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: E-mail yutaka@hiroshima-u.ac.jp

:TEL +81-82-424-7440

ABSTRACT

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

INTRODUCTION

Transcription by RNA polymerase II (Pol II), synthesis process of mRNA from DNA, consists of three biological processes: initiation, elongation, and termination. In higher eukaryotes, two transcription processes (initiation and elongation) and their epigenetic regulation, including DNA methylation and histone modifications, have been extensively studied (1-4). In transcription termination, it is also well known that most protein-coding genes in eukaryotes possess a polyadenylation signal (PAS) located at the end of the 3'-UTR, and a cleavage and polyadenylation (CPA) complex is assembled around the PAS sequence of nascent pre-mRNA, which directs pol II transcription termination (5,6). However, in contrast to transcription initiation and elongation, the epigenetic regulation of transcription termination has remained largely unstudied to date.

In eukaryotes ranging from plants to humans, DNA methylation, one of epigenetic regulations, is well characterized as a repressive epigenetic marker and mainly occurs at CpG dinucleotides (7-9). Advances in sequencing technology have resulted in genomic methylation profiling at single-base resolution revealing that the global DNA methylation (5-methylcytosine, 5mC) is seen in mammals, except for short CpG-rich regions known as CpG islands (7-9). Studies of methylome profiling have shown that DNA methylation patterns dynamically change during development and disease and that these patterns are linked to gene expression in mammals (10,11). Genome-wide

methylome analysis revealed that the DNA methylation levels around the promoter regions, including the transcription start site (TSS), are negatively correlated with gene expression (12,13). Conversely, DNA methylation levels in the gene bodies are positively correlated with gene expression (14,15). Although two possible mechanisms, regulation of splicing and prevention of cryptic promoters, have been proposed for methylation function in the gene body (14,15), the significance of DNA methylation around transcription termination sites (TTSs; cleavage/polyadenylation sites) remains unresolved.

The DNA methylation patterns are well known to be established by the DNA methyltransferases (Dnmts), which are categorized into two types in mammals: maintenance Dnmt (Dnmt1) and *de novo* Dnmts (Dnmt3a and Dnmt3b) (7,16,17). Knock-out mice of both Dnmt3a and Dnmt3b, which die at embryonic day 9.5, fail to start *de novo* methylation after implantation (17), and genetic mutation of Dnmt3a or Dnmt3b leads to various human diseases (11), suggesting the critical role of DNA methylation in development and homeostasis in mammals through the regulation of gene expression. A very recent study using mouse embryonic fibroblasts (MEFs) with genetic ablation of *de novo* Dnmts reported that Dnmt3a functions in *de novo* methylation at both TSS regions (TSS ± 1000 bp) and gene bodies (from TSS regions to TTS) of the Polycomb group (PcG) target developmental genes, whereas Dnmt3b mainly functions in X-chromosome methylation (18). These results indicate that

although the enzymatic functions of both proteins are redundant, genomic target regions are distinct in mammals. Although mammalian Dnmt3a mutant phenotypes and target regions have been reported, the role of methylation during transcription termination by Dnmt3a has not yet been reported.

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

MATERIALS AND METHODS

Zebrafish experiments, ethic statement, and generation of *dnmt3aa*$^{-/-}$ mutant fish using CRISPR/Cas9 system

Adult zebrafish and zebrafish larvae were maintained as previously described (19). All zebrafish experiments were approved from the Hiroshima University Animal Research Committee (Permit Number: F18-2-7).

Genome editing using the CRISPR/Cas9 system was performed as previously described (20). Briefly, the pDR274 vector (Addgene Plasmid 42250) was digested with *Bsa*I, and then annealed oligonucleotides for sgRNAs were cloned into the pDR274 vector. The sequences of the oligonucleotides for sgRNAs are listed in Supplementary Table S1. The sgRNA expression vectors were digested by *Dra*I, and the sgRNAs were synthesized using the MAXIscript™ T7 Kit (Invitrogen, Thermo Fisher Scientific). The synthesized sgRNAs were purified using the mirVana™ miRNA Isolation Kit (Invitrogen, Thermo Fisher Scientific). In addition, the Cas9 expression vector (Addgene Plasmid 51815) was linearized by *Not*I and *Cas9* mRNA was synthesized using the mMESSAGE mMACHINE™ SP6 Kit (Invitrogen, Thermo Fisher Scientific). A mixture of three types of sgRNAs for the zebrafish *dnmt3aa* gene and *Cas9* mRNA of appropriate concentrations was injected at the one-cell stage. To identify the mutations, the target site of the CRISPR/Cas9 system on *dnmt3aa* gene was amplified using TaKaRa Ex Taq® (Takara Bio) with the primers listed in Supplementary

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

Genotyping of MZ*dnmt3aa*$^{-/-}$ mutants

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

Whole-genome bisulfite library preparation and sequencing of zebrafish sample

Zebrafish wild-type (WT) and MZ*dnmt3aa*$^{-/-}$ mutant embryos at 2 days post-fertilization (dpf) were digested with proteinase K and sodium dodecyl sulfate to extract genomic DNA. DNA samples were prepared by mixing zebrafish genomic DNA and unmethylated lambda DNA (Promega) as spike control, at a ratio of 1000:1.

DNA samples (100 ng) were subjected to bisulfite treatment using the EZ DNA Methylation-Gold Kit (Zymo Research). Bisulfite-treated DNA samples (50 ng) were subjected to 10 cycles of PCR using random primers from the TruSeq DNA Methylation Kit (Illumina) to add adapters. Libraries with adapters were sequenced on the HiSeq X Five Sequencing System (Illumina) with 150 bp single-end reads. Bisulfite treatment, library preparation, and sequencing were performed by Takara Bio Inc. (Shiga, Japan).

Whole-genome bisulfite sequencing data analysis of mouse dataset

The publicly available datasets used in this study were from the Gene Expression Omnibus (GEO), and the accession numbers are GSE100957 (J1 mESCs) (21), GSE111172 (MEFs) (18), and GSE122405 (AgRPs) (22). The FASTQ-format read sequences were trimmed of adaptors and low-quality bases were filtered using Trim Galore (version 0.6.4) (https://github.com/FelixKrueger/TrimGalore) with Cutadapt (version 2.10) (23) and FastQC (version 0.11.8) (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). After trimming, the reads were aligned to the mouse genome (mm10) using Bismark (version 0.20.0) (24) and Bowtie (version 1.0.0) (25). The aligned reads were followed by the removal of PCR duplicates using deduplicate_bismark (version 0.20.0) (24), and methylation calls were performed using bismark_methylation_extractor (version 0.20.0) (24). The reads

from samples with multiple sequencing lanes were merged using Samtools (version 1.9) (26) after removing the PCR duplicates. The methylation ratio of CpG sites with a coverage of over 5 reads in WT and *Dnmt3a*$^{-/-}$ was calculated from the methylation call data. The methylation ratio was visualized using Integrative Genomics Viewer (IGV) (version 2.7.2) (27).

Whole-genome bisulfite sequencing data analysis of zebrafish datasets

The read sequences in FASTQ format were first trimmed by removing six bases from random primers of TruSeq DNA Methylation Kit (Illumina) and the last 50 bases of low-quality reads. After trimming, the reads were aligned to the zebrafish genome (danRer10) and lambda phage genome using Bismark (version 0.10.1) (24) and Bowtie (version 1.0.0) (25). Once aligned, reads corresponding to PCR duplicates were removed and methylation calls were performed. The bisulfite conversion rates based on spike control for all samples were over 99%, and the error rates were as follows: WT, 0.3% and MZ*dnmt3aa*$^{-/-}$ mutants, 0.2%. Statistically significant methylated cytosine sites were identified using a binomial distribution with a false discovery rate ≤ 0.05, calculated by the Benjamini-Hochberg (BH) method, and read depth ≥ 5. Data analysis was performed by Takara Bio, Inc. (Shiga, Japan) as previously described (28). Whole genome bisulfite sequencing (WGBS) data were deposited in the GEO under accession number GSE178691. The methylation ratio of CpG sites with a coverage of

5 reads in WT and MZ*dnmt3aa*$^{-/-}$ mutants as calculated from the methylation call data. The methylation ratio was visualized using the IGV (version 2.7.2) (27).

Identification of differentially methylated regions (DMRs)

The methylation call data of mice and zebrafish were used to identify differentially methylated regions (DMRs) using swDMR (version 1.0.0) (29) with the followed settings: cytosine type, CG; window, 500; step size, 50; points (lowest number of selected cytosine type in the window), 3; coverage, 5; fold (methylation level difference), 1.5; diff (value of max-min methylation level), 0.1; p-value, 0.05; fdr, 0.05. Samples were compared using the Fisher's test. DMRs overlapped genomic positions (TSS, gene body, TTS, and intergenic region) were determined using bedtools (version 2.28.0) (30) intersect. The DMRs that overlap both TSS and TTS were included in the DMRs that overlap TTS. Heatmaps, violin plots, scatterplots with marginal distribution, and horizontal bar graphs were generated from swDMR output using Python and R scripts.

RNA-seq library preparation and sequencing of zebrafish sample

Tissue lysates were prepared from 30 pooled 2 dpf WT and MZ*dnmt3aa*$^{-/-}$ mutants by homogenization using TissueLyser LT (QIAGEN) with zirconia beads for 10 s at 50 Hz. RNA was extracted from the lysate using the PureLink® RNA Mini Kit (Thermo Fisher Scientific). RNA quality of all samples was RNA Integrity Number (RIN) >

9.0, using the 4150 TapeStation System (Agilent). Total RNA was treated using the NEBNext® Poly(A) mRNA Magnetic Isolation Module (New England Biolabs) for mRNA isolation. Strand-specific RNA-seq libraries were constructed using the NEBNext® Ultra™ Directional RNA Library Prep Kit for Illumina® (New England Biolabs). The libraries with adapters were sequenced on the Novaseq 6000 (Illumina) with 2 × 150 bp paired-end reads. RNA extraction, mRNA isolation, library preparation, and sequencing were performed by Rhelixa Inc. (Tokyo, Japan).

RNA-seq data analysis for mouse and zebrafish datasets

The publicly available datasets of mice deposited in the GEO and the generated datasets of zebrafish were used in this study. Accession numbers are GSE100957 (J1 mESCs) (21), GSE84165 (MEFs) (31,32), GSE111172 (MEFs) (18), and GSE122405 (AgRPs) (22). RNA-seq data of zebrafish were deposited in the GEO under accession number GSE178691. FASTQ files from samples with multiple sequencing lanes were merged. The FASTQ-format sequence reads were trimmed of adaptors and low-quality bases were filtered using Trim Galore (version 0.6.4) (https://github.com/FelixKrueger/TrimGalore) with Cutadapt (version 1.18 or 2.10) (23) and FastQC (version 0.11.8 or 0.11.9) (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). After trimming, the reads were aligned to the mouse genome (mm10) and the zebrafish genome (danRer10)

using HISAT2 (version 2.1.0) (33). Alignment of zebrafish datasets was performed using the strand-specific option. Aligned files (.bam) were sorted and indexed using Samtools (version 1.9) (26). In addition, the strand-specific aligned files of zebrafish were generated using Samtools (version 1.9) (26). Read coverage was normalized with 1 bin size by counts per million mapped reads (CPM) using the bamCoverage tool from the DeepTools (version 3.3.2) (34) and visualized using the IGV (version 2.7.2) (27). Splice junction data of chimeric transcripts in the aligned files were inspected manually using the IGV (version 2.7.2) (27).

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

Quantification of expression levels was completed using StringTie (version 2.1.2) (35) with NCBI RefSeq annotation. The expression values of each transcript were normalized to the transcripts per million (TPM). Transcripts expressed in WT and *Dnmt3a*$^{-/-}$ mouse cells or MZ*dnmt3aa*$^{-/-}$ mutants were extracted from all transcripts (TPM > 0). In addition, the transcripts expressed in both strains and overlapped hypoDMRs across the TTS were extracted.

All reads that overlapped the exonic or intronic sequences in the aligned files (.bam) were removed using bedtools (version 2.28.0) (30) intersect with the NCBI RefSeq annotation. In addition, the intergenic reads were counted within 3 kb downstream of the TTS using bedtools (version 2.28.0) (30) multicov, except when the

genomic region downstream of the TTS was less than 3 kb. The expression level was calculated using the following equation (TPM 3 kb):

TPM 3 kb = Number of reads aligned in 3kb downstream of TTS of each transcript × 10^6 / Total number of reads aligned in 3kb downstream of TTS.

Normalized read counts were compared in WT and *Dnmt3a*$^{-/-}$ mouse cells or MZ*dnmt3aa*$^{-/-}$ zebrafish mutants as a metagene analysis.

Box plots were generated using R scripts, and statistical significance was determined using Brunner-Munzel test ($p \leq 0.05$).

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

Chimeric transcripts, which were aberrantly fused with exons of downstream genes or intergenic DNA, were manually counted using IGV (27). The expression level was calculated as the counts per million (CPM) mapped reads. Normalized read counts were compared in WT and *Dnmt3a*$^{-/-}$ mouse cells or MZ*dnmt3aa*$^{-/-}$ zebrafish mutants as a metagene analysis. Transcripts whose expression increased downstream of the TTS were used for metagene analysis. Box plots were generated using R scripts. Statistical significance was determined using the Brunner-Munzel test ($p \leq 0.05$).

RNA preparation, quantitative PCR (qPCR), and sequencing for identification of read-through and chimeric transcripts

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

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

Sequencing of chimeric transcripts was performed using MZ*dnmt3aa*$^{-/-}$ mutants. The product of *im:7152937*-intergenic DNA-*mns1* chimeric transcript was amplified using BIOTAQ™ DNA Polymerase (Meridian Bioscience) with the primers listed in Supplementary Table S4. The amplicons were purified using the QIAquick Gel Extraction Kit (QIAGEN) after agarose gel electrophoresis. The purified product was subcloned using a TA-cloning Kit (Thermo Fisher Scientific), and the plasmid DNA

was sequenced using the BigDye™ Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, Thermo Fisher Scientific) and primers listed in Supplementary Table S4.

RESULTS

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

To investigate the functional relevance between DNA methylation and transcription termination, we searched datasets of *Dnmt3a*-deficient mammalian cells that contain both methylome (WGBS) and transcriptome (RNA-seq) data. Five datasets of WGBS (human: one, mouse: four) and ten datasets of RNA-seq (human: four, mouse: six) were obtained from GEO. Four datasets (human: one, mouse: three) contained both WGBS and RNA-seq, and among them we chose three types of mouse datasets (J1 mouse embryonic stem cells, J1 mESCs (21); mouse embryonic fibroblasts, MEFs (18); and Agouti-related protein neurons, AgRPs) (22) because of the differences in their cell differentiation states. We extracted the hypermethylated DMRs (hyperDMRs) and hypoDMRs from these datasets and found that the number of hypoDMRs in *Dnmt3a*$^{-/-}$ J1 mESCs were approximately 50 and 180 times greater than that of *Dnmt3a*$^{-/-}$ MEFs or AgRPs, respectively (Supplementary Figure 1A-C). The hypoDMRs were categorized by genomic position (gene body, intergenic region, and gene body & intergenic region) (Supplementary Figure 2). The number of hypoDMRs gradually decreased in the order of *Dnmt3a*$^{-/-}$ J1 mESCs, MEFs, and AgRPs, whereas their ratios of genomic position were nearly the same among these *Dnmt3a*-deficient cells (Supplementary Figure 2). Next, we selected the hypoDMRs overlapping TTSs

in three types of *Dnmt3a*$^{-/-}$ mouse cells (Figure 1A,B) and found that the number of hypoDMRs overlapping the TTSs in *Dnmt3a*$^{-/-}$ MEFs or AgRPs was less than 1% or 0.4% of that of *Dnmt3a*$^{-/-}$ J1 mESCs, respectively (Figure 1C). These data suggest that DNA methylation regulated by Dnmt3a may be correlated with the differentiation state of mouse cells. Two-dimensional (2D) scatterplots with marginal distribution showed that the hypoDMRs overlapping the TTSs were almost symmetrically distributed across the TTSs in the *Dnmt3a*$^{-/-}$ J1 mESCs and MEFs, whereas the distribution of hypoDMRs was shifted upstream of the TTSs in *Dnmt3a*$^{-/-}$ AgRPs (Figure 1D).

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

To explore the effect of hypomethylation overlapping the TTSs on transcription termination, we analyzed RNA-seq data from three different types of *Dnmt3a*$^{-/-}$ mouse cells. Read counts aligned to 3 kbp of TTSs were normalized by TPM (3 kb) in WT, *Dnmt3a*$^{-/-}$ J1 mESCs, MEFs, or AgRPs. Metagene analysis showed that the read counts downstream of TTSs increased significantly in *Dnmt3a*$^{-/-}$ AgRPs compared to those in WT cells (Figure 2A), and higher expression was observed in more than half of the genes with hypoDMRs in mutant cells (Figure 2B). Conversely, the read counts in *Dnmt3a*$^{-/-}$ J1 mESCs and MEFs were unchanged compared to those in WT cells (Figure 2A).

We then analyzed the content of transcripts downstream of hypomethylated TTSs in the *Dnmt3a*$^{-/-}$ AgRPs and found that the centers of the hypoDMRs, which causes an increase in read counts downstream of TTSs, were distributed close to the TTSs in these mutant cells (Supplementary Figure 3A,B). We further found chimeric transcripts that were aberrantly fused with exons of downstream genes or intergenic DNA. Frequency of chimeric transcripts (about 54.6%) increased significantly in the *Dnmt3a*$^{-/-}$ AgRPs compared to those in WT cells (Figure 2C,D). We showed that splicing of the *Ring Finger 157* (*Rnf157*) gene encoding E3 ubiquitin ligase 2D as an example of chimeric transcripts in the *Dnmt3a*$^{-/-}$ AgRPs, and found an aberrant transcript that connects *Rnf157* exon to the intergenic DNA or *Unkempt family zinc finger* (*Unk*) gene (Supplementary Figure 4). The centers of DMRs causing these chimeric transcripts were distributed close to the TTS (Figure 2E,F). In summary, the increase in the chimeric transcripts in the *Dnmt3a*$^{-/-}$ AgRPs clearly supports our idea that transcription termination is defective and that read-through occurs at hypomethylated TTSs.

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

To explore the correlation between the transcription termination defects and DNA methylation in an animal model, we examined the methylome and transcriptome analysis using the zebrafish as a model. Two zebrafish orthologs (*dnmt3aa* and

dnmt3ab) of mammalian Dnmt3a have been previously identified (36), and the detailed expression profiles of *dnmt3aa* during development and regeneration showed that *dnmt3aa* was expressed in more organs than *dnmt3ab*, and its expression was observed in de-differentiated blastema cells of the regenerating caudal fin (36). Therefore, we attempted to generate the MZ*dnmt3aa*$^{-/-}$ knockout mutants by using CRISPR/Cas9 methods (Figure 3A) and found that they were viable, fertile and showed no specific phenotype at 2 days post fertilization (dpf) (Figure 3B). Analyses of WGBS and RNA-seq were performed at 2 dpf, because most of the organs had already formed at this stage. In MZ*dnmt3aa*$^{-/-}$ mutants, 43.6% DMRs were hypomethylated (Supplementary Figure 5A-C) and the hypoDMRs were categorized by genomic position (gene body, intergenic region, and gene body & intergenic region) (Supplementary Figure 6). We then selected the hypoDMRs overlapping the TTSs, and 27 hypoDMRs were identified overlapping the TTSs (Figure 3C-E). A 2D scatterplot with marginal distribution showed that the hypoDMRs overlapping the TTSs were almost symmetrically distributed across the TTSs in MZ*dnmt3aa*$^{-/-}$ mutants (Figure 3F), as observed in *Dnmt3a*$^{-/-}$ J1 mESCs and MEFs.

Increases in both chimeric and read-through transcripts were identified in the hypoDMRs overlapping TTSs in MZ*dnmt3aa*$^{-/-}$ mutants

We also analyzed the RNA-seq data of MZ*dnmt3aa*$^{-/-}$ mutants and the read counts from genes containing hypoDMRs overlapping the TTSs, which were normalized in the WT and mutants. Metagene analysis showed that the number of transcripts downstream of the TTSs in MZ*dnmt3aa*$^{-/-}$ mutants were unchanged compared to those in WT (Figure 4A). The transcripts that showed an increase (30.3%) were almost symmetrically distributed across the TTSs in the mutants (Figure 4B, Supplementary Figure 7). Next, we have extracted chimeric transcripts fused with exons of downstream genes and intergenic DNA. Normalized read counts of the chimeric transcripts (30%) were also unchanged in MZ*dnmt3aa*$^{-/-}$ mutants compared to those in WT (Figure 4C,D). Three increased chimeric transcripts (*cytokine like 1* (*cytl1*) and *im:7152037*) were found in two different DMRs, which were close to TTSs (Figure 4D-F).

To further analyze the transcription termination defect caused by hypoDMRs overlapping the TTSs, we examined the detailed expression downstream of and across the TTSs of three transcripts that had shown high expression levels downstream of TTSs (*cytl1*, *im:7152037*, and *potassium channel, subfamily K, member 1b* (*kcnk1b*)) by RNA-seq (Supplementary Figure 8,9,10). The expression level of three transcripts in MZ*dnmt3aa*$^{-/-}$ mutants had significantly increased downstream of the TTSs compared to that in WT, as detected by qPCR (Figure 5A,B). We also found a significant increase in the two transcripts (*im:7152037* and *kcnk1b*) across the TTSs (Figure 5C,D),

demonstrating that read-through occurs in these genes. In addition to read-through, chimeric transcripts (*im:7152037*-intermediate DNA-*meiosis-specific nuclear structural 1* (*mns1*)) were detected in *im:7152037* (Supplementary Figure 9,11), and qPCR analysis showed that the expression level of chimeric transcripts between *im:7152037* and intergenic DNA or *mns1* had increased significantly in MZ*dnmt3aa*$^{-/-}$ mutants (Figure 5E). Taken together, the read-through and chimeric transcripts of three genes in MZ*dnmt3aa*$^{-/-}$ mutants, combined with the results from the *Dnmt3a*$^{-/-}$ AgRPs analyses, clearly support our hypothesis that transcription termination is defective in hypomethylated TTSs.

Discussion

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

Our bioinformatic analyses showed that the number of hypoDMRs overlapping TTSs in *Dnmt3a*$^{-/-}$ J1 mESCs was much higher than that in *Dnmt3a*$^{-/-}$ MEFs and AgRPs (Figure 1C). Previous studies have shown that mouse ESCs, such as J1 mESCs, which are cultured in medium containing serum, are globally hypermethylated compared to cells in mouse inner cell mass (37). In other words, *Dnmt3a* knockout J1 ESCs are thought to have a large number of hypoDMRs. The number of hypoDMRs overlapping TTSs in *Dnmt3a*$^{-/-}$ J1 mESCs gradually decreased in the order of *Dnmt3a*$^{-/-}$ J1 mESCs, MEFs, and AgRPs (Figure 1C). This result may be attributed to the fact

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

In this study, we used zebrafish mutants as an animal model because MZ*dnmt3aa*$^{-/-}$ mutants, unlike *Dnmt3a* knockout mice, are viable and fertile. The MZ*dnmt3aa*$^{-/-}$ mutant line can be easily maintained and Dnmt3aa functions in this mutant can be analyzed without maternally expressed *dnmt3aa*. However, normalized read counts downstream of the TTSs and chimeric transcripts were unchanged between MZ*dnmt3aa*$^{-/-}$ mutants and WT (Figure 4). There are two possible explanations for these results: one possibility is that zebrafish *dnmt3ab* may rescue the transcription termination defects in MZ*dnmt3aa*$^{-/-}$ mutants. Another possibility is that cells in the zebrafish body are more heterogeneous and vary greatly compared to cultured cells, because almost all organs have been formed at 2 dpf. Nevertheless, we have successfully detected the read-through and chimeric transcripts of three genes (*cytl1*, *im:7152037*, and *kcnk1b*) in MZ*dnmt3aa*$^{-/-}$ mutants, suggesting that DNA methylation is involved in transcription termination in zebrafish.

Although we have shown the significance of hypomethylation at the TTS region for transcription termination in vertebrates, it is unclear how hypomethylation regulates transcription termination. Despite the relative lack of transcription termination research, a unified allosteric/torpedo model has recently been proposed for the PAS-dependent termination (6,38). In this model, the CPA complex recognizes and

cleaves the nascent transcript when the pol II passes through the PAS; thus, PAS is close to TTS (10-35 bp upstream of the TTS) (39,40). This cleavage by the CPA complex results in pol II pausing via dephosphorylation of the Spt5 elongation factor by a protein phosphatase 1 nuclear targeting subunit (PNUTS) (6). As paused pol II is dissociated from the DNA strand by the function of XRN2 exoribonuclease, which degrades RNA in a 5'→3' direction, the pol II pausing is thought to be a critical step for the transcription termination (5,6). Although the CPA complex mainly binds to nascent pre-mRNA and C-terminal domain (CTD) of pol II, it is possible that DNA hypomethylation of the TTS or PAS genomic region may directly or indirectly suppress the CPA function, leading to loss of pol II pausing. It should be interesting to note that almost all hypoDMRs, which cause chimeric transcripts or increased read counts downstream of the TTSs, overlap with PAS in *Dnmt3a*$^{-/-}$ AgRPs and MZ*dnmt3aa*$^{-/-}$ mutants (Figure 2F, Figure 4F, Supplementary Figure 3B, Supplementary Figure 7B). In PAS-dependent termination, there are two other mechanisms of pol II pausing: heterochromatin- and R-loop-dependent pausing (5). It has been reported previously that the heterochromatin structure downstream of the TTS causes pol II pausing by acting as an obstacle (5). Therefore, hypomethylation in the TTS region may lead to the suppression of heterochromatin formation downstream of TTS. An alternative possibility is R-loop-dependent pausing; R-loops are known as three-stranded structures, which consist of RNA-DNA hybrids, and function in various nuclear processes, such as

transcription, DNA replication, and DNA repair (41,42). As recent studies have shown that pol II pausing is induced by R-loop formation at the TTS, it is worth analyzing the relationship between R-loop formation and DNA methylation. Future molecular biological analyses regarding hypomethylation of the TTS region are needed to improve our understanding of the molecular events underlying transcription termination.

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 statement. None declared.

Figure 1

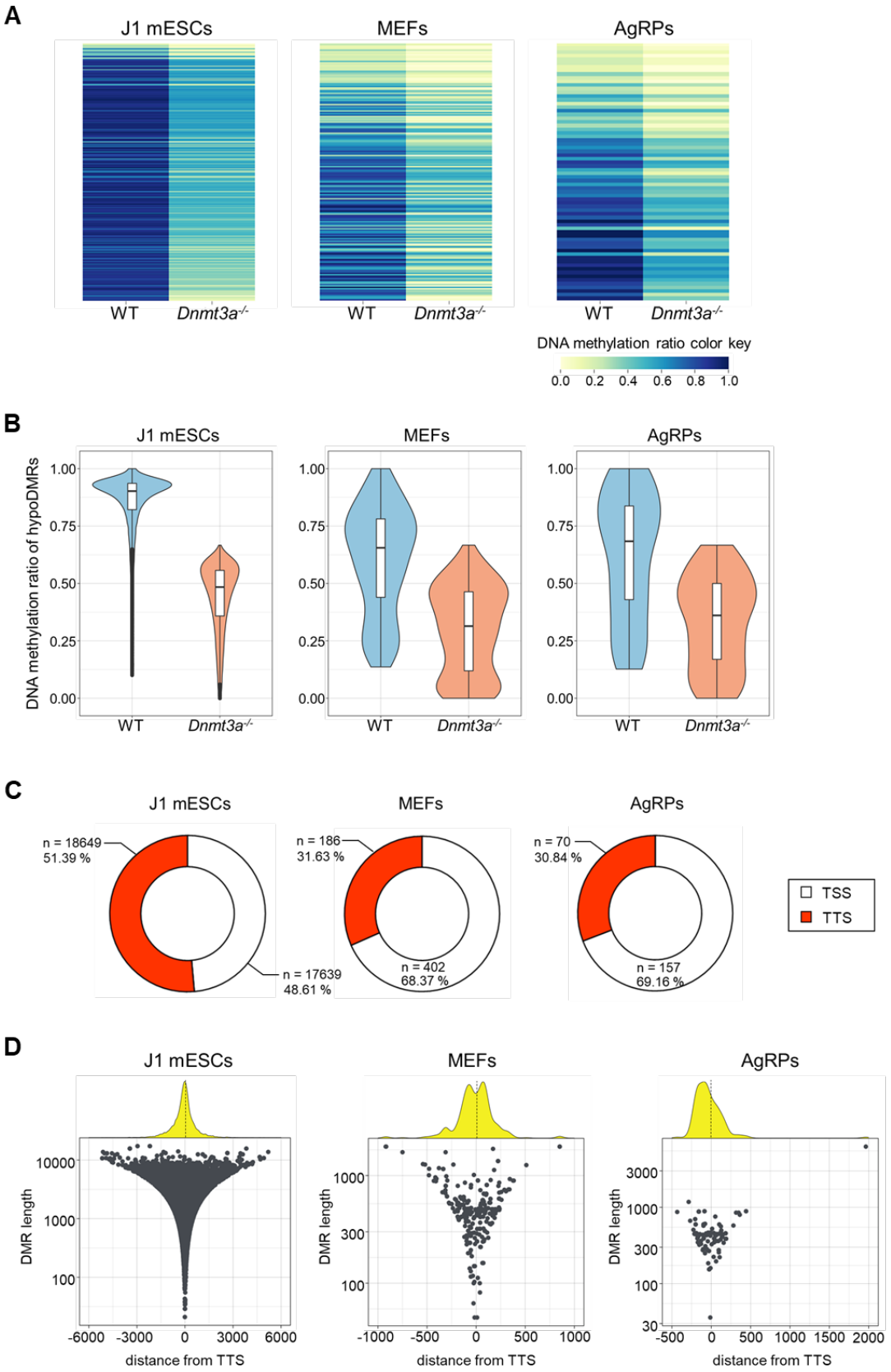
A
J1 mESCs
MEFs
AgRPs
WT
Dnmt3a-/-
DNA methylation ratio color key
0.0 0.2 0.4 0.6 0.8 1.0
B
DNA methylation ratio of hypoDMRs
1.00
0.75
0.50
0.25
0.00
C
n = 18649
51.39 %
n = 17639
48.61 %
n = 186
31.63 %
n = 402
68.37 %
n = 70
30.84 %
n = 157
69.16 %
TSS
TTS
D
DMR length
distance from TTS

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

(A) Heatmap of DNA methylation profiles of hypoDMRs overlapping TTSs in wild-type (WT) and *Dnmt3a*$^{-/-}$ J1 mESCs, MEFs, and AgRPs. Color scale indicates DNA methylation ratio. (B) Violin plots showing DNA methylation ratio of the hypoDMRs overlapping the TTSs in WT and *Dnmt3a*$^{-/-}$ J1 mESCs, MEFs, and AgRPs. (C) Pie charts depicting the number and percentage of hypoDMRs overlapping the TSSs and TTSs in *Dnmt3a*$^{-/-}$ J1 mESCs, MEFs, and AgRPs. (D) 2D scatterplots with marginal distribution comparing hypoDMR length versus distance from TTS in *Dnmt3a*$^{-/-}$ J1 mESCs, MEFs, and AgRPs. Dots represent the center of hypoDMRs.

Figure 2

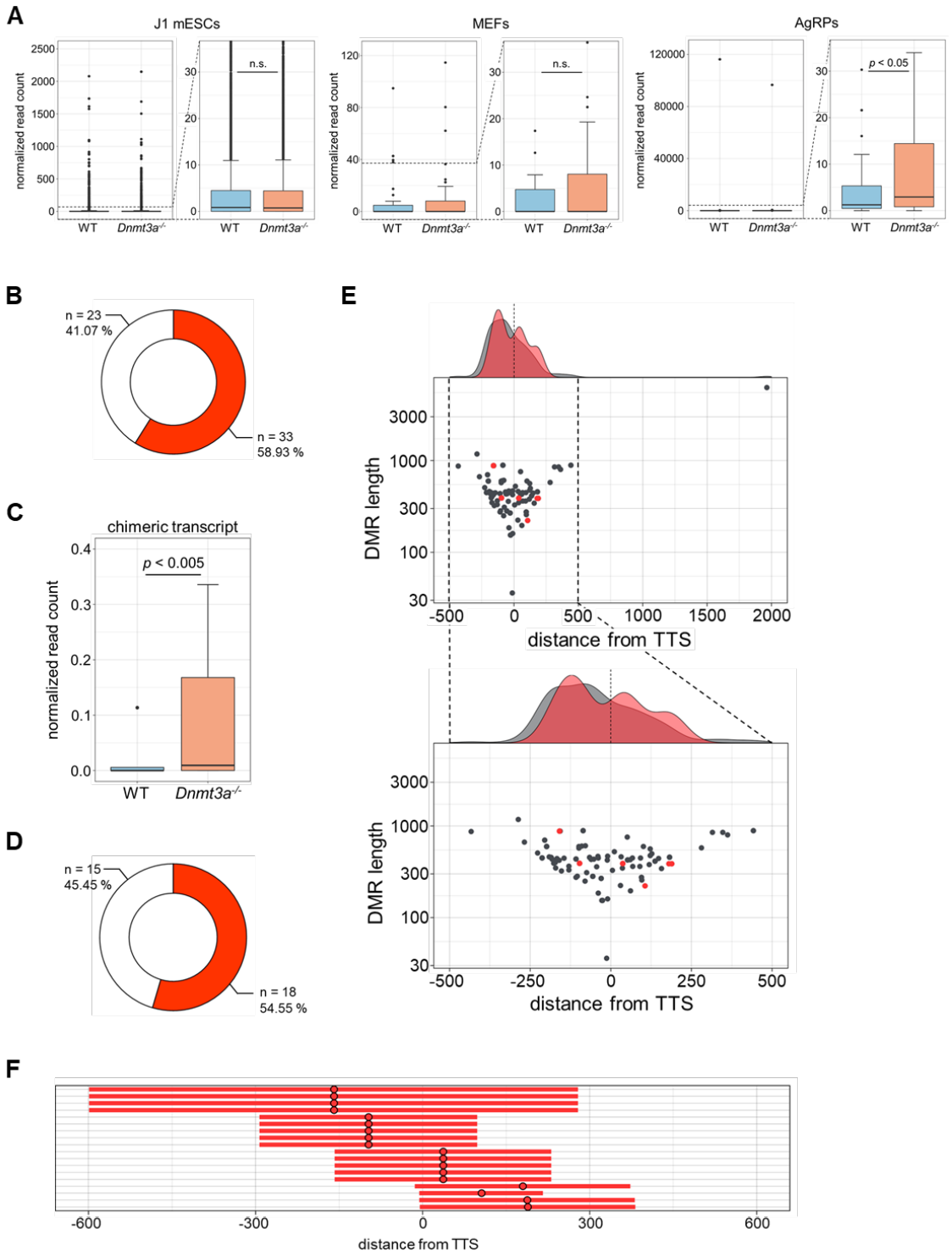
A
J1 mESCs
MEFs
AgRPs
normalized read count
n.s.
p < 0.05
WT
Dnmt3a-/-
B
n = 23
41.07 %
n = 33
58.93 %
C
chimeric transcript
p < 0.005
normalized read count
WT
Dnmt3a-/-
D
n = 15
45.45 %
n = 18
54.55 %
E
DMR length
distance from TTS
F
distance from TTS

Figure 2. Metagene analysis of transcripts in three *Dnmt3a*-deficient mouse cells.

(A) Box plots showing the normalized read counts in WT and *Dnmt3a*$^{-/-}$ mutants (J1 mESCs, MEFs, and AgRPs). Right graphs are enlarged views of the left ones, respectively. (B) A pie chart depicting the number and percentage of transcripts with increased read counts (red) or decreased/unchanged read counts (white) in *Dnmt3a*$^{-/-}$ AgRPs. (C) A box plot showed chimeric transcripts in WT and *Dnmt3a*$^{-/-}$ AgRPs. (D) A pie chart depicting the number and percentage of transcripts with increased chimeric transcripts (red) or decreased/unchanged chimeric transcripts (white) in *Dnmt3a*$^{-/-}$ AgRPs. (E) 2D scatterplots with marginal distribution comparing hypoDMR length versus distance from TTS in *Dnmt3a*$^{-/-}$ AgRPs. Grey and red dots represent the center of hypoDMRs. Red dots represent hypoDMRs, which lead to increased chimeric transcripts. A lower graph is an enlarged view of the upper one. (F) A horizontal bar graph represents the distance from TTS of hypoDMRs, which lead to increased chimeric transcripts, relative to TTS. Red dots and bars represent the center and range of hypoDMRs relative to TTS position, respectively.

Figure 3

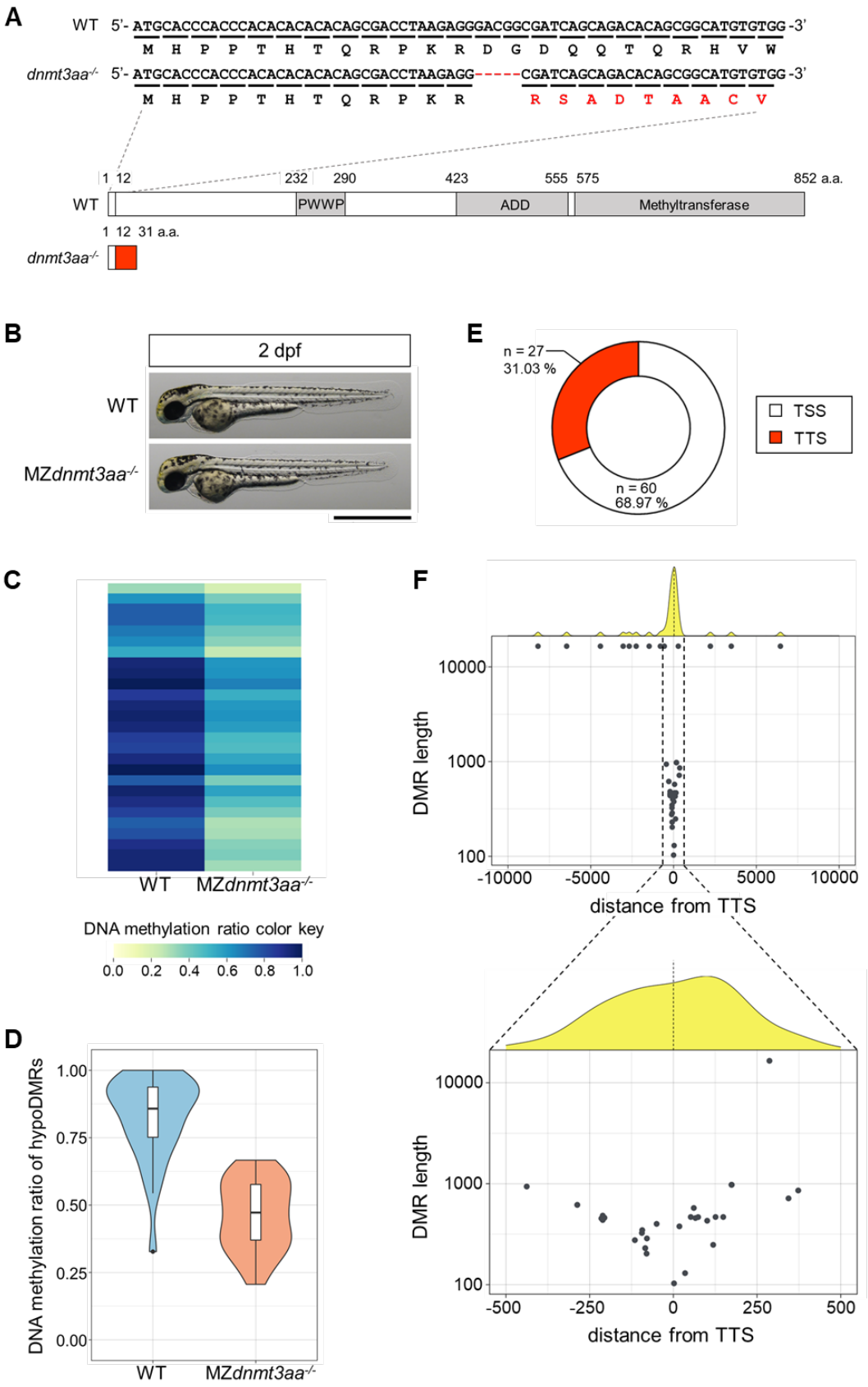
A
WT 5'- ATGCACCCACCCACACACACACAGCGACCTAAGAGGGACGGCGATCAGCAGACACAGCGGCATGTGTGG -3'
M H P P T H T Q R P K R D G D Q Q T Q R H V W
dnmt3aa-/- 5'- ATGCACCCACCCACACACACACAGCGACCTAAGAGG-----CGATCAGCAGACACAGCGGCATGTGTGG -3'
M H P P T H T Q R P K R R S A D T A A C V
1 12 232 290 423 555 575 852 a.a.
WT
PWWP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

Figure 3. Generation of MZ*dnmt3aa*$^{-/-}$ zebrafish mutants and characterization of hypomethylated DMRs.

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

Figure 4

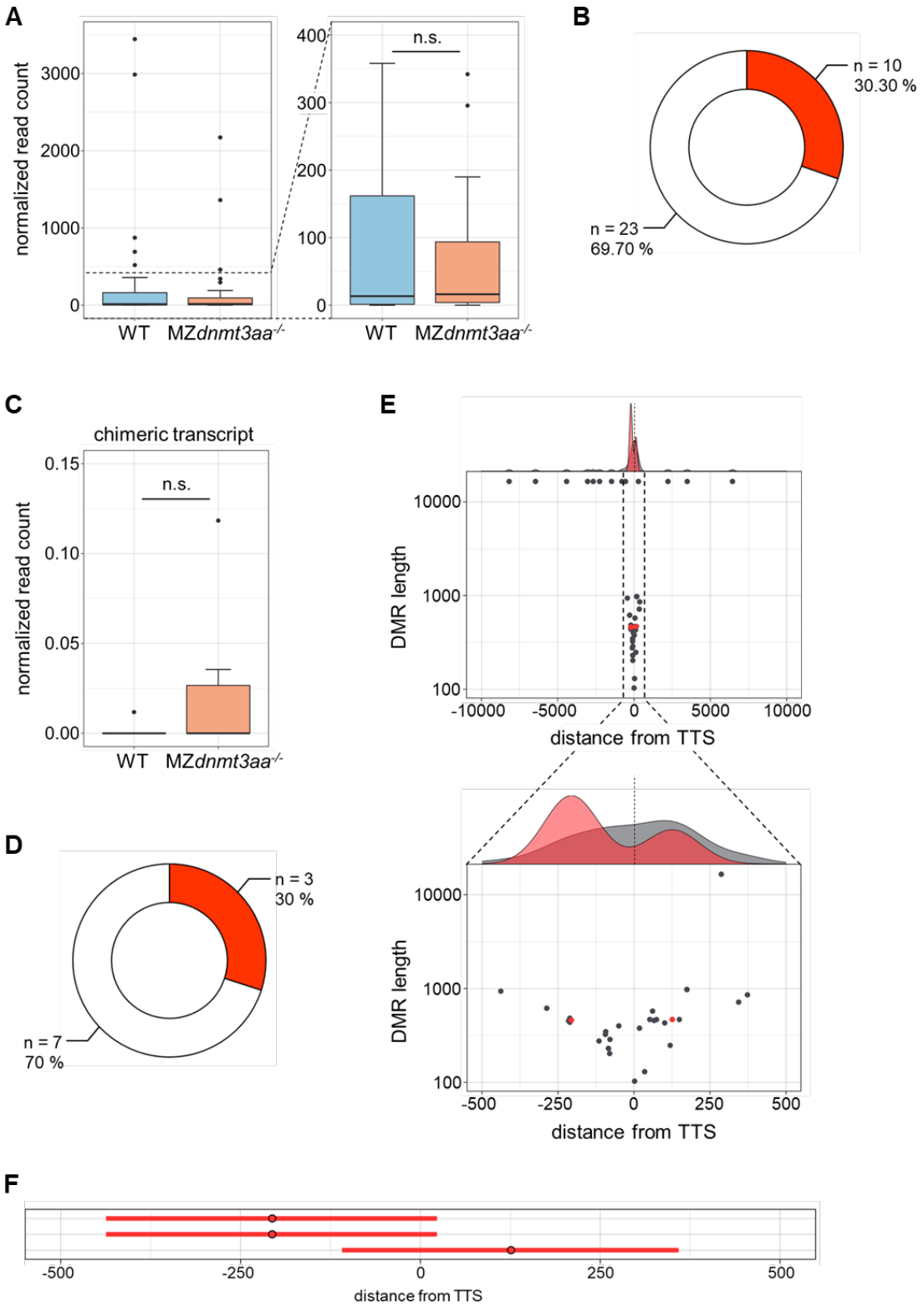
A
normalized read count
3000
2000
1000
0
400
300
200
100
0
n.s.
WT
MZdnmt3aa-/-
B
n = 10
30.30 %
n = 23
69.70 %
C
chimeric transcript
normalized read count
0.15
0.10
0.05
0.00
n.s.
WT
MZdnmt3aa-/-
D
n = 3
30 %
n = 7
70 %
E
DMR length
10000
1000
100
-10000
-5000
0
5000
10000
distance from TTS
-500
-250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

Figure 4. Metagene analysis of transcripts in MZ*dnmt3aa*$^{-/-}$ mutants.

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

Figure 5

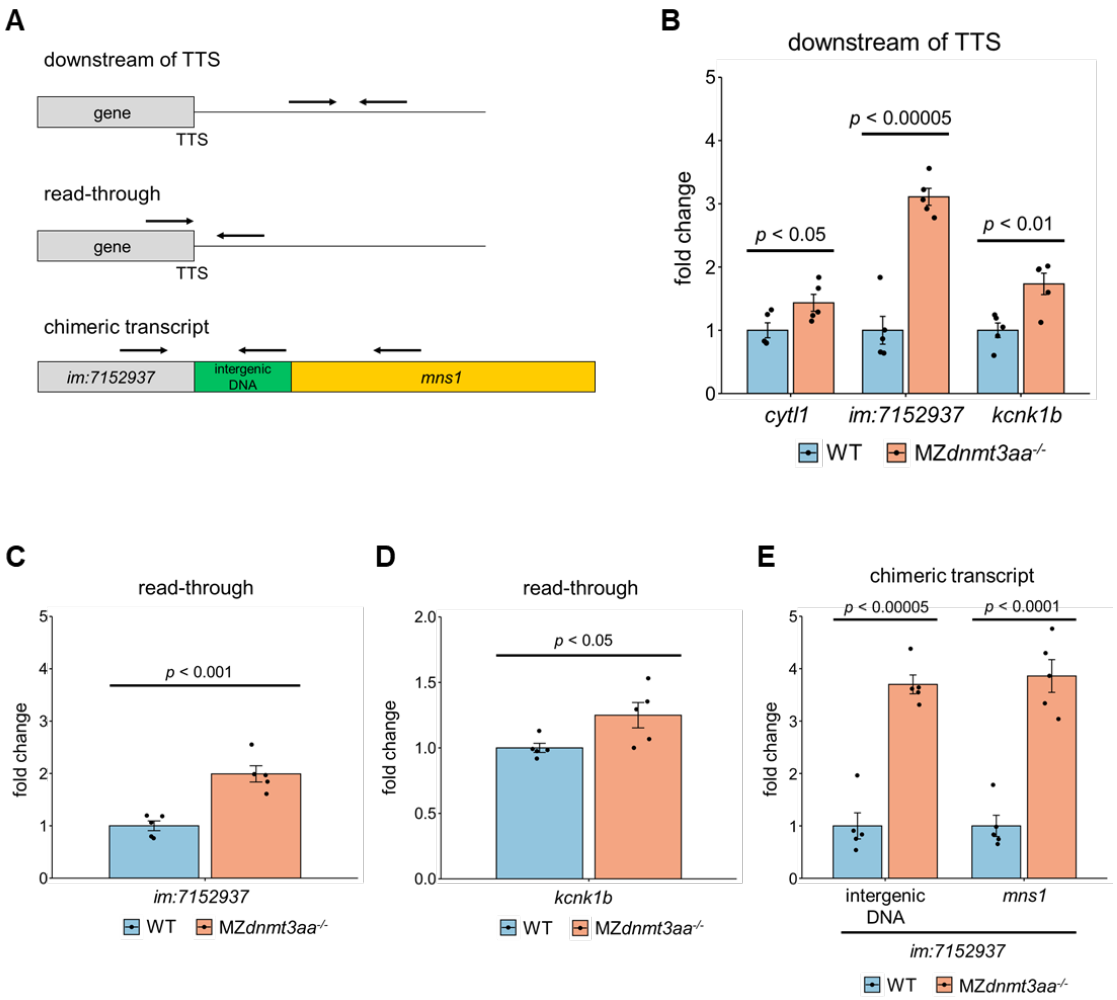
A
downstream of TTS
gene
TTS
read-through
gene
TTS
chimeric transcript
im:7152937
intergenic DNA
mns1
B
downstream of TTS
fold change
p < 0.05
p < 0.00005
p < 0.01
cytl1
im:7152937
kcnk1b
WT
MZdnmt3aa-/-
C
read-through
fold change
p < 0.001
im:7152937
WT
MZdnmt3aa-/-
D
read-through
fold change
p < 0.05
kcnk1b
WT
MZdnmt3aa-/-
E
chimeric transcript
fold change
p < 0.00005
p < 0.0001
intergenic DNA
mns1
im:7152937
WT
MZdnmt3aa-/-

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

(A) Scheme of primer pair positions to detect aberrant transcripts. (B) qRT-PCR analysis for transcripts downstream of TTS in *cytl1*, *im:7152937*, and *kcnk1b* genes. (C, D) qRT-PCR analysis for read-through transcripts in *im:7152937* (C) and *kcnk1b* (D) genes. (E) qRT-PCR analysis for chimeric transcripts between exons in *im:7152937* and intergenic DNA/exon in *mns1*. All qRT-PCR data are presented as means ± SE. The mean expression level of WT is defined as 1.

Figure 6

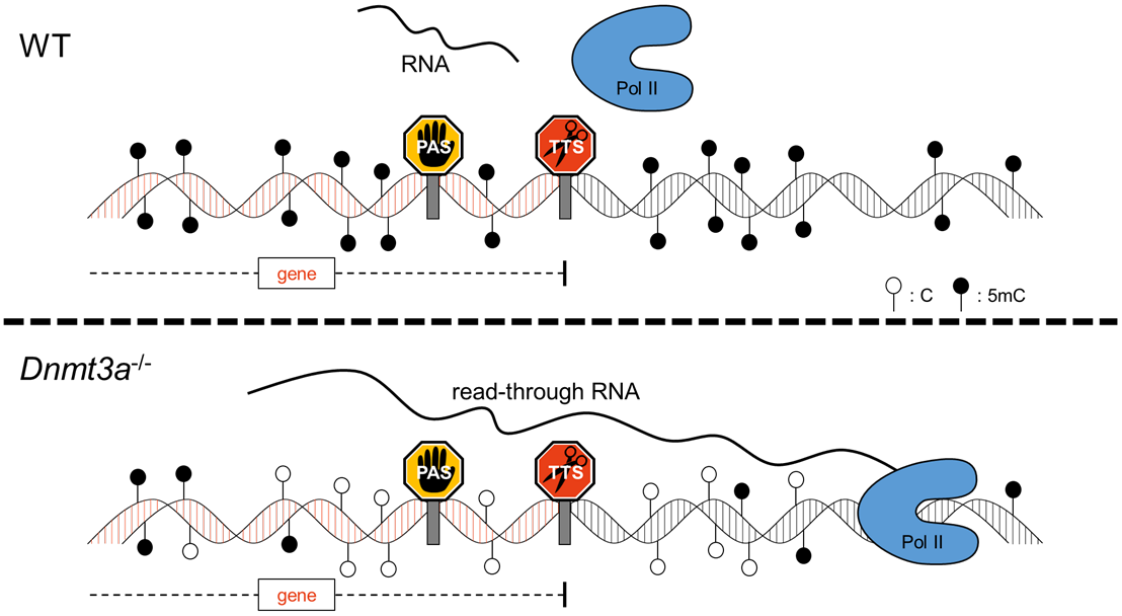

Figure 6. A proposed model of transcription termination.

REFERENCES

1. Lee, T.I. and Young, R.A. (2000) Transcription of eukaryotic protein-coding genes. *Annu Rev Genet*, **34**, 77-137.
2. Jones, P.A. (2012) Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nat Rev Genet*, **13**, 484-492.
3. Gates, L.A., Foulds, C.E. and O'Malley, B.W. (2017) Histone Marks in the 'Driver's Seat': Functional Roles in Steering the Transcription Cycle. *Trends Biochem Sci*, **42**, 977-989.
4. Chen, F.X., Smith, E.R. and Shilatifard, A. (2018) Born to run: control of transcription elongation by RNA polymerase II. *Nat Rev Mol Cell Biol*, **19**, 464-478.
5. Proudfoot, N.J. (2016) Transcriptional termination in mammals: Stopping the RNA polymerase II juggernaut. *Science*, **352**, aad9926.
6. Eaton, J.D. and West, S. (2020) Termination of Transcription by RNA Polymerase II: BOOM! *Trends Genet*, **36**, 664-675.
7. Li, E. and Zhang, Y. (2014) DNA methylation in mammals. *Cold Spring Harb Perspect Biol*, **6**, a019133.
8. Schübeler, D. (2015) Function and information content of DNA methylation. *Nature*, **517**, 321-326.
9. Luo, C., Hajkova, P. and Ecker, J.R. (2018) Dynamic DNA methylation: In the right place at the right time. *Science*, **361**, 1336-1340.
10. Smith, Z.D. and Meissner, A. (2013) DNA methylation: roles in mammalian development. *Nat Rev Genet*, **14**, 204-220.
11. Greenberg, M.V.C. and Bourc'his, D. (2019) The diverse roles of DNA methylation in mammalian development and disease. *Nat Rev Mol Cell Biol*, **20**, 590-607.
12. Rauch, T.A., Wu, X., Zhong, X., Riggs, A.D. and Pfeifer, G.P. (2009) A human B cell methylome at 100-base pair resolution. *Proc Natl Acad Sci U S A*, **106**, 671-678.
13. Ball, M.P., Li, J.B., Gao, Y., Lee, J.H., LeProust, E.M., Park, I.H., Xie, B., Daley, G.Q. and Church, G.M. (2009) Targeted and genome-scale strategies reveal gene-body methylation signatures in human cells. *Nat Biotechnol*, **27**, 361-368.
14. Kulis, M., Queirós, A.C., Beekman, R. and Martín-Subero, J.I. (2013) Intragenic DNA methylation in transcriptional regulation, normal differentiation and cancer. *Biochim Biophys Acta*, **1829**, 1161-1174.
15. Teissandier, A. and Bourc'his, D. (2017) Gene body DNA methylation conspires with H3K36me3 to preclude aberrant transcription. *EMBO J*, **36**, 1471-1473.
16. Li, E., Bestor, T.H. and Jaenisch, R. (1992) Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell*, **69**, 915-926.

17. Okano, M., Bell, D.W., Haber, D.A. and Li, E. (1999) DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell*, **99**, 247-257.
18. Yagi, M., Kabata, M., Tanaka, A., Ukai, T., Ohta, S., Nakabayashi, K., Shimizu, M., Hata, K., Meissner, A., Yamamoto, T. *et al*. (2020) Identification of distinct loci for de novo DNA methylation by DNMT3A and DNMT3B during mammalian development. *Nat Commun*, **11**, 3199.
19. Westerfield, M. (1993) *The zebrafish book : a guide for the laboratory use of zebrafish (Brachydanio rerio)*. M. Westerfield, Eugene, OR.
20. Ansai, S. and Kinoshita, M. (2014) Targeted mutagenesis using CRISPR/Cas system in medaka. *Biol Open*, **3**, 362-371.
21. Gu, T., Lin, X., Cullen, S.M., Luo, M., Jeong, M., Estecio, M., Shen, J., Hardikar, S., Sun, D., Su, J. *et al*. (2018) DNMT3A and TET1 cooperate to regulate promoter epigenetic landscapes in mouse embryonic stem cells. *Genome Biol*, **19**, 88.
22. MacKay, H., Scott, C.A., Duryea, J.D., Baker, M.S., Laritsky, E., Elson, A.E., Garland, T., Fiorotto, M.L., Chen, R., Li, Y. *et al*. (2019) DNA methylation in AgRP neurons regulates voluntary exercise behavior in mice. *Nat Commun*, **10**, 5364.
23. Martin, M. (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. journal*, **17**, 10-12.
24. Krueger, F. and Andrews, S.R. (2011) Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics*, **27**, 1571-1572.
25. Langmead, B., Trapnell, C., Pop, M. and Salzberg, S.L. (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol*, **10**, R25.
26. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R. and Subgroup, G.P.D.P. (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, **25**, 2078-2079.
27. Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G. and Mesirov, J.P. (2011) Integrative genomics viewer. *Nat Biotechnol*, **29**, 24-26.
28. Lister, R., Pelizzola, M., Dowen, R.H., Hawkins, R.D., Hon, G., Tonti-Filippini, J., Nery, J.R., Lee, L., Ye, Z., Ngo, Q.M. *et al*. (2009) Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature*, **462**, 315-322.
29. Wang, Z., Li, X., Jiang, Y., Shao, Q., Liu, Q., Chen, B. and Huang, D. (2015) swDMR: A Sliding Window Approach to Identify Differentially Methylated Regions Based on Whole Genome Bisulfite Sequencing. *PLoS One*, **10**, e0132866.

30. Quinlan, A.R. and Hall, I.M. (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, **26**, 841-842.
31. Yagi, M., Kishigami, S., Tanaka, A., Semi, K., Mizutani, E., Wakayama, S., Wakayama, T., Yamamoto, T. and Yamada, Y. (2017) Derivation of ground-state female ES cells maintaining gamete-derived DNA methylation. *Nature*, **548**, 224-227.
32. Yagi, M., Kabata, M., Ukai, T., Ohta, S., Tanaka, A., Shimada, Y., Sugimoto, M., Araki, K., Okita, K., Woltjen, K. *et al*. (2019) De Novo DNA Methylation at Imprinted Loci during Reprogramming into Naive and Primed Pluripotency. *Stem Cell Reports*, **12**, 1113-1128.
33. Kim, D., Langmead, B. and Salzberg, S.L. (2015) HISAT: a fast spliced aligner with low memory requirements. *Nat Methods*, **12**, 357-360.
34. Ramírez, F., Ryan, D.P., Grüning, B., Bhardwaj, V., Kilpert, F., Richter, A.S., Heyne, S., Dündar, F. and Manke, T. (2016) deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res*, **44**, W160-165.
35. Pertea, M., Pertea, G.M., Antonescu, C.M., Chang, T.C., Mendell, J.T. and Salzberg, S.L. (2015) StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotechnol*, **33**, 290-295.
36. Takayama, K., Shimoda, N., Takanaga, S., Hozumi, S. and Kikuchi, Y. (2014) Expression patterns of dnmt3aa, dnmt3ab, and dnmt4 during development and fin regeneration in zebrafish. *Gene Expr Patterns*, **14**, 105-110.
37. Yagi, M., Yamanaka, S. and Yamada, Y. (2017) Epigenetic foundations of pluripotent stem cells that recapitulate in vivo pluripotency. *Lab Invest*, **97**, 1133-1141.
38. Eaton, J.D., Francis, L., Davidson, L. and West, S. (2020) A unified allosteric/torpedo mechanism for transcriptional termination on human protein-coding genes. *Genes Dev*, **34**, 132-145.
39. Tian, B. and Manley, J.L. (2017) Alternative polyadenylation of mRNA precursors. *Nat Rev Mol Cell Biol*, **18**, 18-30.
40. Nourse, J., Spada, S. and Danckwardt, S. (2020) Emerging Roles of RNA 3'-end Cleavage and Polyadenylation in Pathogenesis, Diagnosis and Therapy of Human Disorders. *Biomolecules*, **10**.
41. Al-Hadid, Q. and Yang, Y. (2016) R-loop: an emerging regulator of chromatin dynamics. *Acta Biochim Biophys Sin (Shanghai)*, **48**, 623-631.
42. Niehrs, C. and Luke, B. (2020) Regulatory R-loops as facilitators of gene expression and genome stability. *Nat Rev Mol Cell Biol*, **21**, 167-178.