Identification of transcription termination defects at DNA hypomethylated transcription termination sites in DNA methyltransferase 3a-deficient vertebrates

Masaki Shirai[1], Takuya Nara[2], Haruko Takahashi[1,2], Kazuya Takayama[1,4], Yuan Chen[2], Yudai Hirose[2,5], Masashi Fujii[2], Akinori Awazu[2], Nobuyoshi Shimoda[3], and Yutaka Kikuchi[1,2,*]

1. Department of Biological Science, Graduate School of Science, Hiroshima University, Kagamiyama 1-3-1, Higashi-Hiroshima, Hiroshima, 739-8526 Japan

2. Graduate School of Integrated Sciences for Life, Hiroshima University, Kagamiyama 1-3-1, Higashi-Hiroshima, Hiroshima 739-8526, Japan.

3. Laboratory of Molecular Analysis, Research and Development Management Center, National Center for Geriatrics and Gerontology, 7-430, Morioka, Obu, Aichi 474-8522, Japan

4. Present address: CRAIF Inc., 3-38-14-3 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan

5. Present address: Ryobi Systems Co., Ltd., 2-7-16 Toyonari, Minami-Ku, Okayama, 700-8504, Japan

*Corresponding Author: Yutaka Kikuchi, yutaka@hiroshima-u.ac.jp

Keywords: Transcription termination; Dnmt3a; DNA methylation; Read-through; Chimeric transcript

22 **Abstract**

23 **CpG methylation in genomic DNA is well known as a repressive epigenetic**
24 **marker in eukaryotic transcription, and DNA methylation of the promoter regions is**
25 **correlated with silencing of gene expression. In contrast to the promoter regions,**
26 **the function of DNA methylation during transcription termination remains to be**
27 **elucidated. A recent study has revealed that mouse DNA methyltransferase 3a**
28 **(Dnmt3a) mainly functions in *de novo* methylation in the promoter and gene body**
29 **regions (including transcription termination sites (TTSs)) during development. To**
30 **investigate the relationship between DNA methylation overlapping the TTSs and**
31 **transcription termination, we employed two strategies: informatic analysis using**
32 **already deposited datasets of *Dnmt3a*$^{-/-}$ mouse cells and the zebrafish model system.**
33 **Bioinformatic analysis using methylome and transcriptome data showed that**
34 **hypomethylated differentially methylated regions overlapping the TTSs were**
35 **associated with increased read counts and chimeric transcripts downstream of TTSs**
36 **in *Dnmt3a*$^{-/-}$ Agouti-related protein neurons, but not in *Dnmt3a*$^{-/-}$ embryonic stem**
37 **cells and mouse embryonic fibroblasts. We experimentally detected increased read-**
38 **through and chimeric transcripts downstream of hypomethylated TTSs in zebrafish**
39 **maternal-zygotic *dnmt3aa*$^{-/-}$ mutant. This study is the first to identify transcription**
40 **termination defects in DNA hypomethylated TTSs in *Dnmt3a*$^{-/-}$ vertebrates.**

41

42 **Introduction**

43 Transcription by RNA polymerase II (pol II), synthesis process of mRNA from
44 DNA, consists of three biological processes: initiation, elongation, and termination. In
45 higher eukaryotes, two transcription processes (initiation and elongation) and their
46 epigenetic regulation, including DNA methylation and histone modifications, have been
47 extensively studied (LEE AND YOUNG 2000; JONES 2012; GATES *et al*. 2017; CHEN *et al*.
48 2018). In transcription termination, it is also well known that most protein-coding
49 genes in eukaryotes possess a polyadenylation signal (PAS) located at the end of the 3'-
50 UTR, and a cleavage and polyadenylation (CPA) complex is assembled around the PAS
51 sequence of nascent pre-mRNA, which directs pol II transcription termination
52 (PROUDFOOT 2016; EATON AND WEST 2020). However, in contrast to transcription
53 initiation and elongation, the epigenetic regulation of transcription termination has
54 remained largely unstudied to date.

55 In eukaryotes ranging from plants to humans, DNA methylation, one of epigenetic
56 regulations, is well characterized as a repressive epigenetic marker and mainly occurs at
57 CpG dinucleotides (LI AND ZHANG 2014; SCHÜBELER 2015; LUO *et al*. 2018).
58 Advances in sequencing technology have resulted in genomic methylation profiling at
59 single-base resolution revealing that the global DNA methylation (5-methylcytosine,
60 5mC) is seen in mammals, except for short CpG-rich regions known as CpG islands (LI
61 AND ZHANG 2014; SCHÜBELER 2015; LUO *et al*. 2018). Studies of methylome

62 profiling have shown that DNA methylation patterns dynamically change during
63 development and disease and that these patterns are linked to gene expression in
64 mammals (SMITH AND MEISSNER 2013; GREENBERG AND BOURC'HIS 2019). Genome-
65 wide methylome analysis revealed that the DNA methylation levels around the
66 promoter regions, including the transcription start site (TSS), are negatively correlated
67 with gene expression (BALL *et al*. 2009; RAUCH *et al*. 2009). Conversely, DNA
68 methylation levels in the gene bodies are positively correlated with gene expression
69 (KULIS *et al*. 2013; TEISSANDIER AND BOURC'HIS 2017). Although two possible
70 mechanisms, regulation of splicing and prevention of cryptic promoters, have been
71 proposed for methylation function in the gene body (KULIS *et al*. 2013; TEISSANDIER
72 AND BOURC'HIS 2017), the significance of DNA methylation around transcription
73 termination sites (TTSs; cleavage/polyadenylation sites) remains unresolved.
74 The DNA methylation patterns are well known to be established by the DNA
75 methyltransferases (Dnmts), which are categorized into two types in mammals:
76 maintenance Dnmt (Dnmt1) and *de novo* Dnmts (Dnmt3a and Dnmt3b) (LI *et al*. 1992;
77 OKANO *et al*. 1999; LI AND ZHANG 2014). Knock-out mice of both Dnmt3a and
78 Dnmt3b, which die at embryonic day 9.5, fail to start *de novo* methylation after
79 implantation (OKANO *et al*. 1999), and genetic mutation of Dnmt3a or Dnmt3b leads to
80 various human diseases (GREENBERG AND BOURC'HIS 2019), suggesting the critical role
81 of DNA methylation in development and homeostasis in mammals through the

82 regulation of gene expression. A very recent study using mouse embryonic fibroblasts
83 (MEFs) with genetic ablation of *de novo* Dnmts reported that Dnmt3a functions in *de*
84 *novo* methylation at both TSS regions (TSS ± 1000 bp) and gene bodies (from TSS
85 regions to TTS) of the Polycomb group (PcG) target developmental genes, whereas
86 Dnmt3b mainly functions in X-chromosome methylation (YAGI *et al*. 2020). These
87 results indicate that although the enzymatic functions of both proteins are redundant,
88 genomic target regions are distinct in mammals. Although mammalian Dnmt3a
89 mutant phenotypes and target regions have been reported, the role of methylation during
90 transcription termination by Dnmt3a has not yet been reported.

91 In the present study, we used bioinformatic analysis and a zebrafish model animal
92 system to explore the function of Dnmt3a-mediated hypomethylation at the TTSs. The
93 bioinformatic analysis showed that hypomethylated differentially methylated regions
94 (hypoDMRs) overlapping the TTSs were associated with increased read counts and
95 chimeric transcripts downstream of TTSs in *Dnmt3a*$^{-/-}$ Agouti-related protein neurons, but
96 not in *Dnmt3a*$^{-/-}$ embryonic stem cells (ESCs) and mouse embryonic fibroblasts (MEFs).
97 Based on bioinformatic analysis, to experimentally confirm the transcription termination
98 defects in *Dnmt3a*$^{-/-}$ vertebrates, we generated zebrafish maternal-zygotic *dnmt3aa*$^{-/-}$
99 (MZ*dnmt3aa*$^{-/-}$) mutants and identified an increase in both read-through and chimeric
100 transcripts in the hypoDMRs overlapping the TTSs in MZ*dnmt3aa*$^{-/-}$ mutant. Taken

101 together, in *Dnmt3a*-deficient vertebrates, we identified transcription termination defects

102 in DNA hypomethylated TTSs.

103

104 **Materials and methods**

105 **Zebrafish experiments, ethic statement, and generation of MZ*dnmt3aa*$^{-/-}$ mutant**

106 **fish using CRISPR/Cas9 system**

107 Adult zebrafish and zebrafish larvae were maintained as previously described

108 (WESTERFIELD 1993). All zebrafish experiments were approved from the Hiroshima

109 University Animal Research Committee (Permit Number: F18-2-7).

110 Genome editing using the CRISPR/Cas9 system was performed as previously

111 described (ANSAI AND KINOSHITA 2014). Briefly, the pDR274 vector (Addgene

112 Plasmid 42250) was digested with *Bsa*I, and then annealed oligonucleotides for

113 sgRNAs were cloned into the pDR274 vector. The sequences of the oligonucleotides

114 for sgRNAs are listed in Supplementary Table S1. The sgRNA expression vectors

115 were digested by *Dra*I, and the sgRNAs were synthesized using the MAXIscript™ T7

116 Kit (Invitrogen, Thermo Fisher Scientific). The synthesized sgRNAs were purified

117 using the mirVana™ miRNA Isolation Kit (Invitrogen, Thermo Fisher Scientific). In

118 addition, the Cas9 expression vector (Addgene Plasmid 51815) was linearized by *Not*I

119 and *Cas9* mRNA was synthesized using the mMESSAGE mMACHINE™ SP6 Kit

120 (Invitrogen, Thermo Fisher Scientific). A mixture of three types of sgRNAs for the

121 zebrafish *dnmt3aa* gene and *Cas9* mRNA of appropriate concentrations was injected at

122 the one-cell stage. To identify the mutations, the target site of the CRISPR/Cas9

123 system on *dnmt3aa* gene was amplified using TaKaRa Ex Taq® (Takara Bio) with the

124 primers listed in Supplementary Table S2. The amplicons were purified using the
125 QIAquick PCR Purification Kit (QIAGEN) and then sequenced using the BigDye™
126 Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, Thermo Fisher Scientific)
127 and primers listed in Supplementary Table S2.

128 The amputated fins of *dnmt3aa*$^{-/-}$ mutant were digested overnight in lysis buffer
129 (100 mM NaCl, 20 mM Tris-HCl pH 8, 50 mM EDTA pH 8, 0.5% SDS and 36 ug/ml
130 Proteinase K) at 55°C. Proteinase K was inactivated at 98°C for 10 min. A
131 heteroduplex mobility assay (HMA) was performed to detect the *dnmt3aa* mutation.
132 The target site of the CRISPR/Cas9 system on exon 1 of the *dnmt3aa* gene was
133 amplified using BIOTAQ™ DNA Polymerase (Meridian Bioscience); the primers used
134 are listed in Supplementary Table S2. Amplicons were analyzed using polyacrylamide
135 gel electrophoresis. MZ*dnmt3aa*$^{-/-}$ mutants were obtained by crossing *dnmt3aa*$^{-/-}$
136 adults.

137

138 **Whole-genome bisulfite library preparation and sequencing of zebrafish samples**

139 Zebrafish wild-type (WT) and MZ*dnmt3aa*$^{-/-}$ mutant embryos at 2 days post-
140 fertilization (dpf) were digested with proteinase K and sodium dodecyl sulfate to extract
141 genomic DNA. DNA samples were prepared by mixing zebrafish genomic DNA and
142 unmethylated lambda DNA (Promega) as spike control, at a ratio of 1000:1. DNA
143 samples (100 ng) were subjected to bisulfite treatment using the EZ DNA Methylation-

144 Gold Kit (Zymo Research). Bisulfite-treated DNA samples (50 ng) were subjected to
145 10 cycles of PCR using random primers from the TruSeq DNA Methylation Kit
146 (Illumina) to add adapters. Libraries with adapters were sequenced on the HiSeq X
147 Five Sequencing System (Illumina) with 150 bp single-end reads. Bisulfite treatment,
148 library preparation, and sequencing were performed by Takara Bio Inc. (Shiga, Japan).
149

150 **Whole-genome bisulfite sequencing (WGBS) data analysis of mouse dataset**

151 The publicly available datasets used in this study were from the Gene Expression
152 Omnibus (GEO), and the accession numbers are GSE100957 (J1 mESCs) (GU *et al.*
153 2018), GSE111172 (MEFs) (YAGI *et al.* 2020), and GSE122405 (AgRPs) (MACKAY *et*
154 *al.* 2019). The FASTQ-format read sequences were trimmed of adaptors and low-
155 quality bases were filtered using Trim Galore (version 0.6.4)
156 (https://github.com/FelixKrueger/TrimGalore) with Cutadapt (version 2.10) (MARTIN
157 2011) and FastQC (version 0.11.8)
158 (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). After trimming, the
159 reads were aligned to the mouse genome (mm10) using Bismark (version 0.20.0)
160 (KRUEGER AND ANDREWS 2011) and Bowtie (version 1.0.0) (LANGMEAD *et al.* 2009).
161 The aligned reads were followed by the removal of PCR duplicates using
162 deduplicate_bismark (version 0.20.0) (KRUEGER AND ANDREWS 2011), and methylation
163 calls were performed using bismark_methylation_extractor (version 0.20.0) (KRUEGER

164 AND ANDREWS 2011). The reads from samples with multiple sequencing lanes were
165 merged using Samtools (version 1.9) (LI *et al*. 2009) after removing the PCR duplicates.
166 The methylation ratio of CpG sites with a coverage of over 5 reads in WT and *Dnmt3a*$^{-/-}$
167 was calculated from the methylation call data. The methylation ratio was visualized
168 using Integrative Genomics Viewer (IGV) (version 2.7.2) (ROBINSON *et al*. 2011).
169

170 **WGBS data analysis of zebrafish datasets**

171 The read sequences in FASTQ format were first trimmed by removing six bases
172 from random primers of TruSeq DNA Methylation Kit (Illumina) and the last 50 bases
173 of low-quality reads. After trimming, the reads were aligned to the zebrafish genome
174 (danRer10) and lambda phage genome using Bismark (version 0.10.1) (KRUEGER AND
175 ANDREWS 2011) and Bowtie (version 1.0.0) (LANGMEAD *et al*. 2009). Once aligned,
176 reads corresponding to PCR duplicates were removed and methylation calls were
177 performed. The bisulfite conversion rates based on spike control for all samples were
178 over 99%, and the error rates were as follows: WT, 0.3% and MZ*dnmt3aa*$^{-/-}$ mutant,
179 0.2%. Statistically significant methylated cytosine sites were identified using a
180 binomial distribution with a false discovery rate ≤ 0.05, calculated by the Benjamini-
181 Hochberg (BH) method, and read depth ≥ 5. Data analysis was performed by Takara
182 Bio, Inc. (Shiga, Japan) as previously described (LISTER *et al*. 2009). WGBS data
183 were deposited in the GEO under accession number GSE178691. The methylation

184 ratio of CpG sites with a coverage of over 5 reads in WT and MZ*dnmt3aa*$^{-/-}$ mutant as
185 calculated from the methylation call data. The methylation ratio was visualized using
186 the IGV (version 2.7.2) (ROBINSON *et al*. 2011).
187
188 **Identification of differentially methylated regions (DMRs)**
189 The methylation call data of mice and zebrafish were used to identify DMRs using
190 swDMR (version 1.0.0) (WANG *et al*. 2015) with the followed settings: cytosine type,
191 CG; window, 500; step size, 50; points (lowest number of selected cytosine type in the
192 window), 3; coverage, 5; fold (methylation level difference), 1.5; diff (value of max-min
193 methylation level), 0.1; p-value, 0.05; fdr, 0.05. Samples were compared using the
194 Fisher's test. DMRs overlapped genomic positions (TSS, gene body, TTS, and
195 intergenic region) were determined using bedtools (version 2.28.0) (QUINLAN AND
196 HALL 2010) intersect. The DMRs that overlap both TSS and TTS were included in the
197 DMRs that overlap TTS. Heatmaps, violin plots, scatterplots with marginal
198 distribution, and horizontal bar graphs were generated from swDMR output using
199 Python and R scripts.
200
201 **RNA sequencing (RNA-seq) library preparation and sequencing of zebrafish**
202 **samples**
203 Tissue lysates were prepared from 30 pooled 2 dpf WT and MZ*dnmt3aa*$^{-/-}$ mutant
204 by homogenization using TissueLyser LT (QIAGEN) with zirconia beads for 10 s at 50

205 Hz. RNA was extracted from the lysate using the PureLink® RNA Mini Kit (Thermo
206 Fisher Scientific). RNA quality of all samples was RNA Integrity Number (RIN) >
207 9.0, using the 4150 TapeStation System (Agilent). Total RNA was treated using the
208 NEBNext® Poly(A) mRNA Magnetic Isolation Module (New England Biolabs) for
209 mRNA isolation. Strand-specific RNA-seq libraries were constructed using the
210 NEBNext® Ultra™ Directional RNA Library Prep Kit for Illumina® (New England
211 Biolabs). The libraries with adapters were sequenced on the Novaseq 6000 (Illumina)
212 with 2 × 150 bp paired-end reads. RNA extraction, mRNA isolation, library
213 preparation, and sequencing were performed by Rhelixa Inc. (Tokyo, Japan).
214

215 **RNA-seq data analysis for mouse and zebrafish datasets**

216 The publicly available datasets of mice deposited in the GEO and the generated
217 datasets of zebrafish were used in this study. Accession numbers are GSE100957 (J1
218 mESCs) (GU *et al*. 2018), GSE84165 (MEFs) (YAGI *et al*. 2017a; YAGI *et al*. 2019),
219 GSE111172 (MEFs) (YAGI *et al*. 2020), and GSE122405 (AgRPs) (MACKAY *et al*.
220 2019). RNA-seq data of zebrafish were deposited in the GEO under accession number
221 GSE178691. FASTQ files from samples with multiple sequencing lanes were merged.
222 The FASTQ-format sequence reads were trimmed of adaptors and low-quality bases
223 were filtered using Trim Galore (version 0.6.4)
224 (https://github.com/FelixKrueger/TrimGalore) with Cutadapt (version 1.18 or 2.10)

225 (MARTIN 2011) and FastQC (version 0.11.8 or 0.11.9)
226 (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). After trimming, the
227 reads were aligned to the mouse genome (mm10) and the zebrafish genome (danRer10)
228 using HISAT2 (version 2.1.0) (KIM *et al*. 2015). Alignment of zebrafish datasets was
229 performed using the strand-specific option. Aligned files (.bam) were sorted and
230 indexed using Samtools (version 1.9) (LI *et al*. 2009). In addition, the strand-specific
231 aligned files of zebrafish were generated using Samtools (version 1.9) (LI *et al*. 2009).
232 Read coverage was normalized with 1 bin size by counts per million mapped reads
233 (CPM) using the bamCoverage tool from the DeepTools (version 3.3.2) (RAMÍREZ *et al*.
234 2016) and visualized using the IGV (version 2.7.2) (ROBINSON *et al*. 2011). Splice
235 junction data of chimeric transcripts in the aligned files were inspected manually using
236 the IGV (version 2.7.2) (ROBINSON *et al*. 2011).
237

238 **Calculation of expression level downstream of TTS from RNA-seq data**

239 Quantification of expression levels was completed using StringTie (version 2.1.2)
240 (PERTEA *et al*. 2015) with NCBI RefSeq annotation. The expression values of each
241 transcript were normalized to the transcripts per million (TPM). Transcripts
242 expressed in WT and *Dnmt3a*$^{-/-}$ mouse cells or MZ*dnmt3aa*$^{-/-}$ mutant were extracted
243 from all transcripts (TPM > 0). In addition, the transcripts expressed in both strains
244 and overlapped hypoDMRs across the TTS were extracted.

245 All reads that overlapped the exonic or intronic sequences in the aligned files
246 (.bam) were removed using bedtools (version 2.28.0) (QUINLAN AND HALL 2010)
247 intersect with the NCBI RefSeq annotation. In addition, the intergenic reads were
248 counted within 3 kb downstream of the TTS using bedtools (version 2.28.0) (QUINLAN
249 AND HALL 2010) multicov, except when the genomic region downstream of the TTS
250 was less than 3 kb. The expression level was calculated using the following equation
251 (TPM 3 kb):
252 TPM 3 kb = Number of reads aligned in 3kb downstream of TTS of each transcript ×
253 10^6 / Total number of reads aligned in 3kb downstream of TTS.
254 Normalized read counts were compared in WT and *Dnmt3a*$^{-/-}$ mouse cells or
255 MZ*dnmt3aa*$^{-/-}$ zebrafish mutants as a metagene analysis.
256 Box plots were generated using R scripts, and statistical significance was
257 determined using Brunner-Munzel test ($p \leq 0.05$).
258

259 **Calculation of expression level of chimeric transcripts from RNA-seq data**

260 Chimeric transcripts, which were aberrantly fused with exons of downstream genes
261 or intergenic DNA, were manually counted using IGV (ROBINSON *et al*. 2011). The
262 expression level was calculated as the counts per million (CPM) mapped reads.
263 Normalized read counts were compared in WT and *Dnmt3a*$^{-/-}$ mouse cells or
264 MZ*dnmt3aa*$^{-/-}$ zebrafish mutant as a metagene analysis. Transcripts whose expression

265 increased downstream of the TTS were used for metagene analysis. Box plots were
266 generated using R scripts. Statistical significance was determined using the Brunner-
267 Munzel test ($p \leq 0.05$).
268

269 **RNA preparation, quantitative PCR (qPCR), and sequencing for identification of**
270 **read-through and chimeric transcripts of zebrafish samples**

271 Total RNA was extracted using TRIzol (Invitrogen, Thermo Fisher Scientific)
272 from six samples of 30 pooled 2 dpf WT and MZ*dnmt3aa*$^{-/-}$ mutant, and the six
273 extracted total RNA samples were combined as one sample. Total RNA (150 ug) was
274 treated using the *Oligotex*™ *-dT30<Super>* mRNA Purification Kit (Takara Bio). The
275 purified poly(A)$^{+}$ RNA was reverse-transcribed with oligo(dT) primer and reverse
276 transcriptase XL (AMV) (Takara Bio).

277 qPCR was performed using the StepOne™ Real-Time PCR System and
278 PowerUp™ SYBR™ Green Master Mix (Applied Biosystems, Thermo Fisher
279 Scientific). Relative expression levels were calculated by the ΔΔCT method, and
280 targets were normalized by the mean value of a control gene, *ubiquitin-conjugating*
281 *enzyme E2A* (*ube2a*). The qPCR primers used are shown in Supplementary Table S3.
282 Bar charts were generated from the mean values and standard error (SE) data using R
283 scripts. Statistical significance was determined using the t-test ($p \leq 0.05$).

284 Sequencing of chimeric transcript was performed using MZ*dnmt3aa*$^{-/-}$ mutant.
285 The product of *im:7152937*-intergenic DNA (iD)- *meiosis-specific nuclear structural 1*
286 (*mns1*) chimeric transcript was amplified using BIOTAQ™ DNA Polymerase (Meridian
287 Bioscience) with the primers listed in Supplementary Table S4. The amplicons were
288 purified using the QIAquick Gel Extraction Kit (QIAGEN) after agarose gel
289 electrophoresis. The purified product was subcloned using a TA-cloning Kit (Thermo
290 Fisher Scientific), and the plasmid DNA was sequenced using the BigDye™ Terminator
291 v3.1 Cycle Sequencing Kit (Applied Biosystems, Thermo Fisher Scientific) and primers
292 listed in Supplementary Table S4.
293

294 **Results**

295 **Genomic characteristics of hypoDMRs overlapping the TTS in *Dnmt3a*-deficient**
296 **mouse cells**

297 To investigate the functional relevance between DNA methylation and
298 transcription termination, we searched datasets of *Dnmt3a*-deficient mammalian cells
299 that contain both methylome (WGBS) and transcriptome (RNA-seq) data. Five
300 datasets of WGBS (human: one, mouse: four) and ten datasets of RNA-seq (human:
301 four, mouse: six) were obtained from GEO. Four datasets (human: one, mouse: three)
302 contained both WGBS and RNA-seq, and among them we chose three types of mouse
303 datasets (J1 mESCs (GU *et al*. 2018); MEFs (YAGI *et al*. 2020); and Agouti-related
304 protein neurons, AgRPs (MACKAY *et al*. 2019)) because of the differences in their cell
305 differentiation states. The outline of bioinformatics analysis and experimental design
306 was shown in Supplementary Figure S1. We extracted the hypermethylated DMRs
307 (hyperDMRs) and hypoDMRs from these datasets and found that the number of
308 hypoDMRs in *Dnmt3a*$^{-/-}$ J1 mESCs were approximately 57 and 181 times greater than
309 that of *Dnmt3a*$^{-/-}$ MEFs and AgRPs, respectively (Supplementary Figure S2, A-C).
310 The hypoDMRs were categorized by genomic position (gene body, intergenic region,
311 and gene body & intergenic region) (Supplementary Figure S3). The number of
312 hypoDMRs gradually decreased in the order of *Dnmt3a*$^{-/-}$ J1 mESCs, MEFs, and
313 AgRPs, whereas their percentages of genomic positions were nearly the same among

314 these *Dnmt3a*-deficient cells (Supplementary Figure S3). Next, we selected the
315 hypoDMRs overlapping TTSs in three types of *Dnmt3a*$^{-/-}$ mouse cells (Figure 1, A and
316 B) and found that the number of hypoDMRs overlapping the TTSs in *Dnmt3a*$^{-/-}$ MEFs
317 or AgRPs was less than 1% or 0.4% of that of *Dnmt3a*$^{-/-}$ J1 mESCs, respectively (Figure
318 1C). These data suggest that DNA methylation regulated by Dnmt3a may be
319 correlated with the differentiation state of mouse cells. Two-dimensional (2D)
320 scatterplots with marginal distribution showed that the hypoDMRs overlapping the
321 TTSs were almost symmetrically distributed across the TTSs in the *Dnmt3a*$^{-/-}$ J1 mESCs
322 and MEFs, whereas the distribution of hypoDMRs was slightly shifted upstream of the
323 TTSs in *Dnmt3a*$^{-/-}$ AgRPs (Figure 1D).
324

325 **HypoDMRs overlapping the TTSs are associated with an increased read counts**
326 **and chimeric transcripts downstream of TTSs in *Dnmt3a*$^{-/-}$ AgRPs**

327 To explore the effect of hypomethylation overlapping the TTSs on transcription
328 termination, we analyzed RNA-seq data from three different types of *Dnmt3a*$^{-/-}$ mouse
329 cells. Read counts aligned in 3 kbp downstream of TTSs were normalized by TPM (3
330 kb) in wild-type (WT) and *Dnmt3a*$^{-/-}$ J1 mESCs, MEFs, and AgRPs. Metagene
331 analysis showed that the read counts downstream of TTSs increased significantly in
332 *Dnmt3a*$^{-/-}$ AgRPs compared to those in WT cells (Figure 2, A-C and Supplementary
333 Figure S4A; *Ring Finger 157* (*Rnf157*) gene as an example of increased read counts),

334 and the higher expression downstream of TTSs was observed in more than half of the
335 genes with hypoDMRs in mutant cells (Figure 2D). The centers of the hypoDMRs
336 which cause an increase in read counts downstream of TTSs were distributed close to
337 the TTSs in *Dnmt3a*$^{-/-}$ AgRPs (Supplementary Figure S5, A, A', and B). Conversely,
338 the read counts in *Dnmt3a*$^{-/-}$ J1 mESCs and MEFs were unchanged compared to those in
339 WT cells (Figure 2, A and B).

340 We then analyzed the content of transcripts downstream of hypomethylated TTSs
341 in the *Dnmt3a*$^{-/-}$ AgRPs, and found chimeric transcripts that were aberrantly fused with
342 exons of downstream genes or intergenic DNA (iD) in *Dnmt3a*$^{-/-}$ AgRPs, as observed in
343 WT cells. Frequency of chimeric transcripts (about 54.6%) increased significantly in
344 the *Dnmt3a*$^{-/-}$ AgRPs compared to those in WT cells (Figure 2, E and F). We showed
345 that splicing of the *Rnf157* as an example of chimeric transcripts in the *Dnmt3a*$^{-/-}$
346 AgRPs, and found aberrant transcripts that connect an *Rnf157* exon either to the
347 intergenic DNA or to *Unkempt family zinc finger* (*Unk*) gene (Supplementary Figure
348 S4B). The centers of DMRs causing these chimeric transcripts were distributed close
349 to the TTS (Figure 2, G, G', and H). In summary, the increase in the chimeric
350 transcripts in the *Dnmt3a*$^{-/-}$ AgRPs clearly shows that transcription termination is
351 defective and that read-through occurs at hypomethylated TTSs.
352

353 **Genomic characteristics of hypoDMRs overlapping the TTSs in MZ*dnmt3aa*$^{-/-}$**
354 **mutant**

355 To explore the correlation between the transcription termination defects and DNA
356 methylation in an animal model, we examined the methylome and transcriptome
357 analysis using the zebrafish as a model. Two zebrafish orthologs (*dnmt3aa* and
358 *dnmt3ab*) of mammalian Dnmt3a have been previously identified (TAKAYAMA *et al*.
359 2014), and the detailed expression profiles of *dnmt3aa* during development and
360 regeneration showed that *dnmt3aa* was maternally expressed and its expression was
361 observed in more organs than *dnmt3ab* (TAKAYAMA *et al*. 2014). Therefore, we
362 attempted to generate the MZ*dnmt3aa*$^{-/-}$ knockout mutant by using CRISPR/Cas9
363 method (Figure 3A) and found that they were viable, fertile and showed no specific
364 phenotype at 2 dpf (Figure 3B). Analyses of WGBS and RNA-seq were performed at
365 2 dpf, because most of the organs had already formed at this stage. In MZ*dnmt3aa*$^{-/-}$
366 mutant, 43.6% DMRs were hypomethylated (Supplementary Figure S6) and the
367 hypoDMRs were categorized by genomic position (gene body, intergenic region, and
368 gene body & intergenic region) (Supplementary Figure S7). We then selected the
369 hypoDMRs overlapping the TTSs, and 27 hypoDMRs were identified overlapping the
370 TTSs (Figure 3, C-E). A 2D scatterplot with marginal distribution showed that the
371 hypoDMRs overlapping the TTSs were almost symmetrically distributed across the

372 TTSs in MZ*dnmt3aa*$^{-/-}$ mutant (Figure 3, F and F'), as observed in *Dnmt3a*$^{-/-}$ J1 mESCs
373 and MEFs.
374
375 **Increases in both chimeric and read-through transcripts were identified in the**
376 **hypoDMRs overlapping TTSs in MZ*dnmt3aa*$^{-/-}$ mutant**
377 We also analyzed the RNA-seq data of MZ*dnmt3aa*$^{-/-}$ mutant and the read counts
378 from genes containing hypoDMRs overlapping the TTSs, which were normalized in the
379 WT and mutant. Metagene analysis showed that the read counts downstream of the
380 TTSs in MZ*dnmt3aa*$^{-/-}$ mutant were unchanged compared to those in WT (Figure 4A).
381 The transcripts that showed an increase (30.3%) were almost symmetrically distributed
382 across the TTSs in the mutant (Figure 4B and Supplementary Figure S8). Next, we
383 have extracted chimeric transcripts fused with exons of downstream genes and/or
384 intergenic DNA (iD). Normalized read counts of the chimeric transcripts (30%) were
385 also unchanged in MZ*dnmt3aa*$^{-/-}$ mutant compared to those in WT (Figure 4, C and D).
386 Three increased chimeric transcripts (*cytokine like 1* (*cytl1*), one; *im:7152937*, two)
387 were found in two different DMRs, which were close to TTSs (Figure 4, E, E', and F
388 and Supplementary Figure S9 and S10).
389 To further analyze the transcription termination defect caused by hypoDMRs
390 overlapping the TTSs, we examined the detailed expression downstream of and across
391 the TTSs of three genes that had shown high expression levels downstream of TTSs

392 (*cytl1*, *im:7152937*, and *potassium channel, subfamily K, member 1b* (*kcnk1b*)) by
393 RNA-seq. The expression level of three transcripts (*cytl1*_doT, *im:7152937*_doT, and
394 *kcnk1b*_doT) in MZ*dnmt3aa*$^{-/-}$ mutant had significantly increased downstream of the
395 TTSs compared to that in WT, as detected by qPCR (Figure 5, A and E). We also
396 found a significant increase in the two transcripts (*im:7152937*_acT and *kcnk1b*_acT)
397 across the TTSs (Figure 5, B, F, and G), demonstrating that read-through occurs in these
398 genes. In addition to read-through, chimeric transcripts (*im:7152937*-iD_cht and
399 *im:7152937*-iD-*meiosis-specific nuclear structural 1* (*mns1*)_cht) were detected in
400 *im:7152937* gene (Supplementary Figure S10 and S11), and qPCR analysis showed that
401 the expression level of chimeric transcripts between *im:7152937* and iD or *mns1* had
402 increased significantly in MZ*dnmt3aa*$^{-/-}$ mutant (Figure 5C, D, and H). Taken
403 together, the read-through and chimeric transcripts of three genes in MZ*dnmt3aa*$^{-/-}$
404 mutant, combined with the results from the *Dnmt3a*$^{-/-}$ AgRPs analyses, clearly support
405 our idea that transcription termination is defective in hypomethylated TTSs.
406

407 **Discussion**

408 Unlike transcription initiation and elongation, the function of DNA methylation
409 during transcription termination remains unclear. To investigate the DNA methylation
410 function overlapping the TTS, we analyzed the impact of hypomethylation on
411 transcription termination in *Dnmt3a*-deficient vertebrates. We showed that read counts
412 downstream of TTS and number of chimeric transcripts with exons of downstream
413 genes or iD had increased significantly in the *Dnmt3a*$^{-/-}$ AgRPs, but not in *Dnmt3a*$^{-/-}$
414 ESCs and MEFs. Moreover, we identified increased read-through and chimeric
415 transcripts in MZ*dnmt3aa*$^{-/-}$ zebrafish mutant. Taken together, we propose our model
416 that DNA hypomethylation of TTSs is associated with the transcription termination
417 defect in *Dnmt3a*$^{-/-}$ vertebrates (*Dnmt3a*$^{-/-}$ AgRPs and MZ*dnmt3aa*$^{-/-}$ zebrafish mutant)
418 (Figure 6).

419 Our bioinformatic analyses showed that the number of hypoDMRs overlapping
420 TTSs in *Dnmt3a*$^{-/-}$ J1 mESCs was much larger than that in *Dnmt3a*$^{-/-}$ MEFs and AgRPs
421 (Figure 1C). Previous studies have shown that mouse ESCs, such as J1 mESCs, which
422 are cultured in medium containing serum, are globally hypermethylated compared to
423 cells in mouse inner cell mass (YAGI *et al*. 2017b). In other words, *Dnmt3a* knockout
424 J1 ESCs are thought to have a large number of hypoDMRs. The number of
425 hypoDMRs overlapping TTSs in *Dnmt3a*$^{-/-}$ J1 mESCs gradually decreased in the order
426 of *Dnmt3a*$^{-/-}$ J1 mESCs, MEFs, and AgRPs (Figure 1C). This result may be attributed

427 to the fact that Dnmt3a regulates genome-wide methylation in ES cells; nevertheless, its
428 role may be restricted as differentiation proceeds.

429 In this study, we used zebrafish mutant as an animal model because MZ*dnmt3aa*$^{-/-}$
430 mutant, unlike *Dnmt3a* knockout mice, are viable and fertile. The MZ*dnmt3aa*$^{-/-}$
431 mutant line can be easily maintained and in this mutant maternal Dnmt3aa functions can
432 be excluded. However, normalized read counts downstream of the TTSs and chimeric
433 transcripts were unchanged between WT and MZ*dnmt3aa*$^{-/-}$ mutant (Figure 4). There
434 are two possible explanations for these results: one possibility is that zebrafish *dnmt3ab*
435 may rescue the transcription termination defects in MZ*dnmt3aa*$^{-/-}$ mutant.
436 Alternatively, it is hard to detect hypoDMRs overlapping TTSs in MZ*dnmt3aa*$^{-/-}$ mutant,
437 because cells in the zebrafish body are more heterogeneous and vary greatly compared
438 to AgRPs. Nevertheless, we have successfully detected the read-through and/or
439 chimeric transcripts of three genes (*cytl1*, *im:7152937*, and *kcnk1b*) in MZ*dnmt3aa*$^{-/-}$
440 mutant, suggesting that DNA methylation is involved in transcription termination in
441 zebrafish.

442 Although we have shown the significance of hypomethylation at the TTS region
443 for transcription termination in vertebrates, it is unclear how hypomethylation regulates
444 transcription termination. Despite the relative lack of transcription termination
445 research, a unified allosteric/torpedo model has recently been proposed for the PAS-
446 dependent termination (EATON *et al*. 2020; EATON AND WEST 2020). In this model,

447 the CPA complex recognizes and cleaves the nascent transcript when the pol II passes
448 through the PAS; thus, PAS is close to TTS (10-35 bp upstream of the TTS) (Figure 6)
449 (TIAN AND MANLEY 2017; NOURSE *et al*. 2020). This cleavage by the CPA complex
450 results in pol II pausing via dephosphorylation of the Spt5 elongation factor by a protein
451 phosphatase 1 nuclear targeting subunit (PNUTS) (EATON AND WEST 2020). As
452 paused pol II is dissociated from the DNA strand by the function of XRN2
453 exoribonuclease, which degrades RNA in a 5'→3' direction, the pol II pausing is
454 thought to be a critical step for the transcription termination (PROUDFOOT 2016; EATON
455 AND WEST 2020). Although the CPA complex mainly binds to nascent pre-mRNA and
456 C-terminal domain (CTD) of pol II, it is possible that DNA hypomethylation of the TTS
457 or PAS genomic region may directly or indirectly suppress the CPA function, leading to
458 loss of pol II pausing. It should be interesting to note that almost all hypoDMRs,
459 which cause chimeric transcripts or increased read counts downstream of the TTSs,
460 overlap with PAS in *Dnmt3a*$^{-/-}$ AgRPs and MZ*dnmt3aa*$^{-/-}$ mutant (Figure 2H, Figure 4F,
461 Supplementary Figure 5B, Supplementary Figure 8B). In PAS-dependent termination,
462 there are two other mechanisms of pol II pausing: heterochromatin- and R-loop-
463 dependent pausing (PROUDFOOT 2016). It has been reported previously that the
464 heterochromatin structure downstream of the TTS causes pol II pausing by acting as an
465 obstacle (PROUDFOOT 2016). Therefore, hypomethylation in the TTS region may lead
466 to the suppression of heterochromatin formation downstream of TTS. An alternative

467 possibility is R-loop-dependent pausing; R-loops are known as three-stranded
468 structures, which consist of RNA-DNA hybrids, and function in various nuclear
469 processes, such as transcription, DNA replication, and DNA repair (AL-HADID AND
470 YANG 2016; NIEHRS AND LUKE 2020). As recent studies have shown that pol II
471 pausing is induced by R-loop formation at the TTS, it is worth analyzing the
472 relationship between R-loop formation and DNA methylation. Future molecular
473 biological analyses regarding hypomethylation of the TTS region are needed to improve
474 our understanding of the molecular events underlying transcription termination.
475

Acknowledgements

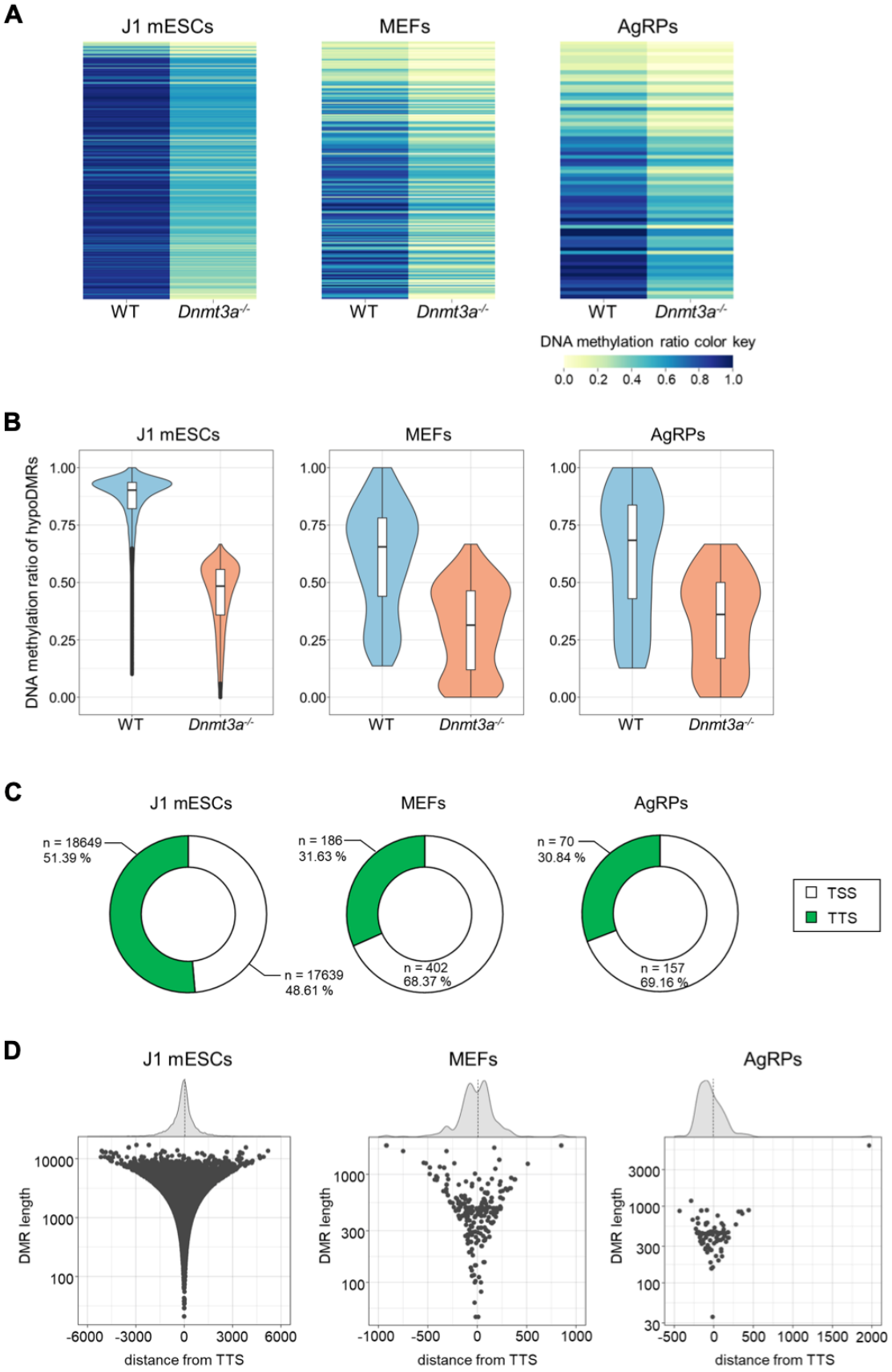
We thank our laboratory members and Shyunya Hozumi for helpful discussion and critical comments; Masato Kinoshita for providing plasmids of CRISPR/Cas9 system; Mingcong Xu and Ikumi Taya for technical assistant. Computations were partially performed on the NIG supercomputer at ROIS National Institute of Genetics.

Funding

This work was supported by JSPS KAKENHI Grant Numbers JP18K06185 and JP21K06123 to Y.K.

Conflict of interest

The authors declare that they have no conflicts of interest with the contents of this article.


A
J1 mESCs
MEFs
AgRPs
WT
Dnmt3a-/-
DNA methylation ratio color key
0.0 0.2 0.4 0.6 0.8 1.0
B
DNA methylation ratio of hypoDMRs
1.00
0.75
0.50
0.25
0.00
C
n = 18649
51.39 %
n = 17639
48.61 %
n = 186
31.63 %
n = 402
68.37 %
n = 70
30.84 %
n = 157
69.16 %
TSS
TTS
D
DMR length
distance from TTS

489

490 **Figure 1. Characterization of hypoDMRs in three *Dnmt3a*-deficient mouse cells.**

491 (A) Heatmaps of DNA methylation profiles of hypoDMRs overlapping TTSs

492 comparing WT with *Dnmt3a*$^{-/-}$ mutants (J1 mESCs, MEFs, and AgRP). Color scale

493 indicates DNA methylation ratio. (B) Violin plots showing DNA methylation ratio of

494 the hypoDMRs overlapping the TTSs comparing WT with *Dnmt3a*$^{-/-}$ mutants (J1

495 mESCs, MEFs, and AgRP). (C) Pie charts depicting the number and percentage of

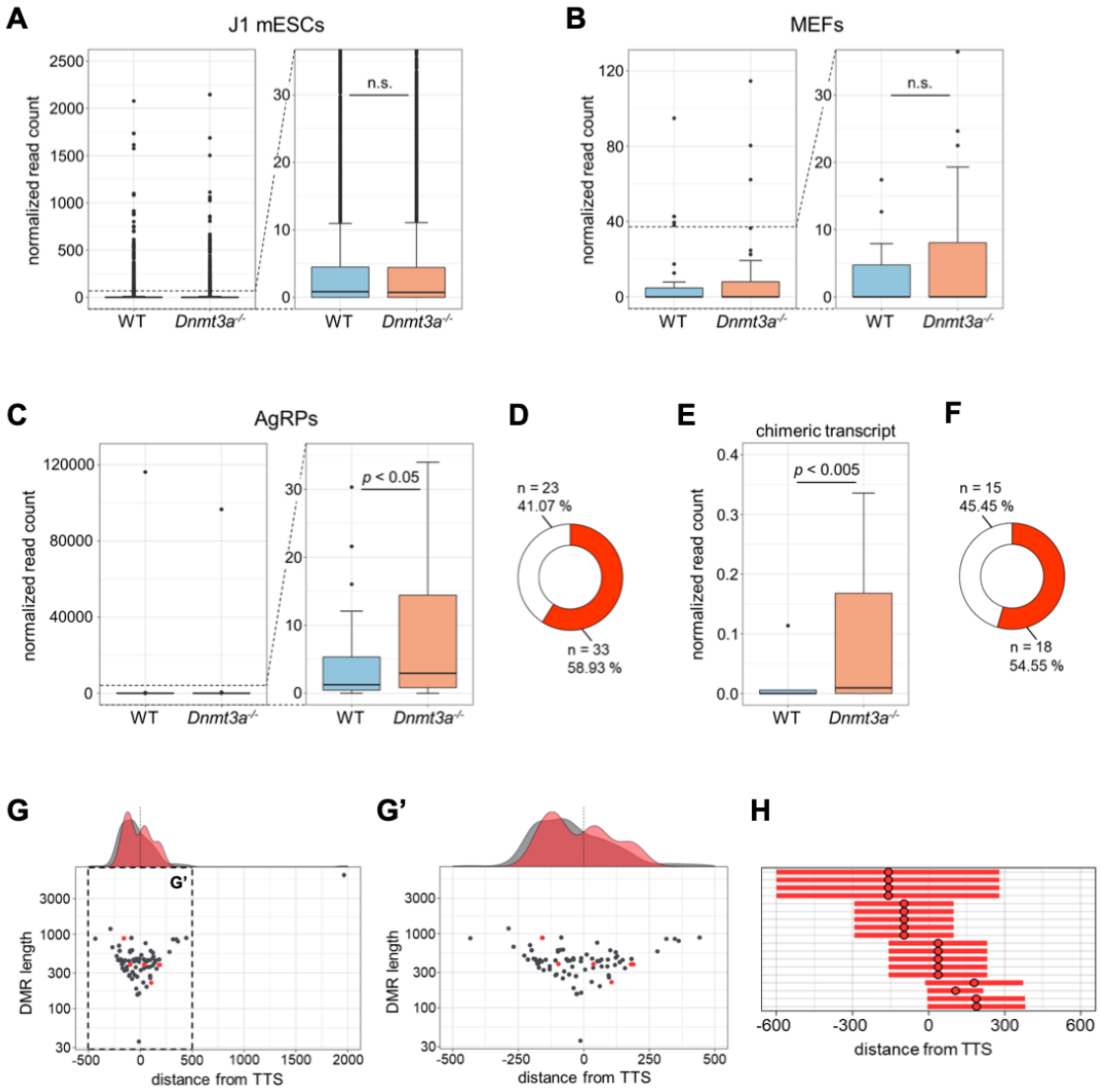
496 hypoDMRs overlapping the TSSs and TTSs in *Dnmt3a*$^{-/-}$ J1 mESCs, MEFs, and AgRPs.

497 (D) 2D scatterplots with marginal distribution comparing hypoDMR length versus

498 distance from TTS in *Dnmt3a*$^{-/-}$ J1 mESCs, MEFs, and AgRPs. Dots represent the

499 center of hypoDMRs.

500

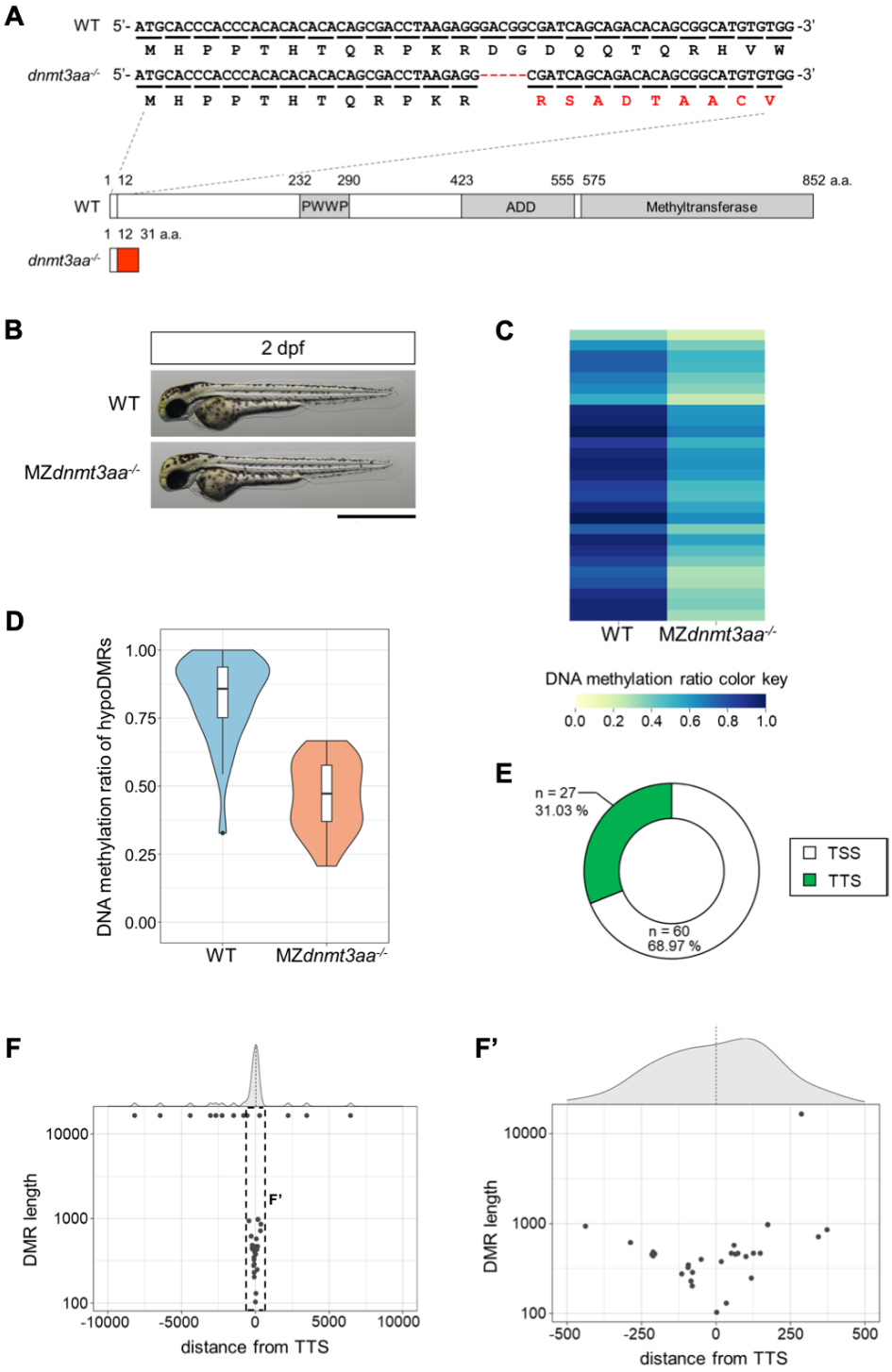

A
J1 mESCs
normalized read count
n.s.
WT
Dnmt3a-/-
B
MEFs
C
AgRPs
p < 0.05
D
n = 23
41.07 %
n = 33
58.93 %
E
chimeric transcript
p < 0.005
F
n = 15
45.45 %
n = 18
54.55 %
G
G'
DMR length
distance from TTS
H

501

502

503

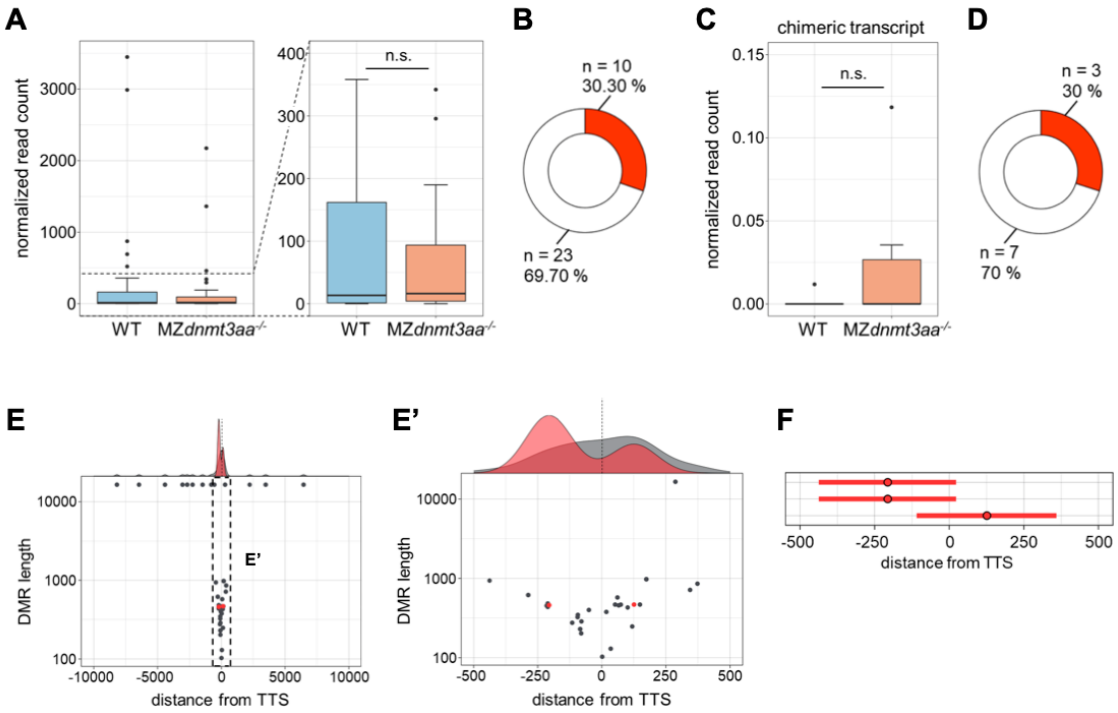
504 **Figure 2. Metagene analysis of transcripts downstream of TTSs in three *Dnmt3a*-**
505 **deficient mouse cells.**
506 (A-C) Box plots showing the normalized read counts in WT and *Dnmt3a*$^{-/-}$ mutants (J1
507 mESCs, MEFs, and AgRPs). Right graphs are enlarged views of the left ones,
508 respectively. (D) A pie chart depicting the number and percentage of transcripts with
509 increased read counts (red) or decreased/unchanged read counts (white) in *Dnmt3a*$^{-/-}$
510 AgRPs. (E) A box plot showed chimeric transcripts in WT and *Dnmt3a*$^{-/-}$ AgRPs.
511 (F) A pie chart depicting the number and percentage of transcripts with increased
512 chimeric transcripts (red) or decreased/unchanged chimeric transcripts (white) in
513 *Dnmt3a*$^{-/-}$ AgRPs. (G, G') 2D scatterplots with marginal distribution comparing
514 hypoDMR length versus distance from TTS in *Dnmt3a*$^{-/-}$ AgRPs. Grey and red dots
515 represent the center of hypoDMRs. Red dots represent hypoDMRs, which lead to
516 increased chimeric transcripts. A right graph (G') is an enlarged view of the left one
517 (G). (H) A horizontal bar graph represents the distance from TTS of hypoDMRs,
518 which lead to increased chimeric transcripts, relative to TTS. Red dots and bars
519 represent the center and range of hypoDMRs relative to TTS position, respectively.
520


A
WT 5'- ATGCACCCACCCACACACACACAGCGACCTAAGAGGGACGGCGATCAGCAGACACAGCGGCATGTGTGG -3'
M H P P T H T Q R P K R D G D Q Q T Q R H V W
dnmt3aa-/- 5'- ATGCACCCACCCACACACACACAGCGACCTAAGAGG-----CGATCAGCAGACACAGCGGCATGTGTGG -3'
M H P P T H T Q R P K R R S A D T A A C V
1 12 232 290 423 555 575 852 a.a.
WT
PWWP
ADD
Methyltransferase
1 12 31 a.a.
dnmt3aa-/-
B
2 dpf
WT
MZdnmt3aa-/-
C
WT MZdnmt3aa-/-
DNA methylation ratio color key
0.0 0.2 0.4 0.6 0.8 1.0
D
DNA methylation ratio of hypoDMRs
1.00
0.75
0.50
0.25
0.00
WT MZdnmt3aa-/-
E
n = 27
31.03 %
n = 60
68.97 %
TSS
TTS
F
DMR length
10000
1000
100
-10000 -5000 0 5000 10000
distance from TTS
F'
F'
DMR length
10000
1000
100
-500 -250 0 250 500
distance from TTS

521

522 **Figure 3. Generation of MZ*dnmt3aa*$^{-/-}$ zebrafish mutant and characterization of**
523 **hypoDMRs.**

524 (A) The *dnmt3aa* genomic DNA sequence and CRISPR/Cas9-mediated indel mutation,
525 with predicted amino acid sequences, in WT and MZ*dnmt3aa*$^{-/-}$ mutant are shown.
526 The MZ*dnmt3aa*$^{-/-}$ mutant has a 5-bp deletion which leads to a frame-shift at amino acid
527 13 position. Schematic diagram shows predicted protein products from the WT and the
528 mutant allele. It is speculated that the deletion mutation results in the production of a
529 truncated protein with additional sequences (red box) at the C-termini. Putative
530 functional domains (PWWP, cysteine-rich ATRX-Dnmt3-Dnmt3L (ADD), and
531 methyltransferase domains) in the WT protein are represented as grey colored boxes.
532 (B) Gross morphology of the MZ*dnmt3aa*$^{-/-}$ mutant is identical to WT at 2 dpf. Scale
533 bar: 1 mm. (C) A heatmap of DNA methylation profiles of hypoDMRs overlapping
534 TTSs comparing WT with MZ*dnmt3aa*$^{-/-}$ mutant. Color scale indicates DNA
535 methylation ratio. (D) A violin plot showing DNA methylation ratio of the
536 hypoDMRs overlapping the TTSs comparing WT with MZ*dnmt3aa*$^{-/-}$ mutant. (E) A
537 pie chart depicting the number and percentage of hypoDMRs overlapping the TSSs and
538 TTSs in MZ*dnmt3aa*$^{-/-}$ mutant. (F, F') 2D scatterplots with marginal distribution
539 comparing hypoDMR length versus distance from TTS in MZ*dnmt3aa*$^{-/-}$ mutant. The
540 right graph (F') is an enlarged view of the left one (F). Grey dots represent the center
541 of hypoDMRs.


A
normalized read count
3000
2000
1000
0
400
300
200
100
0
n.s.
WT
MZdnmt3aa-/-
WT
MZdnmt3aa-/-
B
n = 10
30.30 %
n = 23
69.70 %
C
chimeric transcript
0.15
0.10
0.05
0.00
normalized read count
n.s.
WT
MZdnmt3aa-/-
D
n = 3
30 %
n = 7
70 %
E
DMR length
10000
1000
100
E'
-10000
-5000
0
5000
10000
distance from TTS
E'
DMR length
10000
1000
100
-500
-250
0
250
500
distance from TTS
F
-500
-250
0
250
500
distance from TTS

542

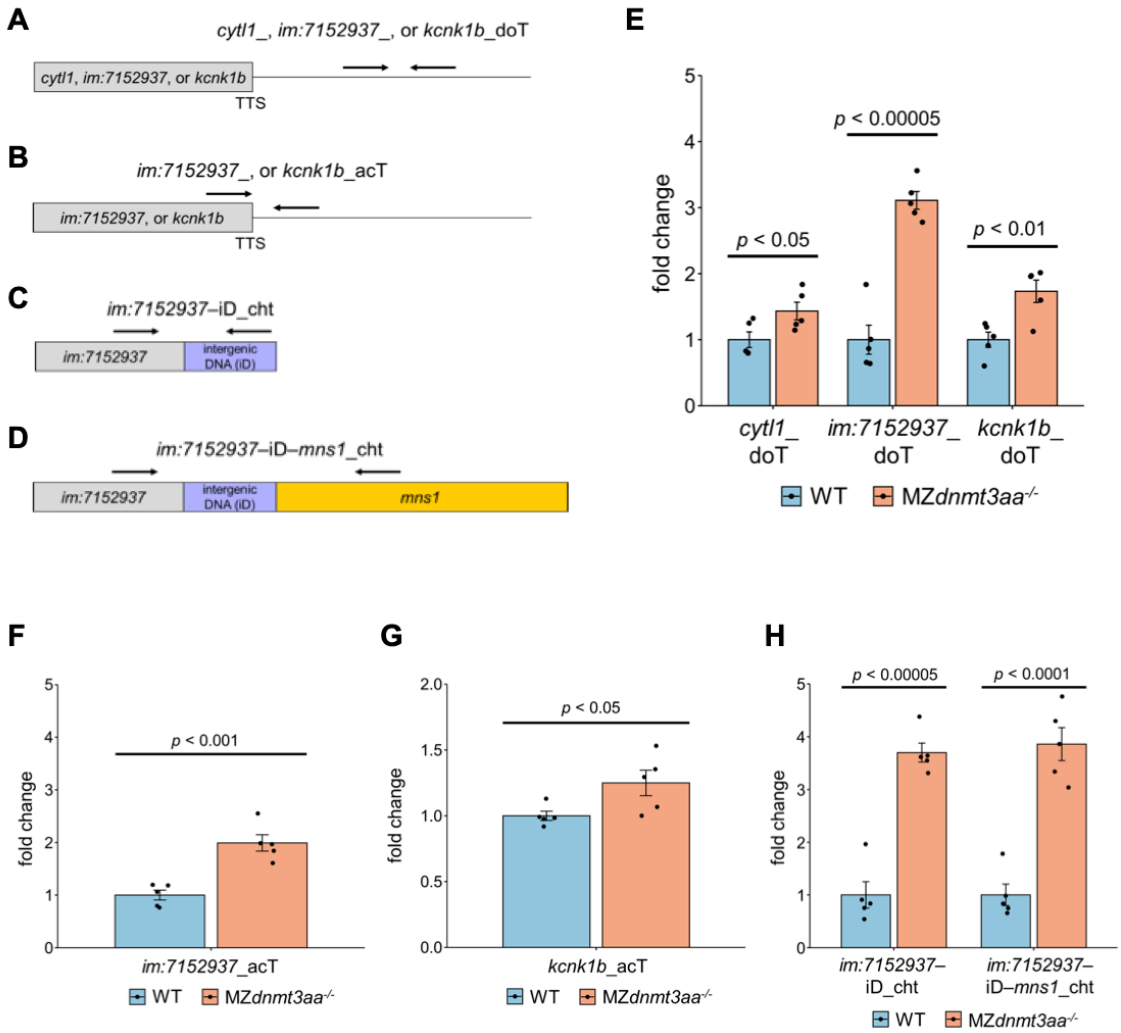
543

544

545 **Figure 4. Metagene analysis of transcripts downstream of TTSs in MZ*dnmt3aa*$^{-/-}$**
546 **mutant.**

547 (A) A box plot showing the normalized read counts in WT and MZ*dnmt3aa*$^{-/-}$ mutant.
548 The right graph is an enlarged view of the left one. (B) A pie chart depicting the
549 number and percentage of transcripts with increased read counts (red) or
550 decreased/unchanged read counts (white) in MZ*dnmt3aa*$^{-/-}$ mutant. (C) A box plot
551 showing chimeric transcripts in WT and MZ*dnmt3aa*$^{-/-}$ mutant. (D) A pie chart
552 depicting the number and percentage of transcripts with increased chimeric transcripts
553 (red) or decreased/unchanged chimeric transcripts (white) in MZ*dnmt3aa*$^{-/-}$ mutant. (E,
554 E') 2D scatterplots with marginal distribution comparing hypoDMR length versus
555 distance from TTS in MZ*dnmt3aa*$^{-/-}$ mutant. Grey and red dots represent the center of
556 hypoDMRs. Red dots represent hypoDMRs, which lead to increased chimeric
557 transcripts. The right graph (E') is an enlarged view of the left one (E). (F) A
558 horizontal bar graph represents the distance from TTS of hypoDMRs, which lead to
559 increased chimeric transcripts, relative to TTS. Red dots and bars represent the center
560 and range of hypoDMRs relative to TTS position, respectively.

561


A
cytl1_, im:7152937_, or kcnk1b_doT
cytl1, im:7152937, or kcnk1b
TTS
B
im:7152937_, or kcnk1b_acT
im:7152937, or kcnk1b
TTS
C
im:7152937–iD_cht
im:7152937
intergenic DNA (iD)
D
im:7152937–iD–mns1_cht
im:7152937
intergenic DNA (iD)
mns1
E
fold change
p < 0.05
p < 0.00005
p < 0.01
cytl1_ doT
im:7152937_ doT
kcnk1b_ doT
WT
MZdnmt3aa-/-
F
fold change
p < 0.001
im:7152937_acT
WT
MZdnmt3aa-/-
G
fold change
p < 0.05
kcnk1b_acT
WT
MZdnmt3aa-/-
H
fold change
p < 0.00005
p < 0.0001
im:7152937– iD_cht
im:7152937– iD–mns1_cht
WT
MZdnmt3aa-/-

562

563

564

565 **Figure 5. Identification of read-through and chimeric transcripts in MZ*dnmt3aa*$^{-/-}$**

566 **mutant.**

567 (A-D) Scheme of primer pair positions to detect aberrant transcripts. (E) qRT-PCR

568 analysis for transcripts downstream of TTS (*cytl1*_doT, *im:7152937*_doT, and *kcnk1b*_

569 doT) in *cytl1*, *im:7152937*, and *kcnk1b* genes. (F, G) qRT-PCR analysis for transcripts

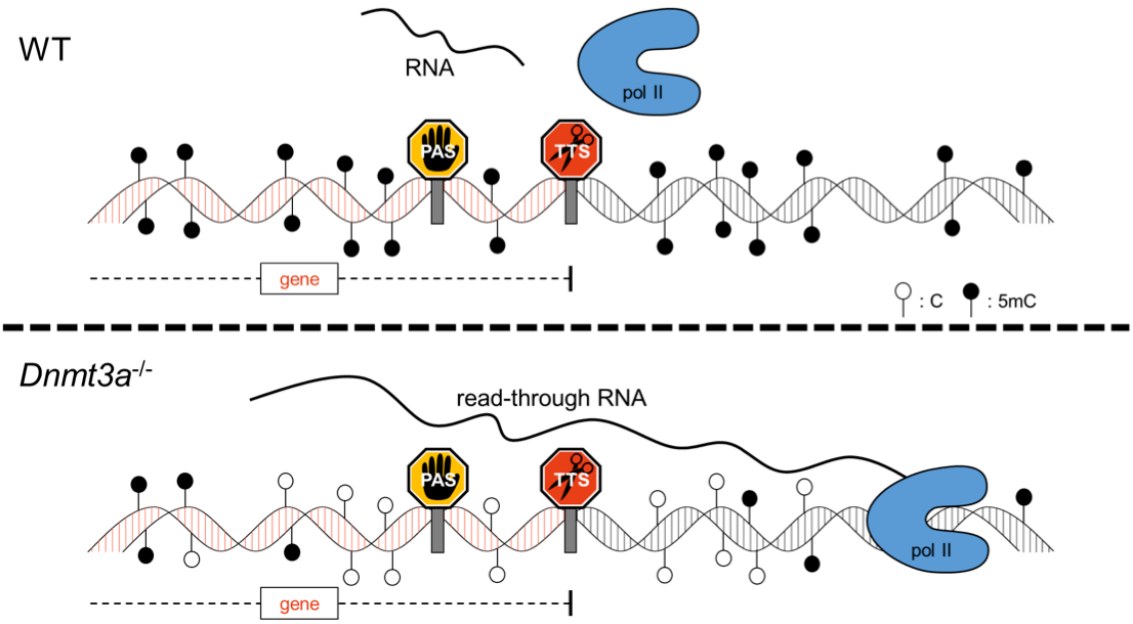
570 across the TTSs (*im:7152937*_acT and *kcnk1b*_acT) in *im:7152937* (F) and *kcnk1b* (G)

571 genes. (H) qRT-PCR analysis for chimeric transcripts between exons in *im:7152937*

572 and iD/exon1 in *mns1* (*im:7152937*-iD_cht and *im:7152937-mns1*_cht). All qRT-PCR

573 data are presented as means ± SE. The mean expression level of WT is defined as 1.

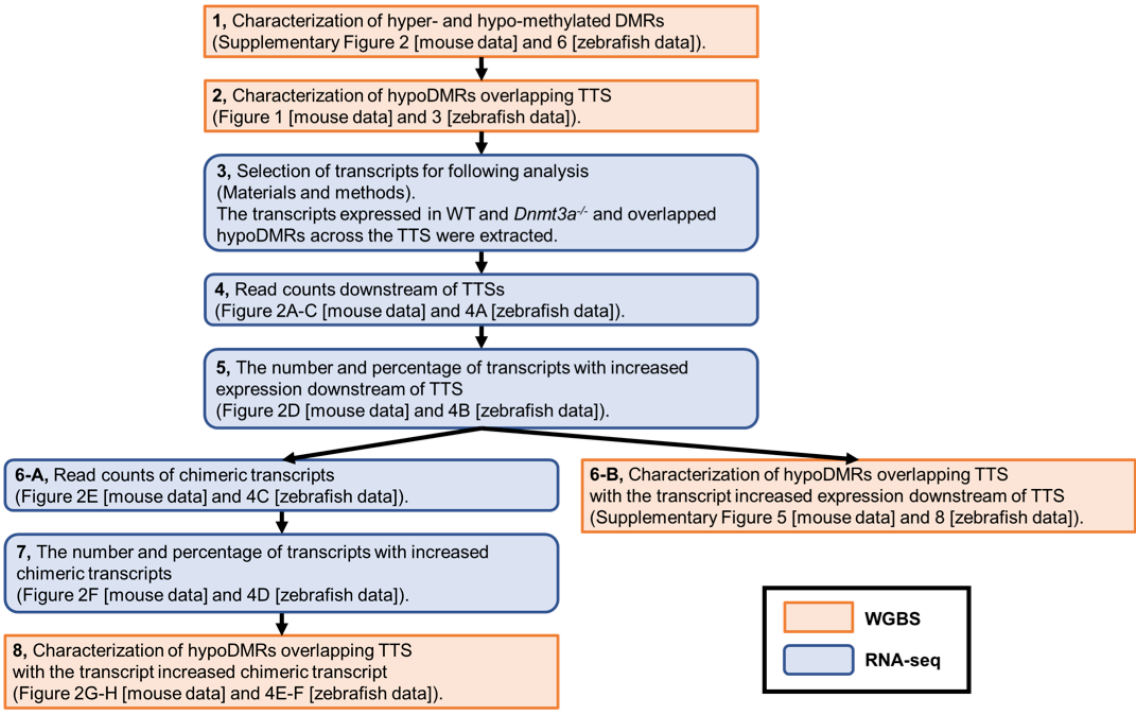
574



575

576 **Figure 6. A proposed model of transcription termination.**

577

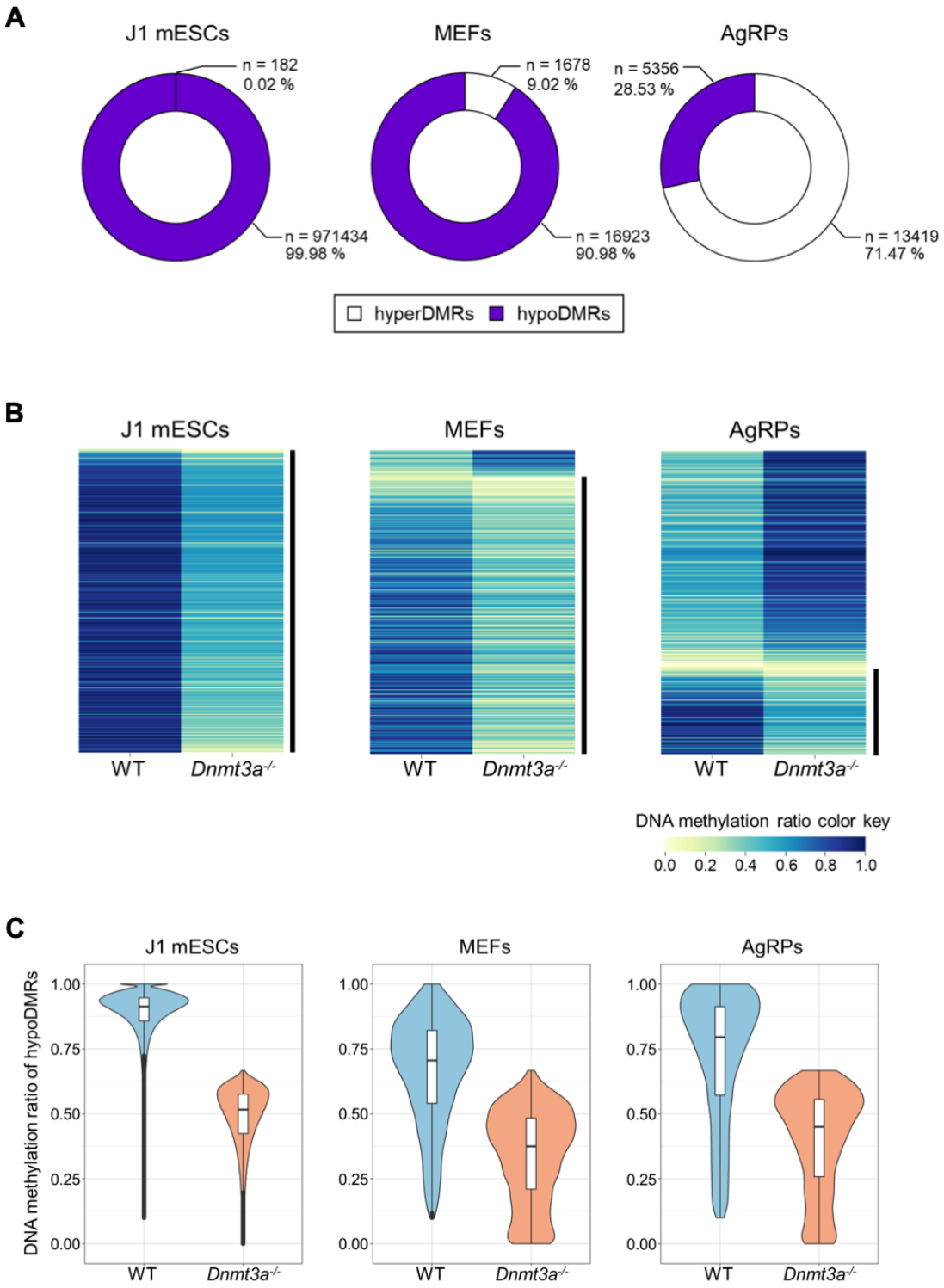


578

579 **Supplementary Figure S1. Flow chart of bioinformatics analysis and experimental**

580 **design.**

581


A
J1 mESCs
n = 182
0.02 %
n = 971434
99.98 %
MEFs
n = 1678
9.02 %
n = 16923
90.98 %
AgRPs
n = 5356
28.53 %
n = 13419
71.47 %
hyperDMRs
hypoDMRs
B
J1 mESCs
MEFs
AgRPs
WT
Dnmt3a-/-
DNA methylation ratio color key
0.0 0.2 0.4 0.6 0.8 1.0
C
J1 mESCs
MEFs
AgRPs
DNA methylation ratio of hypoDMRs
1.00
0.75
0.50
0.25
0.00
WT
Dnmt3a-/-

582

583

584

585 **Supplementary Figure S2. Characterization of hyperDMRs and hypoDMRs in**

586 **three *Dnmt3a*-deficient mouse cells.**

587 (A) Pie charts depicting number and percentage of the hyperDMRs and hypoDMRs

588 comparing WT with *Dnmt3a*$^{-/-}$ mutants (J1 mESCs, MEFs, and AgRPs). (B) Heatmaps

589 of DNA methylation profiles of the DMRs comparing WT with *Dnmt3a*$^{-/-}$ mutants (J1

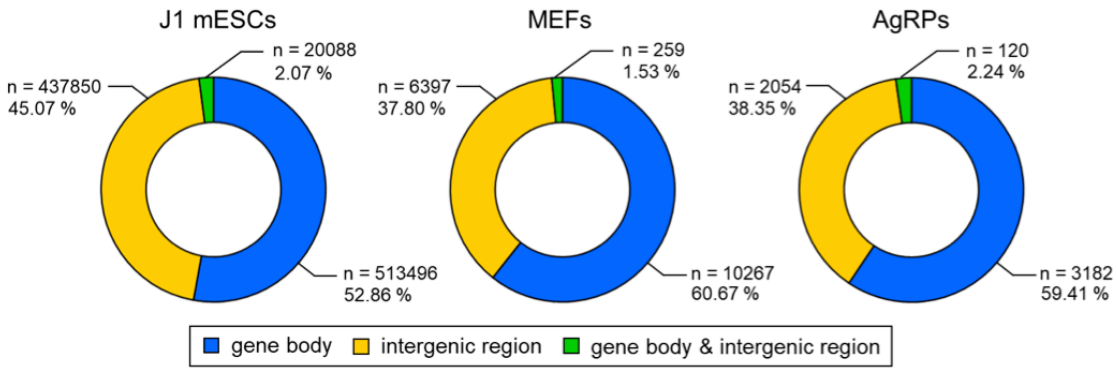
590 mESCs, MEFs, and AgRPs). Color scale indicates DNA methylation ratio.

591 HypoDMRs are shown by vertical black bars. (C) Violin plots showing DNA

592 methylation ratio of the hypoDMRs comparing WT with *Dnmt3a*$^{-/-}$ mutants (J1 mESCs,

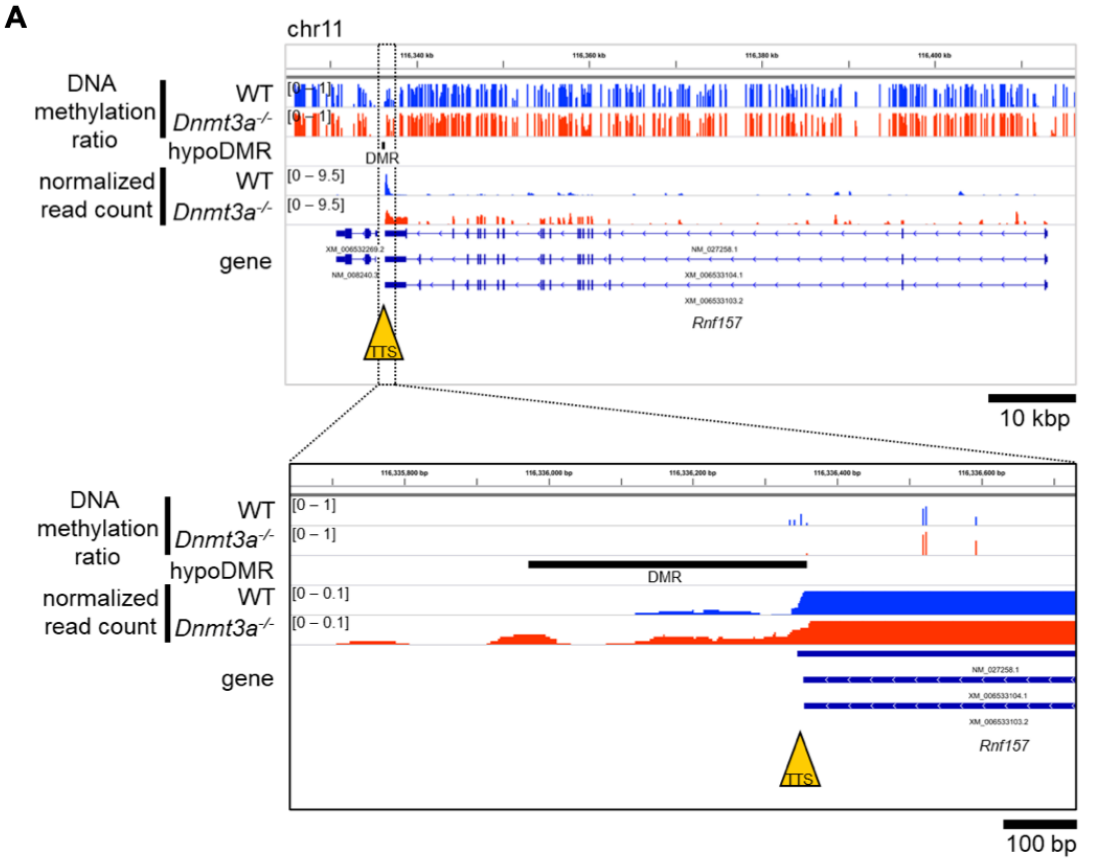
593 MEFs, and AgRPs).

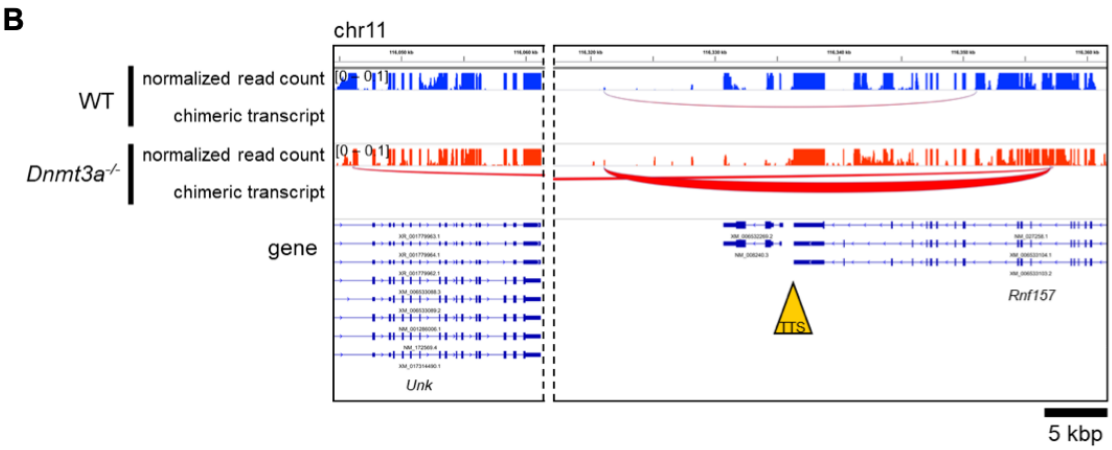
594



Supplementary Figure S3. Genomic position of hypoDMRs in three *Dnmt3a* deficient mouse cells.

Pie charts depicting number and percentage of genomic position (gene body, intergenic region, and gene body & intergenic region) of the hypoDMRs comparing WT with *Dnmt3a*$^{-/-}$ mutants (J1 mESCs, MEFs, and AgRP).


A
chr11
DNA methylation ratio
WT
Dnmt3a-/-
hypoDMR
DMR
normalized read count
WT
Dnmt3a-/-
[0 – 9.5]
gene
Rnf157
TTS
10 kbp
DNA methylation ratio
WT
Dnmt3a-/-
[0 – 1]
hypoDMR
DMR
normalized read count
WT
Dnmt3a-/-
[0 – 0.1]
gene
Rnf157
TTS
100 bp


B
chr11
WT
normalized read count
chimeric transcript
Dnmt3a-/-
normalized read count
chimeric transcript
[0 – 0.1]
gene
Unk
Rnf157
TTS
5 kbp

602

603

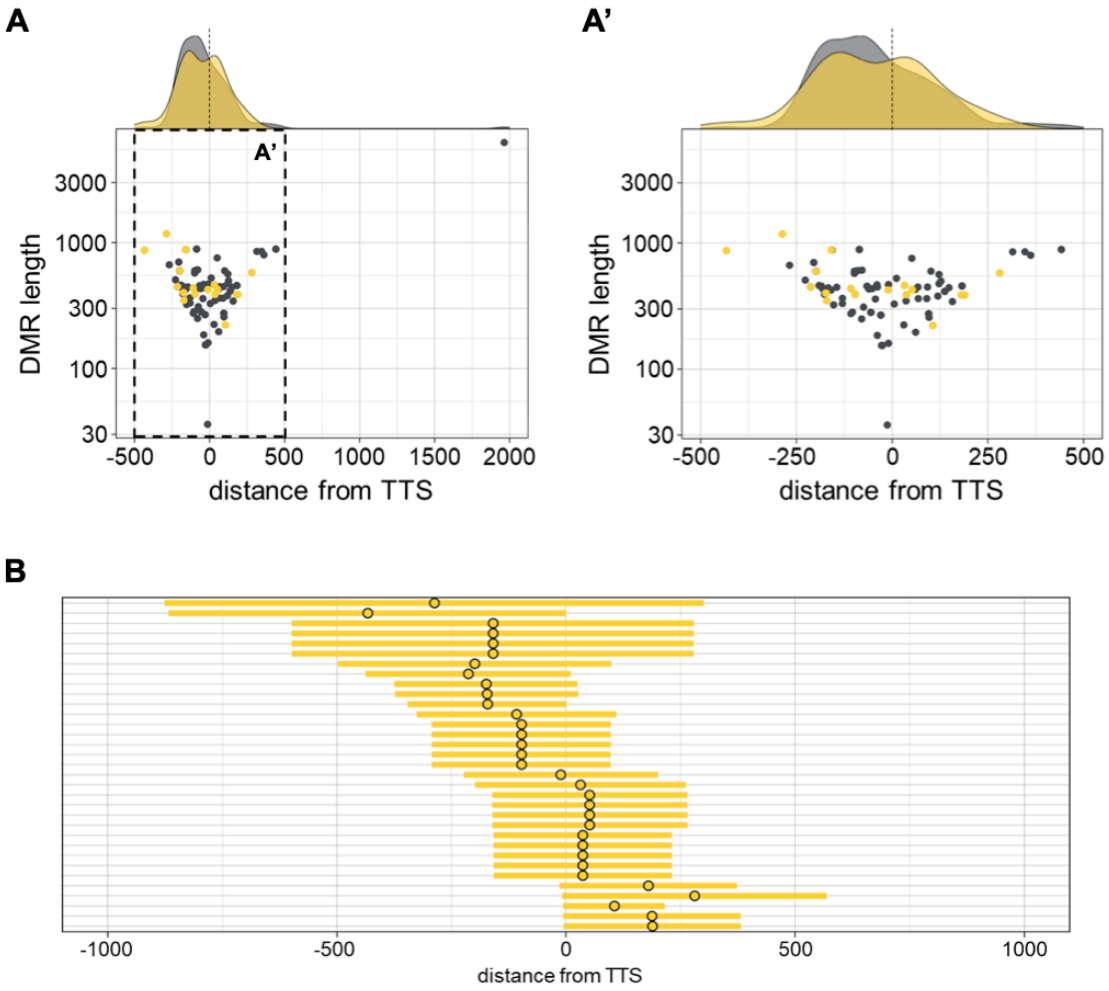
604 **Supplementary Figure S4. Transcription termination defects of *Rnf157* gene in**

605 ***Dnmt3a*$^{-/-}$ AgRPs.**

606 (A, B) Genome browser view for *Rnf157* gene in *Dnmt3a*$^{-/-}$ AgRPs. Normalized read

607 counts (A) and chimeric transcripts (B), as an example, are shown in this Figure.

608


A
A'
DMR length
3000
1000
300
100
30
-500
0
500
1000
1500
2000
distance from TTS
A'
DMR length
3000
1000
300
100
30
-500
-250
0
250
500
distance from TTS
B
-1000
-500
0
500
1000
distance from TTS

609

610

611 **Supplementary Figure S5. Distribution of hypoDMRs which lead to increased read**

612 **counts downstream of TTSs in *Dnmt3a*$^{-/-}$ AgRPs.**

613 (A, A') 2D scatterplots with marginal distribution comparing hypoDMR length versus

614 distance from TTS in *Dnmt3a*$^{-/-}$ AgRPs. Grey and yellow dots represent the center of

615 hypoDMRs. Yellow dots represent the hypoDMRs which lead to increased read

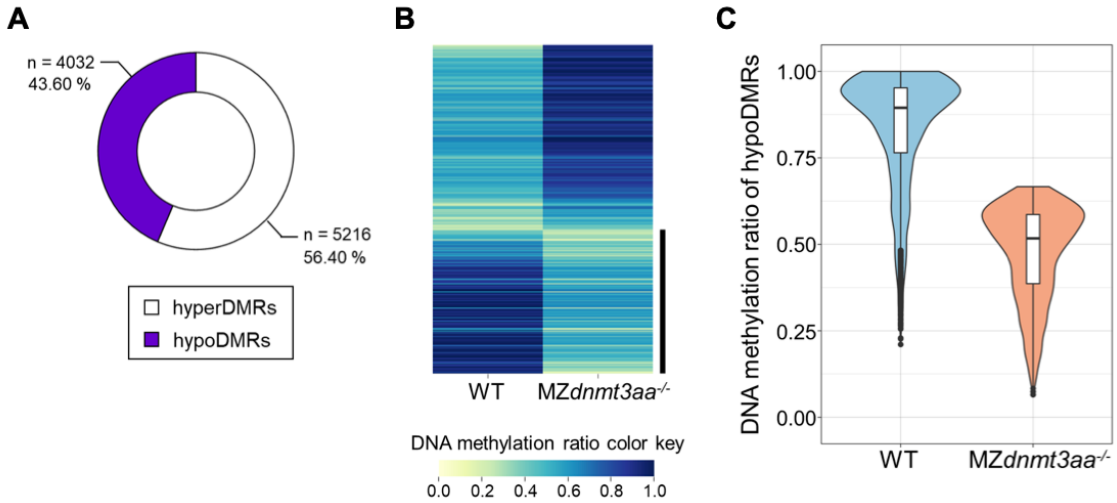
616 counts downstream of TTSs. A right graph (A') is an enlarged view of the left one

617 (A). (B) A horizontal bar graph representing the distribution of hypoDMRs, which

618 lead to increased read counts, relative to TTS. Yellow dots and bars represent the

619 center and range of hypoDMRs, respectively.

620



621

622 **Supplementary Figure S6. Characterization of hyperDMRs and hypoDMRs in**

623 **MZ*dnmt3aa*$^{-/-}$ mutant.**

624 (A) A pie chart depicting number and percentage of the hyperDMRs and hypoDMRs

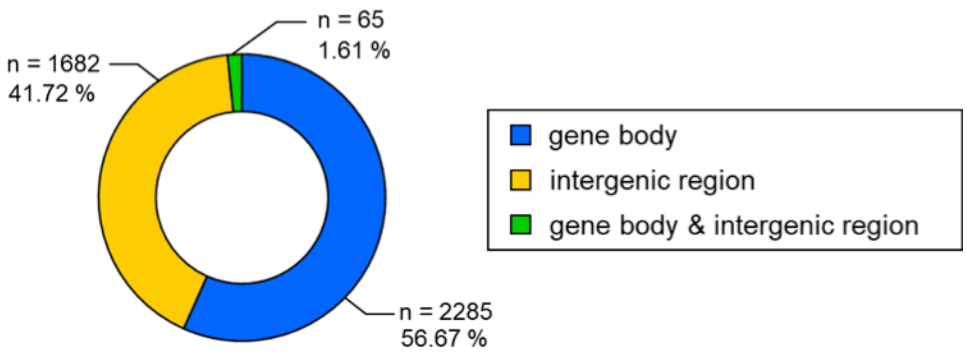
625 comparing WT with MZ*dnmt3aa*$^{-/-}$ mutant. (B) A heatmap of DNA methylation

626 profiles of the DMRs comparing WT with MZ*dnmt3aa*$^{-/-}$ mutant. (C) A violin plot

627 showing DNA methylation ratio of the hypomethylated DMRs comparing WT with

628 MZ*dnmt3aa*$^{-/-}$ mutant.

629



634

635 **Supplementary Figure S7. Genomic position of hypoDMRs in MZ*dnmt3aa*$^{-/-}$**

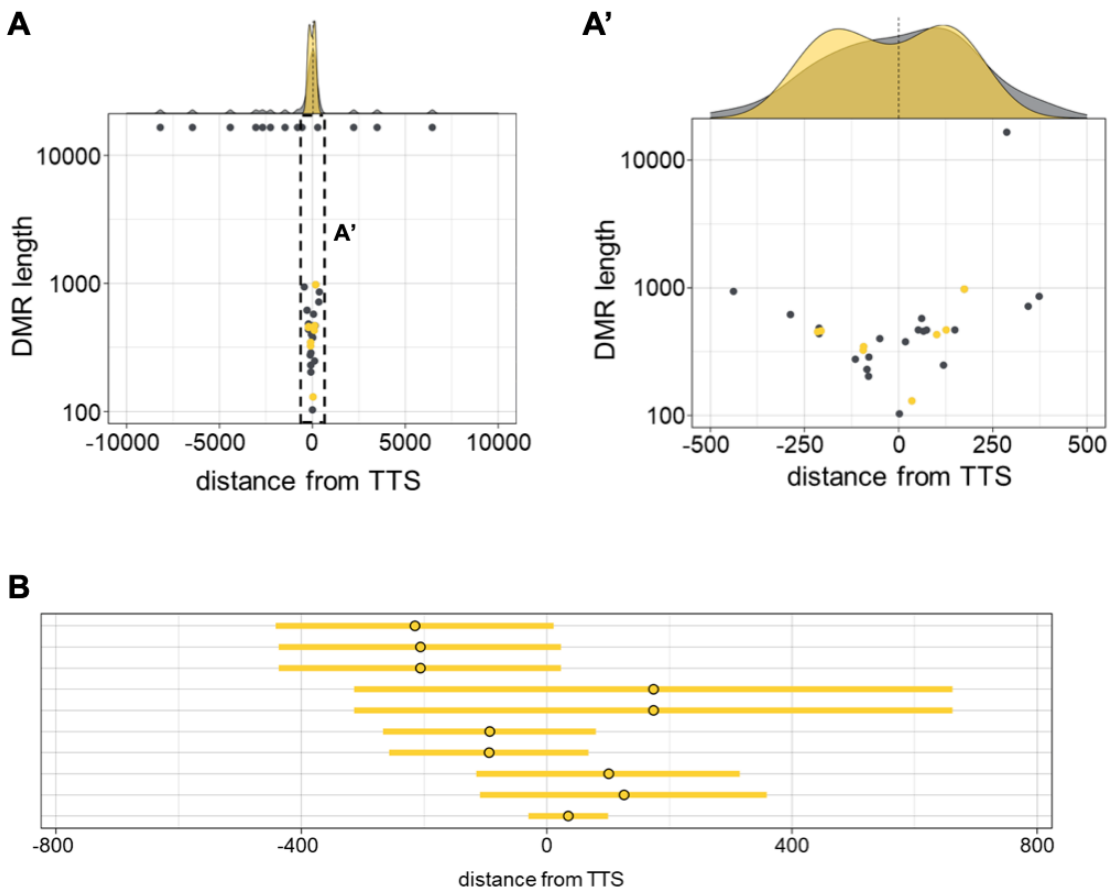
636 **mutant.**

637 A pie chart depicting number and percentage of genomic position (gene body,

638 intergenic region, and gene body & intergenic region) of the hypoDMRs comparing WT

639 with MZ*dnmt3aa*$^{-/-}$ mutant.

640


A
A'
DMR length
10000
1000
100
-10000
-5000
0
5000
10000
distance from TTS
A'
DMR length
10000
1000
100
-500
-250
0
250
500
distance from TTS
B
-800
-400
0
400
800
distance from TTS

641

642

643 **Supplementary Figure S8. Distribution of hypoDMRs which lead to increased read**

644 **counts downstream of TTSs in MZ*dnmt3aa*$^{-/-}$ mutant.**

645 (A, A') 2D scatterplots with marginal distribution comparing hypoDMR length versus

646 distance from TTS in MZ*dnmt3aa*$^{-/-}$ mutant. Grey and yellow dots represent the center

647 of hypoDMRs. Yellow dots represent hypoDMRs which lead to increased read counts

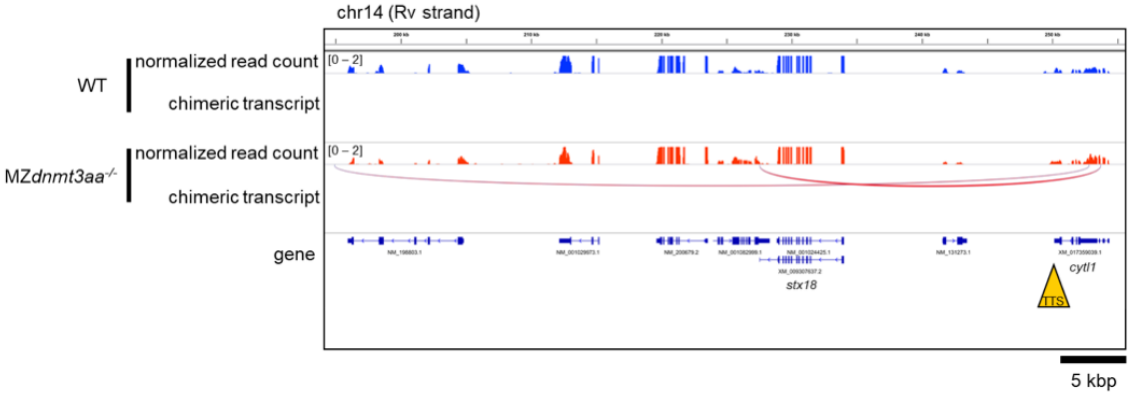
648 downstream of TTSs. A right graph (A') is an enlarged view of the left one (A). (B)

649 A horizontal bar graph represents the distribution of hypoDMRs, which lead to

650 increased read counts, relative to TTS. Yellow dots and bars represent the center and

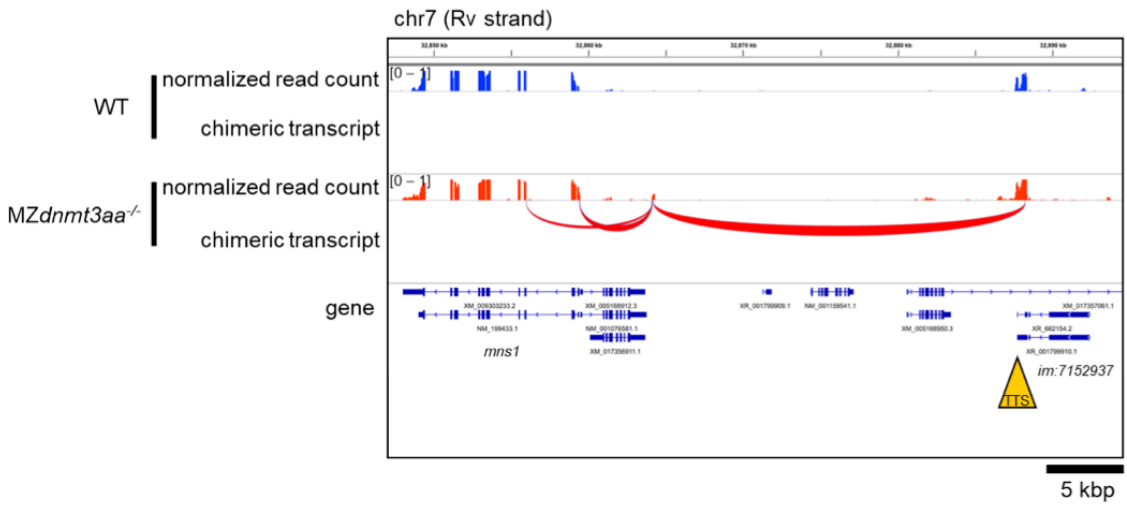
651 range of hypoDMRs, respectively.

652



Supplementary Figure S9. Transcription termination defects of *cytl1* gene in MZ*dnmt3aa*$^{-/-}$ mutant.

Genome browser view for *cytl1* gene in MZ*dnmt3aa*$^{-/-}$ mutant. Chimeric transcripts, as an example, are shown in this Figure.



659

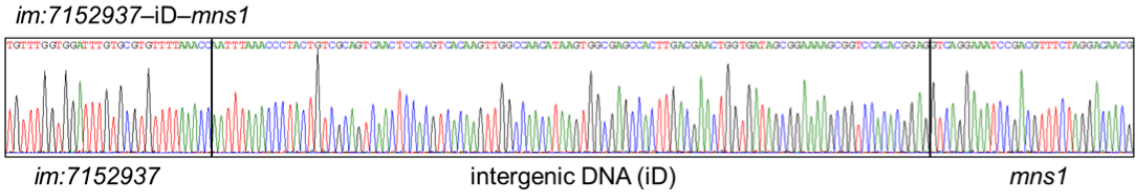
660 **Supplementary Figure S10. Transcription termination defects *im:7152937* in**

661 **MZ*dnmt3aa*$^{-/-}$ mutant.**

662 Genome browser view for *im:7152937* in MZ*dnmt3aa*$^{-/-}$ mutant. Chimeric transcripts,

663 as an example, are shown in this Figure.

664



665

666 **Figure S11. A sequencing chromatogram of chimeric transcripts, *im:7152937*-iD-**

667 ***mns1*.**

668

669 **Supplementary Table S1**

sgRNA ID	oligonucleotide sequence (5' to 3')	gRNA target site (5' to 3')
No.1	TAGGACCCACACACACACAGCG AAACCGCTGTGTGTGTGTGGGT	GTCGCTGTGTGTGTGTGGGTGGG
No.2	TAGGACAGCGACCTAAGAGGGA AAACTCCCTCTTAGGTCGCTGT	ACACAGCGACCTAAGAGGGACGG
No.3	TAGGTTGGATATGGCATTGGAG AAACCTCCAATGCCATATCCAA	GCTCTCCAATGCCATATCCAAGG

670

671

672

673 **Supplementary Table S2**

primer sequence (5' to 3')	objective
Fw: GGTGACTTCACGGATACTACGTC **Rv:** CACATACCCACACATGCCGCTGT	target amplification for sequencing
GTTCAGAGCTGTACGTCTTCCTG	sequencing
Fw: TCAGCATCGAGTCATGTGATGCC **Rv:** CACATACCCACACATGCCGCTGT	genotyping (HMA)

674

675

676 **Supplementary Table S3**

target	primer position	primer sequence (5' to 3')	reference
cytl1	doT	**Fw:** ATCTGCATCTCCGTCTCATGAC **Rv:** AGAAATCCTCCAGTTCGCTCTG	
im:7152937	doT	**Fw:** AACAACGCGATGAAGGGATG **Rv:** GCGATCAGTTGACGTATTGACC	
kcnk1b	doT	**Fw:** TTAATGGCTGGTTCCATGCG **Rv:** CCACAAAGGCTCTAAATGCAAAC	
im:7152937	acT	**Fw:** CACATCTCTATTCACCATTGTGG **Rv:** ATGATCACTGTGGTGCATCG	
kcnk1b	acT	**Fw:** ACTTTAGCCGAGGGTCTTGAAG **Rv:** TCATCTCAATCGGCAAGTGC	
im:7152937–iD	cht	**Fw:** CACGGGCAGAGAAGTTGTG **Rv:** CGCTATCACCAGTTCGTCAA	
im:7152937–iD–*mns1* (exon 1)	cht	**Fw:** CACGGGCAGAGAAGTTGTG **Rv:** TAATACACGGCGCTCAGTTG	
ube2a	mRNA	**Fw:** TGACTGTTGACCCACCTTACAG **Rv:** CAAATAAAAGCAAGTAACCCCG	1

Reference

1. Xu,H., C. Li, Q. Zeng, I. Agrawal, X. Zhu and Z. Gong, 2016 Genome-wide identification of suitable zebrafish Danio rerio reference genes for normalization of gene expression data by RT-qPCR. J Fish Biol 88: 2095–2110.

677

678

679

680 **Supplementary Table S4**

target	primer sequence (5' to 3')	objective
im:7152937–iD–*mns1* (exon 1)	**Fw:** CACGGGCAGAGAAGTTGTG **Rv:** TAATACACGGCGCTCAGTTG	target amplification for sequencing
vector	GGAAACAGCTATGACCATGA	sequencing

681

682

683 **References**

684 Al-Hadid, Q., and Y. Yang, 2016 R-loop: an emerging regulator of chromatin
685 dynamics. Acta Biochim Biophys Sin (Shanghai) 48**:** 623-631.
686 Ansai, S., and M. Kinoshita, 2014 Targeted mutagenesis using CRISPR/Cas system in
687 medaka. Biol Open 3**:** 362-371.
688 Ball, M. P., J. B. Li, Y. Gao, J. H. Lee, E. M. LeProust *et al*., 2009 Targeted and
689 genome-scale strategies reveal gene-body methylation signatures in human cells. Nat
690 Biotechnol 27**:** 361-368.
691 Chen, F. X., E. R. Smith and A. Shilatifard, 2018 Born to run: control of transcription
692 elongation by RNA polymerase II. Nat Rev Mol Cell Biol 19**:** 464-478.
693 Eaton, J. D., L. Francis, L. Davidson and S. West, 2020 A unified allosteric/torpedo
694 mechanism for transcriptional termination on human protein-coding genes. Genes Dev
695 34**:** 132-145.
696 Eaton, J. D., and S. West, 2020 Termination of Transcription by RNA Polymerase II:
697 BOOM! Trends Genet 36**:** 664-675.
698 Gates, L. A., C. E. Foulds and B. W. O'Malley, 2017 Histone Marks in the 'Driver's
699 Seat': Functional Roles in Steering the Transcription Cycle. Trends Biochem Sci 42**:**
700 977-989.
701 Greenberg, M. V. C., and D. Bourc'his, 2019 The diverse roles of DNA methylation in
702 mammalian development and disease. Nat Rev Mol Cell Biol 20**:** 590-607.
703 Gu, T., X. Lin, S. M. Cullen, M. Luo, M. Jeong *et al*., 2018 DNMT3A and TET1
704 cooperate to regulate promoter epigenetic landscapes in mouse embryonic stem cells.
705 Genome Biol 19**:** 88.
706 Jones, P. A., 2012 Functions of DNA methylation: islands, start sites, gene bodies and
707 beyond. Nat Rev Genet 13**:** 484-492.
708 Kim, D., B. Langmead and S. L. Salzberg, 2015 HISAT: a fast spliced aligner with low
709 memory requirements. Nat Methods 12**:** 357-360.
710 Krueger, F., and S. R. Andrews, 2011 Bismark: a flexible aligner and methylation caller
711 for Bisulfite-Seq applications. Bioinformatics 27**:** 1571-1572.
712 Kulis, M., A. C. Queirós, R. Beekman and J. I. Martín-Subero, 2013 Intragenic DNA
713 methylation in transcriptional regulation, normal differentiation and cancer. Biochim
714 Biophys Acta 1829**:** 1161-1174.
715 Langmead, B., C. Trapnell, M. Pop and S. L. Salzberg, 2009 Ultrafast and memory-
716 efficient alignment of short DNA sequences to the human genome. Genome Biol 10**:**
717 R25.
718 Lee, T. I., and R. A. Young, 2000 Transcription of eukaryotic protein-coding genes.
719 Annu Rev Genet 34**:** 77-137.
720 Li, E., T. H. Bestor and R. Jaenisch, 1992 Targeted mutation of the DNA
721 methyltransferase gene results in embryonic lethality. Cell 69**:** 915-926.

722 Li, E., and Y. Zhang, 2014 DNA methylation in mammals. Cold Spring Harb Perspect
723 Biol 6: a019133.
724 Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan *et al.*, 2009 The Sequence
725 Alignment/Map format and SAMtools. Bioinformatics 25: 2078-2079.
726 Lister, R., M. Pelizzola, R. H. Dowen, R. D. Hawkins, G. Hon *et al.*, 2009 Human DNA
727 methylomes at base resolution show widespread epigenomic differences. Nature 462:
728 315-322.
729 Luo, C., P. Hajkova and J. R. Ecker, 2018 Dynamic DNA methylation: In the right
730 place at the right time. Science 361: 1336-1340.
731 MacKay, H., C. A. Scott, J. D. Duryea, M. S. Baker, E. Laritsky *et al.*, 2019 DNA
732 methylation in AgRP neurons regulates voluntary exercise behavior in mice. Nat
733 Commun 10: 5364.
734 Martin, M., 2011 Cutadapt removes adapter sequences from high-throughput
735 sequencing reads. EMBnet. journal 17: 10-12.
736 Niehrs, C., and B. Luke, 2020 Regulatory R-loops as facilitators of gene expression and
737 genome stability. Nat Rev Mol Cell Biol 21: 167-178.
738 Nourse, J., S. Spada and S. Danckwardt, 2020 Emerging Roles of RNA 3'-end Cleavage
739 and Polyadenylation in Pathogenesis, Diagnosis and Therapy of Human Disorders.
740 Biomolecules 10.
741 Okano, M., D. W. Bell, D. A. Haber and E. Li, 1999 DNA methyltransferases Dnmt3a
742 and Dnmt3b are essential for de novo methylation and mammalian development. Cell
743 99: 247-257.
744 Pertea, M., G. M. Pertea, C. M. Antonescu, T. C. Chang, J. T. Mendell *et al.*, 2015
745 StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nat
746 Biotechnol 33: 290-295.
747 Proudfoot, N. J., 2016 Transcriptional termination in mammals: Stopping the RNA
748 polymerase II juggernaut. Science 352: aad9926.
749 Quinlan, A. R., and I. M. Hall, 2010 BEDTools: a flexible suite of utilities for
750 comparing genomic features. Bioinformatics 26: 841-842.
751 Ramírez, F., D. P. Ryan, B. Grüning, V. Bhardwaj, F. Kilpert *et al.*, 2016 deepTools2: a
752 next generation web server for deep-sequencing data analysis. Nucleic Acids Res 44:
753 W160-165.
754 Rauch, T. A., X. Wu, X. Zhong, A. D. Riggs and G. P. Pfeifer, 2009 A human B cell
755 methylome at 100-base pair resolution. Proc Natl Acad Sci U S A 106: 671-678.
756 Robinson, J. T., H. Thorvaldsdóttir, W. Winckler, M. Guttman, E. S. Lander *et al.*, 2011
757 Integrative genomics viewer. Nat Biotechnol 29: 24-26.
758 Schübeler, D., 2015 Function and information content of DNA methylation. Nature
759 517: 321-326.
760 Smith, Z. D., and A. Meissner, 2013 DNA methylation: roles in mammalian
761 development. Nat Rev Genet 14: 204-220.

762 Takayama, K., N. Shimoda, S. Takanaga, S. Hozumi and Y. Kikuchi, 2014 Expression
763 patterns of dnmt3aa, dnmt3ab, and dnmt4 during development and fin regeneration in
764 zebrafish. Gene Expr Patterns 14: 105-110.
765 Teissandier, A., and D. Bourc'his, 2017 Gene body DNA methylation conspires with
766 H3K36me3 to preclude aberrant transcription. EMBO J 36: 1471-1473.
767 Tian, B., and J. L. Manley, 2017 Alternative polyadenylation of mRNA precursors. Nat
768 Rev Mol Cell Biol 18: 18-30.
769 Wang, Z., X. Li, Y. Jiang, Q. Shao, Q. Liu *et al*., 2015 swDMR: A Sliding Window
770 Approach to Identify Differentially Methylated Regions Based on Whole Genome
771 Bisulfite Sequencing. PLoS One 10: e0132866.
772 Westerfield, M., 1993 *The zebrafish book : a guide for the laboratory use of zebrafish*
773 *(Brachydanio rerio)*. M. Westerfield, Eugene, OR.
774 Yagi, M., M. Kabata, A. Tanaka, T. Ukai, S. Ohta *et al*., 2020 Identification of distinct
775 loci for de novo DNA methylation by DNMT3A and DNMT3B during mammalian
776 development. Nat Commun 11: 3199.
777 Yagi, M., M. Kabata, T. Ukai, S. Ohta, A. Tanaka *et al*., 2019 De Novo DNA
778 Methylation at Imprinted Loci during Reprogramming into Naive and Primed
779 Pluripotency. Stem Cell Reports 12: 1113-1128.
780 Yagi, M., S. Kishigami, A. Tanaka, K. Semi, E. Mizutani *et al*., 2017a Derivation of
781 ground-state female ES cells maintaining gamete-derived DNA methylation. Nature
782 548: 224-227.
783 Yagi, M., S. Yamanaka and Y. Yamada, 2017b Epigenetic foundations of pluripotent
784 stem cells that recapitulate in vivo pluripotency. Lab Invest 97: 1133-1141.
785