

## ARTICLE


<https://doi.org/10.1038/s41467-021-22665-4>

OPEN

# DNMT1 reads heterochromatic H4K20me3 to reinforce LINE-1 DNA methylation

Wendan Ren<sup>1,12</sup>, Huitao Fan<sup>2,3,12</sup>, Sara A. Grimm<sup>4</sup>, Jae Jin Kim<sup>5</sup>, Linhui Li<sup>1</sup>, Yiran Guo<sup>2,3</sup>, Christopher James Petell<sup>2,3</sup>, Xiao-Feng Tan<sup>1</sup>, Zhi-Min Zhang<sup>1</sup>, John P. Coan<sup>6</sup>, Jiekai Yin<sup>7</sup>, Dae In Kim<sup>5</sup>, Linfeng Gao<sup>7</sup>, Ling Cai<sup>2,3</sup>, Nelli Khudaverdyan<sup>1</sup>, Burak Çetin<sup>8</sup>, Dinshaw J. Patel<sup>9</sup>, Yinsheng Wang<sup>7,10</sup>, Qiang Cui<sup>11</sup>, Brian D. Strahl<sup>2,3</sup>, Or Gozani<sup>6</sup>, Kyle M. Miller<sup>5</sup>, Seán E. O'Leary<sup>1,8</sup>, Paul A. Wade<sup>4</sup>, Gang Greg Wang<sup>2,3</sup> & Jikui Song<sup>1,7</sup>

DNA methylation and trimethylated histone H4 Lysine 20 (H4K20me3) constitute two important heterochromatin-enriched marks that frequently cooperate in silencing repetitive elements of the mammalian genome. However, it remains elusive how these two chromatin modifications crosstalk. Here, we report that DNA methyltransferase 1 (DNMT1) specifically 'recognizes' H4K20me3 via its first bromo-adjacent-homology domain (DNMT1<sub>BAH1</sub>). Engagement of DNMT1<sub>BAH1</sub>-H4K20me3 ensures heterochromatin targeting of DNMT1 and DNA methylation at LINE-1 retrotransposons, and cooperates with the previously reported readout of histone H3 tail modifications (i.e., H3K9me3 and H3 ubiquitylation) by the RFTS domain to allosterically regulate DNMT1's activity. Interplay between RFTS and BAH1 domains of DNMT1 profoundly impacts DNA methylation at both global and focal levels and genomic resistance to radiation-induced damage. Together, our study establishes a direct link between H4K20me3 and DNA methylation, providing a mechanism in which multivalent recognition of repressive histone modifications by DNMT1 ensures appropriate DNA methylation patterning and genomic stability.

<sup>1</sup>Department of Biochemistry, University of California, Riverside, CA, USA. <sup>2</sup>Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill School of Medicine, Chapel Hill, NC, USA. <sup>3</sup>Department of Biochemistry and Biophysics, University of North Carolina at Chapel Hill School of Medicine, Chapel Hill, NC, USA. <sup>4</sup>Division of Intramural Research, Epigenetics and Stem Cell Biology Laboratory, National Institute of Environmental Health Sciences, Research Triangle Park, Durham, NC, USA. <sup>5</sup>Department of Molecular Biosciences, LIVESTRONG Cancer Institute of the Dell Medical School, Institute for Cellular and Molecular Biology, University of Texas at Austin, Austin, TX, USA. <sup>6</sup>Department of Biology, Stanford University, Stanford, CA, USA. <sup>7</sup>Environmental Toxicology Graduate Program, University of California, Riverside, CA, USA. <sup>8</sup>Cell, Molecular, and Developmental Biology Graduate Program, University of California, Riverside, Riverside, CA, USA. <sup>9</sup>Structural Biology Program, Memorial Sloan Kettering Cancer Center, New York, NY, USA. <sup>10</sup>Department of Chemistry, University of California, Riverside, CA, USA. <sup>11</sup>Departments of Chemistry, Physics and Biomedical Engineering, Boston University, Boston, MA, USA. <sup>12</sup>These authors contributed equally: Wendan Ren, Huitao Fan. ✉email: [greg\\_wang@med.unc.edu](mailto:greg_wang@med.unc.edu); [jikui.song@ucr.edu](mailto:jikui.song@ucr.edu)

The eukaryotic genome is organized into different functional compartments, with the assembly of heterochromatin regulated by both DNA methylation and repressive histone modifications, such as H3 trimethylated at lysine 9 (H3K9me3) and H4 trimethylated at lysine 20 (H4K20me3)<sup>1</sup>. The signaling cascades invoked by these modifications coordinately underpin fundamental biological processes concerning genome compartmentalization, gene silencing, genomic stability, cell differentiation, and development<sup>2,3</sup>. In cancer, depletion of H4K20me3 is closely associated with DNA hypomethylation in repetitive sequences, such as long interspersed nuclear element-1 (LINE-1)<sup>4</sup>, resulting in genomic instability and/or aberrant gene expression<sup>5–7</sup>. Dysregulation of H4K20me3 and DNA methylation has also been associated with neurological and developmental disorders, such as fragile X syndrome<sup>8,9</sup> and Hutchinson–Gilford Progeria Syndrome (HGPS)<sup>10</sup>. However, it remains far from clear how these two gene-repressive epigenetic modifications cooperate in determining specific chromatin states during normal and pathological development.

In mammals, DNA methylation is stably propagated by DNA methyltransferase 1 (DNMT1) during mitotic division<sup>11,12</sup>. DNMT1-mediated DNA methylation maintenance is supported in part by its enzymatic preference for hemimethylated CpG DNA, the so-called maintenance methylation activity<sup>11</sup>. In addition, increasing evidence has suggested a role of de novo methylation activity of DNMT1 in maintaining DNA methylation patterns at H3K9me2/3-enriched<sup>13</sup> or paternal imprinting control regions<sup>14</sup>. DNMT1 contains a C-terminal methyltransferase (MTase) domain, preceded by several regulatory domains including a replication-foci-targeting sequence (RFTS), a CXXC zinc finger domain and a pair of bromo-adjacent homology (BAH) domains (Fig. 1a)<sup>15–18</sup>. Previous studies have demonstrated that the RFTS and CXXC domains serve to “sense” specific chromatin cues, which in turn influences DNA methylation maintenance via allosteric regulations<sup>15,17–21</sup>. For instance, the DNMT1 RFTS domain (DNMT1<sub>RFTS</sub>) recognizes ubiquitinated histone H3 (H3Ub)<sup>22–25</sup> and PCNA-associated factor 15 (PAF15)<sup>26,27</sup> during S phase, which leads to cell cycle-specific chromatin targeting and enzymatic stimulation of DNMT1<sup>23,26,28</sup>. Our recent study further demonstrated that DNMT1<sub>RFTS</sub> also directly recognizes H3K9me3, thereby reinforcing the DNMT1<sub>RFTS</sub>–H3Ub readout for the DNA methylation maintenance at heterochromatic regions<sup>29</sup>. Likewise, the DNMT1 CXXC domain specifically recognizes unmodified CpG DNA to control DNMT1-mediated de novo DNA methylation<sup>15,30</sup>. It has been demonstrated that a ~30-amino acid-long linker connecting the CXXC and BAH domains, termed autoinhibitory linker (Fig. 1a)<sup>15</sup>, plays a pivotal role in the transition of DNMT1 between different conformational and functional states<sup>15,18</sup>. However, the mechanism by which these or other regulatory elements are coordinated in fine-tuning the activity of DNMT1 at discrete chromatin regions remains poorly understood.

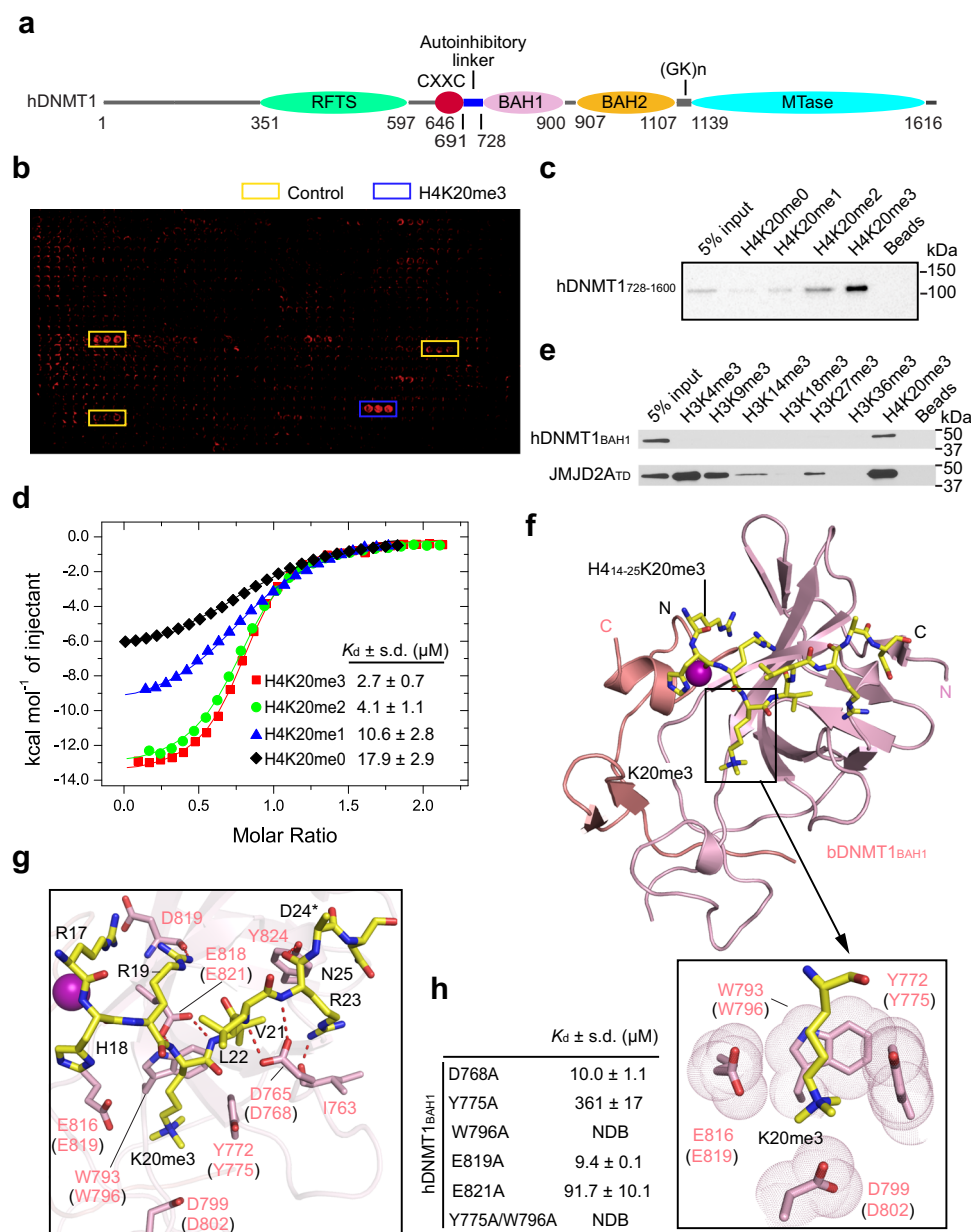
The BAH domain belongs to an evolutionarily conserved class of reader modules for histone and non-histone proteins<sup>31</sup>. Our previous study identified that the BAH domain of origin recognition complex subunit 1 (ORC1<sub>BAH</sub>) specifically recognizes H4K20 dimethylation (H4K20me2), thereby regulating ORC1-chromatin association and DNA replication initiation<sup>32</sup>. More recently, the BAH domains of two homologous plant proteins, EARLY BOLTING IN SHORT DAYS (EBS) and SHORT LIFE (SHL), effector of polycomb repression 1 (EPR-1) in *Neurospora crassa*, as well as BAH domain and coiled-coil containing 1 (BAHCC1) in human cells, have been shown to mediate Polycomb gene repression via specific readout of H3K27 trimethylation (H3K27me3)<sup>4,33–35</sup>. In DNMT1, the two BAH domains are

both associated with the MTase domain to form an integrated structural unit<sup>15</sup>. Although previous studies indicated that these domains play a role in targeting DNMT1 to replication foci<sup>14,36</sup>, their functions remain uncharacterized to date.

To elaborate how the BAH domains regulate DNMT1-mediated DNA methylation, we examined the histone binding activity of DNMT1 BAH domains and its relationship to the chromatin association and enzymatic activity of DNMT1. We identified the first BAH domain of DNMT1 (DNMT1<sub>BAH1</sub>) as a reader for H4K20me3. Strikingly, the H4K20me3 binding by DNMT1<sub>BAH1</sub> causes displacement of the autoinhibitory linker, which frees DNMT1 from the linker-mediated autoinhibition and leads to allosteric stimulation of DNMT1, reminiscent of what was previously observed for the DNMT1<sub>RFTS</sub>–H3K9me3Ub interaction<sup>23,29</sup>. Consistently, single-molecule Förster resonance energy transfer (smFRET) analysis revealed that the DNMT1<sub>RFTS</sub>–H3K9me3Ub and DNMT1<sub>BAH1</sub>–H4K20me3 interactions both trigger fast conformational dynamics of DNMT1, promoting the transition of DNMT1 into an open conformation. In cells, the DNMT1<sub>BAH1</sub> mutation, which is defective in recognition of H4K20me3 and yet hyperactive due to the disrupted association between DNMT1<sub>BAH1</sub> and the autoinhibitory linker, exerts a dual effect on DNA methylation—it causes DNA hypomethylation within the H4K20me3-positive LINE-1, but DNA methylation gains at genomic regions lacking H4K20me3, thereby demonstrating a role for DNMT1<sub>BAH1</sub> in shaping the landscape of DNA methylation. Finally, we also found that RFTS- and BAH1-mediated DNMT1 regulations cooperate to maintain proper levels of DNA methylation and genome stability of cells. Together, this work establishes a direct link between H4K20me3 and DNMT1-mediated DNA methylation, providing mechanistic insights into DNA methylation maintenance at LINE-1 and other heterochromatic regions.

## Results

**The DNMT1<sub>BAH1</sub> module is a reader for H4K20me3.** To explore the putative reader activity of DNMT1’s BAH modules, we performed histone-binding screening using a C-terminal fragment of hDNMT1, comprised of the BAH and MTase domains (residues 728–1600, hDNMT1<sub>728–1600</sub>; Fig. 1a), on a peptide microarray containing 384 unmodified or uniquely modified histone peptides. We found that DNMT1<sub>728–1600</sub> binds to the H4K20me3 peptide with the highest affinity (Fig. 1b). This observation was subsequently confirmed by individual peptide pull-down assays, in which GST-tagged hDNMT1<sub>728–1600</sub> showed a strong preference for H4K20me3 and H4K20me2 over H4K20me1 and H4K20me0 (Fig. 1c). Peptide pull-down and isothermal titration calorimetry (ITC) assays further narrowed down the interaction region to the BAH1 domain (residues 728–900, hDNMT1<sub>BAH1</sub>) (Fig. 1d and Supplementary Fig. 1a and Supplementary Table 1), with preference for H4K20me3 over other lysine-trimethylated histone peptides (Fig. 1e and Supplementary Fig. 1b). Along the line, titration of hDNMT1<sub>BAH1</sub> with the H4<sub>14–25</sub>K20me3 and H4<sub>14–25</sub>K20me2 peptides gave dissociation constants ( $K_d$ ) of 2.7 and 4.1  $\mu$ M, respectively, compared to the  $K_d$  values of 10.6 and 17.9  $\mu$ M determined for hDNMT1<sub>BAH1</sub> with H4<sub>14–25</sub>K20me1 and H4<sub>14–25</sub>K20me0, respectively (Fig. 1d and Supplementary Table 1). The binding affinities of hDNMT1<sub>BAH1</sub> for H4K20me3 and H4K20me2 are comparable with those of previously identified H4K20me3 and H4K20me2 readers, notably the JMJD2A tudor domain (JMJD2A<sub>TD</sub>)<sup>37</sup> and the ORC1<sub>BAH</sub> domain<sup>32</sup>, respectively (Supplementary Fig. 1c, d and Supplementary Table 1). Together, these data establish that the hDNMT1<sub>BAH1</sub> domain recognizes highly methylated H4K20 with a preference for H4K20me3.



**Fig. 1 Biochemical and structural analysis of the DNMT1 BAH1-H4K20me3 interaction.** **a** Domain architecture of human DNMT1 (hDNMT1), with individual domains delimited by residues numbers. **b** Histone peptide array analysis of the interactions between hDNMT1<sub>728-1600</sub> and histone peptides. **c** Peptide pull-down analysis of the interactions of hDNMT1<sub>728-1600</sub> with different H4K20me peptides. The experiment was repeated three times with consistent results. **d** ITC-binding assays for hDNMT1<sub>BAH1</sub>, titrated with different H4K20me peptides. **e** Peptide pull-down analysis of the interactions between hDNMT1<sub>BAH1</sub> and different tri-methylated histone peptides. The experiment was repeated three times with consistent results. **f** Crystal structure of bDNMT1<sub>BAH1</sub> (light pink) and H4<sub>14-25</sub>K20me3 peptide (yellow), with the H4K20me3-binding pocket shown in the expanded view (boxed). The last two  $\beta$ -strands, which were domain-swapped during crystallization, are colored in salmon. The zinc ion is shown as a purple sphere. **g** Close-up view of the intermolecular interactions between bDNMT1<sub>BAH1</sub> and the H4<sub>14-25</sub>K20me3 peptide. The equivalent pocket residues in hDNMT1<sub>BAH1</sub> are labeled in parentheses in (f, g). The hydrogen bonds are shown as dashed lines. \*The side chain of H4 D24, except for the C $\beta$  atom, is untraceable. **h**  $K_d$  values of hDNMT1<sub>BAH1</sub> mutants and H4<sub>14-25</sub>K20me3 peptides, derived from the ITC binding assays. NDB, no detectable binding. Source data are provided as a Source Data file.

**Structural details for the Interaction between DNMT1 BAH1 Domain and H4K20me3.** To understand the structural basis for the DNMT1<sub>BAH1</sub>-H4K20me3 interaction, we determined the crystal structure of the BAH1 domain of bovine DNMT1 (bDNMT1<sub>BAH1</sub>), which shares ~94% sequence identity with hDNMT1<sub>BAH1</sub> (Supplementary Fig. 1e), in complex with the H4<sub>14-25</sub>K20me3 peptide at 2.65 Å resolution (Fig. 1f and Table 1). The structure of bDNMT1<sub>BAH1</sub> reveals a twisted  $\beta$ -barrel resembling hDNMT1<sub>BAH1</sub> (Fig. 1f and Supplementary Fig. 1f,

g). The H4<sub>14-25</sub>K20me3 peptide extends along one end of the  $\beta$ -barrel (Fig. 1f), with the side chain of H4K20me3 inserted into a pocket formed by Y772, W793, D799, and E816 of bDNMT1<sub>BAH1</sub> (Fig. 1f, boxed). Flanking the H4<sub>14-25</sub>K20me3 peptide, the side chains of bDNMT1<sub>BAH1</sub> D765 and E818 each form a bidentate hydrogen bond with the backbone amides of H4, and the backbone of bDNMT1<sub>BAH1</sub> D765 is further hydrogen bonded to the side chain of H4 R23 (Fig. 1g). Additional intermolecular interactions involve bDNMT1<sub>BAH1</sub> I763, D819, and Y824 and H4 R17,

**Table 1** Data collection and refinement statistics.

|                                                     | <b>bDNMT1<sub>BAH1</sub>-H4K20me2<br/>(PDB: 7LMM)</b> | <b>bDNMT1<sub>BAH1</sub>-H4K20me3<br/>(PDB: 7LMK)</b> |
|-----------------------------------------------------|-------------------------------------------------------|-------------------------------------------------------|
| <i>Data collection</i>                              |                                                       |                                                       |
| Space group                                         | P 2 <sub>1</sub> 2 <sub>1</sub>                       | P 2 <sub>1</sub> 2 <sub>1</sub>                       |
| <i>Cell dimensions</i>                              |                                                       |                                                       |
| <i>a</i> , <i>b</i> , <i>c</i> (Å)                  | 70.7, 80.3, 130.4                                     | 71.1, 81.4, 129.6                                     |
| $\alpha$ , $\beta$ , $\gamma$ (°)                   | 90.00, 90.00, 90.00                                   | 90.00, 90.00, 90.00                                   |
| Resolution (Å)                                      | 34.91–2.80(2.90–2.80) <sup>a</sup>                    | 81.37–2.65 (2.71–2.65)                                |
| <i>R</i> <sub>merge</sub>                           | 0.117 (0.968)                                         | 0.073 (1.632)                                         |
| <i>I</i> / $\sigma$ ( <i>I</i> )                    | 20.8 (2.26)                                           | 15.7 (1.0)                                            |
| <i>CC</i> <sub>1/2</sub>                            | 0.998 (0.866)                                         | 0.998 (0.404)                                         |
| Completeness (%)                                    | 99.5 (99.8)                                           | 98.8 (97.6)                                           |
| Redundancy                                          | 12.9 (13.2)                                           | 4.4 (4.4)                                             |
| <i>Refinement</i>                                   |                                                       |                                                       |
| No. of reflections                                  | 18,807                                                | 22,119                                                |
| <i>R</i> <sub>work</sub> / <i>R</i> <sub>free</sub> | 0.207/0.267                                           | 0.206/0.258                                           |
| <i>No. of atoms</i>                                 |                                                       |                                                       |
| Protein                                             | 5168                                                  | 5114                                                  |
| Zn <sup>2+</sup>                                    | 4                                                     | 4                                                     |
| Water                                               | 36                                                    | 29                                                    |
| <i>B</i> factors (Å <sup>2</sup> )                  |                                                       |                                                       |
| Protein                                             | 68.4                                                  | 83.1                                                  |
| Zn <sup>2+</sup>                                    | 43.0                                                  | 68.8                                                  |
| Water                                               | 45.4                                                  | 66.8                                                  |
| <i>R.M.S. deviations</i>                            |                                                       |                                                       |
| Bond lengths (Å)                                    | 0.005                                                 | 0.004                                                 |
| Bond angles (°)                                     | 0.713                                                 | 0.691                                                 |

<sup>a</sup>Values in parentheses are for highest-resolution shell. Each structure was determined using the dataset collected from a single crystal.

R19, and R23, which mediate hydrogen bonding or electrostatic attractions (Fig. 1g).

To elucidate the H4K20me3-binding specificity, we also determined the crystal structure of bDNMT1<sub>BAH1</sub> in complex with the H4<sub>14–25</sub>K20me2 peptide at 2.8 Å resolution (Supplementary Fig. 2a and Table 1). The structure of bDNMT1<sub>BAH1</sub>-H4K20me2 reveals a similar interaction mode as that of bDNMT1<sub>BAH1</sub>-H4K20me3, with the side chain of H4K20me2 surrounded by the same set of H4K20me3-binding residues (Supplementary Fig. 2a, boxed). Distinct from the previously determined structure of the ORC1<sub>BAH</sub>-H4K20me2 complex in which the dimethylammonium group of H4K20me2 forms a hydrogen bond with a surrounding glutamate (ORC1<sub>BAH</sub> E93) (Supplementary Fig. 2b)<sup>32</sup>, the side chain of H4K20me2 in the bDNMT1<sub>BAH1</sub> complex is not within a hydrogen-bond distance with the side-chain carboxylates of D799 and E816 (Supplementary Fig. 2a), thus explaining why hDNMT1<sub>BAH1</sub> shows a relative lower binding affinity for H4K20me2 than for H4K20me3.

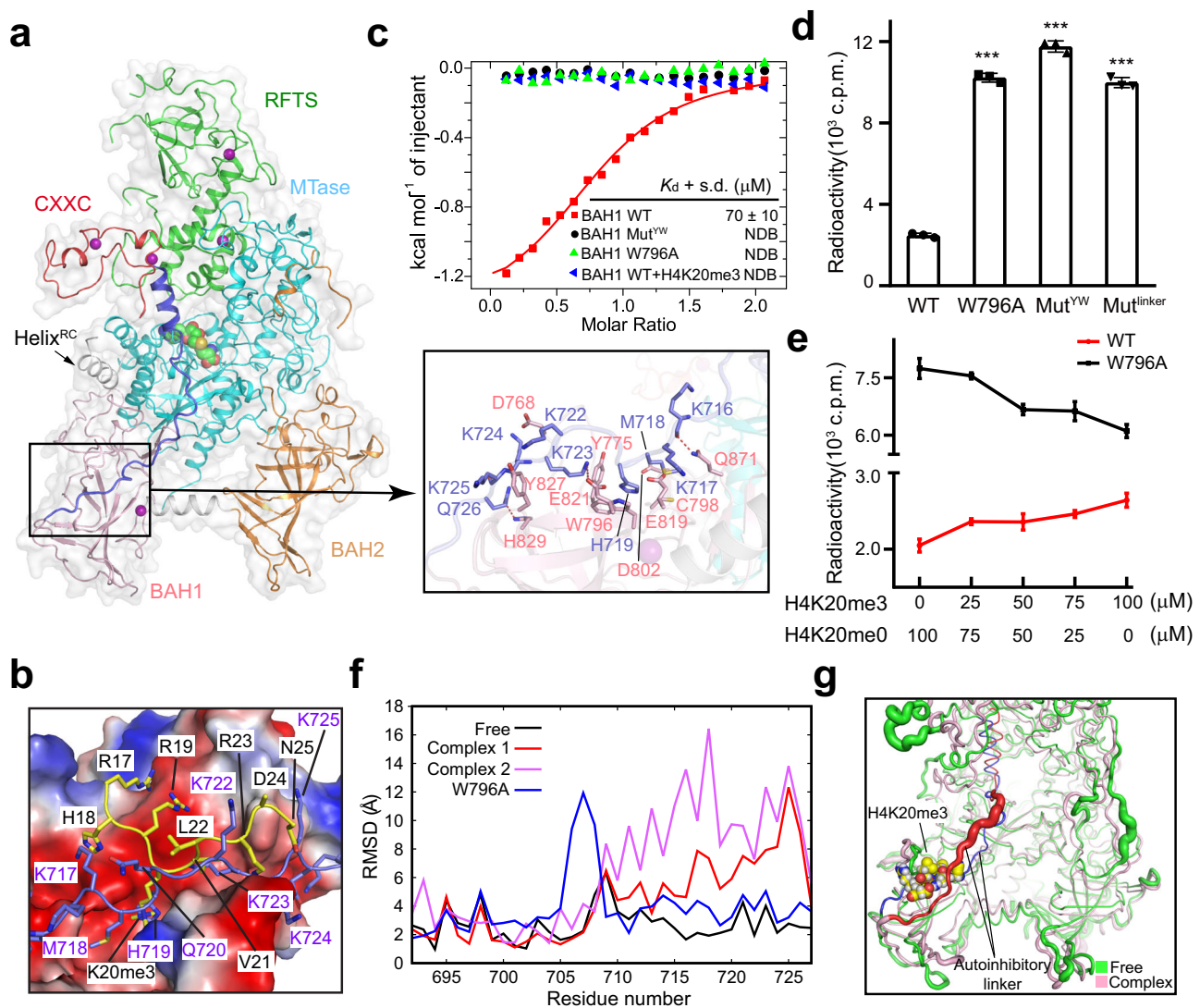
Sequence alignment of the DNMT1 BAH1 domain across different species revealed that the H4K20me2/3-binding sites fall into the most conserved regions (Supplementary Fig. 1e). Mutation of hDNMT1<sub>BAH1</sub> at or near the corresponding H4K20me3-binding pocket residues (D768A, Y775A, E819A, and E821A; Supplementary Fig. 1e) significantly reduced its binding affinity for the H4<sub>14–25</sub>K20me3 or H4<sub>14–25</sub>K20me2 peptide, while introduction of a W796A mutation largely abolishes the hDNMT1<sub>BAH1</sub>-H4<sub>14–25</sub>K20me3 or hDNMT1<sub>BAH1</sub>-H4<sub>14–25</sub>K20me2 binding (Fig. 1h and Supplementary Fig. 2c–l and Supplementary Table 1). These observations suggest that DNMT1 has evolved with a conserved mechanism for H4K20me2/3 recognition.

**H4K20me3 binding displaces the autoinhibitory linker for enzymatic stimulation.** Previously, the autoinhibitory linker (residues 692–727; blue in Fig. 2a) was shown to reinforce the

RFTS-mediated autoinhibition through both stabilization of the RFTS-MTase association and occlusion of the DNA substrate (Fig. 2a and Supplementary Fig. 3a)<sup>17,18,21</sup>. Interestingly, structural comparison of the bDNMT1<sub>BAH1</sub>-H4K20me3 complex with the histone-free hDNMT1<sub>351–1600</sub> (PDB 4WXX)<sup>18</sup> revealed that the H4K20me3-binding site of DNMT1<sub>BAH1</sub> is shielded by the C-terminal half of the autoinhibitory linker in free hDNMT1<sub>351–1600</sub>, via extensive intramolecular interactions (Fig. 2a, b), implying that the H4K20me3 binding to BAH1 would lead to displacement of the autoinhibitory linker, thereby providing a potential mechanism for releasing DNMT1 from autoinhibition. To explore this possibility, we first tested whether the autoinhibitory linker and H4K20me3 compete against each other for BAH1 binding. Using ITC assays, we found that the hDNMT1<sub>BAH1</sub> domain binds to a peptide derived from the autoinhibitory linker with a *K*<sub>d</sub> of ~70 μM (Fig. 2c); this binding was abolished in the presence of a twofold excess of H4K20me3 peptide (Fig. 2c). The ITC assay also reveals that, unlike the hDNMT1<sub>BAH1</sub> domain alone that binds to H4K20me3 peptide with a *K*<sub>d</sub> of 2.7 μM (Fig. 1d), the linker-containing DNMT1<sub>351–1600</sub> only weakly binds to the H4K20me3 peptide, with a *K*<sub>d</sub> of over 300 μM (Supplementary Fig. 3b). These data establish that the competition for BAH1 binding exists between the H4K20me3 peptide and the autoinhibitory linker in a histone-free form of DNMT1.

To elucidate the functional consequence of H4K20me3 binding, we generated mutants of the BAH1 domain, W796A, and Y775A/W796A (hereafter referred to as Mut<sup>YW</sup>), in which the associations of DNMT1<sub>BAH1</sub> with both H4K20me3 and the linker peptides are completely abolished (Fig. 2c and Supplementary Fig. 3b). In addition, we generated a sextuple mutation in the autoinhibitory linker, in which the BAH1-contacting residues M718, H719, K723, K724, K725, and Q726 were each replaced by alanine (hereafter referred to as Mut<sup>linker</sup>), to reduce the association of the autoinhibitory linker with the DNMT1 BAH1 domain without disruption of the H4K20me3-binding





**Fig. 2 H4K20me3 binding allosterically stimulates DNMT1 activity.** **a** Ribbon and surface diagram of the structure of hDNMT1<sub>351-1600</sub> (PDB 4WXX [10.2210/pdb4WXX/pdb]), with individual segments color coded. The RFTS-CXXC linker helix is denoted as Helix<sup>RC</sup>. The detailed interactions between the autoinhibitory linker and the BAH1 domain are shown in the expanded view. The hydrogen bonds are shown as dashed lines. **b** Close-up view of the structural overlay between hDNMT1<sub>351-1600</sub> (PDB 4WXX [10.2210/pdb4WXX/pdb]) and the bDNMT1<sub>BAH1</sub>-H4K20me3 complex, with the autoinhibitory linker and the H4K20me3 peptide colored in slate and yellow, respectively, and the corresponding binding site in hDNMT1<sub>351-1600</sub> shown as electrostatic surface. The bDNMT1<sub>BAH1</sub> domain was removed for clarity. **c** ITC binding analysis of the interaction between hDNMT1<sub>BAH1</sub>, either WT, W796A, or Mut<sup>YW</sup>, and the autoinhibitory linker peptide, in the absence of presence of the H<sub>414-25</sub>K20me3 peptide. **d** DNA methylation activity of hDNMT1<sub>351-1600</sub>, WT, W796A, Mut<sup>YW</sup>, or Mut<sup>linker</sup>. Data are mean  $\pm$  s.d. ( $n = 3$  biological replicates). Statistical analysis for WT vs. mutants used two-tailed Student's  $t$  test. \*\*\* $p < 0.001$ . **e** H4K20me3-dependent DNA methylation activity of hDNMT1<sub>351-1600</sub>, either WT or W796A. The total concentration of H<sub>414-25</sub>K20me0 and H<sub>414-25</sub>K20me3 peptides were maintained at 100  $\mu$ M to eliminate the substrate-competing effects of the H4 peptides. Data are mean  $\pm$  s.d. ( $n = 3$  biological replicates). **f** The RMSD values obtained for the autoinhibitory linker in the structural models of WT (PDB 4WXX [10.2210/pdb4WXX/pdb]) or W796A hDNMT1<sub>351-1600</sub> and the hDNMT1<sub>351-1600</sub>-H4K20me3 complex (complex 1 and complex 2) during 100 ns MD simulations. **g** Sausage view of the RMSF values of the autoinhibitory linker in peptide-free and H4K20me3-bound hDNMT1<sub>351-1600</sub> (complex 2), colored in blue and red, respectively. Source data are provided as a Source Data file.

affinity of the latter. ITC analysis revealed that the Mut<sup>linker</sup> mutation enhanced the hDNMT1<sub>351-1600</sub>-H4K20me3 binding affinity by two to threefold (Supplementary Fig. 3b and Supplementary Table 1), consistent with the notion that the autoinhibitory linker and H4K20me3 compete for BAH1 binding. Furthermore, in vitro DNA methylation assays indicate that W796A, Mut<sup>YW</sup>, and Mut<sup>linker</sup> mutations all lead to enhanced enzymatic activity of DNMT1 by four to fivefold (Fig. 2d), suggesting that disruption of the interactions between the BAH1 domain and the autoinhibitory linker helps relieve DNMT1's autoinhibition. Consistently, under fixed peptide concentration,

incubation with an increasing ratio of H<sub>414-25</sub>K20me3 vs. H<sub>414-25</sub>K20me0 gradually increased the DNA methylation efficiency of wild-type (WT) hDNMT1<sub>351-1600</sub>, but not the W796A mutant (Fig. 2e). The caveat of these studies is that the histone peptides used for the assays may not recapitulate the multivalent DNMT1-nucleosome interaction within the chromatin environment.

To further illustrate the H4K20me3 binding-triggered conformational change, we performed molecular dynamics (MD) simulation studies on hDNMT1<sub>351-1600</sub>, either WT or W796A, and the structural model of the hDNMT1<sub>351-1600</sub>-H4K20me3

complex. The hDNMT1<sub>351–1600</sub>–H4K20me3 complex, simulated under two independent trajectories, gave substantially elevated root-mean-square deviation (RMSD) and root-mean-square fluctuation (RMSF) for the autoinhibitory linker than free-state hDNMT1<sub>351–1600</sub> (Fig. 2f, g and Supplementary Fig. 3c). Likewise, a notable increase in the RMSD and RMSF of the autoinhibitory linker was observed for the W796A mutant (Fig. 2f and Supplementary Fig. 3c), further supporting the notion that linker detachment leads to the enzymatic activation of DNMT1.

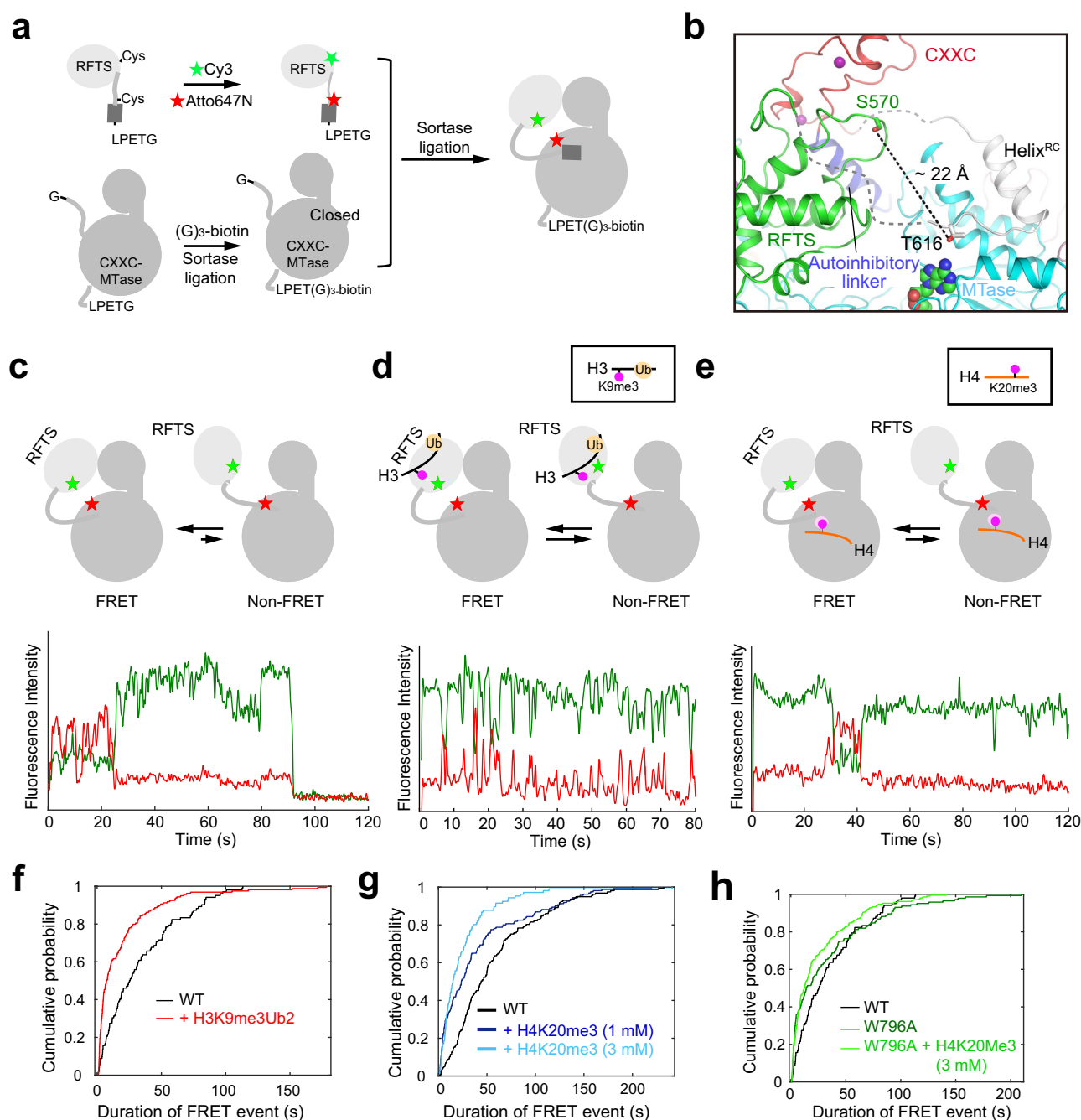
**BAH1–H4K20me3 and RFTS–H3K9me3Ub2 bindings both drive to a conformationally dynamic state of DNMT1.** The BAH1–H4K20me3 binding-mediated allosteric activation of DNMT1 is reminiscent of the previously reported RFTS–H3K9me3Ub2 readout<sup>29</sup>. We therefore asked how the BAH1–H4K20me3 and RFTS–H3K9me3Ub2 interactions crosstalk. Toward this direction, we performed smFRET experiments to interrogate how the RFTS–H3K9me3Ub and BAH1–H4K20me3 interactions impact the conformation of DNMT1. To measure the inter-domain movement between the RFTS and MTase domains, we introduced two mutation sites, one on the RFTS domain (S570C) and the other on the C-terminal end of the RFTS–CXXC linker (T616C), to a hDNMT1 fragment (residues 351–639 followed by a LPETG sequence) for statistical labeling with an equimolar mixture of FRET donor (Cy3) and acceptor (Atto647N) (Fig. 3a). Such a labeling strategy permitted close proximity between the two fluorophores in histone-free hDNMT1, taking advantage of the fact that the RFTS–CXXC linker is C-terminally anchored to the MTase domain through helical packing (Fig. 3b). This hDNMT1 fragment was then sortase-ligated<sup>38</sup> with a C-terminally biotinylated hDNMT1 fragment, creating a Cy3, Atto647N-labeled hDNMT1 fragment encompassing residues 351–1606 (denoted hDNMT1<sup>CY</sup>) for smFRET observation (Fig. 3c–e).

SmFRET between the two fluorophore labels was observed for wild-type hDNMT1<sup>CY</sup>, typically characterized by stable FRET observed from the beginning of the movie until acceptor photobleaching (Fig. 3c), which is consistent with the closed, autoinhibitory structure of the enzyme. Addition of the H3K9me3Ub2 peptide to surface-immobilized hDNMT1<sup>CY</sup> led to a dramatic reduction in the duration of FRET events, with a population of molecules cycling between non-FRET and FRET states (Fig. 3d). In comparison with the peptide-free FRET events, which had a single-exponential distribution of event durations, addition of the peptide therefore led to appearance of an additional, faster, exponential phase (Fig. 3f). These data confirm a previous MD analysis that showed enhanced mobility of the RFTS domain upon the H3Ub2 binding<sup>23</sup>, and suggest that the H3K9me3Ub2 binding promotes DNMT1 transition from a stable, autoinhibitory state to an activated state permitting fast “closed”–“open” conformational interconversion. Intriguingly, the presence of increasing concentrations of H4K20me3 peptide, but not H3K4me3 peptide, leads to a similar reduction in the duration of FRET events (Fig. 3e, g and Supplementary Fig. 3d–f), supporting the model that H4K20me3 binding influences the conformational dynamics of DNMT1. The W796A mutant showed an intrinsically reduced duration of FRET events (Fig. 3h and Supplementary Fig. 3g–i), regardless of the addition of H4K20me3 peptide (Fig. 3h), which reinforces the notion that the hyperactivity of this mutant arises from the detachment of the autoinhibitory linker from the BAH1 domain of DNMT1. Together, these smFRET data support a model where specific interactions, modulated by peptide binding to the RFTS domain

or the BAH1 domain, gate access to a conformationally dynamic, activated state of DNMT1.

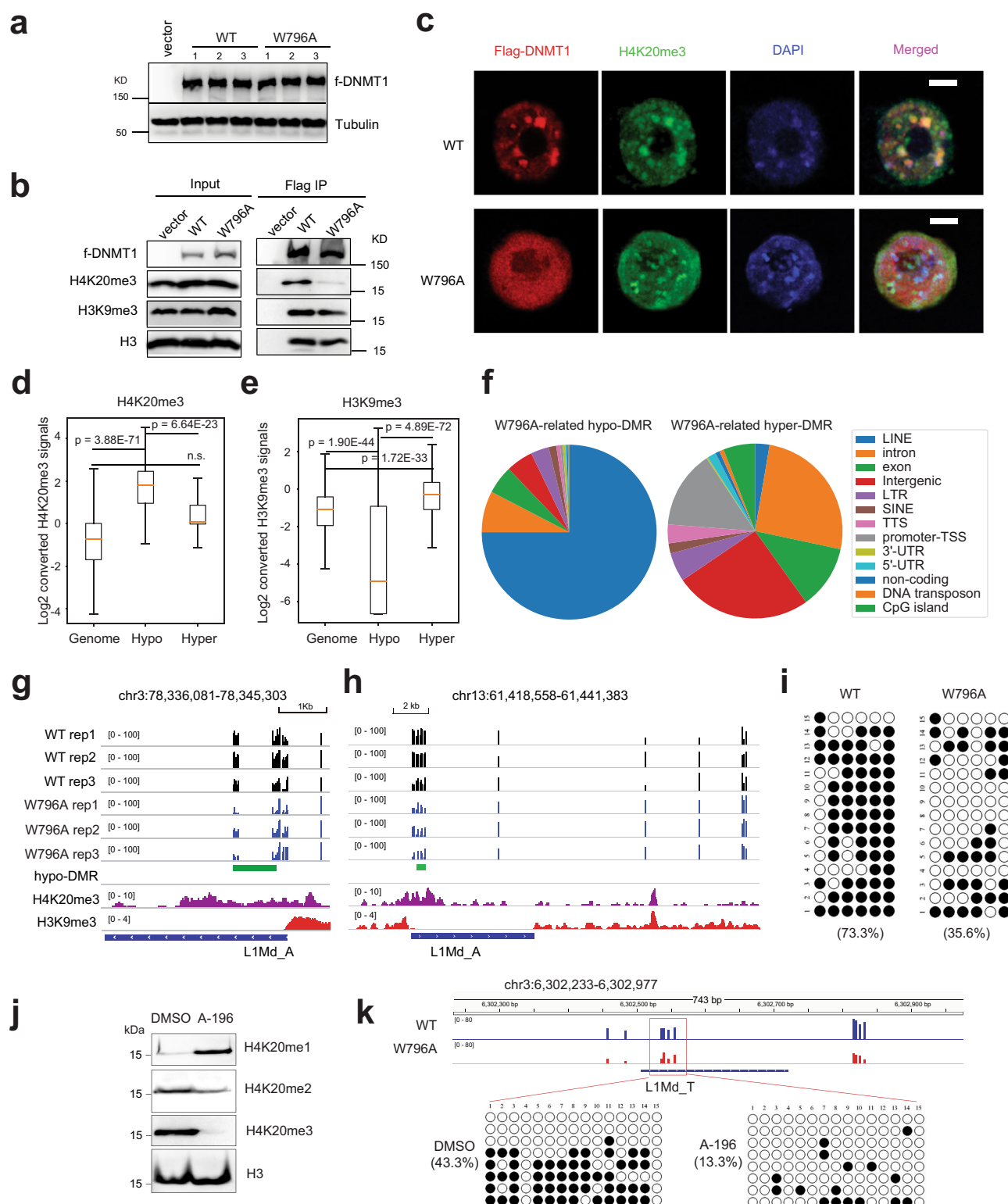
**BAH1–H4K20me3 binding potentiates CpG methylation deposition at H4K20me3-demarcated genomic regions, with the most striking effect seen at LINE-1 elements.** Given that *Dnmt1* deletion is lethal to dividing somatic cells, but not mouse ES cells<sup>39</sup>, we turned to mouse ES cells to examine the role of DNMT1<sub>BAH1</sub> in regulating DNA methylation maintenance. Using our recently reported gene complementation system<sup>29</sup>, we introduced comparable levels of exogenous DNMT1<sup>WT</sup> or a BAH1-defective mutant (DNMT1<sup>W796A</sup>) into *Dnmt1*-knockout mouse embryonic stem cells (1KO-ESCs) (Fig. 4a). In agreement with the above structural and biophysical assays of DNMT1<sub>BAH1</sub>, we found that DNMT1<sup>WT</sup> significantly co-precipitated with H4K20me3 (Fig. 4b), and as previously observed<sup>25</sup>, co-localizes with punctate DAPI-dense heterochromatin foci marked with H4K20me3 (Fig. 4c and Supplementary Fig. 4; upper panels). In contrast, DNMT1<sup>W796A</sup> shows much reduced co-precipitation with H4K20me3 (Fig. 4b) and the reduced H4K20me3 co-staining (Fig. 4c and Supplementary Fig. 4; lower panels).

Further, we carried out genome-wide methylation profiling with enhanced reduced representation bisulfite sequencing (eRRBS), which showed nearly complete bisulfite conversion (99.79–99.83%; Supplementary Fig. 5a), with at least 5-fold coverage for ~5–8 million of CpG sites in all samples (Supplementary Data 1). First, overall DNA methylation levels were found similar among independently derived 1KO-ESC lines with the reconstituted DNMT1<sup>WT</sup> or DNMT1<sup>W796A</sup> (Supplementary Fig. 5b, c and Supplementary Data 1), which was verified by global cytosine methylation quantification by liquid chromatography–mass spectrometry (LC–MS) (Supplementary Fig. 5d). Close inspection of eRRBS data, however, revealed that there was a slight but significant decrease in overall CpG methylation at the H4K20me3-demarcated genomic regions in 1KO-ESCs expressing DNMT1<sup>W796A</sup>, relative to the DNMT1<sup>WT</sup> controls, with the observed decreases being more apparent and significant at regions with the highest H4K20me3 (Supplementary Fig. 5e). In contrast, there was a slight but significant increase of overall CpG methylation at randomized control regions that lack H4K20me3 (Supplementary Fig. 5f, g). We have further defined differentially methylated regions (DMRs), either hypomethylated (hypo-DMR) or hyper-methylated (hyper-DMR), by comparing eRRBS profiles of 1KO-ESCs with DNMT1<sup>W796A</sup> vs. DNMT1<sup>WT</sup>. With two independent DMR calling methods (Supplementary Data 2 and 3), we consistently observed a significant enrichment of the DNMT1<sup>W796A</sup>-associated hypo-DMRs at H4K20me3-marked regions, relative to genome background or DNMT1<sup>W796A</sup>-associated hyper-DMRs (Fig. 4d and Supplementary Fig. 5h). Meanwhile, compared to genome background, the DNMT1<sup>W796A</sup>-associated hypo-DMRs were found significantly depleted from the H3K9me3-marked regions (Fig. 4e and Supplementary Fig. 5i), indicating that different stimulating mechanisms exist to ensure optimal DNA methylation among H3K9me3- and H4K20me3-demarcated heterochromatin. As a control, CUT&RUN of H4K20me3 demonstrated no change of overall H4K20me3 levels in these cells, as exemplified by retrotransposon elements known to be targeted by H4K20me3<sup>40,41</sup> (Supplementary Fig. 5j, k). These genomic profiling results are in agreement with a notion revealed by our in vitro studies that the W796A mutation of BAH1 has dual effects, with one causing the impaired CpG methylations within H4K20me3-marked regions (due to disrupted association of DNMT1 with H4K20me3) and the other leading to a generally



**Fig. 3 smFRET analysis of the conformational dynamics of DNMT1 upon H3K9me3Ub2 or H4K20me3 binding.** **a** Schematic view of the preparation of Cy3/Atto647N-labeled hDNMT1 fragment (hDNMT1<sup>CY</sup>) for smFRET analysis. The Cy3 and Atto647N dyes are shown as green and red stars. The Helix<sup>RC</sup> is shown as cylinder in dark gray. The RFTS domain and the rest of DNMT1 are colored light gray and gray, respectively. **b** Close-up view of the autoinhibited state of hDNMT1<sub>351-1600</sub> (PDB 4WXX [10.2210/pdb4WXX/pdb]), with the distance between S570 and T616 indicated by a dashed line in black. The linker connecting the RFTS and CXXC domains is in gray, with dashed lines indicating disordered regions. **c** (Top) Model for the conformational transition between autoinhibited and activated conformations of peptide-free wild-type hDNMT1<sup>CY</sup>. **d** (Top) Model for the conformational transition between autoinhibited and activated conformations of H3K9me3Ub2-bound hDNMT1<sup>CY</sup>. The H3K9me3 and ubiquitin marks are shown as magenta and wheat spheres, respectively. (Bottom) Representative smFRET trace observed for hDNMT1<sup>CY</sup> in the presence of H3K9me3Ub2. **e** (Top) Model for the conformational transition between autoinhibited and activated conformations of H4K20me3-bound hDNMT1<sup>CY</sup>. The H4K20me3 mark is shown as magenta sphere. (Bottom) Representative smFRET trace observed for hDNMT1<sup>CY</sup> in the presence of H4K20me3. **f** Cumulative probability of the durations of FRET events observed for hDNMT1<sup>CY</sup>, in the absence (black) or presence (red) of H3K9me3Ub2 peptide (97 and 93 molecules, respectively). **g** Cumulative probability of the durations of FRET events observed for hDNMT1<sup>CY</sup>, in the absence (black, 150 molecules) or presence of 1 mM (blue, 119 molecules), or 3 mM (cyan, 88 molecules) H4K20me3 peptide. **h** Cumulative probability of the durations of FRET events observed for peptide-free WT hDNMT1<sup>CY</sup> (black, 97 molecules), compared with W796A hDNMT1<sup>CY</sup> in the absence (dark green, 95 molecules) or presence (light green, 164 molecules) of 3 mM H4K20me3 peptide.





enhanced methylation (due to the reduced association between the mutated BAH1 and the autoinhibitory linker of DNMT1 and hence, hyperactivity of this mutant).

Interestingly, a majority of the DNMT1<sup>W796A</sup>-related hypo-DMRs were found localized within the LINE-1 class of retrotransposon elements (also known as L1; Fig. 4f and Supplementary Fig. 6a), especially the 5' *cis*-regulatory region of a newly evolved, non-truncated L1MdA subfamily (Fig. 4g, h and Supplementary Fig. 6b and Supplementary Data 4). LINE

elements are known to be targeted by H4K20me3 and DNA methylation<sup>40,41</sup>, and in our eRRBS data, had comparable coverage and sequencing depth relative to the Alu/SINE family of repetitive elements and gene-coding regions (Supplementary Fig. 6a, right panels). By individual bisulfite sequencing, we verified decreased CpG methylation at the examined regulatory region of L1 in cells with DNMT1<sup>W796A</sup> relative to DNMT1<sup>WT</sup> (Fig. 4i). Moreover, we used CRISPR/cas9-based gene editing technology to introduce into E14 ESCs a point mutation of



**Fig. 4 The BAH1-H4K20me3 binding regulates DNMT1-mediated methylation at the H4K20me3-enriched genomic regions.** **a** Immunoblots of the indicated Flag-tagged DNMT1 after its stable reconstitution into 1KO-ESCs among the independently derived cell lines. The experiment was repeated three times with consistent results. **b** CoIP (right panels) detecting association of the indicated Flag-tagged DNMT1 with H4K20me3 or histone H3. Left panels are immunoblots of input. The experiment was repeated three times with consistent results. **c** Representative confocal immunofluorescence images revealing localization of the indicated DNMT1 (Flag-tagged, red), H4K20me3-dense loci (green), and chromatin (stained by DAPI, blue) in the 1KO-ESC stable expression lines, synchronized in S phase. Scale bar, 5  $\mu$ m. The experiment was repeated three times with consistent results. **d, e** Box plots showing the levels of H4K20me3 (**d**) or H3K9me3 (**e**) in ESCs, genome-wide (left) or within DMRs found associated with the W796A mutation of DNMT1 relative to WT, either hypo (middle) or hyper-methylated (right) (for panels **d, e**,  $n = 422$  for Hypo, 377 for Hyper and 422 for randomized genome regions as a control). DMR calling is performed using DSS. Box-and whisker plots depict 25–75% in the box, whiskers are 10–90%, and median is indicated. Statistical analysis used two-tailed Student's *t* test. **f** Pie chart showing distribution of the DMRs associated with the W796A-mutated DNMT1 relative to WT, either hypo (left) or hyper-methylated (right), among the indicated genomic feature. DMR calling is performed using DSS. **g, h** Representative IGV views show CpG methylation levels at the called W796A-associated hypo-DMRs (green bar), either located in the chromosome 3 (**g**) or 13 (**h**), among three replicated 1KO-ESC lines with stable expression of WT (black) or the BAH1-mutated (blue) DNMT1. Cytosines covered by at least five reads according to eRRBS data are shown, with each site designated by a vertical line. The bottom panels show IGV view of H4K20me3 (purple) and H3K9me3 (red) in WT ESCs, and position of LINE-1. **i** Bisulfite sequencing analysis of CpG methylation at the regulatory region of L1 in cells transfected with DNMT1<sup>W796A</sup> or DNMT1<sup>WT</sup>. **j** Immunoblotting of H4K20 methylation (H4K20me1/2/3; with H3 as control) in E14 ESCs treated with 10  $\mu$ M A-196, vs. DMSO, for three days. The experiment was repeated twice with consistent results. **k** Site-targeted bisulfite sequencing results (bottom) revealed hypomethylation of a LINE-1 element (L1Md\_T) in E14 ESCs treated with A-196 vs. DMSO. Top panel shows an IGV view of eRRBS data at this L1Md\_T site, with a red box marking the CpG-containing region examined by site-targeted bisulfite sequencing. Source data are provided as a Source Data file.

W799A to endogenous Dnmt1 (equivalent to hDNMT1 W796A used above; Supplementary Fig. 6c, d), and subsequent individual bisulfite sequencing also detected decrease in CpG methylation at the examined L1 element (Supplementary Fig. 6e, f). Next, to further test a requirement of H4K20me3 for CpG methylation of LINE-1, we turned to A-196, a selective inhibitor of H4K20 methyltransferases SUV420H1 and SUV420H2<sup>42</sup>. As expected, A-196 treatment of E14 ESCs led to a global loss of H4K20me3, a significant decrease of H4K20me2 and a concomitant increase in H4K20me1 (Fig. 4j), which was concurrent with CpG hypomethylation of the examined L1 as well (Fig. 4k). Altogether, these results support that DNMT1<sup>BAH1</sup> binding to H4K20me3 potentiates DNMT1-mediated CpG methylation of H4K20me3-demarcated regions, notably those newly evolved non-truncated LINE-1 elements.

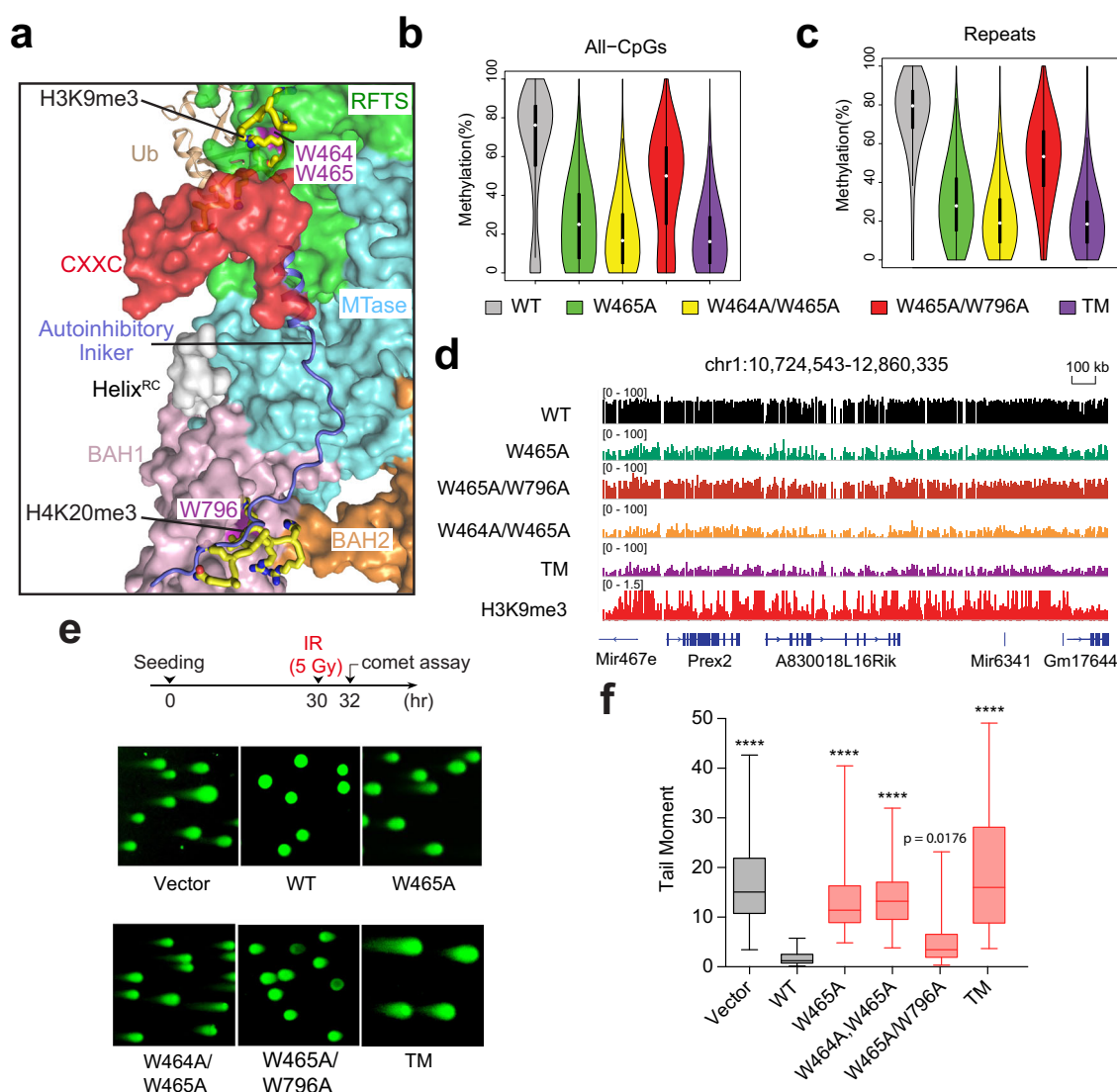
**BAH1 and RFTS regulations cooperate in fine-tuning DNMT1 activity.** Our recent study has identified that W464 and W465 within the DNMT1<sup>RFTS</sup> module are essential for H3K9me3Ub2 recognition (Fig. 5a); introduction of the W464A and W464/W465A mutation leads to severely impaired RFTS–H3K9me3Ub2 binding, and consequently, a large reduction of DNA methylation in cells<sup>29</sup>. To further examine the interplay between the RFTS- and BAH1-mediated regulations of DNMT1, we combined the W796A mutation of DNMT1<sup>BAH1</sup> together with the RFTS-defective mutation, either W465A or W464A/W465A, and subsequently used such RFTS/BAH1 compound mutants for DNMT1 reconstitution in 1KO-ESCs (Supplementary Fig. 7a). Intriguingly, both LC–MS analysis of cytosine methylation (Supplementary Fig. 7b) and eRRBS (Fig. 5b, c and Supplementary Fig. 7c–e) revealed that, relative to the single mutant of RFTS (DNMT1<sup>W465A</sup>)<sup>29</sup>, the compound mutant of RFTS/BAH1 (DNMT1<sup>W465A/W796A</sup>) displayed a markedly enhanced methylation activity in cells, leading to partial but significant rescue of the global CpG hypomethylation phenotype caused by DNMT1<sup>W465A</sup>, including the RFTS-related defects seen at the H3K9me3-marked regions as exemplified by sub-telomeric regions located at the chromosome 1 (Fig. 5d). Given the intrinsic link between the hyperactivity of DNMT1<sup>W796A</sup> and the BAH1-mediated DNMT1 activation, these data therefore show that the BAH1-mediated DNMT1 activation may partially rescue the CpG methylation defects caused by the RFTS dysfunction. Note that such a rescue effect was not observed for the RFTS W464A/W465A mutation in the

resultant triple mutation of DNMT1<sup>W464A/W465A/W796A</sup> (referred to as “TM”; Fig. 5b–d and Supplementary Fig. 7b–d), presumably due to the more severe impairment of chromatin targeting capability associated with this mutant<sup>29</sup>.

**RFTS and BAH1 coordinately regulate DNMT1-mediated genomic stabilization.** We further challenged 1KO-ESC cells, reconstituted with WT or mutant DNMT1, with ionizing radiation (IR) followed by the neutral comet assay (Fig. 5e). First, compared to 1KO-ESC cells rescued with WT DNMT1, those with vector control or a catalytically inactive DNMT1 mutant (C1226S) showed similar hyper-sensitivity to IR (Supplementary Fig. 7f–h), thus confirming a direct link between DNMT1-mediated DNA methylation and genomic stability. We next examined whether the BAH1 and RFTS mutations crosstalk during the IR response. As previously demonstrated<sup>29</sup>, cells with the RFTS W464A/W465A and W465A mutants displayed severe impairment of IR resistance (Fig. 5e, f). The compound mutation of DNMT1<sup>W464A/W465A/W796A</sup> (Fig. 5e, f) led to a stronger phenotype of DSB persistence following IR treatment, in line with severe hypomethylation seen with this TM mutation (Fig. 5b–d). In contrast, while cells with the RFTS single mutant (DNMT1<sup>W465A</sup>) exhibited the impaired genome integrity post-treatment of IR, introduction of an additional BAH1 mutation (DNMT1<sup>W465A/W796A</sup>) significantly rescued this defect (Fig. 5e, f), in line with the rescuing effect of the BAH1 W796A mutation manifested in LC–MS- and eRRBS-based methylome analysis (Fig. 5b–d and Supplementary Fig. 7b). Consistently, cell survival analysis revealed that the DNMT1<sup>W465A/W796A</sup> mutation led to a significant reduction in IR sensitivity in comparison with the DNMT1<sup>W465A</sup> mutation (Supplementary Fig. 7i). Thus, it is evident that DNMT1-mediated DNA methylation is involved in maintenance of genomic stability, the molecular detail of which awaits further investigation.

## Discussion

The spatio-temporal regulation of DNMT1-mediated DNA methylation is essential for the faithful inheritance of DNA methylation patterns during mitotic division. DNA methylation, in cooperation with other gene silencing mechanisms, such as histone modifications, provides a mechanism for the long-term stability of the heterochromatic state. However, how DNA methylation crosstalks with histone modifications remains elusive. Through a set of structural, biochemical, computational, and



**Fig. 5 RFTS and BAH1 domains crosstalk in DNMT1-mediated DNA methylation.** **a** Close-up view of the structurally overlaid hDNMT1<sub>351-1600</sub> (PDB 4WXX [10.2210/pdb4WXX/pdb]), bDNMT1<sub>RFTS</sub> bound to H3K9me3 and ubiquitin (PDB 6PZV [10.2210/pdb6PZV/pdb]), and the bDNMT1<sub>BAH1</sub>-H4K20me3 complexes, with the surface of individual DNMT1 domains color-coded in the same fashion as in Fig. 2a. hDNMT1 W464, W465, and W796 are colored purple. The autoinhibitory linker is colored blue. The H3K9me3 and H4K20me3 peptides are shown as yellow sticks. The ubiquitin molecule is shown in ribbon representation and colored in wheat. The bDNMT1<sub>RFTS</sub> and bDNMT1<sub>BAH1</sub> molecules were removed for clarity. **b, c** Violin plots showing distribution of absolute methylation levels for CpG sites with >5 coverage at all CpG sites (**b**) or those within the repeated genomic sequences (**c**) among the indicated cells samples. ( $n = 5,040,986$  CpG sites in panel (**b**);  $n = 812,251$  repeats in panel (**c**)). White dots are the median and box lines are the first and third quartile of the data. **d** Representative IGV view of CpG methylation at an H3K9me3-marked genomic region located in the chromosome 1 among three replicated 1KO-ESC lines with stable expression of the indicated DNMT1. Cytosines covered by at least ten reads according to eRRBS data (with three replications merged for each group) are shown, with each site designated by a vertical line. **e** Neutral comet assays, experimental scheme (top) and DNA breaks (bottom). **f** Quantified DNA breaks after ionizing radiation (IR) treatment of 1KO-ESC cells reconstituted with vector control or the indicated DNMT1. Box-and-whisker plots in panel depict 25–75% in the box, whiskers are 10–90%, and median is indicated. Data are mean  $\pm$  s.d. from >100 cells ( $n = 3$  biologically independent replicates). One-way ANOVA with post Tukey analysis was used. \*\*\*\* $p < 0.0001$ . The data for vector, WT, W465A, and W464A/W465A were adopted from those published previously<sup>29</sup>.

cellular analyses, this study addresses two critical questions regarding the functional regulation of DNMT1.

First, this study links the repressive histone modification H4K20me3 directly to the reader activity of DNMT1. Here, we show that DNMT1 directly recognizes H4K20me3 through the regulatory domain BAH1, which in turn mediates the chromatin association and enzymatic activation of DNMT1. The BAH1-H4K20me3 interaction clarifies a putative role of the BAH domains of DNMT1 in genomic targeting<sup>14,36</sup>, and provides a crucial safeguard mechanism for efficient DNA methylation at

H4K20me3-enriched regions, in particular the repetitive element LINE-1, thereby providing further insights into the mechanism underlying DNA methylation maintenance and heterochromatin formation. Distinct from the typical Kme3 readout that depends on aromatic or other hydrophobic residues only<sup>43</sup>, the BAH1 domain recognizes H4K20me3 via a pocket formed by mixed aromatic and acidic residues. In the DNA-free state of DNMT1, the H4K20me3-binding site is occluded by the C-terminal portion of the autoinhibitory linker, which likely reinforces the intramolecular RFTS-linker-MTase association (Supplementary Fig. 8a),

resulting in a direct link between the H4K20me3 binding and enzymatic activation of DNMT1. It is worth noting that de novo P767T mutation of DNMT1, located next to the H4K20me3-binding pocket of the BAH1 domain (Supplementary Fig. 8b), is reportedly linked to schizophrenia<sup>44</sup>. How the disease-associated DNMT1 BAH1 mutations affect genomic methylation and disease progression awaits further investigation.

While epigenetic silencing of repeated elements such as endogenous retroviruses (ERVs) and LINE-1 elements is crucial for maintaining integrity of the mammalian genome, the underlying silencing mechanisms are rather complex. Previous studies have pointed to a set of repressive complexes involved in H3K9me3<sup>45–47</sup> or DNA methylation<sup>48,49</sup> pathways. In this work, we uncovered that the interaction between the BAH1 domain of DNMT1 and H4K20me3 is especially important for optimal DNA methylation at the 5' *cis*-regulatory region of the newly evolved, non-truncated LINE-1 subfamily. A recent study also reported the repression of evolutionarily young LINE-1 by N<sup>6</sup>-methyladenine<sup>50</sup>. It is conceivable that complicated interplays exist among H3K9me3, H4K20me3, and DNA methylation by DNMT1 along with cofactors<sup>29</sup> and DNMT3A/3B (in complex with DNMT3L<sup>51,52</sup>). Further investigation is warranted to determine the relative contribution of these epigenetic pathways to repression of different subgroups of ERVs and LINE-1 elements.

The specific interaction between DNMT1 BAH1 and H4K20me3 adds to the growing list of BAH domain-mediated histone readouts, which include the H4K20me2 readout by the BAH domain of ORC1<sup>32</sup> and the H3K27me3 readout by the BAH domains of effector proteins in *Neurospora crassa*, plants and mammals<sup>4,33–35</sup>. Intriguingly, these BAH domains harbor a similar histone-binding surface, but with distinct binding modes (Supplementary Figs. 2b and 8c, d), highlighting the evolutionary divergence of the histone modification-binding mechanisms of this reader module family.

Second, this study delineates how DNMT1 transduces a multitude of environmental cues into its targeting and enzymatic activity. It has been established that Ubiquitin-like, containing PHD and RING Finger domains, 1 (UHRF1) plays a critical role in activating and targeting DNMT1 during S phase: UHRF1-mediated ubiquitylation of PAF15<sup>26,27</sup> and histone H3<sup>22–25</sup> serves to recruit DNMT1 to the replication foci during early and late S phase, respectively. This study demonstrates that the BAH1–H4K20me3 readout cooperates with the RFTS–H3K9me3/H3Ub readout in allosterically stimulating DNMT1, thereby uncovering another regulatory axis in DNA methylation maintenance in heterochromatin domains (Fig. 6). Disruption of these histone-directed

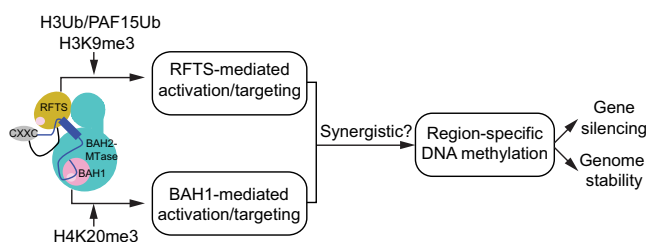
regulations in cells impairs the CpG methylation by DNMT1, leading to an aberrant landscape of DNA methylation and defects in the maintenance of genome stability, a phenotype known to be associated with cancer<sup>53</sup> and developmental disorders<sup>54</sup>. Note that neither RFTS- nor BAH1-mediated DNMT1 activation involves the discrimination of the methylation state of DNA substrates, providing a potential mechanism for compensating the “imprecise” maintenance methylation activity of DNMT1, thereby strengthening the region-specific methylation maintenance<sup>13,55</sup>. It is worth mentioning that maintenance of histone lysine (e.g., H3K9 and H4K20) methylation has been shown to be gradually established following S phase and may differ between parental and newly incorporated histones<sup>56–58</sup>. For instance, the H4K20 of recycled histones is dominantly methylated during replication, whereas the new histones only become methylated in the G2/M phase<sup>56</sup>. In this regard, it is likely that H4K20me3- and H3K9me3-mediated DNMT1 targeting and activation may be initiated by recycled parental nucleosomes and persist beyond S phase, providing a mechanism that cooperates with the S phase-specific, UHRF1-mediated regulation for efficient DNA methylation maintenance. Indeed, recent studies have indicated that the de novo methyltransferase activity of DNMT1, partially mediated by the BAH domains, is important in reinforcing DNA methylation maintenance in imprinting control or H3K9me3-marked regions<sup>13,14</sup>.

Together, this study uncovers a signaling pathway in which a BAH domain harbored within DNMT1 acts to transform H4K20me3-marked chromatin domains, in particular LINE-1 elements, into a more stable, DNA damage resistant, heterochromatic state with optimal levels of DNA methylation.

## Methods

**Plasmids.** The DNMT1 cDNA was purchased from Addgene (cat # 24952). For cellular assays, full-length DNMT1 with an N-terminal 3×Flag tag was inserted into the pPyCAGIP vector<sup>59</sup>, a kind gift of I. Chambers. All of the DNMT1 point mutations were generated by a QuikChange II XL Site-Directed Mutagenesis Kit (Agilent). For all the in vitro assays, DNA encoding the human DNMT1 BAH1 domain (residues 728–900, hDNMT1<sub>BAH1</sub>), the bovine DNMT1 BAH1 domain (residues 725–897, bDNMT1<sub>BAH1</sub>), or the JMJD2A-Tudor domain (amino acids 894–1011, JMJD2A<sub>TD</sub>) was cloned into a modified pRSF-Duet vector, which introduced an N-terminal His<sub>6</sub>-SUMO tag and ULP1 (ubiquitin-like protease 1) cleavage site. To facilitate crystallization, residues 838–858 of bDNMT1<sub>BAH1</sub> were replaced by a GAGSA sequence. For analysis of hDNMT1<sub>351–1600</sub> methylation activity and H4K20me3 binding, the interaction between hDNMT1<sub>BAH1</sub> and the autoinhibitory linker and the interaction between the hDNMT1 BAH2 domain (901–1107, hDNMT1<sub>BAH2</sub>) and H4K20me3, the hDNMT1 constructs were inserted into an in-house expression vector as a His<sub>6</sub>-MBP-tagged form. For peptide binding assay, hDNMT1<sub>BAH1</sub>, hDNMT1 fragment spanning the BAH1, BAH2, and MTase domains (residues 728–1600, hDNMT1<sub>728–1600</sub>) or JMJD2A<sub>TD</sub> (amino acids 895–1011) was cloned into pGEX6P-1 vector. For fluorescence labeling, an hDNMT1 fragment (residues 351–639), containing C409S/C420S/S570C/C580S/T616C mutations and a C-terminal Lys-Pro-Glu-Thr-Gly tail, referred herein to as hDNMT1<sub>RFTS-linker</sub>, was also cloned into the modified pRSF-Duet vector. For sortase ligation with hDNMT1<sub>RFTS-linker</sub>, an hDNMT1 fragment (residues 646–1606) containing a C-terminal Lys-Pro-Glu-Thr-Gly tail was cloned into the in-house expression vector preceded by a His<sub>6</sub>-MBP-tag. All plasmid sequences were verified by sequencing before use, with the residue numerations based on the isoform 1 of DNMT1 that contains 1616 amino acids. The primers used for cloning are listed in Supplementary Table 2.

**Protein purification.** The *E. coli* BL21(DE3) RIL cells (Novagen Inc) were used for protein expression. The cells harboring the expression plasmids were first cultured in LB medium at 37 °C until the OD<sub>600</sub> (optical density at 600 nm) reached 0.6. Totally, 0.1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) was then added to induce the protein expression at 16 °C overnight. After harvesting, the cells were resuspended and lysed in a buffer containing 50 mM Tris-HCl (pH 7.5), 25 mM imidazole, 1 M NaCl, 0.5 mM DTT and 1 mM PMSF. To purify the His<sub>6</sub>-SUMO-tagged hDNMT1<sub>BAH1</sub>, bDNMT1<sub>BAH1</sub> or JMJD2A<sub>TD</sub> protein, the clarified supernatant was applied to a nickel column and the fusion protein was eluted with buffer containing 25 mM Tris-HCl (pH 8.0), 100 mM NaCl and 300 mM imidazole. The eluted protein was then subjected to tag removal by ULP1 cleavage, anion exchange chromatography on a HiTrap Q XL column (GE Healthcare), nickel affinity chromatography, and size-exclusion chromatography on a HiLoad 16/600 Superdex 75 pg column (GE Healthcare) that was pre-equilibrated with buffer containing



**Fig. 6 A model for DNMT1 RFTS- and BAH1-regulated DNA methylation.**

Structural elements of DNMT1, including RFTS, CXXC, autoinhibitory linker and BAH1 are colored in corn, gray, blue and pink, respectively. The BAH2 and MTase domains are colored in cyan. Histone marks (H3K9me3 and H4K20me3) cooperate with UHRF1-generated H3Ub or PAF15Ub in activation and targeting of DNMT1 for optimal DNA methylation at specific genomic loci, which in turn leads to gene repression and genomic stabilization. Note that the H3K9me3Ub and H4K20me3 marks likely regulate the activity of DNMT1 in a synergistic manner.



20 mM Tris-HCl (pH 7.5), 50 mM NaCl and 5 mM DTT. For ITC binding assay, human ORC1<sub>BAH1</sub> (residues 1–185) was purified as described previously<sup>32</sup>. The His<sub>6</sub>-MBP tagged DNMT1 proteins were sequentially purified via Ni<sup>2+</sup> chromatography, ion-exchange chromatography on a Heparin HP (GE Healthcare) or Q HP column (GE Healthcare), TEV protease cleavage for tag removal, a second round of nickel affinity chromatography, and size-exclusion chromatography on a Superdex 200 16/600 column (GE Healthcare). The GST-tagged hDNMT1<sub>BAH1</sub> and JMJD2A<sub>TD</sub> for tri-methylated histone peptides-binding assay was purified through GST-affinity chromatography and eluted in 100 mM Tris-HCl (pH 8.0), 10 mg/mL reduced glutathione (Sigma-Aldrich). Purification of GST-hDNMT1<sub>728–1600</sub> fusion protein also involved ion-exchange chromatography on a Heparin HP column and size exclusion chromatography on a HiLoad 16/600 Superdex 200 pg column (GE Healthcare). The final protein sample was stored in a buffer containing 20 mM Tris-HCl (pH 7.5), 250 mM NaCl, 5% Glycerol, and 5 mM DTT. DNMT1 mutants were introduced by site-directed mutagenesis and purified as that described for wild-type protein. For preparation of H3K9me3Ub2 peptide, the His<sub>6</sub>-SUMO-tagged Ub(G76C) protein was purified through Ni<sup>2+</sup> chromatography, followed by tag removal via ULP1 cleavage and a second step of Ni<sup>2+</sup> chromatography. All purified protein samples were stored at –80 °C before use.

**Chemical modification of histone peptide.** The H3K9me3Ub2 peptide was generated and purified as previously described<sup>29</sup>, following a published protocol<sup>37</sup>. In essence, the H3<sub>1–24</sub>K9me3K18CK23C peptide was synthesized from LifeTein LLC, with an additional C-terminal tyrosine for spectroscopic quantification. To prepare the H3K9me3Ub2 peptide, the Ub(G76C) protein was mixed with the H3<sub>1–24</sub>K9me3K18CK23C peptide in a 4:1 molar ratio in buffer containing 250 mM Tris-HCl (pH 8.6), 8 M urea and 5 mM TCEP, and incubated at room temperature for 30 min. The cross-linker 1,3-dichloroacetone was then added to the reaction mixture with the amount equal to one-half of the total sulfhydryl groups. After 2 h-incubation on ice, the reaction was stopped by 5 mM β-Mercaptoethanol. The H3<sub>1–24</sub>K9me3Ub2 peptide was further purified using a mono S column (GE Healthcare). The caveat of this chemical modification method is that the crosslinking to be introduced is only an analog of the isopeptide bond between Ub and lysine.

**Crystallization and structure determination.** To prepare the bDNMT1<sub>BAH1</sub>-H4<sub>14–25</sub>K20me3 and bDNMT1<sub>BAH1</sub>-H4<sub>14–25</sub>K20me2 complexes, the bDNMT1<sub>BAH1</sub> protein was mixed with the H4<sub>14–25</sub>K20me3 or H4<sub>14–25</sub>K20me2 peptide, each containing an additional C-terminal tyrosine in a 1:3 molar ratio, and incubated on ice for 30 min. The bDNMT1<sub>BAH1</sub>-H4<sub>14–25</sub>K20me3 and bDNMT1<sub>BAH1</sub>-H4<sub>14–25</sub>K20me2 complexes were crystallized in a buffer containing 0.1 M HEPES (pH 7.5) (for bDNMT1<sub>BAH1</sub>-H4<sub>14–25</sub>K20me3) or 0.1 M MES (pH 6.5) (for bDNMT1<sub>BAH1</sub>-H4<sub>14–25</sub>K20me2), 13% PEG1500, 20 mM DTT, and 200 mM L-proline at 4 °C, using the hanging-drop vapor diffusion method. The crystals were soaked in the crystallization buffer supplemented with 20–25% (v/v) glycerol as cryoprotectant before flash frozen in liquid nitrogen. The X-ray diffraction data were collected on the beamline 24-ID-E, NE-CAT at Advanced Photo Source (APS) for bDNMT1<sub>BAH1</sub>-H4K20me3, and beamline BL92 at Stanford Synchrotron Radiation Lightsource (SSRL) for bDNMT1<sub>BAH1</sub>-H4K20me2. The data were indexed, integrated and scaled by HKL2000 program<sup>60</sup> or XDS<sup>61</sup>. The structures were solved by molecular replacement using PHASER<sup>62</sup> with the corresponding domain in the DNMT1 structure (PDB 4WXX [<https://doi.org/10.2210/pdb4WXX/pdb>]) as searching model. Iterative cycles of model rebuilding and refinement were carried out using COOT<sup>63</sup> and PHENIX<sup>64</sup>, respectively. Data collection and structure refinement statistics were summarized in Table 1.

**ITC binding assay.** ITC measurements were performed at 5 °C using a MicroCal iTC200 instrument (GE Healthcare). The peptides, including hDNMT1<sub>714–727</sub>, H4<sub>14–25</sub>K20me0, H4<sub>14–25</sub>K20me1, H4<sub>14–25</sub>K20me2 and H4<sub>14–25</sub>K20me3, were synthesized from LifeTein LLC, with an additional C-terminal tyrosine for spectroscopic quantification. To measure the bindings between the hDNMT1 BAH1 or BAH2 domain samples and H4K20me peptides, 1 mM H4K20me peptide was titrated over 0.1 mM hDNMT1<sub>BAH1</sub> or hDNMT1<sub>BAH2</sub> in buffer containing 25 mM Tris-HCl (pH 7.5), 100 mM NaCl and 2 mM fresh DTT. To measure the bindings between MBP-tagged hDNMT1<sub>351–1600</sub> proteins and H4<sub>14–25</sub>K20me3, 1.44 mM peptide was titrated over 72 μM MBP-hDNMT1<sub>351–1600</sub> in buffer containing 20 mM Tris-HCl (pH 7.5), 100 mM NaCl and 1 mM DTT at 5 °C. The spacing between injections was set to 180 s. To measure the binding between ORC1 BAH domain and H4K20me2 peptide, 0.45 mM peptide was titrated over 64 μM ORC1 BAH domain in buffer containing 20 mM Tris-HCl (pH 7.5), 100 mM NaCl and 1 mM fresh DTT. Analyses of all data were performed with MicroCal Origin software, fitted with single-site binding mode. The ITC binding parameters are summarized in Supplementary Table 1.

**Histone peptide microarrays and pulldowns.** Histone peptide microarrays and in-solution peptide pulldowns were performed as described previously<sup>32,65–67</sup>. To evaluate the bindings of DNMT1 to H4 peptides with different K20 methylation states, 25 pmol of the GST-hDNMT1<sub>728–1600</sub> was incubated with 250 pmol of biotinylated H4<sub>11–27</sub>K4me0, H4<sub>11–27</sub>K20me1, H4<sub>11–27</sub>K20me2, and H4<sub>11–27</sub>K20me3 peptide for 1 h. with rotation at 4 °C in Peptide Pulldown Buffer (50 mM Tris, 0.1%

NP-40, 0.5% bovine serum albumin (BSA), 150 mM NaCl, pH 8.0). After this, 2.5 μL of streptavidin magnetic beads that had been blocked in the same buffer were added to the reaction and allowed to incubate for an additional 1 h. Next, 3 × 5 min washes were performed using the same buffer. Protein was eluted with 1 × Laemmli's SDS Loading Buffer and ran for a Western blot using standard molecular biology techniques. The input lanes represent 1% of the total protein loaded per reaction and each lane account for 10% of the total reaction. The primary antibody used was anti-GST at 1:20,000 (EpiCypher, 13-0022) and the secondary was anti-Rabbit-HRP at 1:20,000 (GE, NA934V). Imaging was performed using a chemiluminescent kit (GE, RPN2232) and BioRad ChemiDoc system. To evaluate the bindings of hDNMT1<sub>BAH1</sub> and JMJD2A tudor domain (JMJD2A<sub>TD</sub>) to various tri-methylated histone peptides, H3<sub>1–21</sub>K4me3, H3<sub>1–21</sub>K9me3, H3<sub>1–21</sub>K14me3, H3<sub>10–27</sub>K18me3, H3<sub>21–44</sub>K27me3, K3<sub>21–44</sub>K36me3, and H4<sub>11–29</sub>K20me3 peptides were synthesized as previously described<sup>68</sup>. One microgram of peptide was incubated with 2 μg of recombinant proteins in the binding buffer containing 50 mM Tris-HCl (pH 7.5), 150 mM NaCl, and 0.1% NP-40. Peptides were pulled down using streptavidin sepharose beads (Amersham) and protein-peptide bindings were detected by western analyses using anti-GST at 1:2500 (custom rabbit polyclonal)<sup>66</sup> and anti-Rabbit-HRP at 1:10,000 (Cell Signaling Technologies).

**DNA methylation kinetics assay.** The DNA methylation assays were performed as previously described<sup>29</sup>. In essence, each reaction mixture contains 0.1 μM hDNMT1<sub>351–1600</sub>, wild type or mutants, 0.5 μM S-adenosyl-L-[methyl-<sup>3</sup>H] methionine (SAM) (Perkin Elmer), 0.4 μM (GT<sup>12</sup>/GAC)<sub>12</sub> hemimethylated DNA duplex, or various amount of histone peptides in 50 mM Tris-HCl (pH 8.0), 7 mM β-ME, 5% glycerol, 100 μg/mL BSA, and 100 mM NaCl. Each reaction lasted 20 min at 37 °C, before being quenched by 2 mM cold SAM. The activity was measured by Beckman LS6500 scintillation counter.

#### Single-molecule Förster resonance energy transfer (smFRET) experiment.

The DNMT1 constructs used for the smFRET experiment was prepared by sortase-mediated ligation of hDNMT1<sub>RFTS-linker</sub> labeled with both Cy3 (GE Healthcare) and Atto647N (Sigma-Aldrich), with the C-terminally biotin-conjugated hDNMT1<sub>646–1606</sub>-LPETG fragment. In detail, the hDNMT1<sub>RFTS-linker</sub> sample was reduced by 10-fold molar excess of TCEP in buffer containing 20 mM Tris-HCl (pH 7.3), 150 mM NaCl and 5% glycerol for over 30 min, before reacting with tenfold molar excess of Cy3 maleimide and fivefold molar excess of Atto 647N maleimide, first at room temperature for 2 h, then at 4 °C overnight in darkness. The unreacted dye was subsequently removed through desalting using a PD-10 column (GE healthcare). Meanwhile, the hDNMT1<sub>646–1606</sub>-LPETG sample was incubated with tenfold molar excess of GGGK<sup>biotin</sup> (K<sup>biotin</sup>, biotinylated lysine) peptide and twofold molar excess of sortase in ligation buffer (20 mM Tris-HCl (pH 7.3), 150 mM NaCl, 5 mM CaCl<sub>2</sub>, 5% glycerol) at room temperature for 4 h, followed by removal of unreacted peptide using a PD-10 column. Both protein samples were stored in buffer containing 20 mM Tris-HCl (pH 7.5), 150 mM NaCl, 5% glycerol and 1 mM β-Mercaptoethanol. Next, ~10 μM Cy3- and Atto647N-labeled hDNMT1<sub>RFTS-linker</sub> was incubated equimolar with biotinylated hDNMT1<sub>646–1606</sub>-LPETG and sortase in the ligation buffer at room temperature for 2 h. The ligated hDNMT1 product was further purified on a HiLoad 16/600 Superdex 200 pg column (GE healthcare). The protein homogeneity was verified by sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE). The protein concentration and the efficiency of dye incorporation were measured using NanoDrop (Thermo Fisher Scientific).

The smFRET experiments were performed at 20 °C. A commercial, biotinylated Pacific Biosciences SMRT cell was incubated with NeutrAvidin (25 μM) (Thermo Fisher Scientific) for 5 min, then unbound NeutrAvidin was washed away. Twenty microlitres of 10 nM hDNMT1<sub>RFTS-linker</sub>-hDNMT1<sub>646–1606</sub>-LPETG ligated sample (hDNMT1<sup>CY</sup>) was then surface-immobilized by incubation for 5 min with the NeutrAvidinylated cell, followed by further washing with buffer (20 mM Tris-HCl (pH 7.5), 150 mM NaCl, 5% glycerol, 1 mM β-Mercaptoethanol) to remove unbound protein. During the final wash, oxygen-scavenging buffer containing 2.5 mM protocatechuic (PCA), 250 nM protocatechuic-3,4-dioxygenase (PCD), 2 mM triplet quencher mixture (TSY, Pacific Biosciences) were included<sup>69,70</sup>. The SMRT cell with immobilized protein was then placed in the RS II instrument (Pacific Biosciences) and imaged by illumination with a 532-nm laser (0.7 W/m<sup>2</sup>). For experiments with histone peptides, the peptide was robotically delivered to the SMRT cell by the instrument at the beginning of the movie. Movies were recorded for 10 min at 10 fps. The delivery mix was made of the final wash buffer supplemented with 2–6 mM H4<sub>11–25</sub>K20me3 (containing a C-terminal tyrosine) or 100 nM H3<sub>1–24</sub>K9me3Ub2 or 2 mM H3<sub>1–15</sub>K4me3 (containing a C-terminal tyrosine).

**smFRET data analysis.** Data processing were performed with in-house MATLAB scripts<sup>70</sup>. Individual single-molecule fluorescence traces were extracted from raw movie data, filtering for those that showed colocalized green and red fluorescence. A substantial fraction of the molecules exhibited significant fluorophore photophysical instability, with significant excursions in the donor intensity on a seconds time scale, including transient periods of complete donor quenching. To minimize the possibility of artifacts, traces were therefore curated manually and their event timings were assigned manually. Traces for assignment were thus selected with the following characteristics: (1) The beginning of all assigned FRET events was either



the beginning of the movie or a frame in which there was an abrupt increase in the red signal with an anticorrelated decrease in the green signal, and (2) the end of all assigned events was an abrupt decrease in the red signal with an anticorrelated increase in the green signal. The resulting FRET event time durations for each condition were then converted to an empirical cumulative distribution function and fit to a single-exponential ( $P(t) = 1 - e^{-\lambda t}$ ) or double-exponential ( $p(t) = A(1 - e^{-\lambda_1 t}) + (1 - A)(1 - e^{-\lambda_2 t})$ ) model as required to maximize the  $R^2$  value from nonlinear least-squares regression.

**MD simulations.** To probe the effect of the H4K20me3 binding and the W796A mutation on the conformation and flexibility of the autoinhibitory linker of hDNMT1, MD simulations were carried out for WT and the W796A mutant of hDNMT1<sub>351–1600</sub>, and a model for the hDNMT1<sub>351–1600</sub>–H4<sub>14–25</sub>K20me3 complex using the OpenMM software package on GPUs. Two independent trajectories were conducted for the complex. As the starting conformation, the crystal structure of hDNMT1<sub>351–1600</sub> (PDB 4WXX [<https://doi.org/10.2210/pdb4WXX/pdb>]) was used for the simulation of WT and the W796A mutant; for the hDNMT1<sub>351–1600</sub>–H4<sub>14–25</sub>K20me3 complex, a structural model was generated based on the structural alignment of hDNMT1<sub>351–1600</sub> with the bDNMT1<sub>BAH1</sub>–H4<sub>14–25</sub>K20me3 complex, with the conformation of residues 712–729 of hDNMT1<sub>351–1600</sub> adjusted using the coot software<sup>63</sup> to accommodate the H4<sub>14–25</sub>K20me3 binding. For each MD simulation, the protein was solvated using a cubic water box such that the protein was at least ten Angstroms away from the box edge. The system was charge neutralized with NaCl counter ions and a salt concentration of 150 mM was maintained during the simulation. The system contained approximately 374,000 atoms. For each system, equilibration of 50 ps was done with the NVT ensemble, while production run of 100 ns was conducted with the NPT ensemble using an isotropic Monte Carlo barostat, a Langevin thermostat and an integration time step of 2 fs. For non-bonded interactions, a force-switch based scheme ( $r_{\text{on}} = 1.0$  nm,  $r_{\text{off}} = 1.2$  nm) was used for van der Waals interactions, and Particle–Mesh–Ewald summation with a grid size of approximately 1 Å was used for treating electrostatic interactions. For analysis, structures were saved every 10 ps. For the computation of root-mean-square-fluctuations, only the last 20 ns of the 100 ns production run was used.

**Cell lines and tissue culture.** The murine Dnmt1-knockout (1KO) ESCs were used as we previously described<sup>29</sup>. In brief, 1KO-ESCs were transfected with the pPyCAGIP empty vector or that carrying WT or mutant DNMT1, followed by drug selection with 1 µg/mL puromycin for over two weeks. Both the pooled stable-expression lines and independent single-cell-derived clonal lines were established and verified (such as immunoblotting of DNMT1) before use.

**Antibodies and Western blotting.** The cultured cells were collected, rinsed in cold phosphate-buffered saline (PBS), and suspended in buffer containing 50 mM Tris-HCl (pH 8.0), 150 mM NaCl and 1% NP-40. After brief sonication and centrifugation, the soluble fractions of cell lysates were mixed with 2× SDS-PAGE loading buffer and boiled for 5 min, followed by loading onto a SDS-PAGE gel for immunoblotting analysis, as previously described<sup>51,71</sup>. Information for the antibodies used in this work is included in the Supplementary Table 3.

**Quantification of 5-methyl-2'-deoxycytidine (5-mdC) in gDNA.** Measurements of the global levels of 5-mdC in cellular DNA were carried out as described previously<sup>29,51</sup>. Briefly, gDNA was digested into mononucleosides by nuclease P1 and alkaline phosphatase. The enzymes in the digestion mixtures were removed by chloroform extraction, and the resulting aqueous layer was dried, reconstituted in water, and subjected to LC–MS/MS analyses on an LTQ XL linear ion trap mass spectrometer for quantifications of 5-mdC and dG. The amounts of 5-mdC and dG (in moles) in the nucleoside mixtures were calculated based on comparisons of their signal intensities with their corresponding stable isotope-labeled standards and calibration curves. The final levels of 5-mdC were calculated as molar ratios of 5-mdC over dG.

**Confocal immunofluorescence (IF).** G1/S phase synchronization of ES cells follows a previously established protocol<sup>29</sup>. In essence, ES cells were treated with thymidine (Sigma T9250) at a final concentration of 2 mM for 16 h, washed twice with pre-warmed PBS, and then grown in fresh ES cell medium. After a 6 h release, thymidine was added to the medium again. Cells were incubated with thymidine for another 16 h, washed, released for 5 h from the thymidine block with the addition of fresh medium and then proceed for IF. The IF was carried out as described before<sup>29,72</sup>.

**Co-immunoprecipitation (CoIP).** Cells were lysed as described<sup>29,72</sup>. Anti-Flag M2-conjugated agarose beads (Sigma) were incubated with the lysates overnight at 4 °C. The beads were then extensively washed and the bound proteins analyzed by western blotting.

**Enhanced reduced representation bisulfite sequencing (eRRBS).** The eRRBS experiment, including construction of libraries and processing and analysis of eRRBS dataset, was performed as described previously<sup>29</sup>. Briefly, genomic DNA (gDNA) mixed with 0.1% of unmethylated lambda DNA (Promega) was digested

with MspI, MseI, and BfaI, and subjected to end repair, A-tailing and ligation to NEBNext Methylated Adapters (NEBNext DNA Library Prep Kit). Next, the DNA product was purified using AMPure beads and subject to bisulfite conversion and library construction using the EpiMark Bisulfite Conversion Kit (NEB cat# E3318), followed by deep sequencing in an Illumina HiSeq 4000 platform with a paired end PE150 cycle (carried out by UNC HTSF Genomic Core). The eRRBS data were then analyzed with FastQC v0.11.2 [<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>] and Bismark v0.18.1<sup>73</sup>. Analysis of the sequences corresponding to phase λ indicated a bisulfite conversion rate of >99%.

**DMR detection.** Counts of methylated and unmethylated bases for cytosines in the CpG context were analyzed with the DSS (v.2.30.1) R/Bioconductor package<sup>74</sup> for the calling of differential methylation loci or regions (DML/DMRs) and their quantification. The  $q$  values under 0.05 and mean differences of methylation greater than 0.15 were used as the cutoff line to define DMRs with DSS. Alternative identification of DML/DMRs was also performed using the R-package methylKit<sup>75</sup> (v.1.8.1) with tiling windows of 500 bp. In the setting of this tool,  $q$  values under 0.01 and mean differences of methylation greater than 0.25 were used as the cutoff line. In both analyses, at least five covered cytosine sites were utilized. After the calling of differential methylation, each locus or region was annotated using Homer<sup>76</sup> (v4.10.3).

**ChIP-seq and data analysis.** ChIP-seq data of H3K9me3 in the E14Tg2a.4 mouse ES cell line was obtained from ENCODE for both raw FASTQ files (ENCODE ENCFF001ZHD, and ENCFF001ZHF) and called peaks (ENCODE ENCFF180LQA). ChIP-seq data of H4K20me3 in the R1 mouse ES cell line was obtained from GEO GSE94086<sup>77</sup> as raw FASTQ files (SRR5198791, SRR5198793). Sequence data was processed as follows: mapping to mm10 via Bowtie2 v2.1.0<sup>78</sup> (parameters –sensitive-local), deduplication using Picard tools v1.110 MarkDuplicates.jar [<http://broadinstitute.github.io/picard/>], then extension of each mapped non-duplicate read to assumed fragment length (200 nt for H3K9me3, 150 nt for H4K20me3). Depth tracks for the ChIP-seq data were generated by BEDTools v2.24.0 ‘genomecov’, followed by bed-GraphToBigWig (UCSC utility script; <http://hgdownload.soe.ucsc.edu/admin/exe/>). Peak calls for H4K20me3 were made via HOMER findPeaks (parameters –region-size 1000 –minDist 2500). Peak sets were filtered to only include peaks on canonical chromosomes (chr1–19,X,Y), resulting in final peak sets of  $N = 123,309$  (H3K9me3) and  $N = 24,373$  (H4K20me3). Peaks were assigned to quartiles of decreasing ChIP-seq signal according to RPKM per peak region, which was calculated based on overlapping read counts as reported by BEDTools v2.24.0 ‘multicov’. Random genomic regions size-matched to each peak set were generated by BEDTools v2.24.0 [see reference above] ‘shuffle’ with parameters –noOverlapping –excl (where the excluded regions are defined as long poly-N stretches and all non-canonical chromosomes). For plotting of ChIP-seq signals, each signal was log<sub>2</sub> transformed after addition of 0.01 as pseudo values.

**CUT&RUN followed by deep sequencing.** CUT&RUN were performed according to manufacturer’s instructions (EpiCypher CUTANA™ pAG-MNase for ChIC/CUT&RUN, Cat# 15-1116). In brief, after washing with CUT&RUN wash buffer (20 mM HEPES pH 7.5, 150 mM NaCl, 0.5 mM spermidine, 1× Roche Complete Protease Inhibitor), a million of cells were first bound to activated ConA beads (Bangs Laboratories, cat# BP531), followed by addition of anti-H4K20me3 antibody (Abcam, ab9053; 1:100 dilution) and cell permeabilization with the digitonin buffer (CUT&RUN wash buffer plus 0.01% digitonin). After washing in the digitonin buffer, samples were incubated with pAG-MNase, followed by additional washes with digitonin buffer. After the final wash, pAG-MNase activation was induced for DNA digestion by suspending cell samples in the pAG-MNase digestion buffer (digitonin buffer plus 2 mM CaCl<sub>2</sub>) and incubation on nutator at 4 °C for 2 h. Solubilized chromatin was released using the stop buffer (340 mM NaCl, 20 mM EDTA, 4 mM EGTA, 50 µg/ml RNase A, 50 µg/ml glycogen) and collected using a PCR cleanup kit (New England BioLabs [NEB] Monarch PCR & DNA Cleanup Kit, cat# T1030). Ten nanogram of the purified CUT&RUN-enriched DNA was used for preparation of multiplexed Illumina libraries using the NEB Ultra II DNA Library Prep Kit according to manufacturer’s instructions (NEB cat#E7103).

**CUT&RUN data analysis.** Raw fastq data were mapped to the reference genome (mm10) using bowtie2.3.5<sup>79</sup>. The non-primary alignment, PCR duplicates, or blacklist regions from aligned data were removed by Samtools (v1.9), Picard MarkDuplicates funtion (ver 2.20.4), and bedtools (v2.28.0), respectively. Peak calling was performed using MACS2 (macs2 callpeak -f BAMPE -g hs/mm –keep-dup 1)<sup>80</sup>. Deeptools (v3.3.0) was used to make bigwig files, heatmaps, and averaged plotting of CUT&RUN signal<sup>81</sup>. Genomic binding profiles were generated using the deepTools “bamCompare” functions. The peaks were annotated in the reference genome (NCBI Reference Sequences [RefSeq]) and then used for heatmap and averaged signal plotting.

**CRISPR/Cas9-based genome editing for site-specifically mutating Dnmt1<sub>BAH1</sub>.** As we recently described<sup>4</sup>, the Alt-R® CRISPR–Cas9 System, together with a single-stranded oligodeoxynucleotide donor (ssODN), was employed to introduce a

W799A point mutation to mouse Dnmt1 gene (equivalent to W796A in the human DNMT1 BAH1 domain) in E14 ESCs. In brief, trans-activating CRISPR RNA (tracrRNA) with ATTO<sup>™</sup> 550, the S.p. HiFi Cas9 Nuclease V3, Electroporation Enhancer, and crRNA targeting the Dnmt1 genomic site to be mutated (with the sequence information of crRNA listed in Supplementary Table 2) were all ordered through IDT Inc. The ssODN was custom-designed with a tool of benchling (<https://benchling.com/crispr>), the sequence information of ssODN listed in Supplementary Table 2) and synthesized as PAGE-purified Ultramer<sup>®</sup> DNA oligonucleotides from IDT Inc. Three phosphorothioate bonds were added at both ends of ssODN to stabilize the donor oligo and make homology-directed repair more efficient. For introducing a Dnmt1 W799A mutation, the codon was mutated from TGG to GCG, which generates a site of MluI enzyme and thus facilitates subsequent cell screening and genotyping by enzyme digestion of PCR products. To make the crRNA:tracrRNA duplex, the crRNA and tracrRNA were mixed in equimolar concentration, followed by heating at 95 °C for 5 min and cooling down to room temperature. The CRISPR-Cas9 ribonucleoprotein (RNP) complex was then prepared by diluting the crRNA:tracrRNA duplex and Cas9 enzyme components in PBS, followed by incubation at room temperature for 15 min. The RNP complex was next mixed with Electroporation Enhancer and ssODN, followed by electroporation-based delivery to E14 cells using the Mouse Embryonic Stem Cell Nucleofector Kit (Lonza). ATTO<sup>™</sup> 550 positive cells were sorted out by FACS (at UNC flow core) after 36 h and split into 96-well plates for genotyping. After genotyping, lines with homozygous mutation were further validated at DNA levels by direct sequencing of PCR products (the sequence information of genotyping primers listed in Supplementary Table 2).

**Chemical compound.** The chemical inhibitor selective for the H4K20 methyltransferases SUV420H1 and SUV420H2, A-196<sup>42</sup> (Sigma, SML1565), was dissolved in dimethylsulfoxide (DMSO) as 5 mM stock solution. A 3-day treatment with 10 μM A-196, compared to DMSO, was used in this study.

**Neutral comet assay.** To perform neutral comet assays for DNMT1 C1226S mutants (Supplementary Fig. 7f–h), mouse DNMT1 knockout cells ( $7 \times 10^4$  cells) complemented by pPyCAGIP empty vector (vector), WT DNMT1, and two different clones of DNMT1 C1226S mutant (clones #1 and #2) were plated on gelatin-coated culture dishes without feeder cells. After 2 h post ionizing irradiation (5 Gy) cells were harvested and mixed with LMAgarose (Trevigen). The LMAgarose mixed samples were placed onto comet assay slides and immersed into comet assay lysis solution (Trevigen) at 4 °C for 1 h. Subsequently, the slides were incubated with TBE buffer (90 mM Tris borate) for 1 h and subjected to electrophoresis at 40 V for 40 min. After electrophoresis, the samples were fixed with 70% ethanol at RT for 30 min and dried at 37 °C for 30 min. DNAs were visualized using SYBR-green (Invitrogen) and imaged using Fluoview FV3000 confocal microscope (Olympus). Images were subsequently analyzed using ImageJ (v.1.53). Statistics and graph were calculated using Prism software (Graphpad v6). Experiments were performed with at least two independent replicates. Neutral comet assays for cells with DNMT1 W465A, W465A/W796A, W464A/W465A, or TM (Fig. 5e, f), were performed as described previously<sup>29</sup>.

**Clonogenic cell survival assay.** Cells were seeded into six-well plates and treated with different dosages of IR using a Faxitron X-ray irradiator. Following IR treatment, cells were incubated for 12 days in tissue culture incubator (37 °C, 5% CO<sub>2</sub>). Cells were washed with PBS and the colonies were stained with 0.5% (w/v) crystal violet and 20% (v/v) ethanol for 30 min at RT. Results were normalized to plating efficiencies of untreated cells for each group.

**Statistics.** The comet assays were performed using one-way ANOVA with post Tukey analysis. The *p* value lower than 0.05 was considered to be statistically significant. For all the other analyses, the two-tailed Student *t* tests were performed to compare distributions between different groups. And the *p* value lower than 0.01 was considered to be statistically significant.

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

## Data availability

The data that support this study are available from the corresponding author upon reasonable request. Coordinates and structure factors for the bDNMT1<sub>BAH1-H414-25</sub>K20me3 and bDNMT1<sub>BAH1-H414-25</sub>K20me2 complexes have been deposited in the Protein Data Bank under accession codes 7LMK [<https://doi.org/10.2210/pdb7LMK/pdb>] and 7LMM [<https://doi.org/10.2210/pdb7LMM/pdb>], respectively. The eRRBS data generated during the course of this study have been deposited in Gene Expression Omnibus (GEO) under accession code GSE145698. Publicly available datasets analyzed for this study: ChIP-seq data of H3K9me3 in the E14Tg2a.4 mouse ES cell line was obtained from ENCODE for both raw FASTQ files (ENCODE ENCFF001ZHD and ENCFF001ZHF) and called peaks (ENCODE ENCFF180LQA). ChIP-seq data of H4K20me3 in the R1 mouse ES cell line was obtained from GEO GSE94086 as raw FASTQ files (SRR5198791 and SRR5198793). Source data are provided with this paper.

Received: 23 December 2020; Accepted: 15 March 2021;

Published online: 03 May 2021

## References

- Richards, E. J. & Elgin, S. C. Epigenetic codes for heterochromatin formation and silencing: rounding up the usual suspects. *Cell* **108**, 489–500 (2002).
- Bird, A. The essentials of DNA methylation. *Cell* **70**, 5–8 (1992).
- Cheng, X. Structural and functional coordination of DNA and histone methylation. *Cold Spring Harbor Perspect. Biol.* **6**, a018747 (2014).
- Fan, H. et al. BAHCC1 binds H3K27me3 via a conserved BAH module to mediate gene silencing and oncogenesis. *Nat. Genet.* **52**, 1384–1396 (2020).
- Eden, A., Gaudet, F., Waghmare, A. & Jaenisch, R. Chromosomal instability and tumors promoted by DNA hypomethylation. *Science* **300**, 455 (2003).
- Feinberg, A. P. & Vogelstein, B. Hypomethylation distinguishes genes of some human cancers from their normal counterparts. *Nature* **301**, 89–92 (1983).
- Fraga, M. F. et al. Loss of acetylation at Lys16 and trimethylation at Lys20 of histone H4 is a common hallmark of human cancer. *Nat. Genet.* **37**, 391–400 (2005).
- Kumari, D. & Usdin, K. The distribution of repressive histone modifications on silenced FMR1 alleles provides clues to the mechanism of gene silencing in fragile X syndrome. *Hum. Mol. Genet.* **19**, 4634–4642 (2010).
- Sutcliffe, J. S. et al. DNA methylation represses FMR-1 transcription in fragile X syndrome. *Hum. Mol. Genet.* **1**, 397–400 (1992).
- Arancio, W., Pizzolanti, G., Genovese, S. I., Pitrone, M. & Giordano, C. Epigenetic involvement in Hutchinson-Gilford progeria syndrome: a mini-review. *Gerontology* **60**, 197–203 (2014).
- Goll, M. G. & Bestor, T. H. Eukaryotic cytosine methyltransferases. *Annu. Rev. Biochem.* **74**, 481–514 (2005).
- Jeltsch, A. On the enzymatic properties of Dnmt1: specificity, processivity, mechanism of linear diffusion and allosteric regulation of the enzyme. *Epigenetics* **1**, 63–66 (2006).
- Wang, Q. et al. Imprecise DNMT1 activity coupled with neighbor-guided correction enables robust yet flexible epigenetic inheritance. *Nat. Genet.* **52**, 828–839 (2020).
- Yarychivska, O., Shahabuddin, Z., Comfort, N., Boulard, M. & Bestor, T. H. BAH domains and a histone-like motif in DNA methyltransferase 1 (DNMT1) regulate de novo and maintenance methylation in vivo. *J. Biol. Chem.* **293**, 19466–19475 (2018).
- Song, J., Rechko, O., Bestor, T. H. & Patel, D. J. Structure of DNMT1-DNA complex reveals a role for autoinhibition in maintenance DNA methylation. *Science* **331**, 1036–1040 (2011).
- Song, J., Teplova, M., Ishibe-Murakami, S. & Patel, D. J. Structure-based mechanistic insights into DNMT1-mediated maintenance DNA methylation. *Science* **335**, 709–712 (2012).
- Takeshita, K. et al. Structural insight into maintenance methylation by mouse DNA methyltransferase 1 (Dnmt1). *Proc. Natl Acad. Sci. USA* **108**, 9055–9059 (2011).
- Zhang, Z. M. et al. Crystal Structure of Human DNA Methyltransferase 1. *J. Mol. Biol.* **427**, 2520–2531 (2015).
- Bashtrykov, P. et al. Targeted mutagenesis results in an activation of DNA methyltransferase 1 and confirms an autoinhibitory role of its RFTS domain. *Chembiochem: a Eur. J. Chem. Biol.* **15**, 743–748 (2014).
- Berkuyrek, A. C. et al. The DNA methyltransferase Dnmt1 directly interacts with the SET and RING finger-associated (SRA) domain of the multifunctional protein Uhrf1 to facilitate accession of the catalytic center to hemi-methylated DNA. *J. Biol. Chem.* **289**, 379–386 (2014).
- Syeda, F. et al. The replication focus targeting sequence (RFTS) domain is a DNA-competitive inhibitor of Dnmt1. *J. Biol. Chem.* **286**, 15344–15351 (2011).
- Harrison, J. S. et al. Hemi-methylated DNA regulates DNA methylation inheritance through allosteric activation of H3 ubiquitylation by UHRF1. *eLife* **5**, e17101 (2016).
- Ishiyama, S. et al. Structure of the Dnmt1 reader module complexed with a unique two-mono-ubiquitin mark on histone H3 reveals the basis for DNA methylation maintenance. *Mol. Cell* **68**, 350–360.e357 (2017).
- Nishiyama, A. et al. Uhrf1-dependent H3K23 ubiquitylation couples maintenance DNA methylation and replication. *Nature* **502**, 249–253 (2013).
- Qin, W. et al. DNA methylation requires a DNMT1 ubiquitin interacting motif (UIM) and histone ubiquitylation. *Cell Res.* **25**, 911–929 (2015).
- Nishiyama, A. et al. Two distinct modes of DNMT1 recruitment ensure stable maintenance DNA methylation. *Nat. Commun.* **11**, 1222 (2020).
- Gonzalez-Magana, A. et al. Double monoubiquitination modifies the molecular recognition properties of p15(PAF) promoting binding to the reader module of Dnmt1. *ACS Chem. Biol.* **14**, 2315–2326 (2019).
- Leonhardt, H., Page, A. W., Wei, H. U. & Bestor, T. H. A targeting sequence directs DNA methyltransferase to sites of DNA replication in mammalian nuclei. *Cell* **71**, 865–873 (1992).

29. Ren, W. et al. Direct readout of heterochromatic H3K9me3 regulates DNMT1-mediated maintenance DNA methylation. *Proc. Natl Acad. Sci. USA* **117**, 18439–18447 (2020).
30. Svedruzic, Z. M. & Reich, N. O. Mechanism of allosteric regulation of Dnmt1's processivity. *Biochemistry* **44**, 14977–14988 (2005).
31. Yang, N. & Xu, R. M. Structure and function of the BAH domain in chromatin biology. *Crit. Rev. Biochem. Mol. Biol.* **48**, 211–221 (2013).
32. Kuo, A. J. et al. The BAH domain of ORC1 links H4K20me2 to DNA replication licensing and Meier-Gorlin syndrome. *Nature* **484**, 115–119 (2012).
33. Li, Z., Fu, X., Wang, Y., Liu, R. & He, Y. Polycomb-mediated gene silencing by the BAH-EMF1 complex in plants. *Nat. Genet.* **50**, 1254–1261 (2018).
34. Wiles, E. T. et al. Evolutionarily ancient BAH-PHD protein mediates Polycomb silencing. *Proc. Natl Acad. Sci. USA* **117**, 11614–11623 (2020).
35. Yang, Z. et al. EBS is a bivalent histone reader that regulates floral phase transition in Arabidopsis. *Nat. Genet.* **50**, 1247–1253 (2018).
36. Liu, Y., Oakeley, E. J., Sun, L. & Jost, J. P. Multiple domains are involved in the targeting of the mouse DNA methyltransferase to the DNA replication foci. *Nucleic Acids Res.* **26**, 1038–1045 (1998).
37. Long, L., Furgason, M. & Yao, T. Generation of nonhydrolyzable ubiquitin-histone mimics. *Methods* **70**, 134–138 (2014).
38. Mao, H., Hart, S. A., Schink, A. & Pollok, B. A. Sortase-mediated protein ligation: a new method for protein engineering. *J. Am. Chem. Soc.* **126**, 2670–2671 (2004).
39. Li, E., Bestor, T. H. & Jaenisch, R. Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell* **69**, 915–926 (1992).
40. Montoya-Durango, D. E. et al. Epigenetic control of mammalian LINE-1 retrotransposon by retinoblastoma proteins. *Mutat. Res.* **665**, 20–28 (2009).
41. Weisenberger, D. J. et al. Analysis of repetitive element DNA methylation by MethyLight. *Nucleic Acids Res.* **33**, 6823–6836 (2005).
42. Bromberg, K. D. et al. The SUV4-20 inhibitor A-196 verifies a role for epigenetics in genomic integrity. *Nat. Chem. Biol.* **13**, 317–324 (2017).
43. Taverna, S. D., Li, H., Ruthenburg, A. J., Allis, C. D. & Patel, D. J. How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers. *Nat. Struct. Mol. Biol.* **14**, 1025–1040 (2007).
44. Gulsuner, S. et al. Spatial and temporal mapping of de novo mutations in schizophrenia to a fetal prefrontal cortical network. *Cell* **154**, 518–529 (2013).
45. Matsui, T. et al. Proviral silencing in embryonic stem cells requires the histone methyltransferase ESET. *Nature* **464**, 927–931 (2010).
46. Bulut-Karslioglu, A. et al. Suv39h-dependent H3K9me3 marks intact retrotransposons and silences LINE elements in mouse embryonic stem cells. *Mol. Cell* **55**, 277–290 (2014).
47. Yang, B. X. et al. Systematic identification of factors for provirus silencing in embryonic stem cells. *Cell* **163**, 230–245 (2015).
48. Deniz, O., Frost, J. M. & Branco, M. R. Regulation of transposable elements by DNA modifications. *Nat. Rev.* **20**, 417–431 (2019).
49. Bourc'his, D. & Bestor, T. H. Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L. *Nature* **431**, 96–99 (2004).
50. Wu, T. P. et al. DNA methylation on N(6)-adenine in mammalian embryonic stem cells. *Nature* **532**, 329–333 (2016).
51. Zhang, Z. M. et al. Structural basis for DNMT3A-mediated de novo DNA methylation. *Nature* **554**, 387–391 (2018).
52. Gao, L. et al. Comprehensive structure-function characterization of DNMT3B and DNMT3A reveals distinctive de novo DNA methylation mechanisms. *Nat. Commun.* **11**, 3355 (2020).
53. Baylin, S. B. DNA methylation and gene silencing in cancer. *Nat. Clin. Pract. Oncol.* **2**, S4–S11 (2005).
54. Grayson, D. R. & Guidotti, A. The dynamics of DNA methylation in schizophrenia and related psychiatric disorders. *Neuropsychopharmacology* **38**, 138–166 (2013).
55. Jones, P. A. & Liang, G. Rethinking how DNA methylation patterns are maintained. *Nat. Rev. Genet.* **10**, 805–811 (2009).
56. Saredi, G. et al. H4K20me0 marks post-replicative chromatin and recruits the TONSL-MMS22L DNA repair complex. *Nature* **534**, 714–718 (2016).
57. Xu, M., Wang, W., Chen, S. & Zhu, B. A model for mitotic inheritance of histone lysine methylation. *EMBO Rep.* **13**, 60–67 (2011).
58. Escobar, T. M. et al. Active and repressed chromatin domains exhibit distinct nucleosome segregation during DNA replication. *Cell* **179**, 953–963 (2019). e911.
59. Chambers, I. et al. Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells. *Cell* **113**, 643–655 (2003).
60. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Method Enzymol.* **276**, 307–326 (1997).
61. Kabsch, W. Xds. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 125–132 (2010).
62. McCoy, A. J. et al. Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
63. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132 (2004).
64. Adams, P. D. et al. PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr. D Biol. Crystallogr.* **58**, 1948–1954 (2002).
65. Petell, C. J., Pham, A. T., Skela, J. & Strahl, B. D. Improved methods for the detection of histone interactions with peptide microarrays. *Sci. Rep.* **9**, 6265 (2019).
66. Chen, S. et al. The PZP domain of AF10 senses unmodified H3K27 to regulate DOT1L-mediated methylation of H3K79. *Mol. Cell* **60**, 319–327 (2015).
67. Shanle, E. K. et al. Histone peptide microarray screen of chromo and Tudor domains defines new histone lysine methylation interactions. *Epigenetics Chromatin* **10**, 12 (2017).
68. Bua, D. J. et al. Epigenome microarray platform for proteome-wide dissection of chromatin-signaling networks. *PLoS ONE* **4**, e6789 (2009).
69. Aitken, C. E., Marshall, R. A. & Puglisi, J. D. An oxygen scavenging system for improvement of dye stability in single-molecule fluorescence experiments. *Biophys. J.* **94**, 1826–1835 (2008).
70. Chen, J. et al. High-throughput platform for real-time monitoring of biological processes by multicolor single-molecule fluorescence. *Proc. Natl Acad. Sci. USA* **111**, 664–669 (2014).
71. Lu, R. et al. Epigenetic perturbations by Arg882-mutated DNMT3A potentiate aberrant stem cell gene-expression program and acute leukemia development. *Cancer Cell* **30**, 92–107 (2016).
72. Cai, L. et al. An H3K36 methylation-engaging tudor motif of polycomb-like proteins mediates PRC2 complex targeting. *Mol. Cell* **49**, 571–582 (2013).
73. Krueger, F. & Andrews, S. R. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* **27**, 1571–1572 (2011).
74. Feng, H., Conneely, K. N. & Wu, H. A Bayesian hierarchical model to detect differentially methylated loci from single nucleotide resolution sequencing data. *Nucleic Acids Res.* **42**, e69 (2014).
75. Akalin, A. et al. methylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles. *Genome Biol.* **13**, R87 (2012).
76. Heinz, S. et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589 (2010).
77. Kidder, B. L., Hu, G., Cui, K. & Zhao, K. SMYD5 regulates H4K20me3-marked heterochromatin to safeguard ES cell self-renewal and prevent spurious differentiation. *Epigenetics Chromatin* **10**, 8 (2017).
78. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
79. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).
80. Zhang, Y. et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).
81. Ramirez, F. et al. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* **44**, W160–W165 (2016).

## Acknowledgements

We would like to thank staff members at the Stanford Synchrotron Radiation Lightsource and at the NE-CAT beamlines (GM124165), Advanced Photo Source (DE-AC02-06CH11357), Argonne National Laboratory for access to X-ray beamlines. We are also grateful for professional support of UNC facilities including Genomics Core, which are partly supported by the UNC Cancer Center Core Support Grant P30-CA016086. This work was supported by NIH grants (1R35GM119721 to J.S., 1R35ES031707 to Y.W., R35GM139569 to O.G., 1R35GM126900 to B.D.S., R00GM111858 to S.O.L., 1R01CA215284 and 1R01CA211336 to G.G.W., and CA198279 and CA201268 to K.M.M.), NIH T32 Training Fellowship for Integrated Training in Cancer Model Systems (T32CA009156) and an American Cancer Society Postdoctoral Fellowship (PF-19-027-01-DMC) to C.J.P., University of California Cancer Research Coordinating Committee (UC CRCC) grant (CRR-20-634140) to J.S., and grants from When Everyone Survives (WES) Leukemia Research Foundation (to G.G.W.) and Gabrielle's Angel Foundation for Cancer Research (to G.G.W.). D.J.P. was supported by a SCOR grant by the Leukemia and Lymphoma Society and by the Memorial Sloan-Kettering Cancer Center Core Grant (P30CA008748). This work was also supported in part by the Intramural Research Program of the National Institute of Environmental Health Sciences, NIH (ES101965 to PAW). J.P.C. is supported in part by T32CA09302. G.G.W. and K.M.M. are American Cancer Society (ACS) Research Scholars, C.J.P. is an ACS postdoctoral fellow, and G.G.W. is a Leukemia & Lymphoma Society (LLS) Scholar.

## Author contributions

W.R., H.F., S.A.G., J.J.K., L.L., C.J.P., X.-F.T., Z.-M.Z., J.P.C., J.Y., Y.G., D.I.K., L.G., S.L., L.C., N.K., B.C., Y.W., Q.C., and S.O.L. performed the experiments. Y.W., B.S., O.G., K.M.M., S.O.L., P.A.W., G.G.W., and J.S. organized the study. D.J.P., O.G., G.G.W., and J.S. conceived the study. G.G.W. and J.S. wrote the paper with input from all the authors.

## Competing interests

O.G. and B.D.S. are co-founders of EpiCypher Inc. G.G.W. received research fund from the Deerfield Management/Pinnacle Hill Company. The remaining authors declare no competing interests.



### Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41467-021-22665-4>.

**Correspondence** and requests for materials should be addressed to G.G.W. or J.S.

**Peer review information** *Nature Communications* thanks the anonymous reviewers 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.



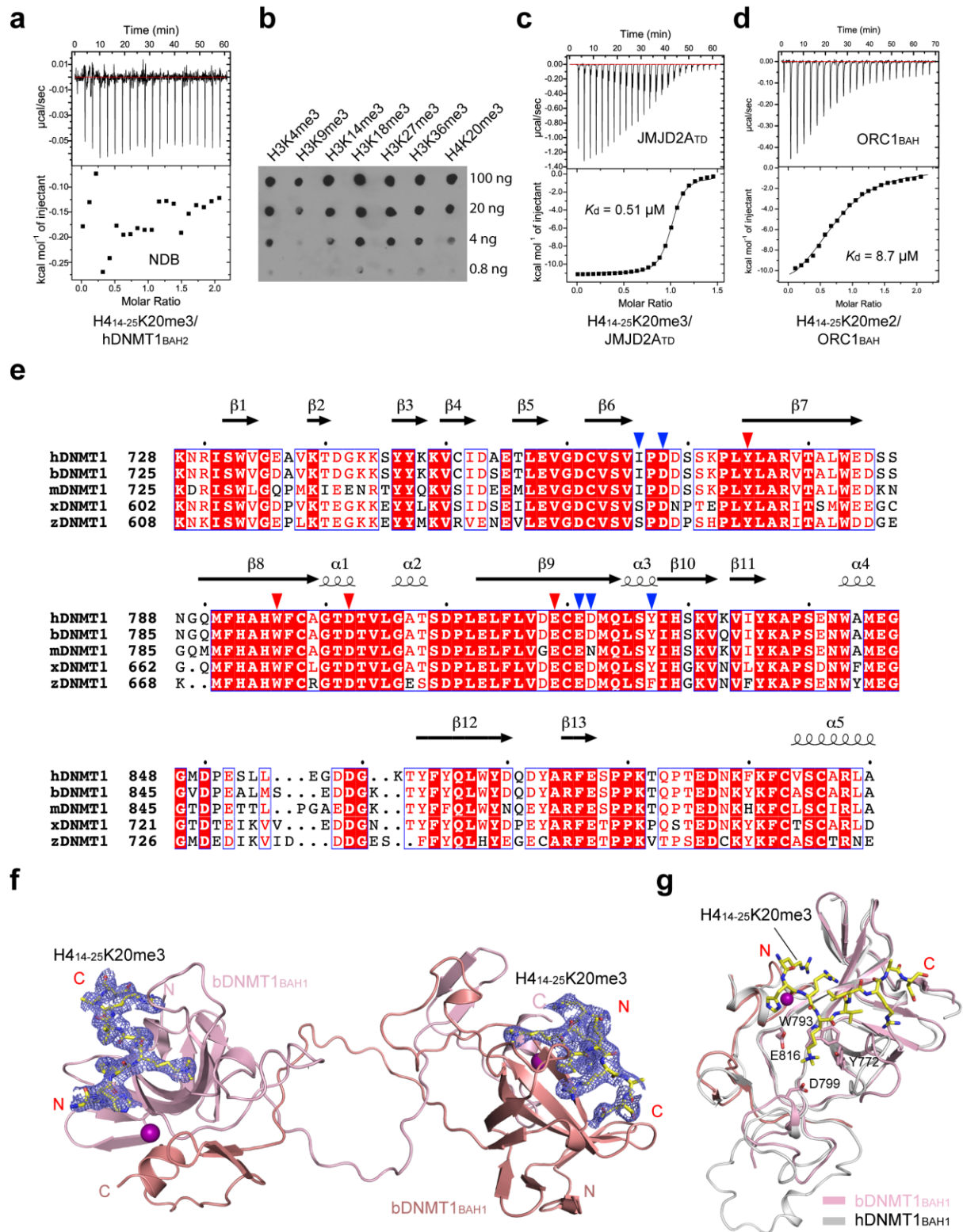
**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the 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) 2021

## **Supplementary information**

### **DNMT1 reads heterochromatic H4K20me3 to reinforce LINE-1 DNA methylation**

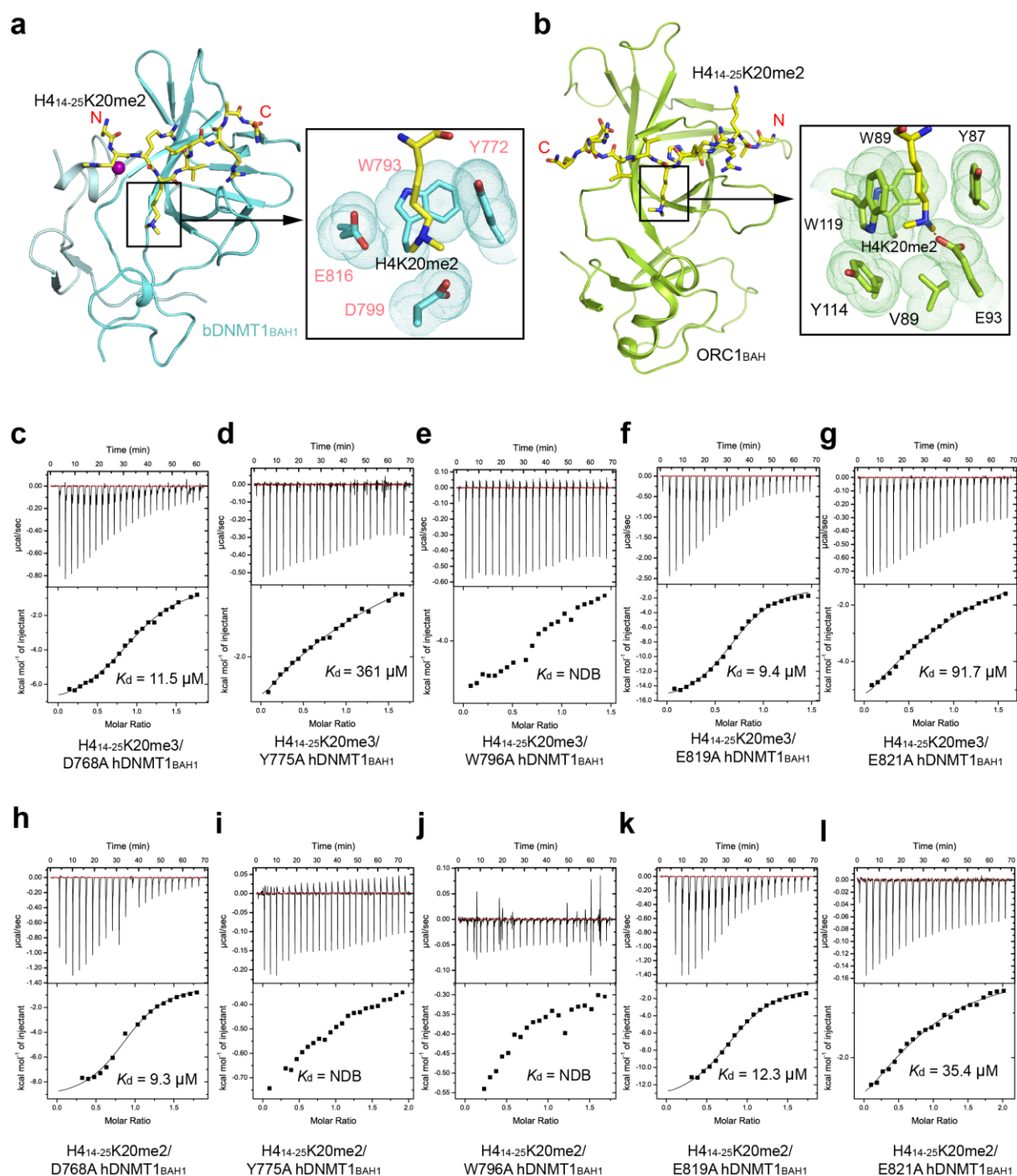
Wendan Ren, Huitao Fan, Sara A Grimm, Jae Jin Kim, Linhui Li, Yiran Guo, Christopher James Petell, Xiao-Feng Tan, Zhi-Min Zhang, John P. Coan, Jiekai Yin, Dae In Kim, Linfeng Gao, Ling Cai, Nelli Khudaverdyan, Burak Çetin, Dinshaw J. Patel, Qiang Cui, Brian D. Strahl, Yinsheng Wang, Or Gozani, Kyle M. Miller, Seán E. O'Leary, Paul A. Wade, Gang Greg Wang, Jikui Song



**Supplementary Figure 1. Biochemical, sequence and structural analyses of the histone binding affinities of the DNMT1 BAH domains. (a)** ITC binding curve for hDNMT1<sub>BAH2</sub> titrated with the H4<sub>14-25</sub>K20me3 peptide. **(b)** Dot plot showing relative concentrations of biotinylated histone peptides used in peptide pull-downs detected

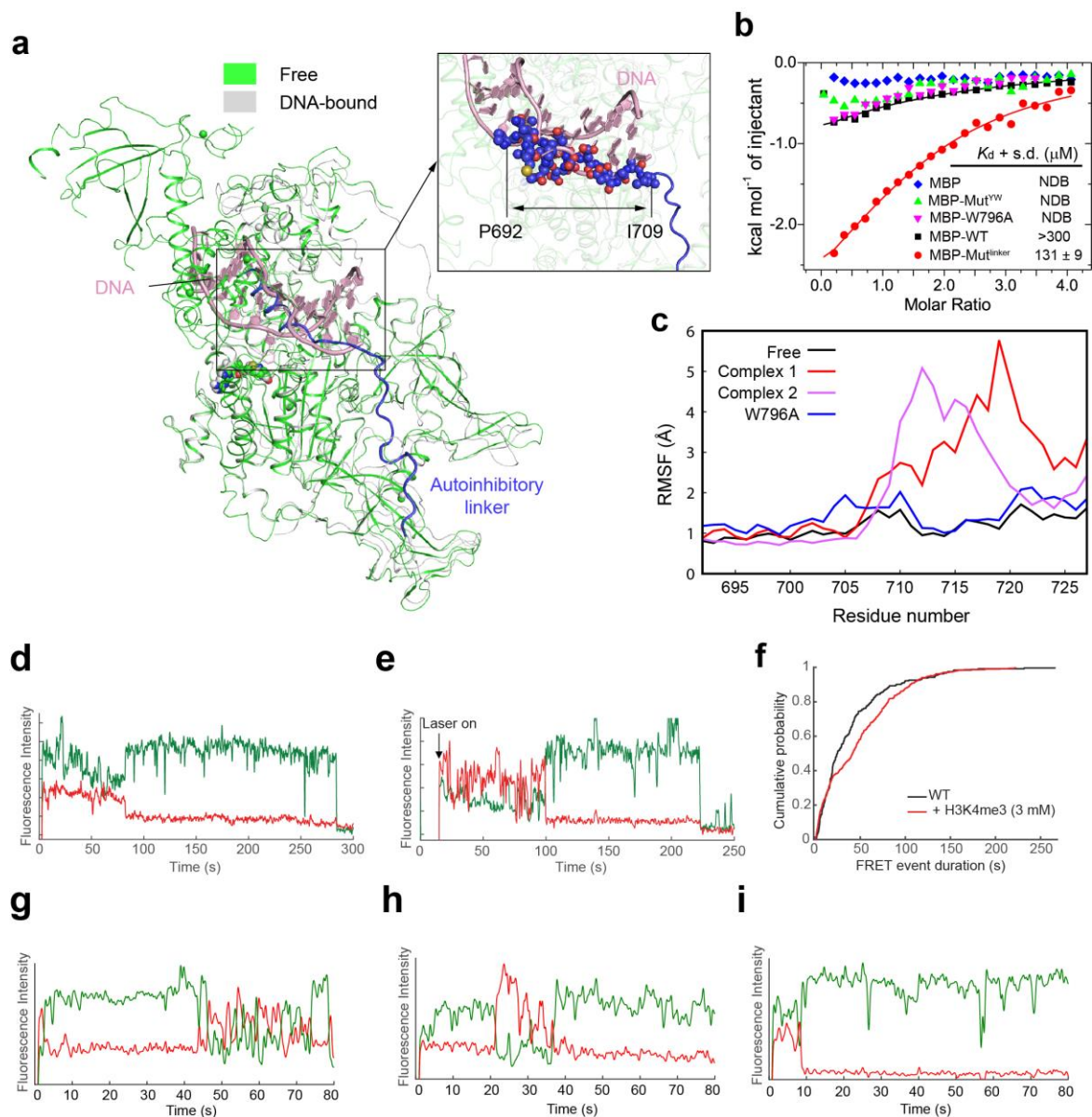


with peroxidase-conjugated streptavidin in Fig. 1e. **(c)** ITC binding curve for JMJD2A<sub>TD</sub> titrated with the H4<sub>14-25</sub>K20me3 peptide. **(d)** ITC binding curve for ORC1<sub>BAH</sub> titrated with the H4<sub>14-25</sub>K20me2 peptide. **(e)** Sequence alignment of the DNMT1 BAH1 domain from human (hDNMT1), bovine (bDNMT1), mouse (mDNMT1), *Xenopus Laevis* (xDNMT1) and Zebrafish (zDNMT1). Strictly conserved residues are colored in white and shaded in red. Similar residues are colored in red. The secondary structures corresponding to hDNMT1 BAH1 are marked on top. The residues forming the H4K20me3-binding pocket are indicated by red arrows. All the other H4-interacting residues are indicated by blue arrows. **(f)** Structure of two bDNMT1<sub>BAH1</sub>-H<sub>14-25</sub>K20me3 complexes engaging domain swapping. Due to crystal packing, the last two  $\beta$ -strands of bDNMT1<sub>BAH1</sub> were swapped with their counterparts from a crystal symmetry-related molecule, resulting in formation of a homodimer in crystals. The two BAH1 domains are colored in light pink and salmon respectively. The H4<sub>14-25</sub>K20me3 peptide is shown as yellow sticks. The Fo-Fc omit map (blue) of the H4<sub>14-25</sub>K20me3 peptide was contoured at the 2.0  $\sigma$  level. The zinc ions are shown as purple spheres. **(g)** Structural superposition of hDNMT1<sub>BAH1</sub> (derived from PDB 4WXX) and the bDNMT1<sub>BAH1</sub>-H<sub>14-25</sub>K20me3 complex. The structural overlay of bDNMT1<sub>BAH1</sub> and (PDB 4WXX) yields an RMSD of 1.0 Å over 130 C $\alpha$  atoms.



**Supplementary Figure 2. Structural and biochemical analysis of the interaction between DNMT1 BAH1 domain and H4K20me2/3.** (a) The crystal structure of bDNMT1<sub>BAH1</sub> (cyan) in complex with the H4<sub>14-25</sub>K20me2 (yellow) peptide, with the H4K20me2-binding pocket shown in expanded view. (b) The crystal structure of ORC1<sub>BAH1</sub> (green) in complex with the H4<sub>14-25</sub>K20me2 peptide (yellow) (PDB 4DOW), with the H4K20me2-binding pocket shown in expanded view. (c-g) ITC binding curves of hDNMT1<sub>BAH1</sub> mutants with the H4<sub>14-25</sub>K20me3 peptide. NDB: no detectable

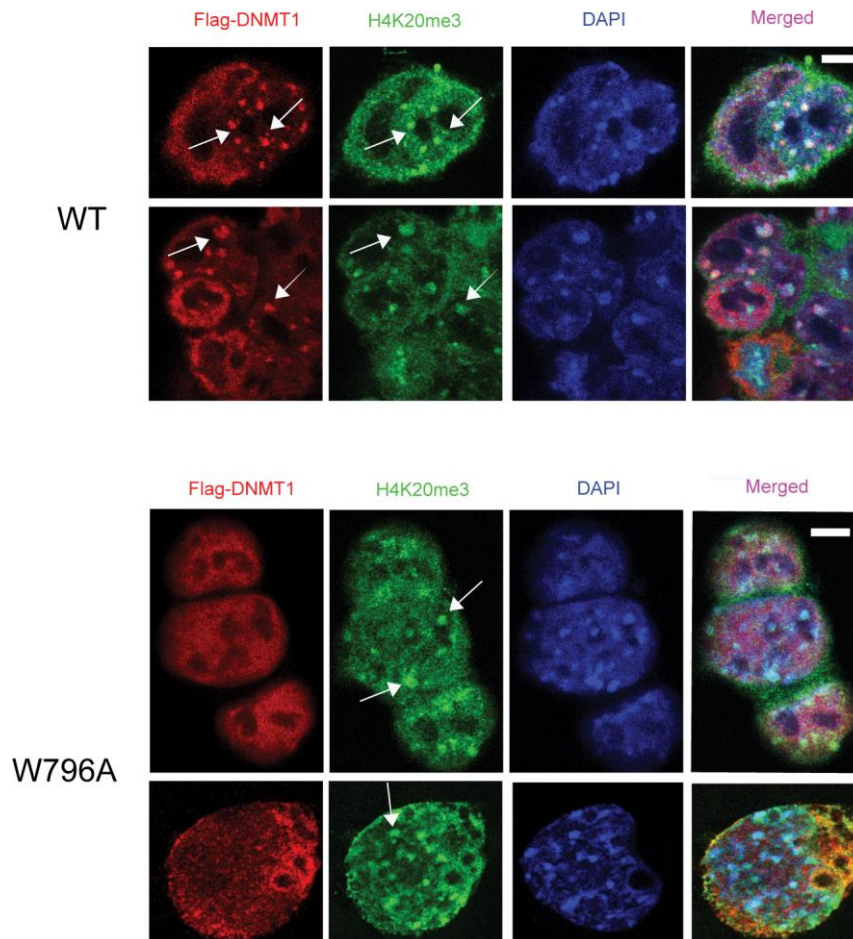
binding. **(h-l)** ITC binding curves of hDNMT1<sub>BAH1</sub> mutants with the H4<sub>14-25</sub>K20me2 peptide. NDB: no detectable binding.



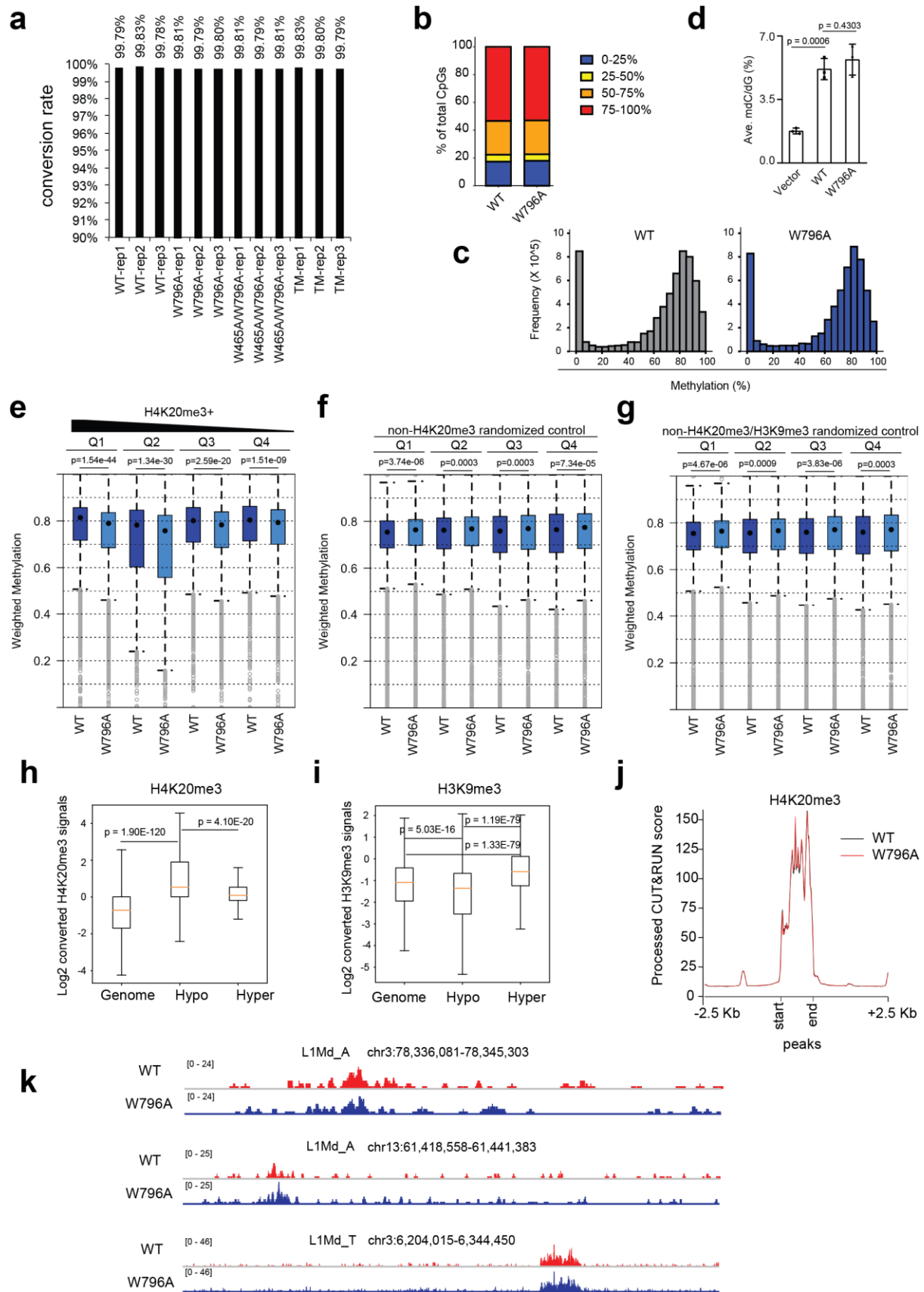
**Supplementary Figure 3. Displacement of the autoinhibitory linker by the DNA or H4K20me3 binding of DNMT1.** (a) Structural overlay DNA-free hDNMT1<sub>351-1600</sub> (PDB 4WXX; green) and the mDNMT1<sub>731-1602</sub>-DNA complex (PDB 4DA4; grey). The autoinhibitory linker of hDNMT1<sub>351-1600</sub> is colored in blue. The mDNMT1<sub>731-1602</sub>-bound DNA duplex is colored in light pink and clashes with residues P692-I709 (blue spheres) on the autoinhibitory linker of hDNMT1<sub>351-1600</sub> in expanded view. Zinc ions and SAH molecules are shown in sphere representation. (b) ITC binding analysis of the interaction between MBP-tagged hDNMT1<sub>351-1600</sub>, WT, W796A, Mut<sup>YW</sup> or Mut<sup>linker</sup>, and the H4<sub>14-25</sub>K20me3 peptide. (c) The RMSF values obtained for the autoinhibitory linker in the structural models of WT (PDB 4WXX) or W796A-mutated hDNMT1<sub>351-1600</sub> and the hDNMT1<sub>351-1600</sub>-H4K20me3 complex (complex 1 and complex 2) during 100 ns MD simulations. (d,e) Representative smFRET traces observed for hDNMT1<sup>CY</sup> in the absence (d) or presence (e) of H3K4me3. (f) Cumulative probability of the



durations of FRET events observed for hDNMT1<sup>CY</sup>, in the absence (black) or presence (red) of H3K4me3 peptide (212 and 151 molecules, respectively). **(g-i)** Representative smFRET traces observed for three population of peptide-free, W796A hDNMT1<sup>CY</sup> showing different FRET states.



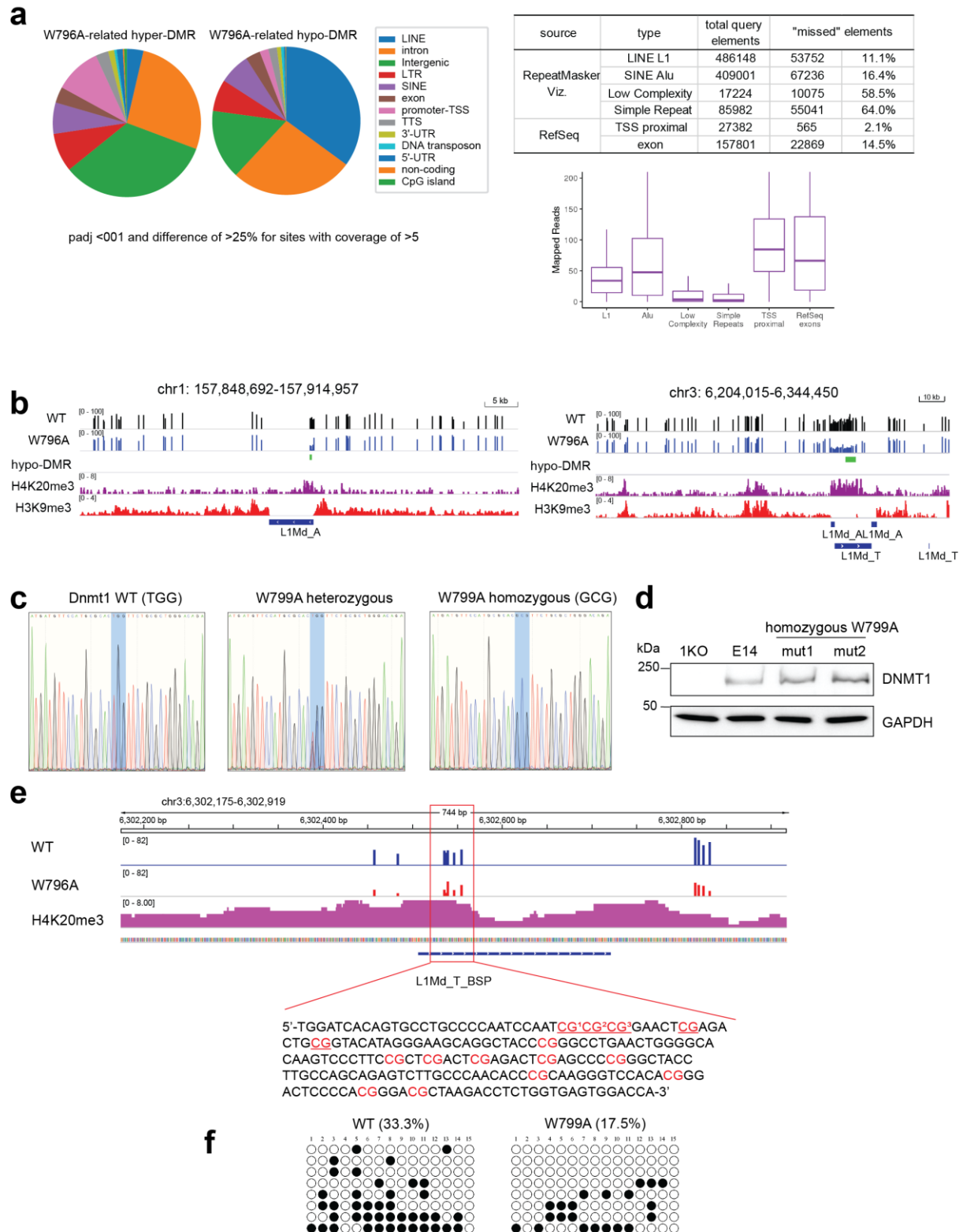
**Supplementary Figure 4. Replicates of IF images in Fig. 4c.** Representative confocal immunofluorescence images revealing localization of the indicated DNMT1 (Flag-tagged, red), H4K20me3 (green) and chromatin (stained by DAPI, blue) in the 1KO-ESC stable expression lines, synchronized in S phase. Scale bar, 5 micrometers. Selected H4K20me3 loci that are colocalized with WT (top) but not W796A (bottom) DNMT1 are indicated by arrows. The experiment was repeated 3 times with consistent results.



**Supplementary Figure 5. eRRBS profiling of DNA methylation in 1KO-ESC cells reconstituted with either WT or BAH1-mutated form of DNMT1.** (a) The rates of bisulfite conversion, labeled on top of columns, for all cytosines in each 1KO-ESC cell sample with expression of the indicated DNMT1, as determined by the unmethylated lambda DNA used as spike-in controls. TM denotes W464A/W465A/W796A. (b, c) Bar plots (b) and distribution plot (c) of absolute methylation levels for CpG sites with >5 coverage among 1KO-ESC lines with stable expression of the indicated DNMT1, as detected by eRRBS. (d) LC-MS analysis of global 5-mdC levels (calculated as 5-mdC/2MS analysis of g on the y axis) in 1KO-ESC lines after stable transduction of empty vector or the indicated DNMT1 (n=3 biological replicates). Data are mean±s.d. The two-tailed Student t-tests were performed to compare distributions between different groups. The data for 1KO and WT were adopted from those published previously<sup>1</sup>. (e) Distribution of weighted methylation scores at the H4K20me3 peaks categorized by quartile (Q1-Q4) of the descending H4K20me3 ChIP-seq signals. Scores are calculated after aggregating data from replicate samples per group. Peaks without mapped reads for a given sample group are ignored, resulting in N=6013 (WT Q1), N=6023 (W796A Q1), N=5908 (WT Q2), N=5910 (W796A Q2), N=5681 (WT Q3), N=5677 (W796A Q3), N=5409 (WT Q4), and N=5408 (W796A Q4). In the box-and-whisker plot, the box depicts the 25th to 75th percentiles, the black dot is the median, the whiskers extend to data points up to 1.5 IQR beyond the box, and open gray circles are data points outside the whisker range. (f, g) Distribution of weighted methylation scores at random genomic regions, which either lack H4K20me3 (f) or lack both H4K20me3 and H3K9me3 (g), which are size-matched to H4K20me3 peaks categorized by quartile. Scores are calculated after aggregating data from replicate samples per group. Random regions without mapped reads for a given sample group are ignored. For non-H4K20me3 randomized regions, N=5927 (WT Q1), N=5911 (W796A Q1), N=5836 (WT Q2), N=5840 (W796A Q2), N=5764 (WT Q3), N=5741 (W796A Q3), N=5588 (WT Q4), and N=5570 (W796A Q4). For random regions without H4K20me3 and H3K9me3, N=5918 (WT Q1), N=5906 (W796A Q1), N=5797 (WT Q2), N=5800 (W796A Q2), N=5709 (WT Q3), N=5718 (W796A Q3), N=5549 (WT Q4), and N=5532 (W796A Q4). In the box-and-whisker plot, the box depicts the 25th to 75th percentiles, the black dot is the median, the whiskers extend to data points up to 1.5 of IQR beyond the box, and open gray circles are data points outside the whisker range. (h, i) Bar plots showing the levels of H4K20me3 (h) or H3K9me3 (i) in ESCs, genome-wide (left) or within DMRs found to be associated with the W796A-mutated DNMT1 relative to WT, either hypo (middle) or hyper-methylated (right; for panels h-i, n = 8590 for Hypo, 4636 for Hyper and 8590 for randomized genome regions as a control). Y-axis, log2 converted ChIP-seq signals. DMR calling is determined by using methylkit. Box-and whisker plots depict 25-75% in the box, whiskers are 10-90%, and median is indicated. Statistical analysis in (e-i) used two-tailed Student's t-test. (j) Normalized H4K20me3 CUT&RUN signals over all called peaks in 1KO-ESCs rescued with exogenous

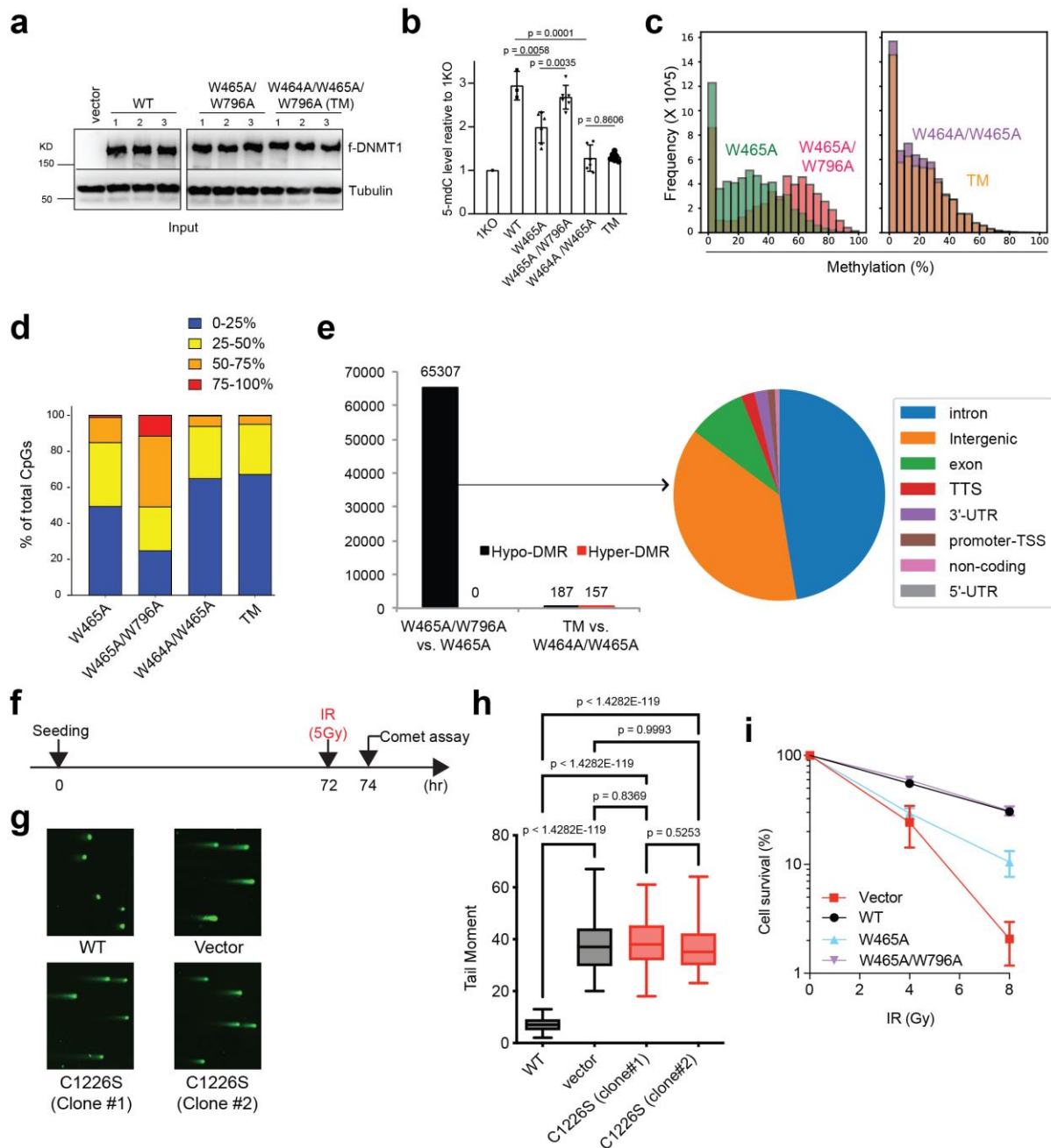


DNMT1, wither WT or W796A-mutated. The start and end points of called H4K20me3 peaks are labeled. **(k)** IGV profiles of H4K20me3 CUT&RUN signals at the indicated retrotransposon elements (the LINE-1 class) in 1KO-ESCs rescued with exogeneous DNMT1, wither WT or W796A-mutated.



**Supplementary Figure 6. Mutational analysis of DNMT1-mediated DNA methylation at LINE1.** (a) Left panels: Pie chart showing distribution of the DMRs associated with the W796A-mutated DNMT1 relative to WT, either hypo (left) or

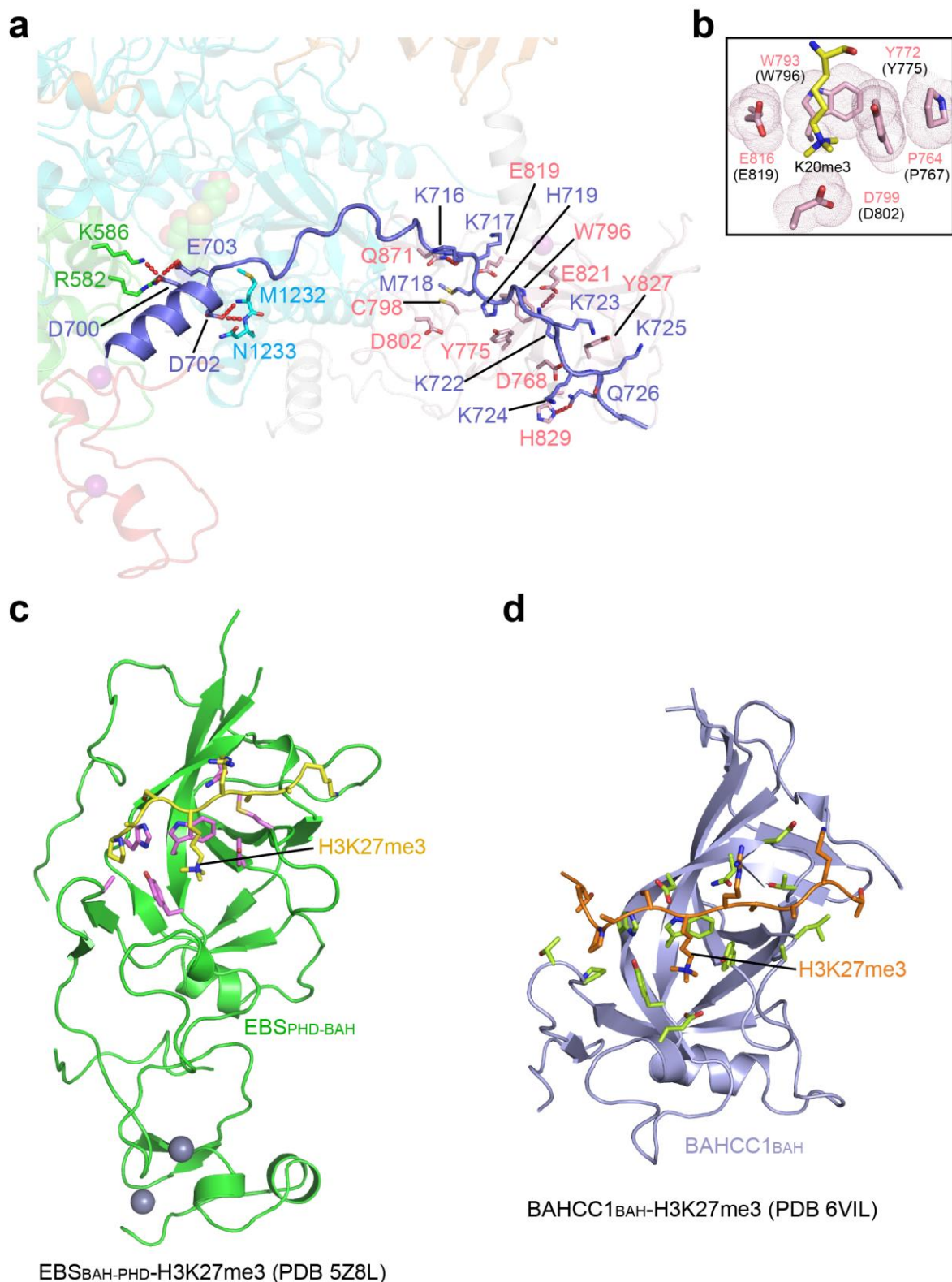
hyper-methylated (right), among the indicated genomic feature. DMR calling is determined by using methylkit. Right panels: a table for calculation of the fraction of “missed/ignored” sequencing reads (top) and boxplot showing the distribution of average mapped read depth (bottom) in each of the indicated group, including the repeated elements (LINE1/L1, SINE/Alu, simple repeat and low complexity repeats), exon and TSS-proximal (1kb upstream of annotated RefSeq TSS). eRRBS data analysis uses a cutoff of read depth of no less than 10 mapped read tags (per group), we define the ignored/missed elements as those that do not have mapped read depth of at least 10 at any coordinate in the element. Box-and whisker plots depict 25-75% in the box, whiskers are 10-90%, and median is indicated. P-value is calculated using logistic regression and then adjusted to padj value using sliding linear model (SLIM) method. **(b)** Representative IGV views show CpG methylation levels at the called W796A-associated hypo-DMRs (green bar), either located in the chromosome 1 or 3, among 1KO-ESC lines with stable expression of WT (black) or the BAH1-mutated (blue) DNMT1. Cytosines covered by at least 5 reads according to eRRBS data are shown, with each site designated by a vertical line. The bottom panels show IGV view of H4K20me3 (purple) and H3K9me3 (red) in WT ESCs, and position of LINE-1. **(c)** Representative Sanger sequencing results verified the Dnmt1 W799A knock-in mutation, either heterozygous or homozygous, in the E14 ESCs. **(d)** Immunoblotting for Dnmt1 in Dnmt1-null ESCs (1KO), WT E14 ESCs, and those with homozygous Dnmt1 W799A mutation. The experiment was repeated 3 times with consistent results. **(e)** IGV views of eRRBS and H4K20me3 profiles at the indicated LINE-1 element L1Md\_T (top), as well as the amplicon sequence (bottom) covering 15 of the highlighted CpG sites as examined by individual site-targeted bisulfite sequencing. **(f)** Site-targeted bisulfite sequencing results showed overall DNA methylation levels at a LINE-1 element, L1Md\_T, in E14 ESCs with either WT Dnmt1 or knockin of homozygous Dnmt1 W799A mutations.



**Supplementary Figure 7. DNMT1 BAH1 mutation crosstalks with the RFTS mutations in CpG methylation and genome stabilization.** (a) Immunoblots of the indicated Flag-tagged DNMT1 after stable reconstitution into the independently derived 1KO-ESC lines. The experiment was repeated 3 times with consistent results. (b) LC-MS analysis of global 5-methyl-2-deoxycytidine (5-mdC) levels (calculated as 5-mdC/2-deoxyguanosine on the y-axis) in 1KO-ESC lines after stable transduction of empty vector or the indicated DNMT1 (n = 3 biological replicates for 1KO and WT; n = 6 biological replicates for W465A, W465A/W796A, W465A/W796A and TM). Data are mean  $\pm$  s.d. Statistical analysis used two-tailed Student's t-test. The data for 1KO, WT, W465A and W464A/W465A were adopted from those published previously<sup>1</sup>. (c, d)



The distribution **(c)** and bar plots **(d)** showing absolute methylation levels of CpG sites with >5 coverage among 1KO-ESC lines with stable expression of the indicated DNMT1, as detected by eRRBS. **(e)** Summary of the total number of DMRs (left), either hyper-DMRs or hypo-DMRs, and pie chart (right) showing genomic distribution of the DMRs associated with BAH1 mutation (W796A) when introduced to the indicated RFTS mutant (either W465A or W464A/W465A) in 1KO-ESC cells. TM denotes W464A/W465A/W796A mutation. **(f)** Experimental scheme of the comet assay for (g,h). **(g)** Representative images of comet assay of DNMT1-/- 1KO-ESC cells complemented with empty vector or DNMT1, either WT or a catalytically dead mutant (C1226S; two different clones #1, and #2). **(h)** Box plot showing the tail moment of 1KO-ESC cells complemented with vector (n=103), WT (n=106), or C1242S (clone #1, n=110; clone #2, n=129). Each bar on the plot indicates the max value, upper quartile, median value, lower quartile, and minimum values, respectively. Plot was created and analyzed using one-way ANOVA with post Tukey analysis. **(i)** Survival of 1KO-ESC cells, reconstituted with vector control or the indicated DNMT1, post-treatment of IR. Data are mean  $\pm$  s.d. (n = 3 biologically independent replicates).



**Supplementary Figure 8. Evolutionarily conserved histone recognition by the BAH domains.** (a) Close-up view of the autoinhibitory linker-mediated intramolecular interaction of DNMT1. The autoinhibitory linker (slate) forms contacts with the BAH1 domain (light pink), as well as the RFTS (green) and MTase (cyan) domains. The

interacting residues are shown in stick representation. The hydrogen bonds are depicted as dashed lines. **(b)** The H4K20me3-binding pocket of bDNMT1<sub>BAH1</sub>, with the equivalent pocket residues in hDNMT1<sub>BAH1</sub> labelled in parentheses. **(c)** Ribbon representation of the crystal structure of plant EBS BAH-PHD (EBS<sub>BAH-PHD</sub>) dual domain bound to a H3K27me3 peptide (PDB 5Z8L). The H3K27me3-binding sites (magenta) are shown in stick representation. **(d)** Ribbon representation of the crystal structure of mouse BAHCC1 BAH (BAHCC1<sub>BAH</sub>) domain bound to a H3K27me3 peptide (PDB 6VIL). The H3K27me3-binding sites (limon) are shown in stick representation.

**Supplementary Table 1. Summary of ITC binding parameters.**

| Protein                                                            | Peptide                      | $K_d$ ( $\mu$ M) | N value         |
|--------------------------------------------------------------------|------------------------------|------------------|-----------------|
| hDNMT1 <sub>BAH1</sub> , WT                                        | H4(14-25) <sup>#</sup>       | 18.8 $\pm$ 2.6   | 0.75 $\pm$ 0.16 |
| hDNMT1 <sub>BAH1</sub> , WT                                        | H4(14-25)K20me <sup>#</sup>  | 11.4 $\pm$ 2.5   | 0.74 $\pm$ 0.16 |
| hDNMT1 <sub>BAH1</sub> , WT                                        | H4(14-25)K20me2 <sup>#</sup> | 4.1 $\pm$ 1.1    | 0.74 $\pm$ 0.05 |
| hDNMT1 <sub>BAH1</sub> , WT                                        | H4(14-25)K20me3 <sup>#</sup> | 2.7 $\pm$ 0.6    | 0.79 $\pm$ 0.01 |
| hDNMT1 <sub>BAH1</sub> , D768A                                     | H4(14-25)K20me3 <sup>#</sup> | 11.5 $\pm$ 2.1   | 0.82 $\pm$ 0.26 |
| hDNMT1 <sub>BAH1</sub> , Y775A                                     | H4(14-25)K20me3              | 361 $\pm$ 17.0   | 1.00*           |
| hDNMT1 <sub>BAH1</sub> , W796A                                     | H4(14-25)K20me3              | NDB              |                 |
| hDNMT1 <sub>BAH1</sub> , E819A                                     | H4(14-25)K20me3              | 9.4 $\pm$ 0.7    | 0.73 $\pm$ 0.01 |
| hDNMT1 <sub>BAH1</sub> , E821A                                     | H4(14-25)K20me3              | 91.7 $\pm$ 12.4  | 0.88 $\pm$ 0.06 |
| hDNMT1 <sub>BAH1</sub> , D768A                                     | H4(14-25)K20me2              | 9.3 $\pm$ 0.1    | 0.92 $\pm$ 0.01 |
| hDNMT1 <sub>BAH1</sub> , Y775A                                     | H4(14-25)K20me2              | NDB              |                 |
| hDNMT1 <sub>BAH1</sub> , W796A                                     | H4(14-25)K20me2              | NDB              |                 |
| hDNMT1 <sub>BAH1</sub> , E819A                                     | H4(14-25)K20me2              | 12.3 $\pm$ 1.0   | 0.87 $\pm$ 0.02 |
| hDNMT1 <sub>BAH1</sub> , E821A                                     | H4(14-25)K20me2 <sup>#</sup> | 35.4 $\pm$ 2.2   | 0.83 $\pm$ 0.03 |
| hDNMT1 <sub>BAH1</sub> , WT                                        | hDNMT1(714-727)              | 70 $\pm$ 10      | 1.00 $\pm$ 0.12 |
| hDNMT1 <sub>BAH1</sub> , W796A                                     | hDNMT1(714-727)              | NDB              |                 |
| hDNMT1 <sub>BAH1</sub> , Mut <sup>YW</sup>                         | hDNMT1(714-727)              | NDB              |                 |
| hDNMT1 <sub>RFTS</sub> , WT+H4K20me3<br>(Mixed in 1:2 molar ratio) | H3(1-22)K9me3                | NDB              |                 |
| MBP                                                                | H4(14-25)K20me3              | NDB              |                 |
| MBP-hDNMT1 <sub>351-1600</sub> , WT                                | H4(14-25)K20me3              | >300             |                 |
| MBP-hDNMT1 <sub>351-1600</sub> , W796A                             | H4(14-25)K20me3              | NDB              |                 |
| MBP-hDNMT1 <sub>351-1600</sub> , Mut <sup>YW</sup>                 | H4(14-25)K20me3              | NDB              |                 |
| MBP-hDNMT1 <sub>351-1600</sub> , Mut <sup>Linker</sup>             | H4(14-25)K20me3 <sup>#</sup> | 122.5 $\pm$ 3.3  | 1.45 $\pm$ 0.26 |
| JMJD2A <sub>TD</sub>                                               | H4(14-25)K20me3 <sup>#</sup> | 0.51 $\pm$ 0.04  | 0.97 $\pm$ 0.04 |
| ORC1 <sub>BAH</sub>                                                | H4(14-25)K20me2 <sup>#</sup> | 8.7 $\pm$ 0.2    | 0.75 $\pm$ 0.01 |

NDB, no detectable binding. \*N value was fixed to 1.00 for curve fitting. <sup>#</sup>The mean value and S.D. were derived from at least two-independent measurements. The rest were derived from single measurements.

**Supplementary Table 2. List of Primers used in this study.**

| Gene_primer                                                                    | Sequence                                                                                                                                                                                                    | Species | Application                                                        |
|--------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------|--------------------------------------------------------------------|
| <b><u>CRISPR/Cas9-sgRNA</u></b>                                                |                                                                                                                                                                                                             |         |                                                                    |
| crRNA_W796A                                                                    | AGATGATGTTCCATGCGCAC                                                                                                                                                                                        | mouse   | CRISPR/Cas9-based editing and generation of a Dnmt1 W799A mutation |
| genotyping_Dnmt1_W796A_for                                                     | GCTCCATCGCTTGCTTTTCAG                                                                                                                                                                                       | mouse   | PCR genotyping for Dnmt1 W799A mutation in the E14 knockin model   |
| genotyping_Dnmt1_W796A_rev                                                     | CCTTGCGTACTCCTGGTTGT                                                                                                                                                                                        |         |                                                                    |
| W796A_ssODN                                                                    | A*C*T*CGCCCACCAGGAACAGTTCC<br>AGGGGGTTCGGAGGTGGCTCCCAAG<br>GACTGTGTCTGTCCCAGCGCAGAA<br>CGCGTGCGCATGGAACATCATCTG<br>ACCATTTTTGTCTTCCCACAGAGCT<br>GTGACCCTGGCTAGATAGAGT*G*G<br>*T (*: phosphorothioate bonds) | mouse   | CRISPR/Cas9-based editing and generation of a Dnmt1 W799A mutation |
| Dnmt1-ScF1                                                                     | TCTAGCCAGGTATGCATGCC                                                                                                                                                                                        | mouse   | Sequencing primer for genotyping of mouse Dnmt1 W799A mutation     |
| Dnmt1-ScR1                                                                     | TGGGACAGAACATAGGGACG                                                                                                                                                                                        |         |                                                                    |
| <b><u>Biochemical Assay for protein interaction and enzymatic activity</u></b> |                                                                                                                                                                                                             |         |                                                                    |
| DNMT1_728_PRS_F_F                                                              | CGCGGCGGTCTCGGATCCAAGAAT<br>CGCATCTCTTGGG                                                                                                                                                                   | human   | Cloning of His-SUMO-tagged hDNMT1 BAH1 domain                      |
| DNMT1_900_PRS_F_R                                                              | CGCGGCGGTCTCCTCGAGTTAAGC<br>CAGACGGGCACAGCTCACACAG                                                                                                                                                          |         |                                                                    |
| DNMT1_D768A_F                                                                  | TTATTCCAGCTGATTCCTCAAAACC<br>GCTGTAT                                                                                                                                                                        | human   | hDNMT1 BAH1 D768A mutation                                         |
| DNMT1_D768A_R                                                                  | AGGAATCAGCTGGAATAACAGAGA<br>CACAGTCC                                                                                                                                                                        |         |                                                                    |
| DNMT1_Y775A_F                                                                  | AACCGCTGGCTCTAGCAAGGGTCA<br>CGGCGCTG                                                                                                                                                                        | human   | hDNMT1 BAH1 Y775A mutation                                         |
| DNMT1_Y775A_R                                                                  | CTTGCTAGAGCCAGCGGTTTTGAG<br>GAATCATC                                                                                                                                                                        |         |                                                                    |
| DNMT1_W796A_F                                                                  | ACGCCCACGCGTTCTGCGCTGGGA<br>CAGACACA                                                                                                                                                                        | human   | hDNMT1 BAH1 W796A mutation                                         |
| DNMT1_W796A_R                                                                  | GCAGAACGCGTGGGCGTGAAACAT<br>CTGCCCGT                                                                                                                                                                        |         |                                                                    |
| DNMT1_E819A_F                                                                  | TGGTGGATGCATGTGAGGACATGC<br>AGCTTTCA                                                                                                                                                                        | human   | hDNMT1 BAH1 E819A mutation                                         |
| DNMT1_E819A_R                                                                  | TCCTCACATGCATCCACCAAGAACA                                                                                                                                                                                   |         |                                                                    |



|                     |                                                  |        |                                                           |
|---------------------|--------------------------------------------------|--------|-----------------------------------------------------------|
|                     | GCTCCAG                                          |        |                                                           |
| DNMT1_E821A_F       | GATGAATGTGCGGACATGCAGCTT<br>TCATATAT             | human  | hDNMT1 BAH1 E821A<br>mutation                             |
| DNMT1_E821A_R       | TGCATGTCCGCACATTCATCCACCA<br>AGAACAG             |        |                                                           |
| DNMT1_728_GST_F     | GGGCCCCCTGGGATCCAAGAATCGC<br>ATCTCTTGGGTCGG      | human  | Cloning of GST-tagged<br>hDNMT1 (728-1600)                |
| DNMT1_1600_GST_R    | GTGGTGGTGCTCGAGTTAGGCTTT<br>GGCCAACATACAAAGC     |        |                                                           |
| hDNMT1-BAH1-F<br>WD | CGATCGAATTCAAGAATCGCATCTC<br>TTGGGTCGG           | human  | Cloning of GST-tagged<br>hDNMT1 BAH1                      |
| hDNMT1-BAH1-REV     | CGATCGTCGACTTAAGCCAGACGG<br>GCA                  |        |                                                           |
| Bovine_DNMT1_725_F  | CCGCCGGATCCAAGAATCGGATCT<br>CTT GGGTTGG          | bovine | Cloning of<br>His-SUMO-tagged<br>bDNMT1 BAH1 domain.      |
| Bovine_DNMT1_897_R  | CCGCCCTCGAGTTAGGCCAGACGT<br>GCACAGCTTGCGC        |        |                                                           |
| DNMT1_MBP_F         | GTACATCCAAGGATCCCCAAGTG<br>CATTCAAGTGCAGGGC      | human  | Cloning of<br>His-MBP-tagged<br>hDNMT1(351-1600).         |
| DNMT1_MBP_R         | GTCCTACAGGCGCGCCTCATTAGG<br>CTTTGGCCAACATACAAAGC |        |                                                           |
| DNMT1_Linker_1F     | GCGGCGCAGGGGAAGGCGGCGGC<br>ACAGAACAAGAATCGCATC   | human  | hDNMT1<br>M718A/H719A/K723A<br>K724A/K725A mutation       |
| DNMT1_Linker_1R     | GCCGCCGCCTTCCCCTGCGCCGCT<br>TTTTTGGGTGACGGCATCTC |        |                                                           |
| DNMT1_Linker_2F     | GGCGGCGGCAGCGAACAAGAATC<br>GCATCTCTTGGG          | human  | hDNMT1<br>M718A/H719A/K723A/K724A/K725A/Q726A<br>mutation |
| DNMT1_Linker_2R     | GATTCTTGTTGCTGCGCCGCCTT<br>CCCCTGCGCCGC          |        |                                                           |
| hDNMT1_BAH2_MBP_F   | GTACATCCAAGGATCCGAGATGAG<br>GCAAAAAGAAATCC       | Human  | Cloning of<br>His-MBP-tagged<br>hDNMT1 BAH2 domain        |
| hDNMT1_BAH2_MBP_R   | GTCCTACAGGCGCGCCTCATCCAG<br>GGCTACGGGCATGGTTGGGA |        |                                                           |
| JMJD2A_PRSF_F       | ACAGATTGGTGGATCCGGGGCCTT<br>GCAAAGCATC           | Human  | Cloning of<br>His-SUMO-tagged<br>JMJD2A Tudor domain.     |
| JMJD2A_PRSF_R       | CTTTACCAGACTCGAGTTAGGGAA<br>GCTCTTCATCCAGTGTG    |        |                                                           |

#### **SmFRET Assay**

|                   |                                                                      |       |                                                                |
|-------------------|----------------------------------------------------------------------|-------|----------------------------------------------------------------|
| DNMT1_LPETG_609_F | GCAGACCATCAGGCATTCTCTGCC<br>GGAGACCGGATAACTCGAGTCTGG<br>TAAAGAAACCGC | human | Generation of<br>His-SUMO-tagged<br>hDNMT1<br>(351-609)-LPETG. |
| DNMT1_LPETG_6     | TGCCTGATGGTCTGCCGCCTCGCC                                             |       |                                                                |

|                        |                                                                 |       |                                                                      |
|------------------------|-----------------------------------------------------------------|-------|----------------------------------------------------------------------|
| 09_R                   | TGGGCTCGCCTCTGTCCCAGCGTG                                        |       |                                                                      |
| DNMT1_C616_F           | ACCAGGGAGAAGGACAGGGGACC<br>CTGCAAAGCCACCACCACCAAGC              | human | Generation of<br>His-SUMO-tagged<br>hDNMT1 (351-639)<br>T616C-LPETG. |
| DNMT1_640_R            | TTATCCGGTCTCCGGCAGATCCTTT<br>TCAATTTG                           |       |                                                                      |
| LPETG_PRSF_F           | CTGCCGGAGACCGGATAACTC                                           | Human | Generation of<br>His-SUMO-tagged<br>hDNMT1 (351-639)<br>T616C-LPETG  |
| DNMT1_612_R            | GTCCTTCTCCCTGGTAGAATGCCTG<br>ATGGTCTGCCG                        |       |                                                                      |
| DNMT1<br>C409S_C420S_F | GTGTACTGTAAGCACGGTCACCTG<br>AGTCCCATCGACACCGGCCTCATC<br>G       | human | hDNMT1 RFTS<br>C409S/C420S mutation                                  |
| DNMT1<br>C409S_C420S_R | GTGCTTACAGTACACACTGAAGCTG<br>GTCAGTTTGTGCTGGGGAAGC              |       |                                                                      |
| DNMT1 C580S_F          | TCCTGACGCCCAGCATGCGGGACC<br>TGATCAAG                            | human | hDNMT1 RFTS C580S<br>mutation                                        |
| DNMT1 C580S_R          | GGTCCCGCATGCTGGGCGTCAGGA<br>AGATGGGC                            |       |                                                                      |
| DNMT1_646_F            | GTACATCCAAGGATCCAACGCCTTT<br>AAGCGCCGGC                         | human | Cloning of<br>His-MBP-tagged<br>hDNMT1<br>(646-1606)-LPETG           |
| DNMT1_<br>1606_LPETG_R | GTCCTACAGGCGCGCCTCATCCGG<br>TCTCCGGCAGAGCTGAGGCACTCT<br>CTCGGGC |       |                                                                      |
| DNMT1 S570C_F          | GAGAGTTATGACGAGGCCGGGGAC<br>TGTGATGAGCAGCCCATCTTCCTG            | human | hDNMT1 RFTS S570C<br>mutation                                        |
| DNMT1 S570C_R          | CAGGAAGATGGGCTGCTCATCACA<br>GTCCCCGGCCTCGTCATAACTCTC            |       |                                                                      |

**Supplementary Table 3. List of antibodies used for cellular analysis**

| Antibody targets                           | Vendor                   | Catalog # | Application (Antibody dilution) |
|--------------------------------------------|--------------------------|-----------|---------------------------------|
| Dnmt1 Antibody (H-12)                      | Santa Cruz Biotechnology | sc-271729 | immunoblot (1:200)              |
| GAPDH (14C10) Rabbit mAb                   | Cell Signaling           | #2118     | immunoblot (1:5000)             |
| Anti-Histone H4 (mono methyl K20) antibody | Abcam                    | ab9051    | immunoblot (1:3000)             |
| Anti-Histone H4 (di methyl K20)            | Abcam                    | ab9052    | immunoblot (1:3000)             |
| Anti-Histone H4 (tri methyl K20) antibody  | Abcam                    | ab9053    | immunoblot (1:3000),<br>CUT&RUN |
| Histone H3 antibody                        | Abcam                    | ab1791    | immunoblot (1:3000)             |
| goat anti-mouse IgG HRP                    | Santa Cruz Biotechnology | sc-2005   | immunoblot (1:5000)             |
| goat anti-rabbit IgG HRP                   | Santa Cruz Biotechnology | sc-2004   | immunoblot (1:5000)             |

## References

1. Ren, W. *et al.* Direct readout of heterochromatic H3K9me3 regulates DNMT1-mediated maintenance DNA methylation. *Proc Natl Acad Sci U S A* **117**, 18439-18447 (2020).