

1 **The Acute Myeloid Leukemia variant DNMT3A Arg882His is a DNMT3B-like**
2 **enzyme**

3 Allison B. Norvil¹, Lama AlAbdi¹, Bigang Liu², Nicole E. Forstoffer¹, Amie R. Michie¹,
4 Taiping Chen², Humaira Gowher*¹

5 ¹Department of Biochemistry, Purdue University, West Lafayette, Indiana 47907

6 ²Department of Epigenetics and Molecular Carcinogenesis, Division of Basic Sciences, The
7 University of Texas MD Anderson Cancer Center, Houston, Texas 77054

8 *Corresponding author: hgowher@purdue.edu

9

10 Keywords: AML, DNMT3A, DNMT3B, DNA methylation, flanking sequence preference, enzymatic
11 activity, kinetic mechanism

12

13

14

15

16

17

18

19 **Abstract**

20 Mutations in DNMT3A, particularly the Arg882His substitution is highly prevalent in
21 acute myeloid leukemia. Although the reduced activity of DNMT3A Arg882His variant
22 alters DNA methylation, the underlying cause of its oncogenic effect is not fully
23 understood. Our data show that DNMT3A Arg882His variant acquires CpG flanking
24 sequence preference highly similar to that of DNMT3B. Interestingly, a similar substrate
25 preference was observed in DNMT3A WT enzyme upon the loss of cooperative kinetic
26 mechanism. We tested if DNMT3A Arg882His could preferably methylate DNMT3B-
27 specific target sites. Rescue experiments in *Dnmt3a/3b* double knockout mouse
28 embryonic stem cells show that the corresponding Arg878His mutation in mouse
29 DNMT3A severely impairs its ability to methylate major satellite DNA, a DNMT3A-
30 preferred target, but has no overt effect on the ability to methylate minor satellite DNA, a
31 DNMT3B-preferred target. Our data suggest that methylation of DNMT3B target sites by
32 DNMT3A Arg882His variant could contribute to its oncogenic potential.

33 **Introduction**

34 DNA methylation in mammals is critical for development and maintenance of the
35 somatic cell state^{1,2}. It has diverse functions including regulation of gene expression and
36 silencing of repetitive elements³⁻⁵. In mammals, CpG methylation is established and
37 maintained by DNA methyltransferases (DNMTs)⁶. DNMT1 largely functions as a
38 maintenance methyltransferase by copying the methylation pattern from parent to
39 daughter strand during DNA replication^{7,8}. The DNMT3 family includes two active
40 homologs DNMT3A and DNMT3B, and an inactive homolog DNMT3L^{9,10}. These

41 enzymes perform *de novo* DNA methylation, which predominantly takes place during
42 early embryogenesis and pluripotent stem cell differentiation^{9,11,12}. Whereas DNMT3A is
43 expressed ubiquitously, DNMT3B is highly expressed during early embryogenesis and
44 largely silenced in somatic cell types¹³⁻¹⁵. Despite having a high sequence similarity
45 (>40%), DNMT3A and DNMT3B have distinct preferences for some target sites,
46 although many sites can be methylated redundantly¹⁶. This is represented by a
47 preference of DNMT3A for major satellite repeats, whereas DNMT3B preferentially
48 targets minor satellite repeats^{9,11}.

49 Mutations in DNMT3A and DNMT3B have been identified in several diseases. Germline
50 transmitted mutations of DNMT3A and DNMT3B cause Tatton-Brown-Rahman
51 syndrome and ICF (Immunodeficiency, Centromeric instability, and Facial anomalies)
52 syndrome, respectively^{9,17-20}. Somatic mutations in DNMT3A are commonly found in
53 patients with acute myeloid leukemia (AML) and other hematologic neoplasms^{21,22}.
54 Detailed studies revealed that ~20% of AML cases have heterozygous DNMT3A
55 mutations, with the majority (~60-70%) carrying the Arg882His mutation, which cause
56 genome-wide hypomethylation^{21,23-25}. Given that genetic knockout of one copy of
57 DNMT3A exhibits no obvious phenotype, the heterozygous DNMT3A Arg882His
58 mutation was suggested to have a dominant negative effect^{9,21}. The observation that the
59 expression of the murine DNMT3A Arg878His variant (equivalent to Arg882His in
60 human DNMT3A) in mouse embryonic stem cells (mESCs) causes genome-wide loss of
61 DNA methylation further supported the dominant negative activity of this variant^{26,27}. In
62 contrast, DNMT3B mutations have not been identified in cancer. However aberrant
63 overexpression of DNMT3B is highly tumorigenic, including in AML^{14,28-30}.

64 Overexpression of DNMT3B in AML leads to disease prognosis similar to that in
65 patients with DNMT3A Arg882His mutation^{29,30}.

66 The effect of Arg882His in the activity of DNMT3A could be predicted from the crystal
67 structure of the DNMT3A catalytic domain (DNMT3A-C) showing that it forms
68 homodimers and tetramers through two interaction surfaces^{31,32}. Several AML-
69 associated DNMT3A mutations, including Arg882His, which are present at or close to
70 the protein-protein interface disrupt tetramer formation and lead to reduced catalytic
71 activity *in vitro*^{21,27,33-36}. Further, DNMT3A-C tetramers can oligomerize on DNA, forming
72 nucleoprotein filaments^{37,38}. This oligomerization allows the enzyme to bind and
73 methylate multiple CpG sites in a cooperative manner, thus increasing its activity³⁹. Our
74 previous biochemical studies show that in DNMT3A-C, the Arg882His substitution
75 results in loss of cooperativity potentially causing a decrease in its catalytic activity⁴⁰.

76 Interestingly, in DNMT3B, which is a non-cooperative enzyme, the homologous
77 mutation, Arg829His, has no effect on its catalytic activity *in vitro*⁴⁰. A recent co-crystal
78 structure of the DNMT3A-C/DNA complex shows that Arg882 interacts with the
79 phosphate backbone of the DNA at the N+3 position downstream of the CpG site⁴¹. The
80 DNMT3A-C Arg882His variant was shown to have a bias for G at N+3 when compared
81 to the wild-type (WT) enzyme⁴².

82 In this study, we tested whether this flanking sequence preference was directly due to
83 the substitution of Arg with His or caused by loss of cooperative mechanism in the
84 variant enzyme. Our data show that all variants of Arg882 found in AML patients have
85 low catalytic activity and lack a cooperative mechanism. We further show that in the
86 absence of cooperative mechanism at low enzyme concentration, the DNMT3A-C WT

87 enzyme prefers G at the N+3 position, similar to that of DNMT3A-C Arg882His variant.
88 Based on previous observations that DNMT3B-C is a non-cooperative enzyme, we
89 performed a comparative analysis of its substrate preference with that of the DNMT3A-
90 C WT and the Arg882His variant. DNA methylation levels at 56 CpGs were rank
91 ordered to compute consensus motifs that are preferred by these enzymes.
92 Interestingly, our data show that the Arg882His variant and DNMT3B-C have a similar
93 preference for nucleotides at N+1, 2 and 3 positions flanking the CpG site. These data
94 strongly support that the gain of flanking sequence preference is due to loss of
95 cooperative mechanism, and suggest that the DNMT3A-C Arg882His variant could
96 methylate DNMT3B-C preferred sites. We tested this “gain of function” prediction by
97 expressing WT mouse DNMT3A or the Arg878His (corresponding to Arg882His in
98 human DNMT3A) in *Dnmt3a/3b* double knockout (DKO) mESCs. Our data show that
99 whereas the DNMT3A Arg878His variant failed to rescue methylation at the major
100 satellite repeats (DNMT3A preferred target sites), its ability to methylate the minor
101 satellite repeats (DNMT3B preferred target sites), was comparable to that of DNMT3A
102 WT enzyme. Further, our analysis of the CpG flanking sequences of the satellite
103 repeats show that whereas major satellite repeats have an A or T at the N+3 position,
104 the minor satellite repeats are enriched in G at N+3 position. This observation provides
105 the previously unknown explanation for the substrate specificity of DNMT3A and
106 DNMT3B for major and minor satellite repeats, respectively. Taken together, our data
107 provide novel mechanistic insights into DNMT3A and DNMT3B substrate specificities
108 that could influence the oncogenic potential of these enzymes. We suggest that in

109 leukemia, DNMT3A Arg882His substitution establishes a DNMT3B-like activity, the
110 tumorigenic properties of which are exploited by cancer cells.

111 **Results**

112 **The cooperative kinetic mechanism is absent in DNMT3A Arg882 variants**

113 The catalytic properties of the DNMT3A-C Arg882His variant have been previously
114 characterized *in vitro* showing that the mutation reduces the activity and attenuates
115 tetramerization of the enzyme^{21,27,33}. DNMT3A-C WT methylates neighboring CpGs on
116 DNA by a cooperative kinetic mechanism, in which the DNA bound tetramer promotes
117 successive binding events through protein-protein interaction³⁹. Our previous studies
118 also show that the DNMT3A-C Arg882His variant fails to methylate DNA by a
119 cooperative mechanism, which could be due to impaired tetramerization^{26,33,40}. It is not
120 clear, however, whether this is due to the loss of Arg or gain of His at this position.
121 Based on the observation that many AML patients have mutations that lead to
122 substitution of Arg882 to Cys or Ser, we asked if these variant enzymes could methylate
123 multiple CpGs on a single DNA molecule in a cooperative manner. His-tagged
124 recombinant DNMT3A-C (catalytic domain) WT and Arg882 variants were produced in
125 *E. coli* and purified using Ni-NTA affinity chromatography to about 90 – 95% purity
126 (Supplementary Fig. 1a)⁴³. The catalytic activity of DNMT3A-C Arg882 variants was
127 compared to the WT enzyme by performing DNA methylation assays using a 30-bp
128 substrate containing one CpG site. Consistent with previous reports, a 60 - 80% loss of
129 catalytic activity was observed in all variants compared to WT enzyme (Fig.
130 1a)^{27,33,35,36,40,42}. Cooperativity is observed as an exponential relationship between

131 catalytic activity enzyme concentrations shown for DNMT3A-C WT. We tested the
132 cooperative mechanism of the Arg882 variants using the pUC19 plasmid. All Arg882
133 variant enzymes failed to methylate the substrate in a cooperative fashion (Fig. 1b).
134 From these data, we conclude that the Arg882 residue plays a key role in the
135 cooperative mechanism of DNMT3A-C.

136 **Loss of cooperativity modulates flanking sequence preference of DNMT3A**

137 A recent study of the crystal structure of DNMT3A shows that the Arg882 residue
138 interacts with the phosphate backbone of the nucleotide at N+3 position (N=CpG) and
139 contributes to DNA binding⁴¹. It was shown recently that the Arg882His has a
140 preference for G at the N+3 position compared to the WT enzyme⁴². Given that the
141 Arg882 residue is also necessary for the cooperative kinetic mechanism of DNMT3A,
142 we tested the relationship between loss of cooperative mechanism and flanking
143 sequence preference. We performed *in vitro* methylation of a 509-bp DNA substrate
144 containing 56 CpG sites using the WT and Arg882His variant enzymes (Supplementary
145 Fig. 2). DNA methylation was quantified by bisulfite conversion and high throughput
146 sequencing. Our data confirmed the previously reported G preference at N+3 for
147 Arg882His variant compared to the WT enzyme (Fig. 2a). We have previously shown
148 that at a low concentration (0.25 μ M), DNMT3A-C does not multimerize on the 509-bp
149 DNA substrate and therefore cannot methylate multiple CpGs using cooperative
150 mechanism, which is observed at a higher concentration (1 μ M) of the enzyme⁴⁰. DNA
151 methylation assays were performed using 0.25 μ M and 1 μ M enzyme, and the flanking
152 sequence preferences were compared. The preference of sites by DNMT3A-C at 0.25
153 μ M over 1 μ M is represented by fold change greater than 1 (Fig. 2b, Supplementary

154 Table 1). Strikingly, 3 out of 4 sites preferred by DNMT3A-C at 0.25 μ M have G at N+3
155 position, which is similar to the flanking sequence preference of the DNMT3A-C
156 Arg882His variant. When compared to the relative preference at the N+1 and N+2
157 positions, interestingly, the similarity was specifically observed at N+3 position,
158 suggesting the role of cooperativity in modulating the interaction of the Arg882 residue
159 with DNA.

160 **DNMT3B has a flanking sequence preference similar to DNMT3A Arg882His**

161 DNMT3B is a homolog of DNMT3A that is frequently overexpressed in tumors, including
162 AML^{14,29}. Similar to DNMT3A-C Arg882His variant, DNMT3B-C functions as a non-
163 cooperative enzyme⁴⁰. Therefore, we asked if DNMT3B-C and the DNMT3A-C
164 Arg882His variant have a similar flanking sequence preference. Although the nucleotide
165 preference of DNMT3B is reported for N+1 position, the extended flanking sequence
166 preference has not been thoroughly evaluated⁴⁴. Using recombinant DNMT3B-C WT
167 and the Arg829His variant (homologous to DNMT3A Arg882His), we performed *in vitro*
168 methylation assays using the 509-bp DNA substrate. We compared the preferred sites
169 of DNMT3B-C WT with DNMT3A-C WT, DNMT3B-C WT with DNMT3A-C Arg882His,
170 and DNMT3B-C Arg829His with WT. The preferred sites of DNMT3B-C WT compared
171 to DNMT3A-C WT showed pronounced increase at 11 sites, of which 8 had G at N+3
172 position (Fig. 3a). Strikingly, many of these sites also overlapped with those preferred
173 by the DNMT3A-C Arg882His enzyme. This was confirmed by a direct comparison of
174 the preferred sites of DNMT3B-C WT and the DNMT3A Arg882His enzyme, which
175 showed only 4 sites preferred by DNMT3B-C over DNMT3A-C Arg882His (Fig. 3b).
176 Interestingly, 3 out of the 4 sites carry G at N+3 position. The data also showed only a

177 few sites that were preferred by DNMT3B-C the Arg829His variant compared to WT,
178 suggesting that the Arg829 mutation has little or no effect on DNMT3B-C activity and
179 flanking sequence preference (Fig. 3c). Unlike DNMT3A, DNMT3B methylates DNA
180 using a processive kinetic mechanism⁴⁰. Our data confirmed that, similar to DNMT3A-C
181 WT, the Arg882His variant is a non-processive enzyme indicating that the Arg882His
182 mutation only affects its flanking sequence preference (Supplementary Fig. 3). Taken
183 together, these data suggest that the Arg882His substitution alters the specificity of
184 DNMT3A to be similar to that of DNMT3B.

185 **The DNMT3A Arg882His variant acquires DNMT3B-like substrate preference**

186 The methylation assays to determine the flanking sequence preference described
187 above were performed for 10 minutes, which represents initial enzyme kinetics. To
188 evaluate the substrate preference during multiple turnovers, the methylation assays
189 were carried out for 30 and 60 minutes, allowing the enzyme kinetics to enter the steady
190 state. Comparing site preference for DNMT3A-C Arg882His to DNMT3A-C WT, the data
191 show progressive loss of preference for G at N+3 position from 30 to 60 minutes (Fig.
192 4a, b). This is expected under conditions where the enzyme, after methylating the
193 preferred sites during the initial reaction, methylates other sites under multiple turnover
194 conditions. A similar comparison of site preference of DNMT3B-C WT over DNMT3A-C
195 WT or Arg882His variant, however, shows that the preference for G at N+3 position is
196 maintained during 30- and 60-minute time points (Fig. 4b, c, e, f). This indicates that
197 whereas DNMT3B-C has a strong intrinsic specificity for G at the N+3 position and
198 methylates other sites at very low frequencies, the preference for this site is acquired by
199 the DNMT3A-C Arg882His variant enzyme.

200 **Temporal change in flanking sequence preference by WT and variant DNMT3**
201 **enzymes**

202 We extended the analysis of the flanking sequence preference to determine the
203 preferred trinucleotides at the N+1/2/3 and N-1/2/3 positions by DNMT3A-C, DNMT3B-
204 C and the variant enzymes. Based on the occurrence of different sites, the methylation
205 at a site was calculated as a ratio of observed to expected fractional methylation.

206 Comparing the flanking sequence preference of DNMT3A-C WT, DNMT3A-C
207 Arg882His, DNMT3B-C WT, and DNMT3B-C Arg829His, the data showed that whereas
208 DNMT3A-C Arg882His, DNMT3B-C WT, and DNMT3B-C Arg829His cluster together at
209 21 out of the 64-nucleotide combinations, DNMT3A-C WT shows an opposite and a
210 distinct preference (Fig. 5a, Supplementary Fig. 5a, Supplementary Table 2).

211 Based on the top 10 preferred sites, we used WebLogo application⁴⁵ to determine the
212 consensus sequence logo flanking the CpG sites for each enzyme. The analysis was
213 performed for methylation data collected at 10, 30, and 60 min time points to monitor
214 the temporal order of site preference by these enzymes (Fig. 5b). As expected, the data
215 again show a dramatic difference in flanking sequence preference between DNMT3A-C
216 and DNMT3B-C. Whereas DNMT3A-C WT shows a preference for A or T at the N+1/2/3
217 positions, DNMT3B-C prefers G and C. The preferred sequence of the DNMT3A-C
218 Arg882His variant is strikingly similar to that of DNMT3B-C, particularly at N+1, where it
219 loses preference for T, and at N+2, where it gains a strong preference for C. The sites
220 with A at the N+1 position are most preferred by DNMT3A-C WT and are methylated
221 within the first 10 minutes, whereas sites with T at the N+1 position are methylated at 30
222 and 60 minutes (Fig. 5b, Supplementary Fig. 5a). This is consistent with previous

223 studies which show DNMT3A to prefer T at N +1 from CpG site⁴⁴. DNMT3A-C WT
224 enzyme has least preference for sites with G at N+1 position, which is just the opposite
225 for DNMT3B-C and DNMT3A-C Arg882His (Supplementary Fig. 5b, c). Both DNMT3B-
226 C and DNMT3A-C Arg882His enzymes strongly prefer sites with C at N+2 position and
227 methylate sites with A and T at N+1 position slowly and with a weaker preference.
228 Comparing preference of all four enzymes for nucleotides at position N-1/3 shows a
229 weak or no preference, whereas at N-2, all of them show a strong preference for C and
230 T (Fig. 5b). This is in agreement with crystal structure data showing fewer interactions
231 between DNMT3A and the nucleotides upstream of CpG site⁴¹. Taken together, these
232 data show that DNMT3A-C Arg882His has a similar nucleotide preference as DNMT3B-
233 C WT, not only at N+3 but also at positions N+1 and N+2 from the CpG site.

234 We next selected CpG sites that had G at the N+3 position and generated the preferred
235 flanking sequence logo using the WebLogo application⁴⁵. A comparison between the
236 consensus sequences again show a striking similarity between DNMT3A-C Arg882His,
237 DNMT3B-C, and DNMT3B-C Arg829His, with G at N+1 position, which is the most
238 preferred nucleotide. In contrast, G at N+1 is least preferred by DNMT3A-C WT
239 enzyme. The consensus flanking sequence of DNMT3A-C WT had a T at this site. This
240 uncovered the importance of the nucleotide at the N+1 position, which can affect the
241 interaction of DNMT3 enzymes with DNA (Fig. 5c). We therefore tested if T at the N+1
242 position affects the preference for G at the N+3 position of DNMT3B-C and DNMT3A-C
243 Arg882His enzymes, by using 30-bp oligonucleotide substrates which contain a central
244 CpG site and varying combinations of nucleotides at N+2/3 or N-2/3 positions. The
245 positions N+1 and N-1 were held constant with T and A respectively (Supplementary

246 Table 3). Methylation assays using radiolabeled AdoMet were performed for 10 minutes
247 and total methylation was measured. Our data show that the preference of DNMT3B-C
248 and the DNMT3A-C Arg882 variant for G at N+3 position is lost (Supplementary Fig. 4a-
249 d). DNMT3A-C and its variants show rather strong preference for sites with A at the N+2
250 position, whereas DNMT3B-C shows a weak preference for sites with A or C occupying
251 the N+2 position (Supplementary Fig. 4e). Interestingly, DNMT3B-C Arg829His showed
252 reduced activity when compared to the WT enzyme, indicating an adverse effect of T at
253 N+1 position on its activity (Supplementary Fig. 4f). These data confirm our previous
254 observations that the interaction of DNMT3A-C Arg882His at the N+3 position is
255 strongly influenced by the nucleotide at the N+1 position.

256 **DNMT3A-C Arg882His and DNMT3B-C preferably methylate the same CpG site in**
257 **the *Meis1* enhancer**

258 We next tested if the change in flanking sequence preference of DNMT3A-C Arg882His
259 affected methylation of the regions that are known to be spuriously hypomethylated in
260 AML patients⁴⁶. The *Meis1* gene is expressed during development and regulates
261 leukemogenesis and hematopoiesis by promoting self-renewal of progenitor-like
262 cells^{47,48}. The enhancer of *Meis1* is methylated by DNMT3A during normal
263 hematopoietic stem cell (HSC) differentiation, whereas in AML patients expressing the
264 DNMT3A Arg882His variant, this region is largely hypomethylated^{46,49}. A 1-kb region of
265 the *Meis* enhancer was used as a substrate for methylation reactions, and methylation
266 was analyzed by bisulfite sequencing (Supplementary Fig. 6a). The average
267 methylation at all CpG sites showed an expected high level of methylation by DNMT3A-
268 C WT compared to DNMT3A-C Arg882His and DNMT3B-C WT, and Arg882 variants

269 showed a loss of cooperativity on this substrate *in vitro* (Supplementary Fig. 6b, c).
270 Next, we computed the flanking sequence preference of these enzymes as described
271 above. The data showed a strong preference of DNMT3B-C for one site compared to
272 DNMT3A-C, which had G at N+1 position (Fig. 6a, Supplementary Fig. 6a).
273 Interestingly, this site was also the most preferred site by DNMT3A-C Arg882His
274 compared to DNMT3A-C WT (Fig. 6b). A comparison between DNMT3B-C and
275 DNMT3A-C Arg882His confirmed that DNMT3B-C has a higher preference for this site
276 (Fig. 6c). Although there are 3 sites in this substrate with G at N+3 position, these sites
277 have either T, C, or A at N+1 position, which are weakly preferred by DNMT3B-C and
278 the DNMT3A-C Arg882His variant. These data again confirm that DNMT3A-C
279 Arg882His has acquired a substrate preference that is similar to that of DNMT3B-C. The
280 data also show that the nucleotide at the N+1 position has substantial effect on the
281 flanking sequence preference of DNMT3 enzymes.

282 **Mouse DNMT3A Arg878His retains its activity for minor satellite DNA in mESCs**

283 While DNMT3A and DNMT3B redundantly methylate many genomic regions in cells,
284 they also have preferred and specific targets^{9,11,50}. For example, in murine cells,
285 DNMT3A preferentially methylates the major satellite repeats in pericentric regions,
286 whereas DNMT3B preferentially methylates the minor satellite repeats in centromeric
287 regions^{9,11}. Based on our observations, we predicted that DNMT3A Arg882His would
288 prefer DNMT3B-specific targets. To test the idea, we carried out rescue experiments in
289 late-passage *Dnmt3a/3b* DKO mESCs, which show severe loss of global DNA
290 methylation, including at the major and minor satellite repeats¹¹. mESCs express two
291 major DNMT3A isoforms, DNMT3A1 (full length) and DNMT3A2 (a shorter form that

292 lacks the N terminus of DNMT3A1), with both showing identical activity^{11,26,51}. We
293 transfected *Dnmt3a/3b* DKO mESCs with plasmid vectors and generated stable lines
294 expressing mouse DNMT3A1 WT, DNMT3A1 Arg878His, DNMT3B1 WT, or DNMT3B1
295 Arg829His (catalytically inactive DNMT3A1 and DNMT3B1, with their PC motif in the
296 catalytic center being mutated⁵², were included as negative controls) (Fig. 7a). The
297 genomic DNA from these cell lines was harvested, and DNA methylation at the major
298 and minor satellite repeats was analyzed by digestion with methylation-sensitive
299 restriction enzymes followed by Southern blot. Consistent with our previous results²⁶,
300 the ability of DNMT3A1 to rescue methylation at the major satellite repeats is severely
301 impaired with the Arg878His substitution (Fig. 7b). However, at the minor satellite
302 repeats, which are largely methylated by DNMT3B, DNMT3A Arg878His rescues DNA
303 methylation to levels similar to the DNMT3A WT enzyme (Fig. 7c). These data
304 demonstrate that the DNMT3A AML mutant specifically retains its ability to methylate
305 DNMT3B preferred target sites, while losing preference for sites methylated by
306 DNMT3A.

307 To test if the preference was driven by a potential sequence bias in the major or minor
308 satellite repeats, we analyzed the flanking sequences around CpG sites and computed
309 the consensus sequence logo using WebLogo application⁵⁰ (Fig. 7d). The analysis
310 shows that the major satellite repeats are enriched with CpG sites carrying T at N+1 and
311 A at N+3, and are depleted in CpG sites carrying G at the N+1 and N+3 positions.
312 However, the minor satellite repeats have a high percentage of sites with G at the N+1
313 and N+3 positions, which are highly preferred by DNMT3B as well as DNMT3A
314 Arg882His. These data confirm that DNMT3A Arg878His acquires catalytic properties

315 similar to DNMT3B, which may allow it to target DNMT3B specific sites in somatic cells
316 and contribute to cancer development.

317 Our data also show that unlike the WT enzyme, the DNMT3B Arg829His variant was
318 unable to rescue methylation of the major satellite repeats and only partially rescued
319 methylation of the minor satellite repeats (Fig. 7b, c). This is in agreement with our
320 observation that substrates with T at the N+1 position are strongly disfavored by the
321 DNMT3B-C Arg829His variant (Supplementary Fig. 4f). Specifically, about half of the
322 CpG sites in both the major and minor satellite repeats have T at the N+1 position,
323 which could explain the Southern blot results.

324 Taken together, we propose a model in which substrate specificity and kinetic
325 mechanism together regulate the DNA methylation levels of various genomic regions. At
326 repetitive elements where CpG content is intermediate to high, DNMT3A WT enzyme
327 acts cooperatively to methylate multiple CpGs at its target sites where it prefers A and T
328 at N+3 position. Loss of cooperativity in the DNMT3A WT or DNMT3A Arg882His
329 variant modifies its specificity to be similar to that of DNMT3B where it prefers G at N+3
330 (Fig. 8). Whereas DNMT3B methylates its targets in a processive manner, DNMT3A WT
331 and Arg882His methylate these sites distributively explaining the lower activity of these
332 enzymes at these sites.

333 **Discussion**

334 Despite numerous studies addressing the biological roles of DNMT3A and DNMT3B in
335 development and diseases, the differences and similarities in their kinetic mechanisms
336 remain poorly understood. Germline mutations in DNMT3A and DNMT3B have

337 deleterious effects and are associated with congenital diseases^{17,18}. In AML and other
338 types of leukemia, the majority of somatic mutations in *DNMT3A* affects Arg882, mostly
339 leading to an Arg-to-His substitution²¹. Because of its high prevalence (~20% in AML)
340 and early occurrence during disease development, the Arg882His variant is considered
341 a founder mutation⁵³. Therefore, the variant enzyme has been the subject of many
342 studies. Through these studies, the Arg882His substitution was shown to alter DNA
343 binding properties, attenuate tetramerization, disrupt cooperativity, and change the
344 flanking sequence preference of the DNMT3A enzyme^{22,27,33,40,42}. In this study, we
345 show that disruption of cooperativity alters the flanking sequence preference of
346 DNMT3A-C, suggesting that the gain of flanking sequence preference of the DNMT3A-
347 C Arg882His variant is the consequence of losing the cooperative kinetic mechanism.
348 We systematically computed the temporal order of site preference for the DNMT3A-C,
349 DNMT3A-C Arg882His, DNMT3B-C, and DNMT3B-C Arg829His enzymes. The most
350 important finding in this study is the discovery that the altered flanking sequence
351 preference of the DNMT3A-C Arg882His variant is nearly identical to the preferred
352 substrate sequence of DNMT3B-C. Consequently, we predicted that the DNMT3A
353 Arg882His variant would potentially methylate DNMT3B-specific targets. Our study
354 using *in vitro* bisulfite sequencing and rescue experiments by stable expression of
355 proteins in *Dnmt3a/3b* DKO mESCs cells provide strong evidence supporting this
356 prediction. The data revealed that, although the DNMT3A Arg882His variant has little
357 activity in methylating the major satellite repeats, it's ability to methylate the minor
358 satellite repeats is almost fully retained. An analysis of the CpG flanking sequence
359 showed an overrepresentation of G at the N+1 and N+3 positions in the minor satellite

360 repeats that is preferred by DNMT3B and the DNMT3A Arg882His variant. The major
361 satellite repeats have high percentages of T and A at the N+1 and N+3 positions,
362 respectively, which are disfavored by DNMT3B WT and the Arg829His variant.

363 The observations from this study reveal how distinct kinetic properties of DNMT3A and
364 DNMT3B influence the selection of their genomic target sites. Given that under
365 conditions where DNMT3A WT functions as non-cooperative enzyme, its flanking
366 sequence preference is similar to DNMT3A Arg878His variant, we propose that
367 DNMT3A methylates minor satellite repeats using non-cooperative kinetic mechanism.
368 This prediction also explains the rationale behind the lower activity of DNMT3A at minor
369 satellite repeats compared to that of DNMT3B, which methylates its target sites using
370 processive kinetic mechanism. Similarly, although DNMT3A Arg878His variant prefers
371 DNMT3B sites, it methylates these sites using a non-cooperative distributive
372 mechanism explaining an incomplete rescue by the variant enzyme compared to
373 DNMT3B. A high activity of DNMT3A at the major satellite repeats could be justified by
374 the cooperative mechanism used to methylate multiple CpG sites.

375 These observations also suggest that at genomic regions with sparse or dispersed CpG
376 sites, in the absence of cooperativity, sites with preferred flanking sequence (such as
377 N+1,3=G) could be methylated at higher levels both by DNMT3A and DNMT3B.

378 However, except during early embryogenesis, the tissue specific expression of
379 DNMT3A and DNMT3B ensures that these enzymes methylate distinct regions, besides
380 having many common target sites. The importance of this regulation is highlighted by
381 aberrant expression of DNMT3B in various types of cancer, including AML^{14,28-30}. The
382 effect of DNMT3B overexpression in AML is similar to that of DNMT3A Arg882His

383 mutation, leading to an increase in stemness, downregulation in apoptotic genes, and
384 poor patient prognosis^{27,29}. Our data here reveal a mechanism by which the DNMT3A
385 Arg882His variant acts like the DNMT3B enzyme, thus providing a mechanistic
386 explanation to above observations in AML patients. We suggest that the oncogenic
387 potential of DNMT3A Arg882His variant may not be only due to its lower activity causing
388 DNA hypomethylation in AML cells, but also due to the gain of DNMT3B-like activity
389 generating aberrant patterns of DNA methylation. Our observations provide novel
390 insights into consequences of cancer-causing mutations on enzymatic activity of
391 DNMT3 enzymes.

392 **Methods**

393 **Protein Purification**

394 Human DNMT3A-C WT, DNMT3A-C Arg882His, DNMT3A-C Arg882Cys, and
395 DNMT3A-C Arg882Ser, and mouse DNMT3B-C WT and DNMT3B-C Arg829His in
396 pET28a(+) with a 6X His tag, were expressed and purified using affinity chromatography
397 as described⁴⁰. Briefly, transformed BL21 (DE3) pLys cells were induced with 1 mM
398 IPTG at OD₆₀₀ 0.3 and expressed for 2 h at 32°C. Harvested cells were washed with
399 STE buffer (10 mM Tris-HCl (pH 8.0), 0.1 mM EDTA, 0.1 M NaCl), and resuspended in
400 Buffer A (20 mM Potassium Phosphate (pH 7.5), 0.5 M NaCl, 10% (v/v) glycerol, 1 mM
401 EDTA, 0.1 mM DTT, 40 mM imidazole). Cells were disrupted by sonication, followed by
402 removal of cell debris by centrifugation. Clarified lysate was incubated with 0.4 mL Ni-
403 NTA agarose for 3 h at 4°C. The protein bound slurry was packed in a 2 ml Biorad
404 column and washed with 150 ml Buffer A. Protein was eluted using 200 mM imidazole

405 in Buffer A at pH 7.5, then stored in 20 mM HEPES pH 7.5, 40 mM KCl, 1 mM EDTA,
406 0.2 mM DTT and 20% (v/v) glycerol at -80°C. The purity and integrity of recombinant
407 proteins were checked by SDS-PAGE gel.

408 **DNA Methylation assays using radiolabeled AdoMet**

409 Radioactive methylation assays to determine kinetic parameters of recombinant
410 enzymes were performed using ³H-labelled S-Adenosylmethionine (AdoMet) as a
411 methyl group donor and biotinylated oligonucleotides of varying sizes bound on avidin-
412 coated high-binding Elisa plates (Corning) as described⁵⁴. The DNA methylation
413 reactions were carried out using either 250 nM 30-bp/32-bp DNA substrate in
414 methylation buffer (20 mM HEPES pH 7.5, 50 mM KCl, and 1 mM EDTA, 5 µg/ml BSA).
415 The methylation reaction included 0.76 µM [methyl³H] AdoMet (PerkinElmer Life
416 Sciences). Storage buffer was added to compensate for the different enzyme volumes
417 in all reactions. Incorporated radioactivity was quantified by scintillation counting.

418 **Processivity Assay**

419 Methylation kinetic analyses were performed using two enzyme concentrations and
420 short oligonucleotide 30-bp and 32-bp substrates with 1 and 2 CpG sites, respectively.
421 Low enzyme concentrations relative to DNA substrate concentrations were used to
422 ensure that the reaction occurred under multiple turnover conditions. Each DNA
423 substrate was used at 250 nM, and a 1:1 ratio of labelled and unlabeled AdoMet (final
424 concentration 1.5 µM) was used.

425 **Cooperativity Assays**

426 To examine cooperativity, increasing concentrations of enzyme were pre-incubated with
427 DNA substrate for 10 minutes at room temperature prior to the addition of AdoMet to
428 start the reaction. AdoMet was a mixture of unlabeled and 0.76 μM $^3\text{[H]}$ labeled AdoMet
429 to yield a final concentration of 2 μM . Methylation assays were performed using 100 ng
430 of an unmethylated pUC19 plasmid purified from *dam*⁻ /*dcm*⁻ *E.coli* strain (C2925I, NEB)
431 or a 1-kb fragment containing 14 CpG sites amplified from the *Meis1* enhancer were
432 used as DNA substrates for filter binding assays. Briefly, 10 μl reaction mix was spotted
433 on 0.5 in DE81 filter that was then washed 5 times in 0.2 M Ammonium Bicarbonate
434 (NH_4HCO_3), washed 2-3 times with 100% ethanol, and air dried. Incorporated
435 radioactivity was quantified by scintillation counting⁵⁵.

436 **Flanking sequence preference**

437 *In vitro* DNA methylation reactions were carried out using 100ng of a 509-bp DNA
438 fragment amplified from the *Suwh1* promoter, or 100 ng of a 1-kb DNA fragment
439 amplified from the *Meis1* enhancer region. Methylation reactions were carried out in
440 methylation buffer (20 mM HEPES pH 7.5, 50 mM KCl, 1 mM EDTA, and 0.05 mg/ml
441 BSA using varying concentrations of each enzyme at 37°C. Samples were taken at 10,
442 30, or 60 minutes and reaction was stopped by freeze/thaw. DNA methylation was
443 analyzed by bisulfite sequencing as described below.

444 Flanking sequence preference was also measured using short oligonucleotides and
445 DNA methylation assays using radiolabeled AdoMet as described above. Sixteen
446 different 30-bp substrates were used with varying combinations of the second and third
447 nucleotide around the CpG site on either side (Supplementary Table 3).

448 **Bisulfite sequencing**

449 Bisulfite conversion was performed using EpiTect Fast Bisulfite Conversion Kit
450 (Qiagen, 59802) according to the manufacturer's protocol. PCR amplifications were
451 performed with primers as described (Supplementary Fig. 2A-B)³⁹. The pooled samples
452 were sequenced using NGS on Wide-Seq platform. The reads were assembled and
453 analyzed by Bismark and Bowtie2. Methylated and unmethylated CpGs for each target
454 were quantified, averaged, and presented as percent CpG methylation.

455 **Rescue experiments in mESCs**

456 WT (J1) and *Dnmt3a/3b* DKO mESCs were cultured on gelatin-coated petri dishes in
457 mESC medium (DMEM supplemented with 15% fetal bovine serum, 0.1 mM
458 nonessential amino acids, 0.1 mM β -mercaptoethanol, 50 U/mL penicillin, 50 μ g/mL
459 streptomycin, and 10^3 U/mL leukemia inhibitory factor)⁹. For the generation of stable
460 clones expressing DNMT3A or DNMT3B proteins, mESCs were transfected with
461 plasmid vectors using Lipofectamine 2000 (Invitrogen) and then seeded at low density
462 on dishes coated with feeder cells, selected with 6 μ g/mL of Blasticidin S HCl (Gibco)
463 for 7-10 days, and individual colonies were picked. Southern blot analysis of DNA
464 methylation at the major and minor satellite repeats were performed as previously
465 described^{11,26}.

466 **Data Analysis**

467 *Comparative flanking sequence preference analysis* - To compare the flanking
468 sequence preference of different enzymes, we first calculated average percent

469 methylation of all CpG sites in the substrate. The fractional variance (v) at each CpG
470 site was calculated by dividing in the percent methylation of each site by average
471 methylation. The preference for a site by an enzyme (B) over enzyme (A) was
472 calculated as relative change $(v_B - v_A) / v_A$, Positive values indicate a preference of
473 enzyme B for a site and values greater than or equal to 1 were considered significant
474 given the preference is more than 2 fold (Supplementary Table 1).

475 To determine the fractional distribution of nucleotides at preferred sites the occurrence
476 of each nucleotide at N+1/2/3 positions was calculated as a fraction. At each position,
477 the number of times each nucleotide occurred was divided by the total number of
478 preferred sites to compute the fractional occurrence.

479 *Individual flanking sequence preference analysis* – Analysis of the bisulfite sequencing
480 data was performed to determine optimal flanking sequence preferred by each enzyme.
481 To take into account the uneven distribution of nucleotides flanking the CpG site, the
482 occurrence of each nucleotide at the N+1/2/3 positions (p_{1n} , p_{2n} , and p_{3n}) was
483 calculated by dividing the number of times it occurred by the total number of CpG sites
484 ($s(p_{1n})$, $s(p_{2n})$, $s(p_{3n})$). The expected occurrence of a trinucleotide set (O) was
485 computed by multiplying the occurrence of three nucleotides as described by equation
486 (1). This created an expected value, which would predict the probability at which this
487 trinucleotide set would be methylated if there were no flanking sequence preference by
488 the enzyme. With the data obtained from bisulfite sequencing, the fractional methylation
489 (f) for each CpG site was calculated by dividing percent methylation at a site by sum of
490 percent methylation at all sites as described by equation (2). Fractional methylation was
491 sorted by nucleotides at position N+1/2/3 ($f(p_{1n})$, $f(p_{2n})$, and $f(p_{3n})$) and summed for

492 each nucleotide at a particular position ($\Sigma f(p1_n)$, $\Sigma f(p2_n)$, and $\Sigma f(p3_n)$). The methylation
493 for the flanking trinucleotide set (M) was calculated then multiplying the summed value
494 for a nucleotide at the three positions as described by equation (3). This gives us the
495 value for observed methylation value at a CpG with a specific flanking trinucleotide. The
496 preference for a flanking trinucleotide by an enzyme was calculated by determining the
497 fold change between the observed and expected values as described by equation (4)
498 (Supplementary Table 2).

499 *Equations*

500 v = fractional variance

501 f = fractional methylation

502 n = any nucleotide

503 A,C,G,T = specific nucleotides

504 s = fraction (number of sites out / 56)

505 $p1$ = nucleotide at position N+1

506 $p2$ = nucleotide at position N+2

507 $p3$ = nucleotide at position N+3

508 M = observed methylation based on fractional methylation

509 O = normalized occurrence of each site based on their frequency in the DNA substrate

510 Σ = sum

$$511 \quad O = s(p1_n) * s(p2_n) * s(p3_n) \quad (1)$$

512
$$f = \frac{\% \text{Methylation at each CpG}}{\text{Sum of all methylation}} \quad (2)$$

513
$$M = \sum f(p1_n) * \sum f(p2_n) * \sum f(p3_n) \quad (3)$$

514
$$\text{Relative Change} = \frac{M-O}{O} \quad (4)$$

515 *Consensus sequence analysis of major and minor satellite repeats* - From the mm9
516 genome, three regions, chr9: 3000466 – 3028144, chr9: 3033472 – 3037264, and chr2:
517 98506820 – 98507474, were used for the sequences of the major satellite repeats. The
518 minor satellite repeat sequences were obtained from chr2:98505036-98505275 and
519 chr2:98506495-98506615. Additionally, more sequences were obtained from GenBank
520 and used for analysis (accession no. X14462.1, X14463.1, X14464.1, X14465.1,
521 X14466.1, X14468.1, X14469.1, X14470.1). Consensus sequences of the CpG sites
522 and flanking regions in the major and minor satellite repeats were built using WebLogo
523 (<https://weblogo.berkeley.edu/logo.cgi>).

524 *Line and Bar graphs*- Data were analyzed using the Prism software (GraphPad). For
525 cooperativity graphs, data were fit with nonlinear, second order polynomial regression
526 curves. Errors were calculated as standard error of the mean (SEM) for two to four
527 independent experiments, as described in the figure legends.

528 **Acknowledgements**

529 We are thankful to Gowher lab members for discussions. This work was supported by
530 grants from NIH (R01GM118654-01 to H.G. and R01AI12140301A1 to T.C.) and NSF
531 (1716678 to H.G.).

532 **Author Contributions**

533 A.B.N. and H.G. conceived the project and designed the experiments. A.B.N., L.A.,
534 A.R.M., and N.E.F. purified recombinant proteins and performed radiolabeled methylation
535 assays. A.B.N. performed in vitro methylation assays and bisulfite conversion. A.B.N. and
536 H.G. analyzed the data, generated figures, and wrote the manuscript. B.L. performed the
537 rescue experiments in mESCs. T.C. supervised the cellular work and also participated in
538 study design and manuscript writing.

539 **Conflict of Interest**

540 The authors declare no conflict of interest.

541 REFERENCES

- 542 1 Reik, W., Dean, W. & Walter, J. Epigenetic reprogramming in mammalian development. *Science*
543 **293**, 1089-1093, doi:10.1126/science.1063443 (2001).
- 544 2 Smith, Z. D. & Meissner, A. DNA methylation: roles in mammalian development. *Nature reviews.*
545 *Genetics* **14**, 204-220, doi:10.1038/nrg3354 (2013).
- 546 3 Jaenisch, R. & Jahner, D. Methylation, Expression and Chromosomal Position of Genes in
547 Mammals. *Biochim Biophys Acta* **782**, 1-9, doi:Doi 10.1016/0167-4781(84)90099-X (1984).
- 548 4 Surani, M. A. Imprinting and the initiation of gene silencing in the germ line. *Cell* **93**, 309-312
549 (1998).
- 550 5 Ng, H. H. & Bird, A. DNA methylation and chromatin modification. *Current opinion in genetics &*
551 *development* **9**, 158-163 (1999).
- 552 6 Lyko, F. The DNA methyltransferase family: a versatile toolkit for epigenetic regulation. *Nature*
553 *reviews. Genetics* **19**, 81-92, doi:10.1038/nrg.2017.80 (2018).
- 554 7 Bestor, T., Laudano, A., Mattaliano, R. & Ingram, V. Cloning and sequencing of a cDNA encoding
555 DNA methyltransferase of mouse cells. The carboxyl-terminal domain of the mammalian
556 enzymes is related to bacterial restriction methyltransferases. *Journal of molecular biology* **203**,
557 971-983 (1988).
- 558 8 Yoder, J. A., Soman, N. S., Verdine, G. L. & Bestor, T. H. DNA (cytosine-5)-methyltransferases in
559 mouse cells and tissues. Studies with a mechanism-based probe. *Journal of molecular biology*
560 **270**, 385-395, doi:10.1006/jmbi.1997.1125 (1997).
- 561 9 Okano, M., Bell, D. W., Haber, D. A. & Li, E. DNA methyltransferases Dnmt3a and Dnmt3b are
562 essential for de novo methylation and mammalian development. *Cell* **99**, 247-257 (1999).
- 563 10 Aapola, U. *et al.* Isolation and initial characterization of a novel zinc finger gene, DNMT3L, on
564 21q22.3, related to the cytosine-5-methyltransferase 3 gene family. *Genomics* **65**, 293-298,
565 doi:10.1006/geno.2000.6168 (2000).
- 566 11 Chen, T., Ueda, Y., Dodge, J. E., Wang, Z. & Li, E. Establishment and maintenance of genomic
567 methylation patterns in mouse embryonic stem cells by Dnmt3a and Dnmt3b. *Molecular and*
568 *cellular biology* **23**, 5594-5605, doi:10.1128/mcb.23.16.5594-5605.2003 (2003).
- 569 12 Gujar, H., Weisenberger, D. J. & Liang, G. The Roles of Human DNA Methyltransferases and Their
570 Isoforms in Shaping the Epigenome. *Genes* **10**, doi:10.3390/genes10020172 (2019).
- 571 13 Watanabe, D., Suetake, I., Tada, T. & Tajima, S. Stage- and cell-specific expression of Dnmt3a
572 and Dnmt3b during embryogenesis. *Mechanisms of development* **118**, 187-190 (2002).
- 573 14 Linhart, H. G. *et al.* Dnmt3b promotes tumorigenesis in vivo by gene-specific de novo
574 methylation and transcriptional silencing. *Genes & development* **21**, 3110-3122,
575 doi:10.1101/gad.1594007 (2007).
- 576 15 Chen, B. F. & Chan, W. Y. The de novo DNA methyltransferase DNMT3A in development and
577 cancer. *Epigenetics* **9**, 669-677, doi:10.4161/epi.28324 (2014).
- 578 16 Schubeler, D. Function and information content of DNA methylation. *Nature* **517**, 321-326,
579 doi:10.1038/nature14192 (2015).
- 580 17 Norvil, A. B., Saha, D., Saleem Dar, M. & Gowher, H. Effect of Disease-Associated Germline
581 Mutations on Structure Function Relationship of DNA Methyltransferases. *Genes* **10**,
582 doi:10.3390/genes10050369 (2019).
- 583 18 Ehrlich, M. The ICF syndrome, a DNA methyltransferase 3B deficiency and immunodeficiency
584 disease. *Clinical immunology* **109**, 17-28 (2003).
- 585 19 Xu, G. L. *et al.* Chromosome instability and immunodeficiency syndrome caused by mutations in
586 a DNA methyltransferase gene. *Nature* **402**, 187-191, doi:10.1038/46052 (1999).

- 587 20 Hansen, R. S. *et al.* The DNMT3B DNA methyltransferase gene is mutated in the ICF
588 immunodeficiency syndrome. *Proceedings of the National Academy of Sciences of the United*
589 *States of America* **96**, 14412-14417, doi:10.1073/pnas.96.25.14412 (1999).
- 590 21 Ley, T. J. *et al.* DNMT3A mutations in acute myeloid leukemia. *The New England journal of*
591 *medicine* **363**, 2424-2433, doi:10.1056/NEJMoa1005143 (2010).
- 592 22 Brunetti, L., Gundry, M. C. & Goodell, M. A. DNMT3A in Leukemia. *Cold Spring Harbor*
593 *perspectives in medicine* **7**, doi:10.1101/cshperspect.a030320 (2017).
- 594 23 Cancer Genome Atlas Research, N. *et al.* Genomic and epigenomic landscapes of adult de novo
595 acute myeloid leukemia. *The New England journal of medicine* **368**, 2059-2074,
596 doi:10.1056/NEJMoa1301689 (2013).
- 597 24 Qu, Y. *et al.* Differential methylation in CN-AML preferentially targets non-CGI regions and is
598 dictated by DNMT3A mutational status and associated with predominant hypomethylation of
599 HOX genes. *Epigenetics* **9**, 1108-1119, doi:10.4161/epi.29315 (2014).
- 600 25 Yang, L. *et al.* DNMT3A Loss Drives Enhancer Hypomethylation in FLT3-ITD-Associated
601 Leukemias. *Cancer cell* **29**, 922-934, doi:10.1016/j.ccell.2016.05.003 (2016).
- 602 26 Kim, S. J. *et al.* A DNMT3A mutation common in AML exhibits dominant-negative effects in
603 murine ES cells. *Blood* **122**, 4086-4089, doi:10.1182/blood-2013-02-483487 (2013).
- 604 27 Russler-Germain, D. A. *et al.* The R882H DNMT3A mutation associated with AML dominantly
605 inhibits wild-type DNMT3A by blocking its ability to form active tetramers. *Cancer cell* **25**, 442-
606 454, doi:10.1016/j.ccr.2014.02.010 (2014).
- 607 28 Gagliardi, M., Strazzullo, M. & Matarazzo, M. R. DNMT3B Functions: Novel Insights From Human
608 Disease. *Frontiers in cell and developmental biology* **6**, 140, doi:10.3389/fcell.2018.00140 (2018).
- 609 29 Niederwieser, C. *et al.* Prognostic and biologic significance of DNMT3B expression in older
610 patients with cytogenetically normal primary acute myeloid leukemia. *Leukemia* **29**, 567-575,
611 doi:10.1038/leu.2014.267 (2015).
- 612 30 Hayette, S. *et al.* High DNA methyltransferase DNMT3B levels: a poor prognostic marker in acute
613 myeloid leukemia. *PloS one* **7**, e51527, doi:10.1371/journal.pone.0051527 (2012).
- 614 31 Jia, D., Jurkowska, R. Z., Zhang, X., Jeltsch, A. & Cheng, X. Structure of Dnmt3a bound to Dnmt3L
615 suggests a model for de novo DNA methylation. *Nature* **449**, 248-251, doi:10.1038/nature06146
616 (2007).
- 617 32 Guo, X. *et al.* Structural insight into autoinhibition and histone H3-induced activation of
618 DNMT3A. *Nature* **517**, 640-644, doi:10.1038/nature13899 (2015).
- 619 33 Holz-Schietinger, C., Matje, D. M. & Reich, N. O. Mutations in DNA methyltransferase (DNMT3A)
620 observed in acute myeloid leukemia patients disrupt processive methylation. *The Journal of*
621 *biological chemistry* **287**, 30941-30951, doi:10.1074/jbc.M112.366625 (2012).
- 622 34 Sandoval, J. E., Huang, Y. H., Muise, A., Goodell, M. A. & Reich, N. O. Mutations in the DNMT3A
623 DNA methyltransferase in acute myeloid leukemia patients cause both loss and gain of function
624 and differential regulation by protein partners. *The Journal of biological chemistry* **294**, 4898-
625 4910, doi:10.1074/jbc.RA118.006795 (2019).
- 626 35 Yan, X. J. *et al.* Exome sequencing identifies somatic mutations of DNA methyltransferase gene
627 DNMT3A in acute monocytic leukemia. *Nature genetics* **43**, 309-315, doi:10.1038/ng.788 (2011).
- 628 36 Yamashita, Y. *et al.* Array-based genomic resequencing of human leukemia. *Oncogene* **29**, 3723-
629 3731, doi:10.1038/onc.2010.117 (2010).
- 630 37 Jurkowska, R. Z. *et al.* Formation of nucleoprotein filaments by mammalian DNA
631 methyltransferase Dnmt3a in complex with regulator Dnmt3L. *Nucleic acids research* **36**, 6656-
632 6663, doi:10.1093/nar/gkn747 (2008).

- 633 38 Rajavelu, A., Jurkowska, R. Z., Fritz, J. & Jeltsch, A. Function and disruption of DNA
634 methyltransferase 3a cooperative DNA binding and nucleoprotein filament formation. *Nucleic*
635 *acids research* **40**, 569-580, doi:10.1093/nar/gkr753 (2012).
- 636 39 Emperle, M., Rajavelu, A., Reinhardt, R., Jurkowska, R. Z. & Jeltsch, A. Cooperative DNA binding
637 and protein/DNA fiber formation increases the activity of the Dnmt3a DNA methyltransferase.
638 *The Journal of biological chemistry* **289**, 29602-29613, doi:10.1074/jbc.M114.572032 (2014).
- 639 40 Norvil, A. B. *et al.* Dnmt3b Methylates DNA by a Noncooperative Mechanism, and Its Activity Is
640 Unaffected by Manipulations at the Predicted Dimer Interface. *Biochemistry* **57**, 4312-4324,
641 doi:10.1021/acs.biochem.6b00964 (2018).
- 642 41 Zhang, Z. M. *et al.* Structural basis for DNMT3A-mediated de novo DNA methylation. *Nature*
643 **554**, 387-391, doi:10.1038/nature25477 (2018).
- 644 42 Emperle, M. *et al.* The DNMT3A R882H mutant displays altered flanking sequence preferences.
645 *Nucleic acids research* **46**, 3130-3139, doi:10.1093/nar/gky168 (2018).
- 646 43 Gowher, H. & Jeltsch, A. Molecular enzymology of the catalytic domains of the Dnmt3a and
647 Dnmt3b DNA methyltransferases. *The Journal of biological chemistry* **277**, 20409-20414,
648 doi:10.1074/jbc.M202148200 (2002).
- 649 44 Handa, V. & Jeltsch, A. Profound flanking sequence preference of Dnmt3a and Dnmt3b
650 mammalian DNA methyltransferases shape the human epigenome. *Journal of molecular biology*
651 **348**, 1103-1112, doi:10.1016/j.jmb.2005.02.044 (2005).
- 652 45 Crooks, G. E., Hon, G., Chandonia, J. M. & Brenner, S. E. WebLogo: a sequence logo generator.
653 *Genome Res* **14**, 1188-1190, doi:10.1101/gr.849004 (2004).
- 654 46 Lu, R. *et al.* Epigenetic Perturbations by Arg882-Mutated DNMT3A Potentiate Aberrant Stem Cell
655 Gene-Expression Program and Acute Leukemia Development. *Cancer cell* **30**, 92-107,
656 doi:10.1016/j.ccell.2016.05.008 (2016).
- 657 47 Mercader, N. *et al.* Conserved regulation of proximodistal limb axis development by Meis1/Hth.
658 *Nature* **402**, 425-429, doi:10.1038/46580 (1999).
- 659 48 Cai, M. *et al.* Dual actions of Meis1 inhibit erythroid progenitor development and sustain general
660 hematopoietic cell proliferation. *Blood* **120**, 335-346, doi:10.1182/blood-2012-01-403139
661 (2012).
- 662 49 Ferreira, H. J. *et al.* DNMT3A mutations mediate the epigenetic reactivation of the leukemogenic
663 factor MEIS1 in acute myeloid leukemia. *Oncogene* **35**, 3079-3082, doi:10.1038/onc.2015.359
664 (2016).
- 665 50 Challen, G. A. *et al.* Dnmt3a and Dnmt3b have overlapping and distinct functions in
666 hematopoietic stem cells. *Cell stem cell* **15**, 350-364, doi:10.1016/j.stem.2014.06.018 (2014).
- 667 51 Chen, T., Ueda, Y., Xie, S. & Li, E. A novel Dnmt3a isoform produced from an alternative
668 promoter localizes to euchromatin and its expression correlates with active de novo
669 methylation. *The Journal of biological chemistry* **277**, 38746-38754,
670 doi:10.1074/jbc.M205312200 (2002).
- 671 52 Chen, T., Tsujimoto, N. & Li, E. The PWWP domain of Dnmt3a and Dnmt3b is required for
672 directing DNA methylation to the major satellite repeats at pericentric heterochromatin.
673 *Molecular and cellular biology* **24**, 9048-9058, doi:10.1128/MCB.24.20.9048-9058.2004 (2004).
- 674 53 Mayle, A. *et al.* Dnmt3a loss predisposes murine hematopoietic stem cells to malignant
675 transformation. *Blood* **125**, 629-638, doi:10.1182/blood-2014-08-594648 (2015).
- 676 54 Roth, M. & Jeltsch, A. Biotin-avidin microplate assay for the quantitative analysis of enzymatic
677 methylation of DNA by DNA methyltransferases. *Biological chemistry* **381**, 269-272,
678 doi:10.1515/BC.2000.035 (2000).

679 55 Hemeon, I., Gutierrez, J. A., Ho, M. C. & Schramm, V. L. Characterizing DNA methyltransferases
680 with an ultrasensitive luciferase-linked continuous assay. *Analytical chemistry* **83**, 4996-5004,
681 doi:10.1021/ac200816m (2011).

682

683

684

685

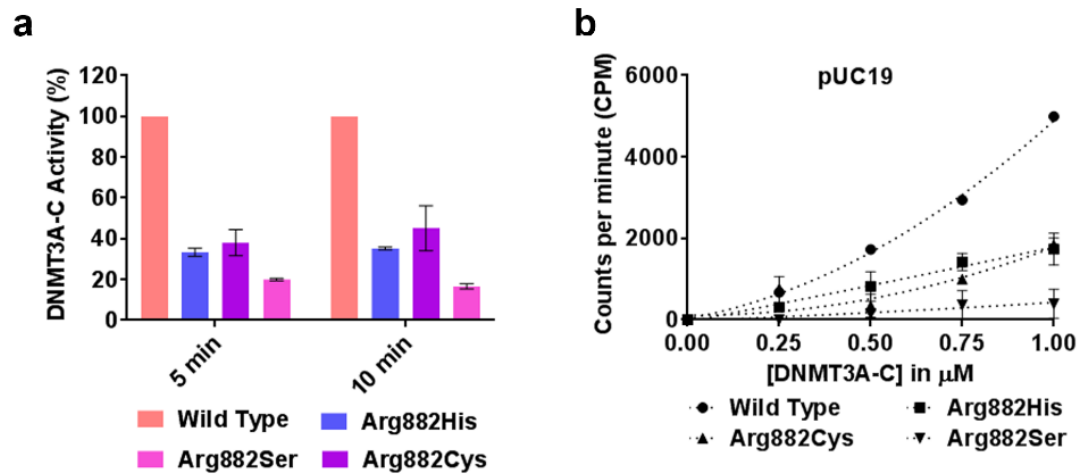
686

687

FIGURES AND FIGURE LEGENDS

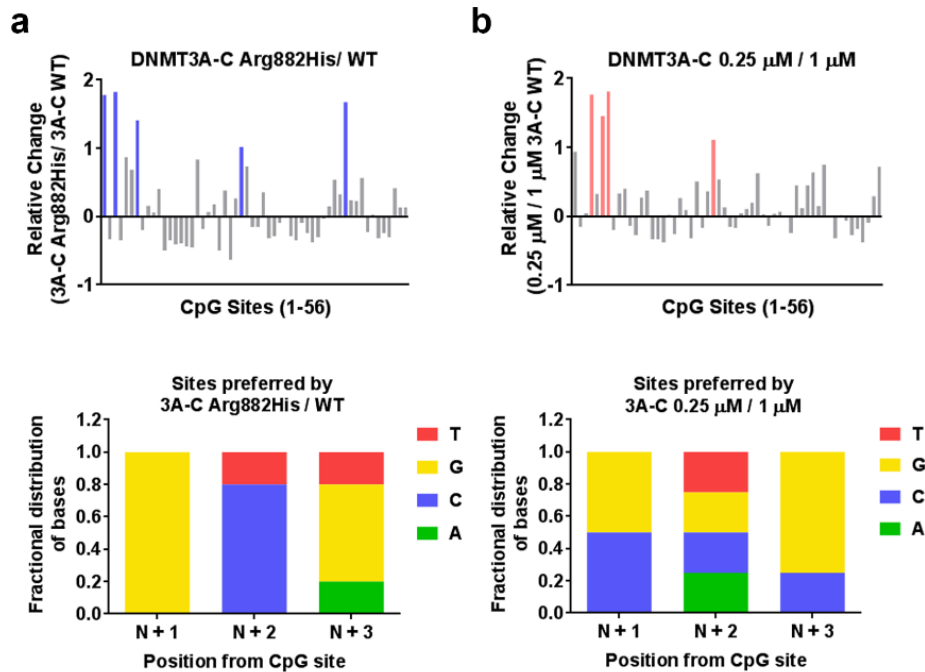
688

689 **FIGURE 1**



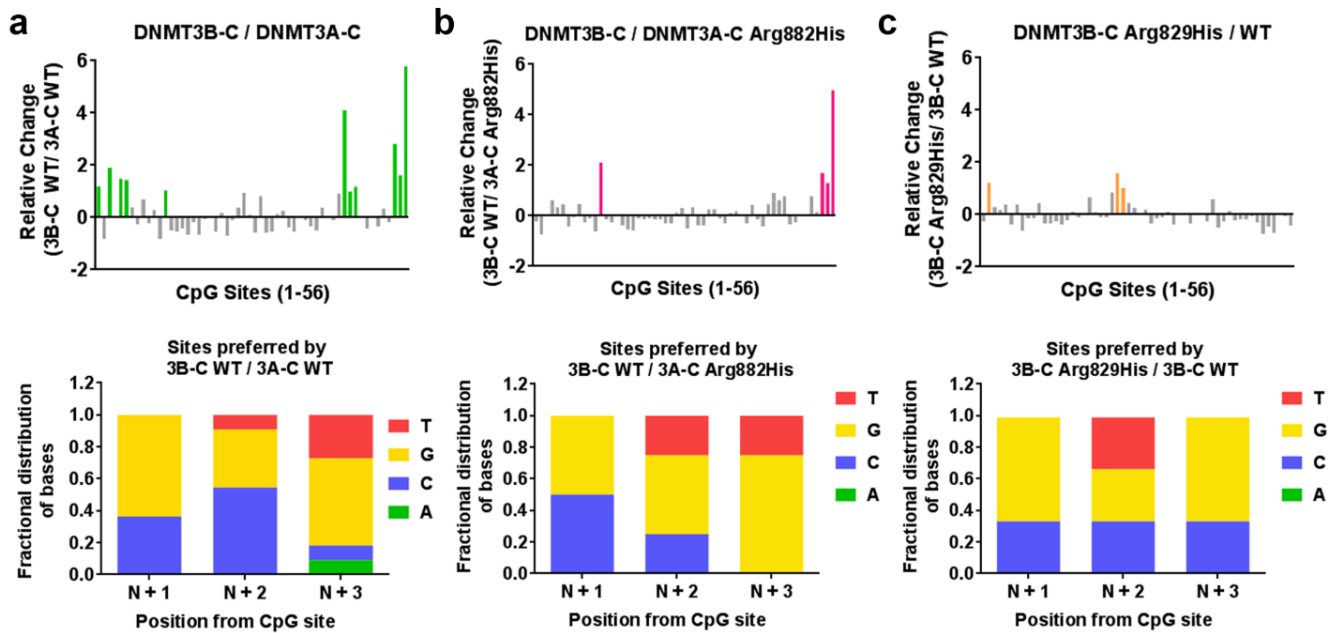
690 Relative activity and kinetic mechanism of DNMT3A-C WT and Arg882 variants. **a**
691 Methylation activity of 1 μM DNMT3A-C WT and Arg882 variants was measured for 5
692 and 10 minutes using ^3H labelled S-Adenosylmethionine (AdoMet). The transfer of
693 radiolabeled $-\text{CH}_3$ group to DNA was measured as counts per minute (CPM) using the
694 MicroBeta scintillation counter. **b** Methylation activity of DNMT3A-C WT and Arg882
695 variants was measured for 10 minutes using 100 ngs of pUC19 plasmid as a substrate,
696 at concentrations of enzymes varying from 0.25 to 1 μM . The enzymes were pre-
697 incubated with DNA for 10 minutes at room temperature and the reaction was initiated
698 by addition of AdoMet. Each data point is an average and standard error of the mean
699 ($n \geq 3$ independent experiments). The data shows reduced activity and loss of
700 cooperativity for all the variant enzymes.

701 **FIGURE 2**



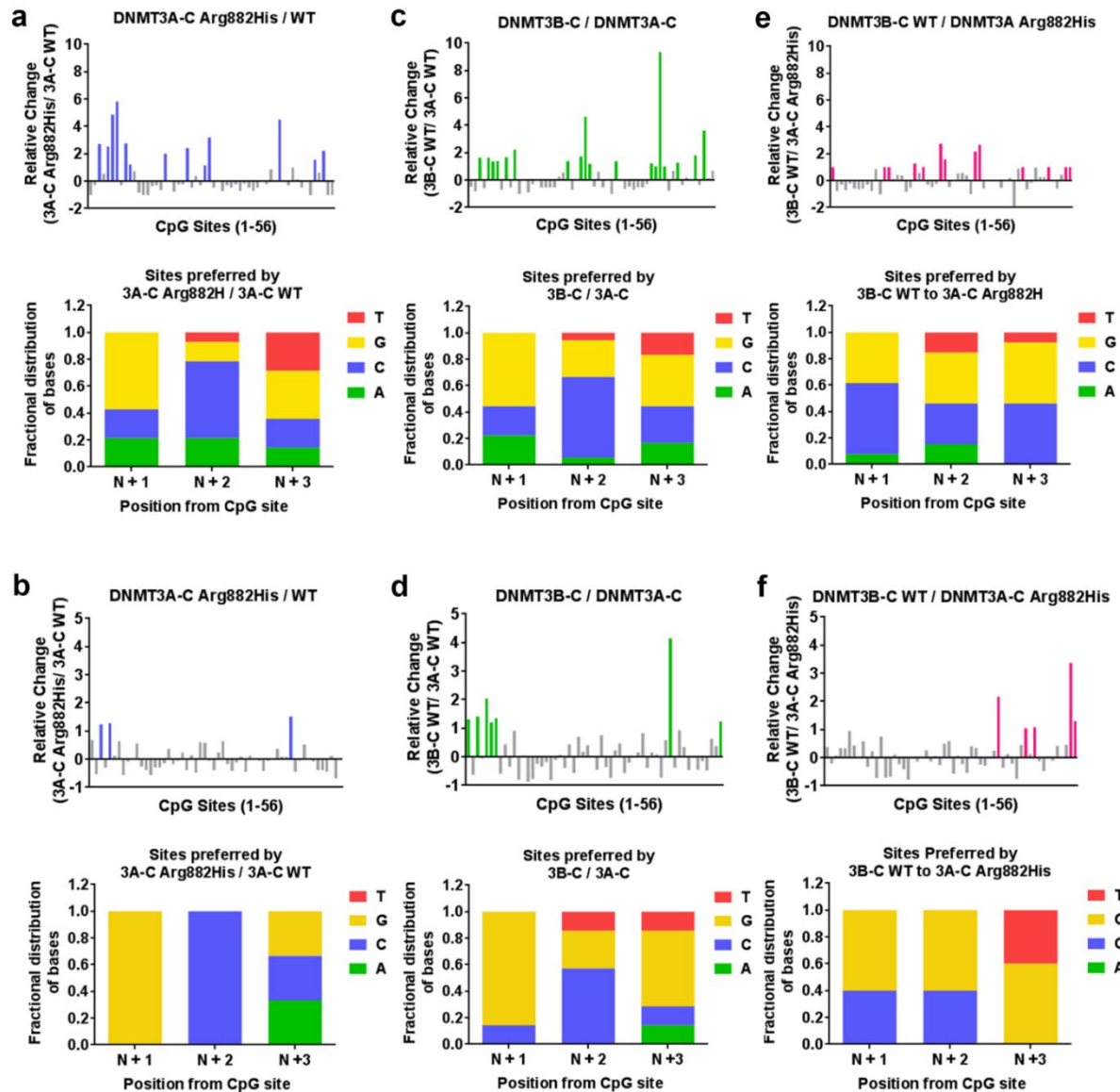
702 Effect of cooperativity on flanking sequence preference by DNMT3A-C. **a, b** DNA
 703 methylation of the 56 CpG sites in the 509-bp substrate was analyzed using bisulfite
 704 sequencing. Methylation reaction was carried out using 1 μ M DNMT3A-C WT, and 1 μ M
 705 DNMT3A-C Arg882His in **a**, and 0.25 μ M and 1 μ M DNMT3A-C WT in **b**. *Top panel*
 706 shows relative preference calculated for each CpG site (1-56) by the DNMT3A-C
 707 Arg882His compared to DNMT3A-C WT in **a**, and 0.25 μ M compared to 1 μ M of
 708 DNMT3A-C WT in **b**. Relative preference of 1 is equal to a 2-fold change, and is
 709 represented by blue and light pink bars, respectively. *Bottom panel* shows the fractional
 710 distribution of nucleotides at the preferred sites. Each bar represents nucleotides at
 711 positions N+1/2/3 respectively from the CpG site. The data show DNMT3A-C WT at low
 712 concentrations have flanking sequence preference similar to DNMT3A-C Arg882His
 713 variant.

714 **FIGURE 3**



715 Comparative analysis of the flanking sequence preferences of DNMT3A-C Arg882His
 716 and DNMT3B-C WT. **a-c** DNA methylation of the 56 CpG sites in the 509-bp substrate
 717 was analyzed using bisulfite sequencing. Methylation reaction was carried out using 1
 718 μ M enzyme for 10 minutes. *Top panels* show relative preference calculated for each
 719 CpG site (1-56) by DNMT3B-C WT compared to DNMT3A-C WT in **a**, DNMT3B-C WT
 720 compared to DNMT3A-C Arg882His in **b**, and DNMT3B-C Arg829His compared to
 721 DNMT3B-C WT in **c**. Relative preference of 1 is equal to a 2-fold change, and is
 722 represented by green, light pink and orange bars, respectively. *Bottom panels* show the
 723 distribution of nucleotides at positions N+1/2/3 respectively from the CpG site at 11
 724 preferred sites by DNMT3B-C WT in **a**, 4 preferred sites by DNMT3B-C WT in **b**, or the
 725 3 sites preferred by DNMT3B-C Arg829His in **c**. The data show that substrate
 726 preference of DNMT3A-C Arg882His variant is similar to that of DNMT3B enzyme.

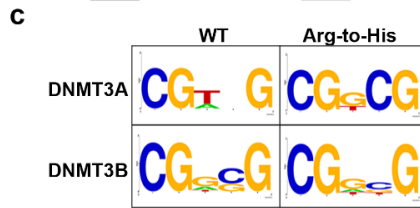
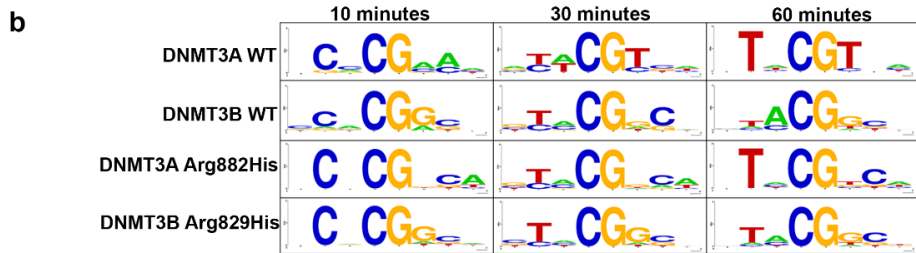
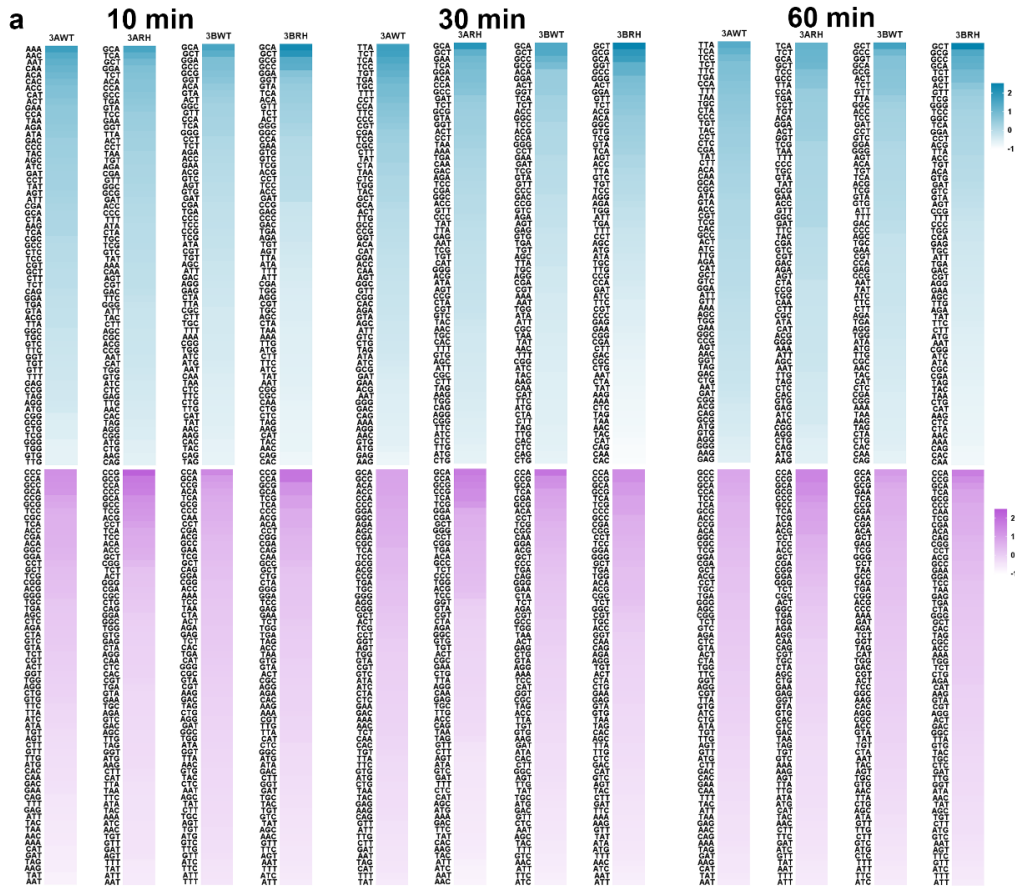
727 **FIGURE 4**



728 Flanking sequence preference at steady state kinetics. DNA methylation of the 56 CpG
 729 sites in the 509-bp substrate was analyzed using bisulfite sequencing. Methylation
 730 reactions were carried out using 1 μ M enzyme for 30 minutes in **a, c, e**, and 60 minutes
 731 in **b, d, f**. *Top panels* show the relative preference calculated for each CpG site (1-56)
 732 by DNMT3A-C WT compared to the DNMT3A-C Arg882His variant in **a, b**, DNMT3B-C

733 WT compared to DNMT3A-C WT in **c, d**, and DNMT3B-C WT compared to DNMT3A-C
734 Arg882His in **e, f**. Relative preference of 1 is equal to a 2-fold change, and is
735 represented by blue, green and pink bars, respectively. *Bottom panels* show the
736 fractional distribution of nucleotides at the preferred sites. Each bar represents
737 nucleotides at positions N+1/2/3 respectively from the CpG site. The data show that the
738 similarity in flanking sequence preference between DNMT3A-C Arg882His and
739 DNMT3B-C is maintained during steady state kinetics.

740 **FIGURE 5**



741 Trinucleotide sequence flanking CpG preferred by WT and variant DNMT3 enzymes. **a**
 742 Heat-map showing the preference of different trinucleotide sets by DNMT3A-C WT
 743 (3AWT), DNMT3A-C Arg882His (3ARH), DNMT3B-C WT (3BWT), or DNMT3B-C

744 Arg829His (3BRH) at 10, 30, and 60 minutes. Sequences with values greater or equal
745 to 1 are considered as preferred. Upper panels in blue represent the flanking sequence
746 preference at the N +1/2/3 positions. Lower panels in purple represent the flanking
747 sequence preference at the N-1/2/3 positions. **b** Consensus sequence generated by
748 WebLogo from top ten preferred sequences by each enzyme as methylation reaction
749 proceeds from 10 to 60 minutes. **c** Consensus flanking sequence for methylated sites
750 with a G at the N+3 position. The data show that for the DNMT3A-C Arg882His variant
751 the nucleotide preference at N+1 is lost and gained at N+3, while at the N+2 position it
752 switches from an A to a C, making it very similar to the flanking sequence preference of
753 DNMT3B-C. Similarly, the methylated sites with G at the N+3 position have inner
754 flanking sequence similar between DNMT3A-C Arg882His and DNMT3B-C, whereas it
755 is different for DNMT3A-C WT enzyme

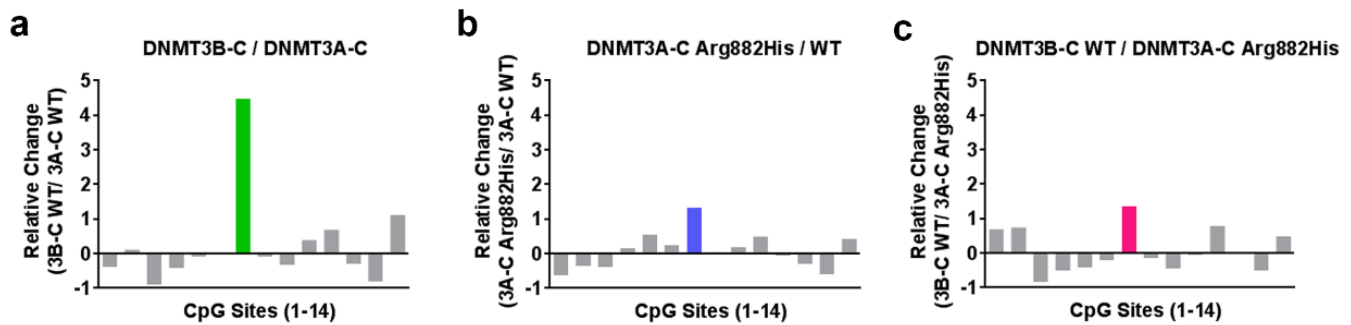
756

757

758

759

760 **FIGURE 6**



761 Relative activity and site preference of DNMT3 WT and mutant enzymes on *Meis1*
762 enhancer substrate. A 1-kb *Meis1* enhancer region was used as substrate for *in vitro*
763 methylation reactions by DNMT3A-C WT, the Arg882His variant, and DNMT3B-C WT.
764 DNA methylation of the 14 CpG sites was analyzed using bisulfite sequencing. Relative
765 preference for each of the 14 CpG sites was calculated for DNMT3B-C compared to
766 DNMT3A-C in **a**, DNMT3A-C Arg882His compared to DNMT3A-C WT in **b**, and
767 DNMT3B-C WT compared to DNMT3A-C Arg882His in **c**. Relative preference of 1 is
768 equal to a 2-fold change, and is represented by green, blue and pink bars, respectively..
769 DNMT3A-C Arg882His and DNMT3B-C prefer to methylate same site in this substrate,
770 however DNMT3B-C shows a very strong preference compared to the DNMT3A-C
771 Arg882His variant.

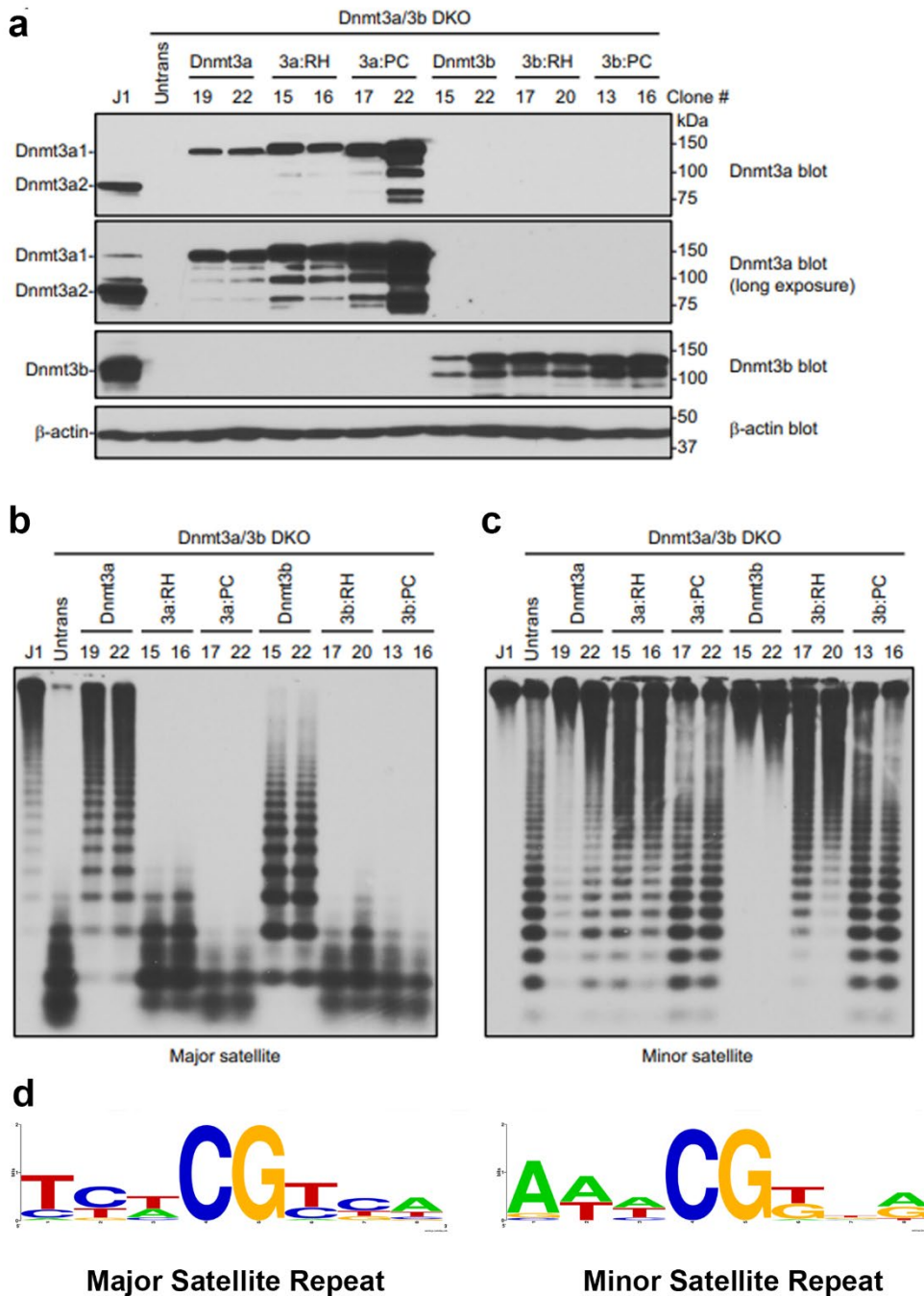
772

773

774

775

776 **FIGURE 7**



777 Rescue of DNA methylation at the major and minor satellite repeats in *Dnmt3a/3b* DKO
 778 mESCs. **a** *Dnmt3a/3b* DKO mESCs were transfected with plasmids encoding mouse
 779 DNMT3A1 WT, DNMT3A1 Arg878His (3A1:RH), DNMT3A Pro705ValCys706Asp

780 (3A1:PC), DNMT3B1 WT, DNMT3B1 Arg829His (3B1:RH), or DNMT3B1
781 Pro656GlyCys657Thr (3B1:PC), and stable clones were derived. Total cell lysates were
782 used to analyze the expression of DNMT3A or DNMT3B proteins by western blot with
783 anti-DNMT3A, anti-DNMT3B, and anti- β -Actin antibodies. A long exposure of the
784 DNMT3A blot is included to show endogenous DNMT3A1 in WT (J1) mESCs. Note that
785 stable clones showing similar expression levels to those of endogenous DNMT3A or
786 DNMT3B were used for the experiments. **b, c** DNA methylation was analyzed by
787 Southern blot. Genomic DNA was digested with *Mae*II (major satellite repeats) or *Hpa*II
788 (minor satellite repeats), and probed for the major **b** or minor **c** satellite repeats. J1
789 (WT) and untransfected DKO mESCs were used as controls. The numbers on the top
790 indicate clone#. Complete digestion due to low or no DNA methylation is indicated by
791 low molecular weight bands as seen in untransfected DKOs, and high molecular weight
792 bands as seen in J1 indicate high DNA methylation and protection from digestion.
793 Comparing the activity of DNMT3A1 clones 19/22 with 3A1:RH clones 15/16 at major
794 and minor satellite repeats shows that 15/16 methylate minor repeats similar to 19/22
795 whereas at major repeats the activity of 15/16 is severely impaired. **d** Consensus
796 sequence of the nucleotides flanking the CpG site in either the major or minor satellite
797 repeats, created using WebLogo shows high prevalence G at N+1 and N+3 positions at
798 minor repeats.

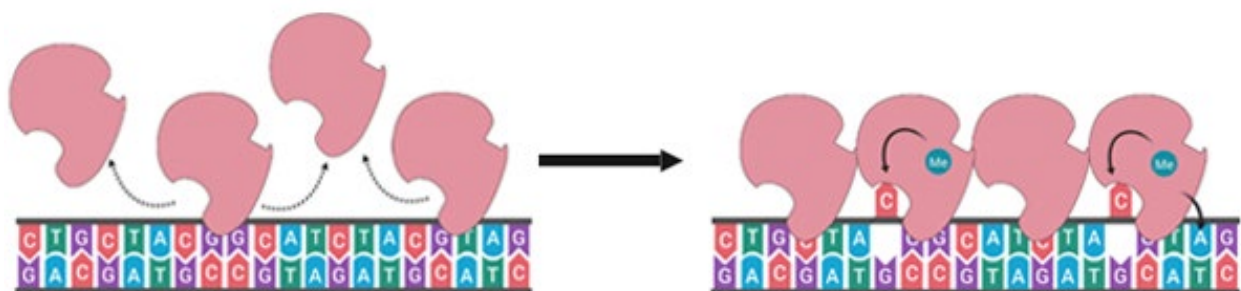
799

800

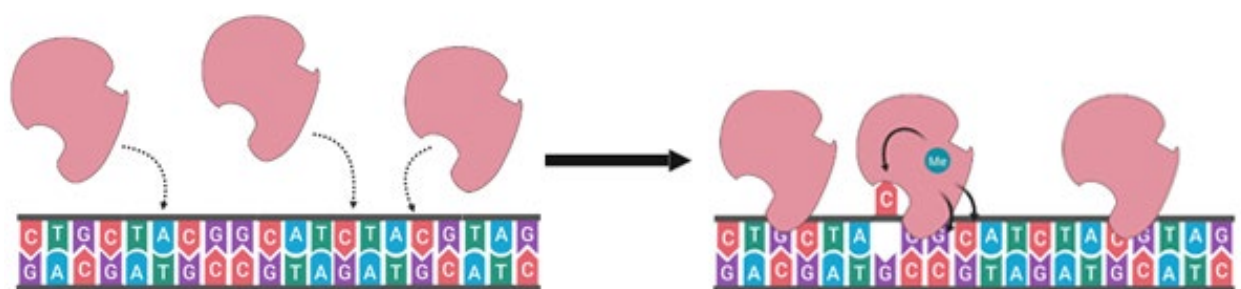
801

802 **FIGURE 8**

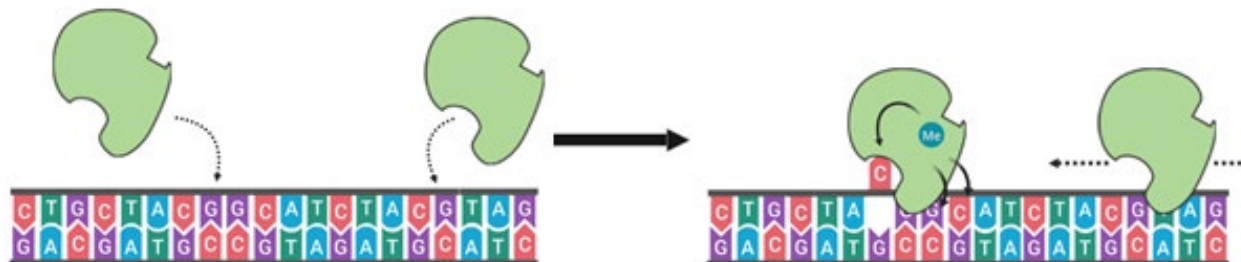
DNMT3A WT: Cooperative



DNMT3A Arg882His: Non-cooperative



DNMT3B: Processive



803
804 Model showing the different kinetic mechanisms of DNMT3A and DNMT3B and their
805 influence their flanking sequence preferences. At high concentrations, DNMT3A
806 methylates multiple CpG sites rapidly by using cooperative mechanism, and has no
807 strong preference at N+1 and N+3, and a minor preference for A at N+2. However, in
808 absence of cooperative kinetic mechanism at low concentrations of DNMT3A enzyme,
809 for the Arg882His variant, and DNMT3B, the flanking sequence preference for G at the
810 N+1 and C at N+2 position is observed.

811

812

813

814