs, M0290) in 50 µL 10 mM Tris-HCl (pH 7.6) reaction volume at 37 °C for 4 h. The digestive enzyme in the samples were removed by ultrafiltration with 3 kD ultrafiltration tubes (Pall Corporation, Port Washington), and then subjected to UHPLC-MS/MS analysis.

### UHPLC-MS/MS

The UHPLC-MS/MS analysis was performed on an ultra-high performance LC system (Agilent 1290 II) coupled with G6495 mass spectrometer. 5 µL sample was injected in to the UHPLC system and the nucleosides were separation by C18 column (Zorbax Eclipse Plus C18, 2.0 mm×100 mm i.d, 1.8 µm particle size, Agilent Technologies, Palo Alto, CA). All the targets were separated in HPLC condition with mobile phase consisted of two solvents: 0.1% formic acid in water (A), and 100% in methanol (B). Set up following separation condition with the gradient elution: 0–2 min, 5% B; 2–4 min, 30% B; 4–6 min, 30% B, 6–9 min, 5% B, and the flow rate was 0.25 mL/min. For the mass spectrometry parameters, the fragmentation voltage was set at 90 V; using nitrogen for nebulization and desolvation, and the flow rate and temperature of gas were set at 9.0 L/min and 300 °C, respectively; collision gas used high purity of nitrogen with 99.999%. The multiple reaction monitoring (MRM) mode was chosen to quantify all the targets: m/z 292 → 150 for m<sup>6</sup>Am, m/z 282 → 150 for m<sup>6</sup>A and m<sup>1</sup>A; m/z 282 → 136 for Am, m/z 266 → 150 for m<sup>8</sup>A. The collision voltage of all targets was set to 5 eV. All the sample were repeated at three times.

### RNA sequencing

Total RNA was extracted from 14-days seedlings using an RNAprep pure Plant Kit (Tiangen Biotech) under the guidance of manufacturer's instructions. A strand-specific RNA-seq library was constructed according to the KAPA stranded mRNA-seq Kit instructions (Illumina® Platforms). There were two replicates for WT and *mett14-1*.

Raw RNA-seq paired-end reads were filtered to remove the adapters and reads length less than 20 bp and reads with low sequencing base quality score using Cutadapt (version 3.5)<sup>45</sup>. The filter reads were mapped to the *Arabidopsis thaliana* genome (TAIR10) by Hisat2 (version 2.2.1)<sup>46</sup>. High mapping quality reads were obtained using SAMtools (version 1.13)<sup>47</sup>. The track files which fragments were normalized as fragments per bin (bin size = 10 bp) per million mapped fragments were obtained using bamCoverage tool in deepTools (version 3.5.1)<sup>48</sup> and viewed with IGV (version 2.11.2)<sup>49</sup>. FeatureCounts (version 2.0.1) was used to calculate the read number for each gene<sup>50</sup>. DESeq2 (version 1.32.0) was used to identify differentially expressed genes (DEGs)<sup>51</sup>. Compared to the mutant and wild-type, genes that showed  $|\log_2\text{foldchange}| \geq \log_2(1.5)$  and  $p$  value < 0.05 were considered to be DEGs. Gene ontology (GO) enrichment analysis was carried out with DAVID<sup>52</sup> and ClusterProfiler<sup>53</sup>. Different alternatively splicing genes between the mutant and the wild-type (threshold of  $|\Delta\psi| > 0.1$ ,  $p$  value < 0.05) were identified by rMATS (version 4.1.1)<sup>54</sup>.

### Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this article.

### Data availability

The atomic coordinates of the X-ray crystal structures have been deposited to the Protein Data Bank with following accession codes 7CVA for apo-METTL4, 7CV7 for METTL4-SAM, 7CV9 for METTL4-SAH, 7CV8 for METTL4-SFG, and 7CV6 for METTL4-Am-SAH. The RNA-sequencing data reported in this study have been deposited in NCBI's Gene Expression Omnibus with the accession number GEO Series accession number GSE190241. Source data are provided with this paper.

## References

1. Frye, M., Harada, B. T., Behm, M. & He, C. RNA modifications modulate gene expression during development. *Science* **361**, 1346–1349 (2018).
2. 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).
3. Yue, Y., Liu, J. & He, C. RNA N6-methyladenosine methylation in post-transcriptional gene expression regulation. *Genes Dev.* **29**, 1343–1355 (2015).
4. Shimba, S., Bokar, J. A., Rottman, F. & Reddy, R. Accurate and efficient N-6-adenosine methylation in spliceosomal U6 small nuclear RNA by HeLa cell extract in vitro. *Nucleic Acids Res.* **23**, 2421–2426 (1995).
5. Yang, D. et al. N6-Methyladenosine modification of lincRNA 1281 is critically required for mESC differentiation potential. *Nucleic Acids Res.* **46**, 3906–3920 (2018).
6. Xiao, S. et al. The RNA N(6)-methyladenosine modification landscape of human fetal tissues. *Nat. Cell Biol.* **21**, 651–661 (2019).
7. Sergiev, P. V., Aleksashin, N. A., Chugunova, A. A., Polikanov, Y. S. & Dontsova, O. A. Structural and evolutionary insights into ribosomal RNA methylation. *Nat. Chem. Biol.* **14**, 226–235 (2018).
8. Roundtree, I. A., Evans, M. E., Pan, T. & He, C. Dynamic RNA modifications in gene expression regulation. *Cell* **169**, 1187–1200 (2017).
9. Shi, H., Wei, J. & He, C. Where, when, and how: context-dependent functions of RNA methylation writers, readers, and erasers. *Mol. Cell* **74**, 640–650 (2019).
10. Wu, Y., Zhou, C. & Yuan, Q. Role of DNA and RNA N6-adenine methylation in regulating stem cell fate. *Curr. Stem Cell Res. Ther.* **13**, 31–38 (2018).
11. Vu, L. P., Cheng, Y. & Kharas, M. G. The biology of m(6)A RNA methylation in normal and malignant hematopoiesis. *Cancer Discov.* **9**, 25–33 (2019).
12. Deng, X., Su, R., Stanford, S. & Chen, J. Critical enzymatic functions of FTO in obesity and cancer. *Front Endocrinol.* **9**, 396 (2018).
13. Dai, D., Wang, H., Zhu, L., Jin, H. & Wang, X. N6-methyladenosine links RNA metabolism to cancer progression. *Cell Death Dis.* **9**, 124 (2018).
14. Chen, B., Li, Y., Song, R., Xue, C. & Xu, F. Functions of RNA N6-methyladenosine modification in cancer progression. *Mol. Biol. Rep.* **46**, 2567–2575 (2019).
15. Wang, X. et al. Structural basis of N(6)-adenosine methylation by the METTL3-METTL14 complex. *Nature* **534**, 575–578 (2016).
16. Mendel, M. et al. Methylation of structured RNA by the m(6)A writer METTL16 is essential for mouse embryonic development. *Mol. Cell* **71**, 986–1000 e1011 (2018).
17. Pinto, R. et al. The human methyltransferase ZCCHC4 catalyses N6-methyladenosine modification of 28S ribosomal RNA. *Nucleic Acids Res.* **48**, 830–846 (2020).
18. van Tran, N. et al. The human 18S rRNA m6A methyltransferase METTL5 is stabilized by TRMT112. *Nucleic Acids Res.* **47**, 7719–7733 (2019).
19. Dominissini, D. et al. Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. *Nature* **485**, 201–206 (2012).
20. Meyer, K. D. et al. Comprehensive analysis of mRNA methylation reveals enrichment in 3' UTRs and near stop codons. *Cell* **149**, 1635–1646 (2012).
21. Linder, B. et al. Single-nucleotide-resolution mapping of m6A and m6Am throughout the transcriptome. *Nat. Methods* **12**, 767–772 (2015).
22. Bujnicki, J. M., Feder, M., Radlinska, M. & Blumenthal, R. M. 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 (2002).
23. Chen, H. et al. METTL4 is an snRNA m(6)Am methyltransferase that regulates RNA splicing. *Cell Res.* **30**, 544–547 (2020).
24. Goh, Y. T., Koh, C. W. Q., Sim, D. Y., Roca, X. & Goh, W. S. S. METTL4 catalyzes m6Am methylation in U2 snRNA to regulate pre-mRNA splicing. *Nucleic Acids Res.* **48**, 9250–9261 (2020).
25. Gu, L. et al. CG14906 (mettl4) mediates m(6)A methylation of U2 snRNA in *Drosophila*. *Cell Discov.* **6**, 44 (2020).
26. Shen, C. et al. DNA N<sup>6</sup>-methyldeoxyadenosine in mammals and human disease. *Trends Genet.* **38**, 454–467 (2022).
27. Kweon, S. M. et al. An adversarial DNA N(6)-methyladenine-sensor network preserves polycomb silencing. *Mol. Cell* **74**, 1138–1147.e1136 (2019).
28. Zhang, Z. et al. Regulation of adipocyte differentiation by METTL4, a 6 mA methylase. *Sci. Rep.* **10**, 8285 (2020).
29. Hao, Z. et al. N(6)-deoxyadenosine methylation in mammalian mitochondrial DNA. *Mol. Cell* **78**, 382–395 e388 (2020).
30. Wang, P., Doxtader, K. A. & Nam, Y. Structural basis for cooperative function of Mettl3 and Mettl14 methyltransferases. *Mol. Cell* **63**, 306–317 (2016).
31. Sledz, P. & Jinek, M. Structural insights into the molecular mechanism of the m(6)A writer complex. *Elife* **5**, e18434 (2016).
32. Iyer, L. M., Zhang, D. & Aravind, L. Adenine methylation in eukaryotes: Apprehending the complex evolutionary history and functional potential of an epigenetic modification. *Bioessays* **38**, 27–40 (2016).
33. Huang, Z. P., Zhou, H. & Qu, L. H. Maintaining a conserved methylation in plant and insect U2 snRNA through compensatory mutation by nucleotide insertion. *IUBMB Life* **57**, 693–699 (2005).
34. Liu, J. et al. A METTL3-METTL14 complex mediates mammalian nuclear RNA N6-adenosine methylation. *Nat. Chem. Biol.* **10**, 93–95 (2014).
35. Goedecke, K., Pignot, M., Goody, R. S., Scheidig, A. J. & Weinhold, E. Structure of the N6-adenine DNA methyltransferase M.TaqI in complex with DNA and a cofactor analog. *Nat. Struct. Biol.* **8**, 121–125 (2001).
36. Ren, W. et al. Structure and regulation of ZCCHC4 in m(6)A-methylation of 28S rRNA. *Nat. Commun.* **10**, 5042 (2019).
37. Doxtader, K. A. et al. Structural basis for regulation of METTL16, an S-adenosylmethionine homeostasis factor. *Mol. Cell* **71**, 1001–1011.e1004 (2018).
38. Akichika, S. et al. Cap-specific terminal N(6)-methylation of RNA by an RNA polymerase II-associated methyltransferase. *Science* **363**, eaav0080 (2019).
39. Yankova, E. et al. Small-molecule inhibition of METTL3 as a strategy against myeloid leukaemia. *Nature* **593**, 597–601 (2021).
40. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. Sect. D-Biol. Crystallogr.* **60**, 2126–2132 (2004).
41. Potterton, E., Briggs, P., Turkenburg, M. & Dodson, E. A graphical user interface to the CCP4 program suite. *Acta Crystallogr. Sect. D-Biol. Crystallogr.* **59**, 1131–1137 (2003).
42. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213–221 (2010).
43. Ashkenazy, H. et al. ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Res.* **44**, W344–W350 (2016).
44. Laskowski, R. A. & Swindells, M. B. LigPlot+: multiple ligand-protein interaction diagrams for drug discovery. *J. Chem. Inf. Model* **51**, 2778–2786 (2011).
45. Kechin, A., Boyarskikh, U., Kel, A. & Filipenko, M. cutPrimers: a new tool for accurate cutting of primers from reads of targeted next generation sequencing. *J. Comput. Biol.* **24**, 1138–1143 (2017).
46. Kim, D., Langmead, B. & Salzberg, S. L. HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods* **12**, 357–360 (2015).

47. Li, H. et al. The sequence alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
48. Ramirez, F. et al. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* **44**, W160–W165 (2016).
49. Thorvaldsdottir, H., Robinson, J. T. & Mesirov, J. P. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief. Bioinform.* **14**, 178–192 (2013).
50. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).
51. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
52. Dennis, G. Jr et al. DAVID: database for annotation, visualization, and integrated discovery. *Genome Biol.* **4**, P3 (2003).
53. Yu, G., Wang, L. G., Han, Y. & He, Q. Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* **16**, 284–287 (2012).
54. Shen, S. et al. rMATS: robust and flexible detection of differential alternative splicing from replicate RNA-Seq data. *Proc. Natl Acad. Sci. USA* **111**, E5593–E5601 (2014).

## Acknowledgements

We thank the staff at the Beamlines BL17U, BL18U, and BL19U of SSRF for assistance with data collections, and Y. Gu at the Beamline BL18U of SSRF for help with data processing. We thank C. Cao at Shanghai Institute of Organic Chemistry and H. Liu for the help with synthesis and purification of RNA substrate for crystallization. This work was supported by the National Natural Science Foundation of China (31971130 and 31230041 to J.Ma, 31930017 to A.D.). W.H.S.'s collaboration work was conducted within the context of the International Associated Laboratory on Plant Epigenome Research (LIA PER). Y.S. is an American Cancer Society Research Professor.

## Author contributions

Q.L. and J.Ma conceived the study. Q.L. expressed, purified Arabidopsis METTL4 proteins, and grew crystals, J.Mo. and Z.L. performed U2 snRNA purification and the UHPLC-MS/MS analysis, H.C. performed the in vitro enzymatic assay with help from Y.Z. and X.S., Q.L. and Z.H. synthesized the RNA substrates, purified mutants and performed in vitro binding assay. W.X. analyzed the RNA-seq data and identified global alternative splicing events. Q.L. and J.Ma collected X-ray diffraction data and solved the structures. B.W., J.W., K.D., M.P., Y.L., and T.L. extracted RNAs, identified plant phenotypes and constructed RNA-seq libraries, plant expression constructs and transgenic plants. Q.L., J.Mo, H.C., and J.Ma wrote the manuscript with input from H.W., A.D., Y.S., and W.S. J.Ma

supervised the structural and in vitro biochemical study. H.W. supervised the UHPLC-MS/MS analysis. A.D. supervised the plant study. Y.S. supervised in vitro enzymatic assays.

## Competing interests

Y.S. is a co-founder and member of the Scientific Advisory Board of K36 Therapeutics. Y.S. is also a member of the Scientific Advisory Board of EPICRISPR BIOTECHNOLOGIES, INC, the College of Life Sciences, West Lake University, and a member of the MD Anderson External Advisory Board. Y.S. is a scientific consultant for CBio-X Holdings, Inc., and holds equity in Imago Biosciences, Active Motif, K36 Therapeutics and Epicrispr Biotechnologies, Inc. All other authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41467-022-33277-x>.

**Correspondence** and requests for materials should be addressed to Yang Shi, Aiwu Dong, Hailin Wang or Jinbiao Ma.

**Peer review information** *Nature Communications* thanks Zofia Szweykowska-Kulinska, Kozo Tomita and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022