bioRxiv preprint doi: https://doi.org/10.1101/2020.03.26.009977; this version posted March 28, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

# Maternal DNMT3A-dependent de novo methylation of the zygotic paternal genome inhibits gene expression in the early embryo

Running title: Zygotic de novo DNAme inhibits paternal expression

Julien Richard Albert[1], Wan Kin Au Yeung[2], Keisuke Toriyama[2], Hisato Kobayashi[3,4], Ryutaro Hirasawa[5], Julie Brind'Amour[1], Aaron Bogutz[1], Hiroyuki Sasaki[2] & Matthew Lorincz*[1]

[1] Department of Medical Genetics, University of British Columbia

[2] Division of Epigenomics and Development, Medical Institute of Bioregulation, Kyushu University, Fukuoka, Japan

[3] NODAI Genome Research Center, Tokyo University of Agriculture, Tokyo, Japan

[4] Present address: Department of Embryology, Nara Medical University, Nara, Japan

[5] Japan Agency for Medical Research and Development

* Corresponding author: mlorincz@mail.ubc.ca

Key words: DNA methylation, *Dnmt3a*, zygote, allelic transcription, histone modification

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.26.009977; this version posted March 28, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

## ABSTRACT

De novo DNA methylation (DNAme) during mammalian spermatogenesis yields a densely methylated genome, with the exception of CpG islands (CGIs), which are hypomethylated in sperm. Following fertilization, the paternal genome undergoes widespread DNAme loss before the first S-phase. Paradoxically, recent mass spectrometry analysis revealed that a low level of de novo DNAme occurs exclusively on the zygotic paternal genome. However, the loci involved and impact on genic transcription was not addressed. Here, we employ allele-specific analysis of whole-genome bisulphite sequencing (WGBS) data and show that a number of genomic regions, including several dozen CGI promoters, are de novo methylated on the paternal genome in 2-cell embryos. A subset of these promoters maintains DNAme through development to the blastocyst stage. Consistent with zygotic paternal DNAme acquisition (PDA), many of these loci are hypermethylated in androgenetic blastocysts but hypomethylated in parthenogenetic blastocysts. Strikingly, PDA is lost following maternal deletion of *Dnmt3a*. Furthermore, a subset of promoters showing PDA which are normally transcribed from the paternal allele in blastocysts show premature transcription at the 4-cell stage in maternal *Dnmt3a* knockout embryos. These observations uncover an unexpected role for maternal DNMT3A activity in post-fertilization epigenetic reprogramming and transcriptional silencing of the paternal genome.

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.26.009977; this version posted March 28, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

## INTRODUCTION

Male germ cell development in mammals is characterized by widespread de novo DNA methylation (DNAme) and compaction of DNA via histone-to-protamine exchange[1]. As DNAme is largely maintained throughout spermatogenesis, the genomes of mature spermatozoa harbor characteristically high levels of DNAme[2]. Following fertilization, the paternal genome undergoes another profound change in chromatin state, including replacement of protamines with histones and a global reduction in DNAme before the first S-phase[3,4]. Subsequently, DNAme levels on both parental genomes are progressively reduced with each DNA replication cycle[5]. While passive demethylation in the early embryo is likely explained by sequestration of the maintenance DNA methyltransferase DNMT1 in the cytoplasm[6-8], the mechanism of active demethylation remains controversial. Indeed, while TET3-mediated oxidation followed by base excision repair has been implicated in this process, a TET-independent mechanism is also clearly involved[9-15]. Regardless, whole-genome bisulphite sequencing (WGBS) analyses reveal that DNAme levels on both parental genomes reach a low point in inner cell mass (ICM) cells of embryonic day 3.5 (E3.5) mouse blastocysts, followed by widespread de novo DNAme during post-implantation development[16-18]. This wave of genome-wide demethylation followed by remethylation is conserved in human embryonic development, albeit with slower kinetics[19-21]. Notably, disruption of the machinery required for the establishment or maintenance of DNAme result in infertility and/or embryonic lethality in mice, revealing the importance of DNAme homeostasis in early mammalian development[7,8,13,22-27].

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.26.009977; this version posted March 28, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

CpG islands (CGIs), short stretches of CpG-rich regions, are an exception to the widespread DNAme observed in spermatozoa. Indeed, the vast majority of CGIs, which encompass the promoter regions of most housekeeping genes, are hypomethylated in mature male and female germ cells, as well as in early embryonic development[18,28-33]. In addition, CGIs retain nucleosomes in spermatozoa, foregoing the exchange for protamines[28-30]. These retained nucleosomes include canonical H3 or its variant H3.3, which are enriched for di- and/or tri-methylation on lysine 4 (H3K4me2/3)[30,34,35]. Although the extent of histone post-translational modification (PTM) maintenance following fertilization is controversial[34,36,37], the retention and/or rapid deposition of H3K4 methylation at CGI promoters may protect these regions against de novo DNAme[38] in the developing male germline as well as the early embryo. Indeed, most CGI promoters are enriched for H3K4me3 and remain hypomethylated on both parental genomes throughout early embryonic development and in adult tissues[16,39]. H3K4me2/3 also likely facilitate the initiation of transcription from the paternal genome during zygotic gene activation (ZGA)[28,34], which marks the transition between oocyte and embryonic transcriptional programmes[20]. As in sperm, CGIs in oocytes are generally hypomethylated and harbor nucleosomes enriched for H3K4me3[36] and/or H3K27me3[40,41]. However, a subset of CGIs, are de novo methylated in growing oocytes by DNMT3A, which is highly expressed in oocytes[42-44]. Paradoxically, despite widespread DNAme loss from the paternal genome in mouse zygotes, maternal DNMT3A is also clearly detected in the paternal pronucleus at this stage[14,27] and ongoing *de novo* DNAme is required for maintaining paternal allele DNAme of the paternally imprinted *H19* locus in early embryos[45]. Further evidence that the zygotic paternal genome is subject to *de novo* DNAme was demonstrated in a recent study

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.26.009977; this version posted March 28, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

employing immunofluorescence (IF) and ultrasensitive liquid chromatography/mass spectrometry[14]. However, the genomic regions subject to such paternal DNAme acquisition (PDA) in the early embryo and the relevance of this phenomenon to transcriptional regulation of the paternal genome has not been systematically addressed.

To determine which loci gain DNAme immediately following fertilization, we carried out an allele-specific analysis of WGBS data from 2-cell (2C) F1 hybrid embryos[16] and identified specific genomic regions, including CGI promoters, that show PDA. Corroborating these enigmatic findings, we observe PDA of an overlapping set of CGI promoters in androgenetic but not in parthenogenetic blastocysts. Allele-specific analysis of ChIP-seq data from 2C embryos reveals that PDA is accompanied by loss of H3K4me3 over the same regions, indicating that such DNAme may inhibit transcription from the paternal allele in early embryonic development. Indeed, we show that PDA is lost in the absence of maternal DNMT3A and a subset of hypomethylated genes are concomitantly upregulated specifically from the paternal allele in the 4C embryo. Taken together, these experiments reveal that beyond its role in maternal imprinting, DNMT3A methylates a subset of genes on the paternal genome in the zygote, inhibiting their expression in preimplantation development.

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.26.009977; this version posted March 28, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

## RESULTS

### The paternal genome undergoes de novo DNAme following fertilization

To trace parent-specific DNAme levels following fertilization and throughout embryonic development with single-nucleotide resolution, we first processed publicly available WGBS data derived from primordial germ cells (PGCs), spermatozoa and MII oocytes[2,31,46]. We then applied our recently developed allele-specific pipeline MEA[47] to WGBS data generated from 2C (55X coverage) as well as 4C, ICM, E6.5 and E7.5 F1 hybrid embryos[16]. This integrated analysis yielded female/maternal and male/paternal DNAme profiles through fertilization and early development (**Fig. S1a-b**). Consistent with previous IF data[3,4], comparison of DNAme levels in mature gametes and 2C embryos reveals an overall decrease on the maternal and paternal genomes of 8% and 43%, respectively (**Fig. S1c**). Surprisingly however, coincident with global DNAme loss across the paternal genome, robust DNAme gain (defined as an increase of ≥30%) was detected at ~4% of all hypomethylated regions (≤20%) in sperm, totalling 1.4 Mbp of the mappable genome (**Fig. 1a**). De novo DNAme of the paternal genome is consistent with the rapid translocation of maternal DNMT3A into the zygotic paternal pronucleus (**Fig. S1d**), as reported previously[14,27]. Remarkably, regions showing clear evidence of paternal DNAme acquisition (PDA) are enriched over annotated genic transcription start sites (TSSs) ($\chi^2$ test p=1.3E-211 **Fig. 1b**), particularly over CpG rich promoters ($\chi^2$ test p=0.002, **Fig. 1c**). Furthermore, a subset of TSSs showing PDA maintain such DNAme on the paternal genome to the blastocyst stage **(Fig. 1d)**. While PDA is not restricted to CpG-rich regions, we focused our analyses on CGI promoters, as DNAme is reported to have the strongest impact on transcription of this class of promoters[48,49].

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.26.009977; this version posted March 28, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

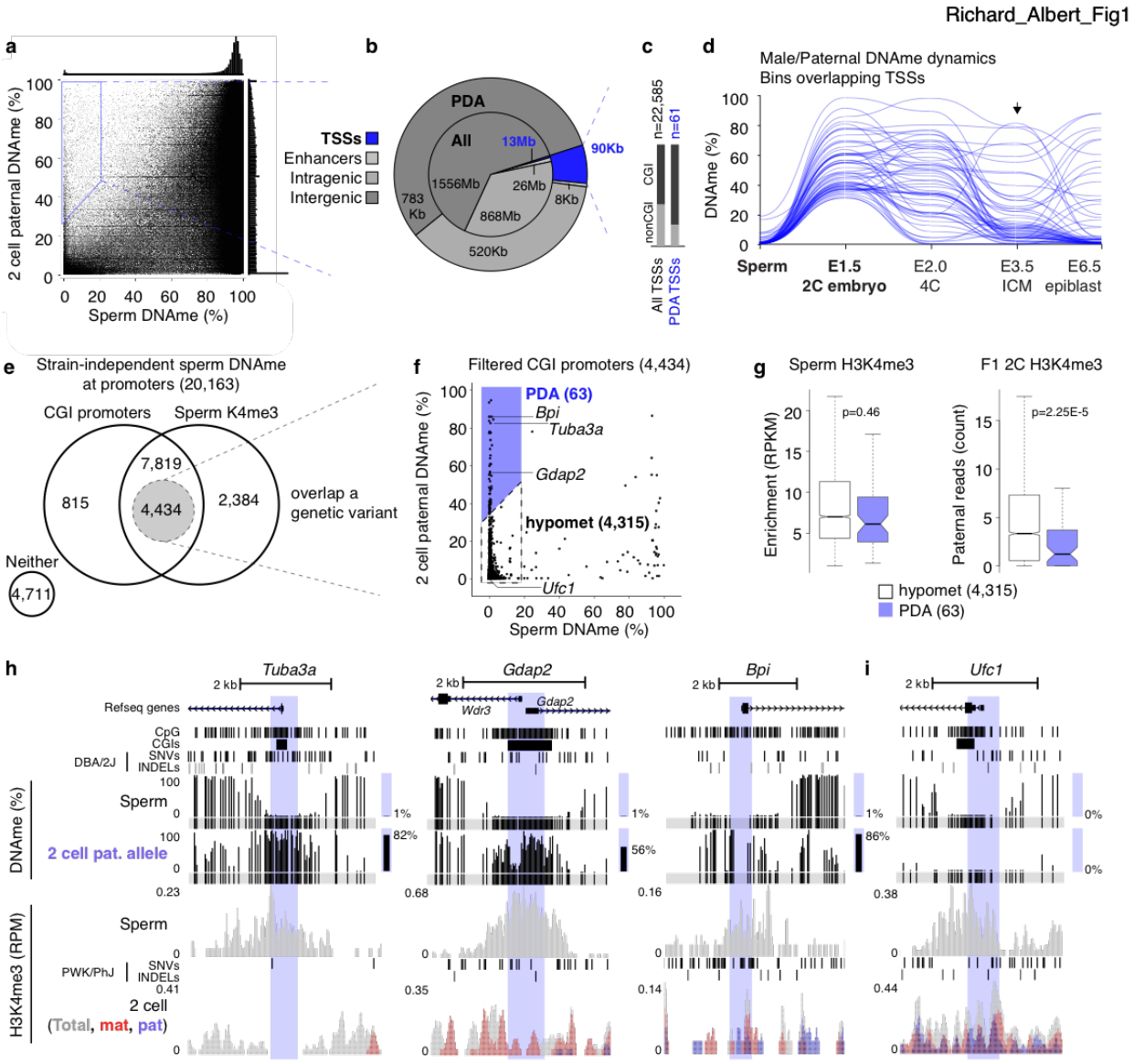


**Figure 1.** Paternal DNAme acquisition (PDA) at CpG-rich promoters following fertilization. **(a)** Male/paternal DNAme levels in sperm and 2C F1 hybrid embryos. Each data point represents the mean DNAme level over a 600bp genomic bin overlapping at least 2 informative CpGs (covered by 5X reads and separated by >1 sequencing read length) in both datasets (n=3,611,409). Bins showing PDA (≥30% gain in paternal DNAme) are highlighted. Histograms depicting the distribution of DNAme levels in both datasets are included. **(b)** Pie chart illustrating the genomic lengths of TSSs, enhancers, gene bodies (intragenic) and intergenic loci: genome-wide (inner circle) and over regions that show PDA (outer circle). **(c)** The proportion of CpG island promoters in annotated autosomal TSSs versus those that undergo PDA is shown. **(d)** Parallel coordinate plot illustrating male/paternal DNAme dynamics over TSSs that show PDA in

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.26.009977; this version posted March 28, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

sperm and the developing embryo (n=61) **(e)** Venn diagram including all annotated autosomal promoters with sperm DNAme levels consistent between C57BL/6J and DBA/2J strains. The proportion of such promoters with high/intermediate CpG density (CGIs) and enriched for H3K4me3 in sperm is depicted. The subset overlapping a genetic variant and at least 2 informative CpGs in sperm and 2C WGBS datasets are highlighted. **(f)** CGI promoters showing PDA (≥30% gain in paternal DNAme) from sperm to the 2C stage (n=63) versus persistent hypomethylation (hypomet, n=4,315) are shown. **(g)** Distribution of H3K4me3 levels at PDA or hypomet CGI promoters is shown for sperm (left) and on the paternal allele in 2C embryos (right). **(h-i)** UCSC genome browser screenshots of the promoter regions of *Tuba3a*, *Gdap2*, *Bpi* and *Ufc1*. TSS regions (+/- 300bp) are highlighted in blue and the location of informative CpGs (5X coverage) for each WGBS dataset are highlighted in grey below each WGBS dataset. Mean DNAme levels over highlighted regions are indicated to the right of each DNAme genomic track. The genomic locations for NCBI Refseq genes, all CpG dinucleotides, CGIs and genetic variants used in our allele-specific analyses (SNVs and INDELs) are also included. 2C H3K4me3 data is represented as a composite track containing total (allele-agnostic, grey), maternal (red) and paternal (blue) genomic tracks.

To generate a curated list of CpG-rich hypomethylated TSSs in sperm, we categorized promoters by CpG density (high, intermediate and low), as described for the human genome[49]. As expected, DNAme levels and CpG density are anti-correlated (**Fig. S2a**). To minimize the potential confounding effects of strain-specific differences when comparing DNAme levels of parental genomes, we next determined the variation in DNAme levels in C57BL/6J versus DBA/2J sperm using published WGBS datasets[2,16]. Methylation profiles from these strains show a strong correlation ($r^2$=0.98, **Fig. S2b**), with 20,163 promoters showing consistent DNAme levels (high or low) in

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.26.009977; this version posted March 28, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

both strains. Promoters showing strain-specific hypomethylation (185 in DBA/2J and 82 in C57BL/6J) were excluded from further analysis. Taking advantage of the fact that DNAme and H3K4me3 are anticorrelated in soma and germ cells, we further refined our list of CGI promoter TSSs using publicly available H3K4me3 ChIP-seq data from C57BL/6J sperm[30]. As expected, H3K4me3 enrichment levels show a positive correlation with CpG density and negative correlation with DNAme (**Fig. S2a**). Selecting TSSs that show H3K4me3-enrichment (RPKM ≥1) and intermediate to high CpG density (CGI promoters, CpG ratio ≥0.12) yielded a list of 12,253 CGI promoters, the vast majority of which are hypomethylated in sperm of both strains (mean DNAme; C57BL/6J=2.1%, DBA/2J=2.0%). Of these, 4,434 harbor a SNV or INDEL and had sufficient WGBS coverage (5X allele-specific read coverage over ≥2 CpGs separated by >1 sequencing read length) to score DNAme levels in both sperm and the paternal genome of C57BL/6J x DBA/2J F1 2C embryos (**Fig. 1e**). To unequivocally identify targets of post-fertilization de novo DNAme, we focussed on these CGI promoters.

While the vast majority of CGI promoters hypomethylated in sperm maintain low paternal DNAme levels in 2C embryos, 63 showed clear evidence of PDA (defined as a ≥30% gain, **Fig. 1f and Supplemental Table 1**). Notably, lower levels of H3K4me3 in sperm are unlikely to explain their propensity to gain DNAme, as the distribution of H3K4me3 levels in sperm was not significantly different between these CGIs and those that showed no gain in DNAme (p=0.46, **Fig. 1g**). However, relative to CGIs that remain unmethylated, PDA CGI promoters show a significantly greater decrease of H3K4me3 levels over the paternal allele in 2C embryos (p=2.25E-5). Though we cannot discriminate between active demethylation of H3K4 and histone H3 turnover, these

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.26.009977; this version posted March 28, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

results reveal that de novo DNAme of CGIs on the paternal genome is generally accompanied by a reduction of H3K4me3. For example, *Tuba3a*, *Gdap2* and *Bpi* gain DNAme across 28, 58 and 7 CpGs proximal to their TSSs (+/- 300bp), respectively, coincident with loss of H3K4me3 on the paternal allele (**Fig. 1h**), while the control gene *Ufc1* displays persistent DNA hypomethylation and H3K4me3 enrichment on the paternal genome both pre- and post-fertilization (**Fig. 1i**). Surprisingly, only 11 of the PDA genes are methylated (≥20% DNAme) in MII oocytes and none gain DNAme on the maternal genome in 2C embryos (**Fig. S1e**). In contrast, maternally imprinted genes, such as *Impact* and *Snurf*, show maternal allele-specific DNA hypermethylation in the ICM and paternal allele-specific enrichment of H3K4me3 throughout early embryonic development, as expected (**Fig. S2c**). Taken together, these analyses reveal that a specific subset of CGI promoters are de novo DNA methylated exclusively on the paternal genome following fertilization, concomitant with reduced H3K4me3.

**DNAme at many PDA loci is maintained through the blastocyst stage**

To determine whether DNAme at PDA sites persists through the wave of global DNAme erasure, we scored paternal DNAme levels at these loci using WGBS data from F1 hybrid ICM cells[16]. Of 63 PDA loci, 53 had sufficient allele-specific WGBS coverage to ascertain paternal DNAme levels in ICM cells. Relative to CGI promoters that remain hypomethylated following fertilization, those that show PDA retain higher DNAme on the paternal allele in ICM cells (p=4.97E-6, **Fig. 2a**), albeit at lower levels than observed in 2C embryos. In contrast, 19 CGI promoters that show PDA, including *Gdap2* (**Fig. 2b**), are hypomethylated (mean <5%) in blastocysts (**Supplemental Table 1**). Thus, while DNAme at a subset of CGI promoters showing PDA is transient, others either resist

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.26.009977; this version posted March 28, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

DNA demethylation or are reiteratively de novo methylated in early embryonic development.

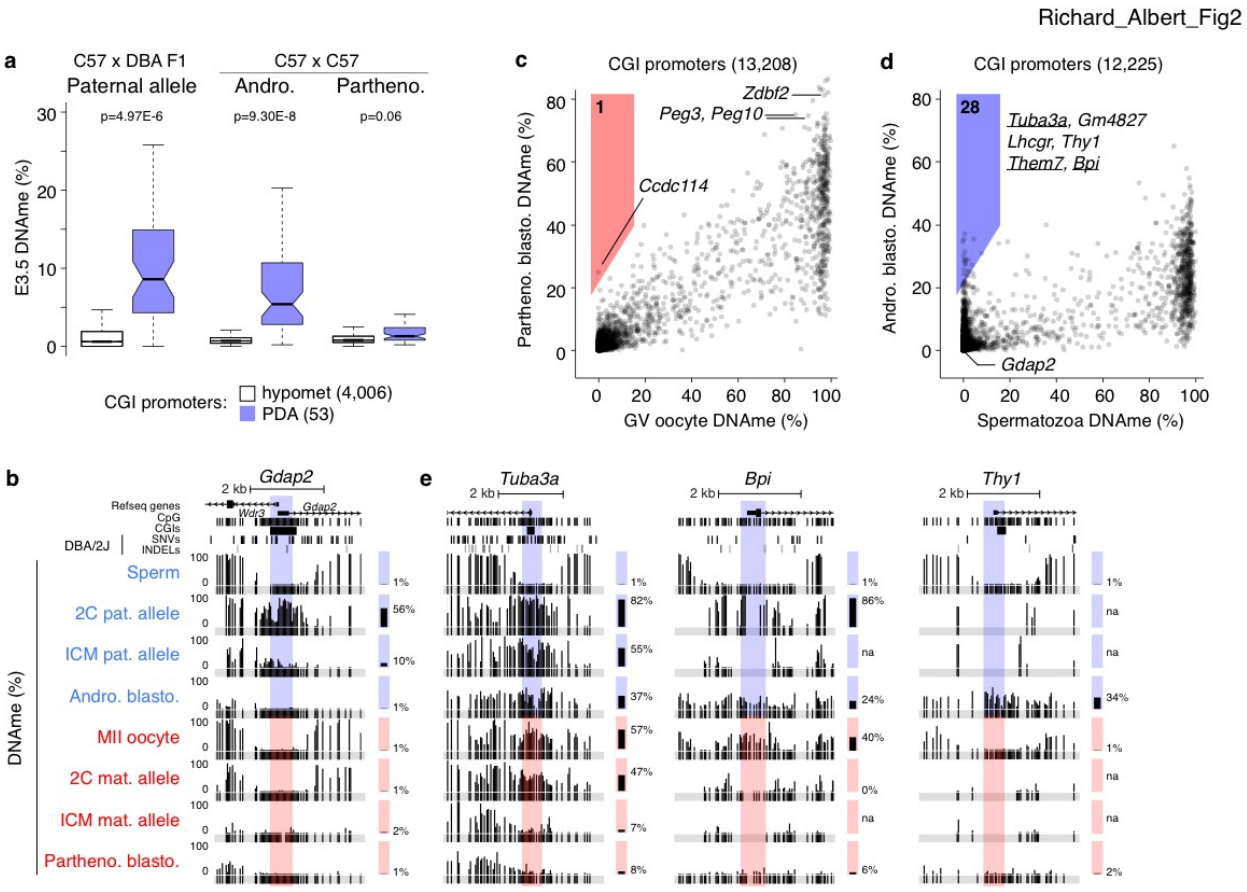


**Figure 2.** Paternal DNAme levels at many PDA sites are maintained in normal and androgenetic blastocysts. **(a)** Distribution of CGI promoter DNAme levels in embryonic day 3.5 C57 x DBA F1 ICM, androgenetic blastocysts and parthenogenetic blastocysts. Outliers not shown. **(b)** Screenshot of the *Gdap2/Wdr3*, CGI promoter, presented as in Fig. 1. **(c-d)** 2D scatterplots depicting DNAme levels over individual CGI promoters in E3.5 **(c)** parthenogenetic blastocysts versus GV oocyte (data excludes polar bodies; n=13,208) and **(d)** androgenetic blastocysts versus sperm (n=12,225). Promoters that gain ≥20% DNAme are highlighted and genes previously identified using our allele-specific DNAme analysis pipeline are underlined. **(e)** Screenshots of the *Tuba3a*, *Bpi* and *Thy1* CGI promoters.

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.26.009977; this version posted March 28, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

To confirm the persistence of paternal allele-specific DNAme and expand upon the number of loci at which PDA is likely occurring in the early embryo, we isolated genomic DNA from isogenic (C57BL/6NJcl) androgenetic blastocysts, which contain only paternally inherited genomes, and conducted WGBS. Compared to CGI promoters that remain hypomethylated in 2C embryos, those that show PDA exhibited a significantly greater level of DNAme in such bipaternal blastocysts (p=9.30E-8, **Fig. 2a**). Furthermore, the majority of CGI promoters that show persistence of DNAme (≥20%) on the paternal allele in ICM cells, including 8 PDA genes, show ≥10% DNAme in androgenetic blastocysts (**Fig. S2d**). In contrast, analysis of our previously published WGBS data from parthenogenetic blastocysts[43], in which both genomes are maternally derived, reveals that DNAme remains low at all of the CGI promoters showing zygotic PDA (**Fig. 2a**), with the exception of 6 of the 11 that are already hypermethylated in MII oocytes (**Supplemental Table 1**). Thus, loci showing PDA are de novo methylated in the early embryo exclusively when at least one paternal genome is present.

Given that androgenetic and parthenogenetic blastocysts are uniparental, we were able to extend our parental genome-specific DNAme analysis from 18 to 90% of all annotated autosomal CGI promoters. As expected, maternally imprinted CGI promoters, such as *Peg10* and *Zdbf2*, which are hypermethylated exclusively on the maternal allele in normal embryos, remain hypermethylated in parthenogenetic blastocysts (**Fig. 2c**). However, only one CGI promoter (*Ccdc114*) shows a ≥20% DNAme gain in these cells relative to germinal vesicle (GV) oocytes. In contrast, 28 CGI promoters show such a gain in androgenetic blastocysts relative to sperm, confirming that the paternal genome is the preferred target for such post-fertilization de novo

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.26.009977; this version posted March 28, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

DNAme (**Fig. 2d and Supplemental Table 1**). While 17 of these loci (including *Thy1*) do not harbor a genetic variant and could therefore not be assessed in the F1 hybrid dataset, six, including *Tuba3a, Bpi, Them7, Shisa7, Syn3 and A230077H06Rik*, were also identified as PDA genes in our allele-specific analysis of F1 hybrid embryos (**Fig. 2e**) and four (*H1fnt, Dbx2, Tbx4* and *Prss39*) showed a gain in DNAme of 9-25%, just below our original threshold of >30%. These results indicate that PDA occurs independent of DBA/2J-specific variants and is thus a *bona fide* parent-of-origin effect at these loci.

**Relationship between histone PTMs and PDA**

To determine whether promoter regions showing PDA share a common DNA motif that may render them susceptible to de novo DNAme, we performed motif discovery using HOMER[50] and MEME[51]. However, we did not detect any common DNA motif around the TSSs (+/- 300 bp) of genes that show PDA. Therefore, we focussed on the relationship between histone PTMs and paternal DNAme acquisition and/or subsequent DNAme maintenance at these sites. We analyzed published H3K4me2[34], H3K4me3[30,36], H3K9me3[17], H3K27me3[30,40] and H3K36me3[37] ChIP-seq data from sperm, oocytes, zygote (1C), 2C and ICM cells using ChromHMM[52]. No single histone PTM or combination thereof in sperm predicted PDA at CGI promoters versus those that remain hypomethylated (**Fig. S3a-b**). However, a striking enrichment of H3K9me3 was observed in the zygote and 2C stage at PDA sites (**Fig. S3b**) and those that show persistence of DNAme to the ICM show an even greater enrichment of H3K9me3 in the zygote and 2C embryo (**Fig. S3c**). Indeed, the paternal allele of many CGI promoters show a positive association between the levels of H3K9me3 at the 2C stage and

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.26.009977; this version posted March 28, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

DNAme in the ICM (**Fig. S3d**), including that of the PDA gene *Tuba3a* (**Fig. S3d-e**). Thus, this PTM may protect regions showing PDA against loss of paternal DNAme in the preimplantation embryo, consistent with previous reports indicating that H3K9me3 plays a role in promoting DNMT1-dependent maintenance of DNAme[53,54].

**Most PDA target genes are expressed during spermatogenesis and silenced in the early embryo**

To determine whether PDA is associated with transcriptional inactivation, we analyzed existing RNA sequencing (RNA-seq) datasets from multiple stages in spermatogenesis as well as F1 hybrid preimplantation embryos[55,56]. While 59 of the 63 genes that show PDA are expressed during at least one stage of spermatogenesis, only 23 (including *Tuba3a* and *Gdap2*) are transcribed from the paternal allele in the early embryo (**Fig. 3a**). Nevertheless, DNAme persists at a subset of these genes through the blastocyst stage, albeit at intermediate levels, indicating that promoter DNAme status per se is not predictive of transcription from the paternal allele in the blastocyst.

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.26.009977; this version posted March 28, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

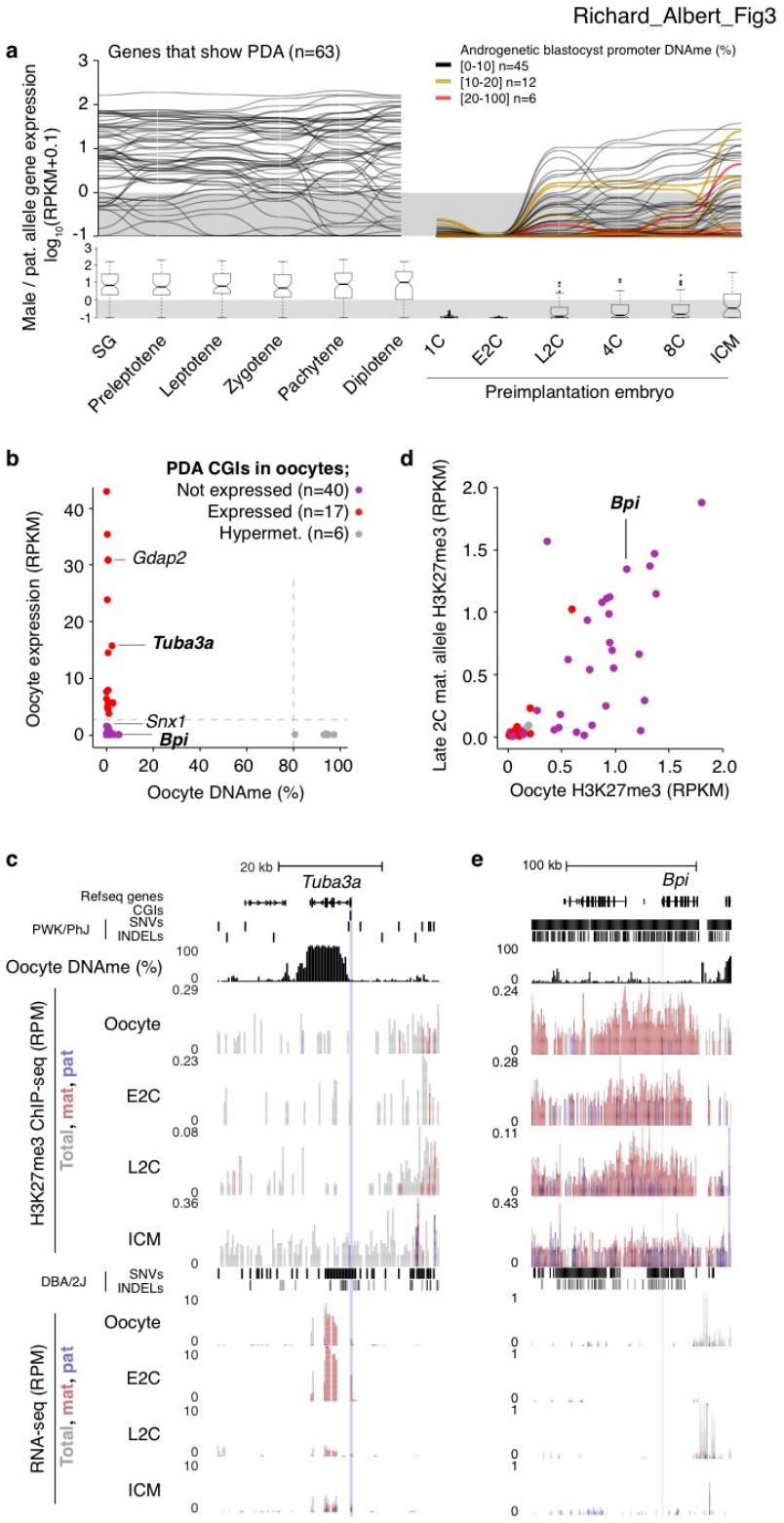


**Figure 3.** PDA genes not expressed in oocytes are enriched for repressive histone marks. **(a)** Parallel coordinate plot showing the temporal expression pattern of genes that show PDA in the male germline (SG to diplotene) and from the paternal allele the early embryo (1C to ICM). SG: spermatogonia. Lines representing paternal allele

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.26.009977; this version posted March 28, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

expression in the preimplantation embryo are colour coded based on CGI promoter DNAme levels observed in androgenetic blastocysts and the threshold for scoring genes as expressed (RPKM≥1) is indicated. The distribution of RPKM values for each developmental stage are shown as boxplots below. 6 PDA genes do not harbour exonic SNVs/INDELs and therefore could not be ascertained for paternal allele expression in F1 hybrid datasets (1C to ICM). **(b)** Mean DNAme (%) and expression (RPKM) levels in GV oocytes (data excludes polar bodies) over CGI promoters that show PDA. Hypermethylation and expression cut-off values (80 and 2, respectively) are indicated by dotted lines. Coloured dots are defined as hypomethylated in GV oocytes. **(c,e)** Genome browser screenshot of the *Tuba3a* and *Bpi* loci, including MII oocyte DNAme, H3K27me3 ChIP-seq and RNA-seq tracks, presented as in Fig. 1. **(d)** Enrichment of H3K27me3 flanking the same CGI promoters (+/-10kb) as in **b** in MII oocytes and the maternal allele of 2-cell embryos.

A similar analysis of the relationship between DNAme and expression in GV oocytes revealed that 57 of the 63 PDA loci are hypomethylated at their promoters. Surprisingly, only 17 of these hypomethylated genes, including *Tuba3a* and *Gdap2*, are transcribed at this stage (**Fig. 3b-c**). To determine whether the inactive genes harbor histone marks associated with transcriptional repression, we integrated previously published H3K27me3 and H3K9me3 ChIP-seq datasets derived from oocytes and F1 hybrid embryos[17,40]. Consistent with a role for H3K9me3 in promoting DNAme maintenance, all 6 hypermethylated PDA loci are enriched for this histone PTM in oocytes (**Fig. S4a-b**). In contrast, 23 of the 40 silent hypomethylated CGI promoters are embedded within H3K27me3-enriched domains (TSS +/-10kb, RPKM ≥0.5), 17 of which are maintained on the maternal allele at least to the late 2C stage (**Fig. 3d**). The TSS of *Bpi* for example is embedded within an extended H3K27me3 domain in oocytes which

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.26.009977; this version posted March 28, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

persists to the 2C stage (**Fig. 3e**). Notably, H3K27me3 was recently implicated in transcriptional silencing of genes subject to non-canonical maternal imprinting in the mouse[41,57]. These results reveal that at the 2C stage, the paternal and maternal alleles of a subset of PDA loci are distinctly marked by DNAme and H3K27me3, respectively.

**Maternal DNMT3A is required for PDA**

As DNMT3A is required for de novo DNAme in the female germline[31] and maternal DNMT3A persists in the zygote (**Fig. S1d**), we next wished to determine whether maternal DNMT3A is responsible for PDA. We crossed oocyte-specific homozygous *Dnmt3a* knock-out (matKO) C57BL/6J females with wild-type (WT) DBA/2J males and performed WGBS on F1 hybrids at the early-mid 2C stage (**Fig. 4a**, 5X mean sequencing coverage). As expected, global maternal DNAme levels in 2C matKO embryos mirror those in *Dnmt3a* matKO GV oocytes (**Fig. S5a**). Furthermore, while the paternal gametic differentially methylated regions (gDMRs) *H19*, *Dlk1-Gtl2* and *Rasgrf1* show no decrease, DNAme at maternal gDMRs is dramatically reduced in matKO relative to control embryos (**Fig. 4b**). Importantly, a positive correlation was observed when comparing paternal DNAme levels of CGI promoters with sufficient allele-specific coverage in our control dataset with published 2C WGBS data[16], with the majority of PDA genes consistently showing relatively high DNAme on the paternal allele (**Fig. S5b**). In the absence of maternal DNMT3A however, DNAme is lost on the paternal allele of all 23 CGI promoters showing PDA for which allelic methylation can be deduced (**Fig. 4c**). At the *Tuba3a*, *Gdap2* and *Snx1* promoters for example, PDA is lost in matKO embryos across all CpGs in the promoter region (**Fig. 4d**). While the remaining 40 PDA loci lack allele-specific coverage in our matKO samples, analysis of

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.26.009977; this version posted March 28, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

total DNAme levels reveals that 33 are clearly hypomethylated (<4%) (**Supplemental Table 1**), as shown for *Bpi* (**Fig. 4d**). Taken together, these results demonstrate that DNAme acquisition on the paternal allele immediately following fertilization is dependent upon maternal DNMT3A.

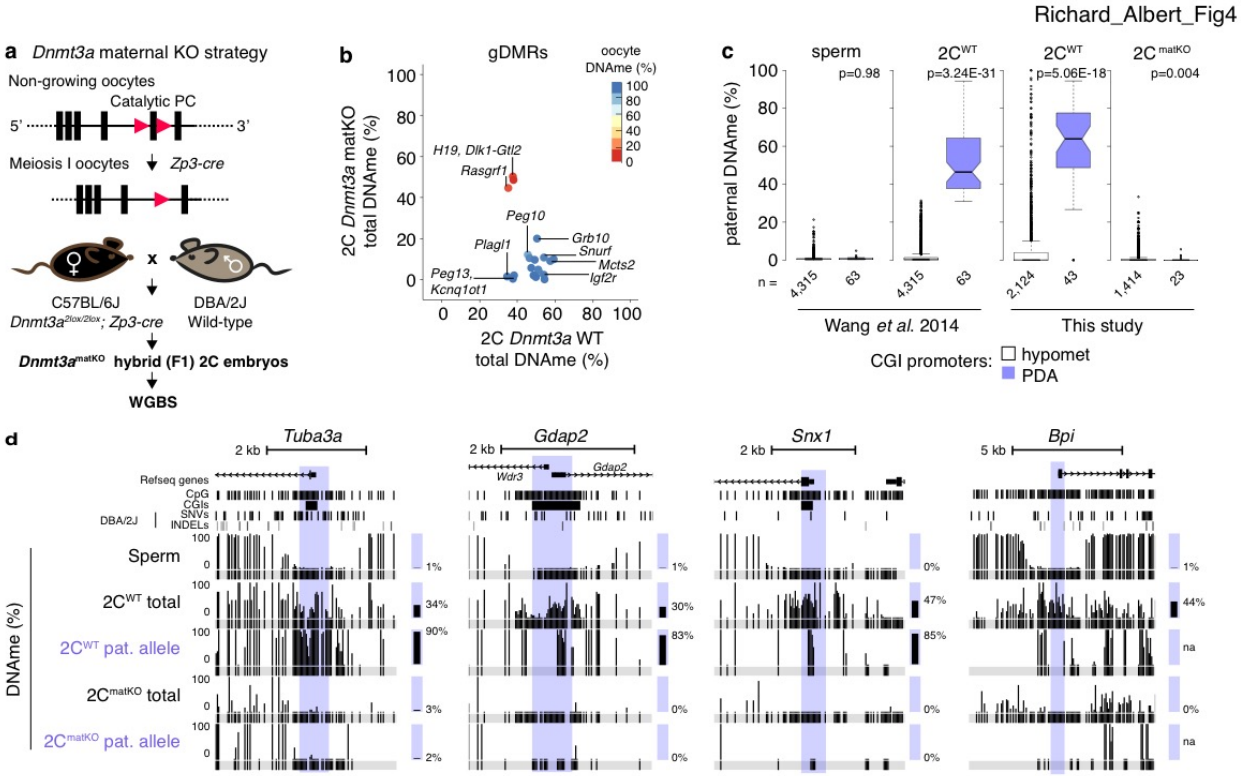


**Figure 4.** Paternal DNAme acquisition is mediated by maternal DNMT3A. **(a)** Maternal *Dnmt3a* knock-out (KO) strategy, including Flox sites (red triangles) and deleted exon 18. Oocytes from C57BL/6J matKO, mice were *in vitro* fertilized using sperm from wild-type (WT) DBA/2J males and PBAT data was generated from independent biological replicates (5X combined sequencing coverage). **(b)** Scatterplot showing the average total (allele-agnostic) DNAme levels over gDMRs overlapping at least 2 informative CpGs in *Dnmt3a* WT and matKO 2C embryos, with heat map of DNAme levels in GV oocytes. **(c)** Paternal DNAme level distribution over CGI promoters in sperm (left), normal 2C embryos (middle), and matKO 2C embryos (right). The number of CGI promoters represented is indicated below each plot. **(d)** Screenshots of the *Tuba3a*, *Gdap2/Wdr3*, *Snx1* and *Bpi* loci. Total (allele-agnostic) and paternal allele-specific DNAme levels are shown. na: no information available.

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.26.009977; this version posted March 28, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

**Loss of PDA results in ectopic expression from the paternal allele**

Hypermethylation of CGI promoters is associated with transcriptional silencing[48,49]. To determine whether DNAme of genes showing PDA impacts their expression specifically from the paternal allele, we conducted strand-specific RNA-seq on early-mid 2C as well as 4C and blastocyst-stage F1 matKO embryos (**Fig. 5a and Fig. S6a**). A strong correlation was observed between matKO and WT 2C, 4C and blastocyst-stage embryos as well as previously published WT transcriptomes of the same developmental stages (**Fig. S6b**), indicating that maternal depletion of DNMT3A does not disrupt global transcriptional programming. This is consistent with previous observations that *Dnmt3a* matKO embryos develop normally until E9.5[26,58]. We next determined whether genes known to be regulated by maternally established genomic imprints show loss of imprinting in *Dnmt3a* matKO embryos. Consistent with their loss of DNAme (**Fig. 4b**), *Mcts2*, *Plagl1*, *Snurf*, *Peg10* & *Zdbf2* imprinted genes were significantly upregulated (≥2-fold change, Wald Chi-squared test, Benjamini-Hochberg adjusted p-value ≤0.1) from the maternal allele (**Fig. S6c-e**). Thus, DNMT3A expression in the oocyte is required for maternal allele-specific transcriptional silencing of a subset of genes in the preimplantation embryo.

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.26.009977; this version posted March 28, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

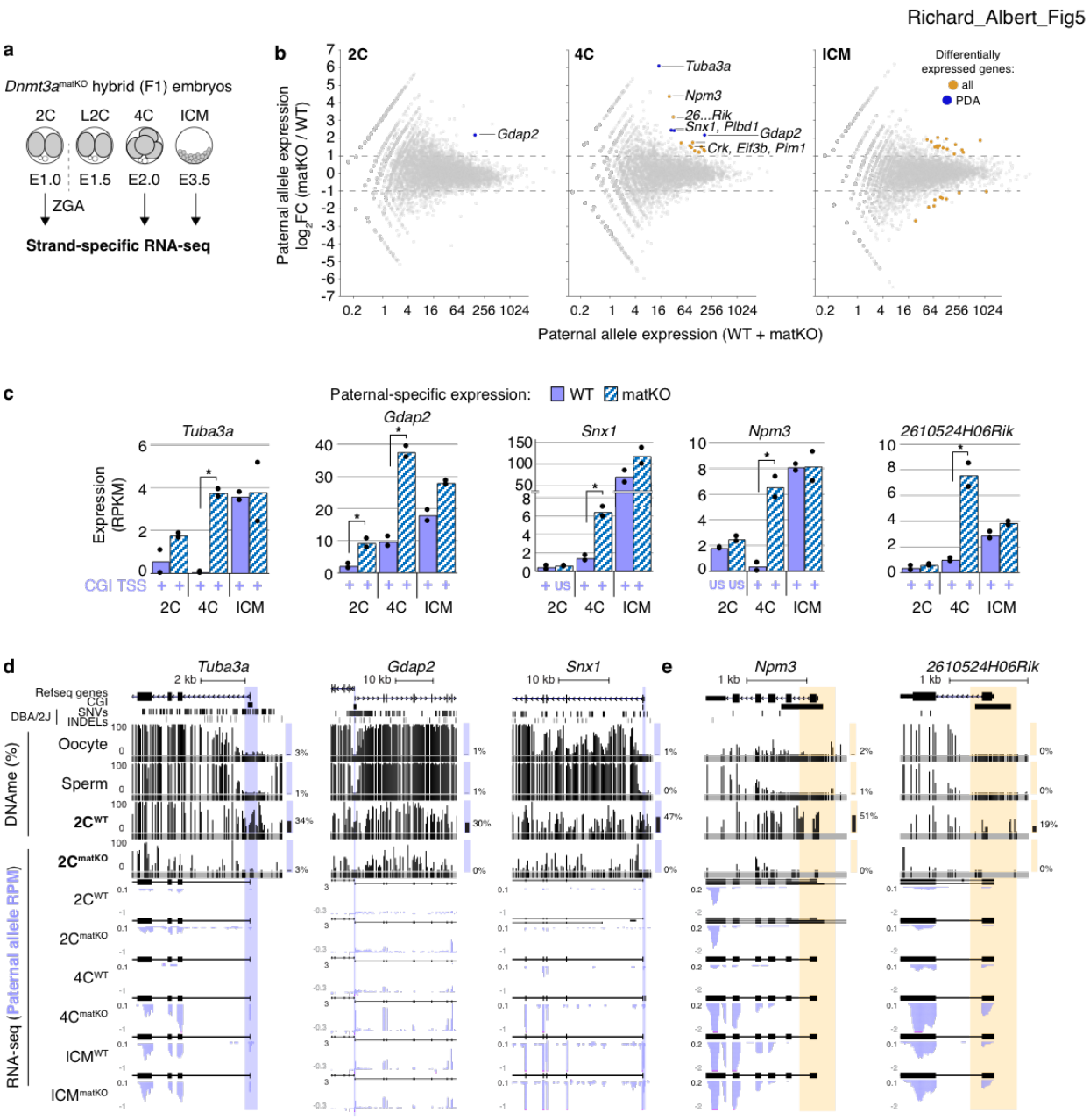


**Figure 5.** Impact of maternal DNMT3A deletion on expression from the paternal allele. **(a)** RNA-seq libraries were generated in biological duplicates for wild-type and *Dnmt3a* matKO early F1 2C, 4C and ICM embryos. **(b)** Scatterplots showing average versus differential paternal-allele expression for 2C and 4C embryos as well as ICM cells. Statistically significant differentially expressed genes (≥2-fold upregulation, Benjamini-Hochberg adjusted P-value ≤0.1) are highlighted in orange, and those that show PDA at their CGI promoters are highlighted in blue. **(c)** Bar graphs illustrating differential expression of select genes from the paternal genome in WT and matKO embryos. Each bar represents the mean expression value of biological replicates. Statistically

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.26.009977; this version posted March 28, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

significant upregulation (≥2-fold, Benjamini-Hochberg adjusted P-value ≤0.1) is indicated by an asterisk. Whether transcription of each gene initiates from the CGI promoter is indicated below each bar. US: upstream. **(d-e)** Screenshots of *Tuba3a, Gdap2, Snx1, Npm3* and *2610524H06Rik,* as presented in Fig. 1. RNA-seq data is represented as a composite track of biological replicates. *De novo* transcript assembly is shown above each RNAseq dataset. Total DNAme levels in adult gametes and 2C WT and *Dnmt3a* matKO embryos are included. Yellow highlights indicate the TSS (+/- 300 bp) of genes that do not contain DBA/2J SNVs or INDELs and therefore could not be analyzed for allele-specific DNAme state.

A similar analysis of genes upregulated from the paternal allele in 2C matKO embryos yielded only *Gdap2, a* PDA gene, as significantly upregulated (**Fig. 5b-c**). Further, of the 16 significantly upregulated genes in 4C matKO embryos, four: *Gdap2*, *Snx1*, *Tuba3a* and *Plbd1*, are PDA genes. In contrast, none of the 15 genes upregulated in ICM are PDA genes and thus likely represent either loci that are methylated post-fertilization on distal regulatory elements or indirect effects of maternal DNMT3A loss. Indeed, 5 of these genes are also significantly upregulated from the maternal allele (**Supplemental Table 1**). Importantly, transcription of all upregulated PDA genes initiates from the aberrantly hypomethylated CGI promoter, excluding alternative promoter usage as an explanation for the increased expression from the paternal allele (**Fig. 5c-d**). Furthermore, no significant upregulation of PDA genes was observed from the maternal allele, consistent with DNAme-independent silencing at this stage (**Fig. 3d**), or when total (allele-agnostic) transcript levels were analyzed (**Fig. S7a-c**). Thus, upregulation of PDA genes occurs exclusively from the paternal allele, raising the question, can expression data be exploited to identify additional PDA genes?

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.26.009977; this version posted March 28, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Due to the density of naturally occurring polymorphisms and depth of WGBS sequencing coverage, paternal DNAme dynamics could be measured in our earlier F1 hybrid analysis over only 4,434 of the 12,253 filtered autosomal CGI promoters (**Fig. 1e**). Of the additional 7,724 autosomal genes with CGI promoters that include at least 1 exonic genetic variant between parental strains and are expressed from the paternal allele in our RNA-seq data, 5 were significantly upregulated from the paternal genome in matKO 4C embryos, including *Npm3*, *Eif3b*, *2610524H06Rik*, *Crk* & *Pim1* (**Fig. 5b-c**). Importantly, as for *bona fide* PDA loci, none of these genes show a change in expression from the maternal allele (**Fig. S7b,d**), nor are they upregulated from the paternal allele in matKO ICM (**Fig. 5b-c**). Furthermore, analysis of total DNAme levels clearly shows that these loci are methylated in WT but not in matKO 2C embryos, indicating that these CGI promoters are direct targets of maternal DNMT3A (**Fig. 5e**). While the lack of genetic variants precludes the measurement of allele-specific DNAme over these regions, analysis of allele-agnostic DNAme levels in WT 2C embryos revealed that 4 of these 5 CGI promoters show a ≥18% gain in DNAme relative to their methylation state in sperm and oocyte (**Supplemental Table 1**). Thus, in addition to the PDA genes described above, these 4 loci are likely de novo methylated on the paternal allele in 2C embryos, and repressed by this mark in 4C embryos. Taken together, these results reveal that the absence of maternal DNMT3A in the early embryo leads to failure of de novo DNAme on the paternal allele of a subset of genes and, in turn, their ectopic expression.

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.26.009977; this version posted March 28, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

## DISCUSSION

Using allele-specific analyses of early embryos, we determined that 4% of all hypomethylated regions in sperm, including at least 63 CGI promoters, are de novo methylated on the paternal genome by the 2C stage. An Independent analysis of androgenetic blastocysts revealed 86 CGI promoters that show a ≥10% gain in DNAme relative to sperm, 15 of which overlap with the PDA genes identified in normal 2C embryos (**Supplemental Table 1**). These observations are particularly surprising, given that the hypermethylated sperm genome is globally demethylated in the mouse zygote. Indeed, following polyspermic fertilization, where methylation of the maternal genome remains relatively constant, up to five paternal genomes are demethylated in the zygote[4]. Although zygotic de novo DNAme of the maternal genome has been reported for the TKZ751 transgene and *Igf2*[59], ETn retroelements[60] and H3K9me2-enriched regions[61], to our knowledge, this is the first report to characterize the specific regions of the paternal genome subject to de novo DNA methylation in the mouse zygote beyond the *H19* gDMR[45,62,63].

To measure DNAme and chromatin dynamics at an allele-specific level in the early embryo, earlier studies relied on IF-based assays, which are inherently of low resolution and therefore uninformative with respect to specific genomic loci[3,4,10]. Further, while such studies revealed chromosome-scale dynamics in the early embryo, dissecting the interplay between local DNAme, histone PTMs and transcription is not possible using IF. Our integrated allele-specific analysis reveals that loss of H3K4me3 and PDA occur shortly following fertilization, consistent with the observation that DNMT3A recognizes the unmethylated state of H3K4[38,64]. Notably, persistence of H3K4

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.26.009977; this version posted March 28, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

methylation on the paternal genome was previously reported to play an important role in gene regulation in 2C embryos[34]. Our results reveal that at a distinct subset of loci, loss of H3K4 methylation may be a pre-requisite for de novo DNAme, and in turn transcriptional silencing.

How DNMT3A is targeted to PDA loci remains to be determined. While we did not uncover any common DNA motifs, it is possible that multiple different DNA binding factors bind to such regions to promote PDA. Interestingly, all but 3 PDA sites concomitantly gain H3K9me3 (RPKM ≥1) in the zygote, raising the possibility that this mark may promote and/or maintain acquired DNAme. As KRAB-ZFPs have been shown to promote H3K9me3 deposition and potentially de novo DNAme at gDMRs[65-67], repetitive elements[68] and genic promoters[69], binding of yet to be characterized KRAB-ZFPs, in complex with TRIM28, may be responsible for sequence-specific targeting of DNMT3A to regions showing PDA. Alternatively, previous studies revealed that a specific set of CpG-rich germline gene promoters harbor E2F6 and E-box motifs that promote binding of the non-canonical PRC1 complex PRC1.6 and de novo DNAme in post-implantation embryos[70,71]. Furthermore, such de novo DNAme occurs in conjunction with H3K9me3 acquisition[72,73]. However, de novo DNAme at these germline gene promoters is DNMT3B-dependent and the E2F6 motif is present in the promoter region of only 4 of the loci showing PDA. Additional studies will be required to determine whether specific transcription factors and/or chromatin features are required for de novo DNA methylation of the paternal genome.

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.26.009977; this version posted March 28, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

*Dnmt3a* maternal KO embryos die during post-implantation development at around E9.5-10.5[26,58]. While it is tempting to speculate that aberrant expression of PDA genes from the paternal allele plays a role, our temporal analysis reveals that this is a transient phenomenon, with mRNA levels of these genes indistinguishable from wild type by the blastocyst stage. Alternatively, aberrant expression of imprinted genes as a consequence of failure of de novo DNAme in the oocyte may be responsible for such embryonic lethality[23,74]. However, we find that several maternally imprinted alleles are already expressed from the hypomethylated maternal allele in *Dnmt3a* matKO ICM cells (~E3.5). Thus, as for genes showing PDA, aberrant expression of at least a subset of maternally imprinted genes occurs well before the gross phenotypic effects are manifest. In summary, this study reveals that maternal DNMT3A is required for de novo DNAme of specific regions on the paternal genome immediately following fertilization, and in turn silencing of a subset of methylated genes on the paternal allele in early mouse embryonic development. Whether maternal DNMT3A acts on the paternal genome following fertilization in other mammals, and what effect such DNAme has on transcriptional regulation of target genes remains to be determined.

## MATERIALS AND METHODS

**Ethical approval for animal work.** All animal experiments were performed under the ethical guidelines of Kyushu University and Tokyo University of Agriculture.

**Isolation of androgenetic blastocyst.** Diploid androgenones were prepared as previously described[75]. Oocyte and spermatozoa were isolated from B6D2F1/Jcl and

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.26.009977; this version posted March 28, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

C57BL/6NJcl mice (Clea Japan, Tokyo, Japan), respectively. Briefly, enucleated oocytes were *in vitro* fertilized and zygotes with two male pronuclei were cultured for 4 days in KSOM medium at 37°C and 5% $CO_2$[76]. 5 blastocyst pools were collected in triplicate for WGBS library construction.

***Dnmt3a* maternal KO embryo culture and genotyping.** *Dnmt3a* KO oocytes were generated using *Dnmt3a*$^{2lox}$ and *Zp3*-cre C57BL/6J mice as described previously[26,77]. Superovulation was induced using PMSG/hCG and MII oocytes were collected from oviducts. *Dnmt3a*$^{2lox}$;*Zp3*-cre MII oocytes were artificially inseminated with DBA/2J spermatozoa. Cumulus cells were removed using hyaluronidase after insemination and embryos were cultured in KSOM at 37°C and 5% $CO_2$. Early-mid 2C embryos were collected at 22 hours (WGBS) and 24 hours (RNA-seq), 4C embryos at 36 hours and blastocysts at 96 hours. Zona pellucida and polar bodies of 2C embryos were removed (WGBS). ICM cells were purified by immunodissection using anti-mouse IgG (Cedarlane) and guinea pig complement (Rockland)[78]. Genotyping was performed by PCR using primers for *Dnmt3a* (CTGTGGCATCTCAGGGTGATGAGCA and GCAAACAGACCCAACATGGAACCCT) and the *Zp3*-cre transgene (GCAGAACCTGAAGATGTTCGCGAT and AGGTATCTCTGACCAGAGTCATCC).

**DNMT3A immunofluorescence.** *Dnmt3a*$^{2lox}$;*Zp3*-cre females were mated with JF1 males and IF was performed on one-cell zygotes. DNMT3A was detected using the IMG-268 IMGENEX antibody, and DNA was counterstained using propidium iodide, as described previously[27].

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.26.009977; this version posted March 28, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

**WGBS and RNA-seq library construction and sequencing.** Lysates of androgenetic blastocysts were spiked with 0.1 ng lambda phage DNA and subjected to WGBS library construction according to the PBAT protocol for single-read sequencing[46]. DNA from 20-30 pooled 2C embryos per replicate was purified and spiked with 1% unmethylated lambda phage DNA, and WGBS libraries were generated by PBAT with 4 cycles of library amplification[61]. All WGBS libraries had >99% bisulphite conversion rates. Total RNA was extracted from 20-40 pooled 2C embryos, 5-10 4C embryos and 2-7 blastocysts per replicate using Trizol reagent. Strand-specific RNA-seq libraries were generated using NEBNext: rRNA Depletion Kit, RNA First Strand Synthesis Module, Ultra Directional RNA Second Strand Synthesis Module, and Ultra II DNA Library Prep Kit. Libraries were sequenced on HiSeq 1500 or HiSeq 2500 (WGBS: HCS v2.2.68 and RTA v1.18.66.3)[79]. See **Supplemental Table 2** for full sequencing and alignment statistics.

**NGS data processing.** Reads were trimmed using Trimmomatic v0.32[80] and processed for total and allele-specific alignments using MEA v1.0[47] using default parameters and the mm10 reference genome. 4 bases were removed from the 5' end of PBAT sequences. All publicly available NGS data (**Supplemental Table 2**) was reprocessed as above, with the exception of WGBS datasets from sperm[2], oocytes[31], parthenones[43] and primordial germ cells[46], which were previously processed and filtered using identical parameters as in this study.

**WGBS data analysis.** DNAme levels over individual CpGs with ≥5x coverage (including allele-specific) were scored, with the exception of allele-specific alignments of WGBS

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.26.009977; this version posted March 28, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

datasets generated in this study, for which a ≥1x allele-specific read coverage cutoff was used to score methylation status. DNAme levels were calculated over CGI promoters using Bedops v2.4.27[81] and visualized using VisRseq v0.9.12[82]. Only CGI promoters that overlapped at least 2 informative CpGs separated by the maximum sequencing read length of the library were kept. Genome-wide 2- and 20kb bins were generated using Bedtools, and bins covered by at least 4 CpGs separated by over 1 read length in each dataset were used, and a random subset of 1,000 bins were visualized as parallel coordinate plots using VisRseq. Genome-wide 600bp bins with 100bp overlap were used to measure paternal allele DNAme dynamics in sperm and F1 hybrid 2C embryos. Bins covered by at least 2 CpGs separated by over 1 read length in each dataset were kept. Bins showing <20% DNAme in sperm and a ≥30% gain in 2C embryos were scored as showing PDA. Overlapping PDA regions were subsequently merged using Bedtools. NCBI RefSeq (default in VisRseq) and Ensembl Regulatory features (release 81) annotations were used to identify PDA regions that overlapped TSSs, gene bodies and enhancers.

**ChIP-seq data analysis.** Raw sequencing reads were reprocessed as described above into total and allele-specific genomic tracks. RPKM values were calculated over TSSs (+/- 300bp) using VisRseq on the basis of normalized genomic tracks. ChromHMM v1.12[52] was employed to define distinct chromatin states (LearnModel, k=6) on the basis of filtered BAM files (BinarizeBam) using default parameters.

**RNA-seq data processing.** Raw sequencing reads were reprocessed as described above into total and allele-specific genomic tracks. Gene expression (RPKM) values

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.26.009977; this version posted March 28, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

over genic exons were calculated using VisRseq (NCBI Refseq). Paternal RPKM values for each gene showing PDA was then plotted on a parallel-coordinate plot using VisRseq. Oocyte CGI promoter expression (RPKM) was calculated over TSSs +/-300bp and normalized to total aligned reads. Correlograms were generated using Morpheus (https://software.broadinstitute.org/morpheus) on the basis of log2 transformed values. For genome browser visualization, biological replicates were merged and total (allele-agnostic) and allele-specific genomic tracks were organized into UCSC Track Hubs as described previously[47]. Total, paternal and maternal-specific changes in gene expression were calculated using DESeq2 v1.26.0 with default parameters (FDR=10%)[83]. Genes with a ≥2-fold change in expression and a Benjamini-Hochberg adjusted p-value ≤ 0.1 were considered differentially expressed. Transcription initiation sites for *Dnmt3a* matKO and wild-type embryos were determined using StringTie v1.3.5[84].

**Motif analysis.** Known and *de novo* DNA motif discovery was conducted using HOMER findMotifs.pl v4.11.1[50] and the MEME suite[51]. The sequences of CGI promoters showing PDA were input in fasta format, default parameters were used and CGIs showing persistent paternal hypomethylation (n=4,315) were used as background sequences. Both C57BL/6J and DBA/2J sequences were tested.

**Statistical tests.** T-tests of two samples assuming unequal variances were performed when comparing the distribution of DNAme or H3K4me3 levels between CGI promoters that show PDA or persistent hypomethylation. Chi-squared tests were performed when comparing categorical data.

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.26.009977; this version posted March 28, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

**Data availability.** Datasets generated in this study have been deposited in GEO under the accession number GSE141877. See **Supplemental Table 2** or the full list of data analyzed for this study.

## ACKNOWLEDGEMENTS

We are grateful to Dr. Tomohiro Kono (Tokyo University of Agriculture) and Dr. Takuya Wakai (Okayama University) for technical assistance in mouse androgenone collection. We thank J. Oishi and T. Akinaga (Kyushu University) for their assistance with PBAT sequencing and data collection, Alex Boiselle for help with parallel coordinate plots, and Tiffany Leung for help with DESeq2. We thank members of the Lorincz lab for helpful discussions, and Louis Lefebvre and Sanne Janssen for reading the manuscript. Financial support was from the Canadian Institutes of Health Research (PJT-153049) and the Natural Sciences and Engineering Research Council of Canada (RGPIN-2015-05228) to M.C.L. and a JSPS KAKENHI grant to H.S. (JP18H05214). J.R.A. was a recipient of an NSERC Postgraduate Scholarships-Doctoral Program award (PGSD3–476000-2015) and a Killam Doctoral Scholarship.

## AUTHOR CONTRIBUTIONS

J.R.A. and M.C.L. conceived the project and wrote the paper. W.K.A.U, K.T and H.S. carried out the experiments on *Dnmt3a matKO* mice. H.K. performed the uniparental embryo experiments. R.H. and H.S. performed the zygotic DNMT3A IF analysis. J.R.A. performed bioinformatic analysis with the assistance of J.B.D and A.B.B.

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.26.009977; this version posted March 28, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

## REFERENCES


1. Miller, D., Brinkworth, M. & Iles, D. Paternal DNA packaging in spermatozoa: more than the sum of its parts? DNA, histones, protamines and epigenetics. *Reproduction* **139,** 287–301 (2010).
2. Kubo, N. *et al.* DNA methylation and gene expression dynamics during spermatogonial stem cell differentiation in the early postnatal mouse testis. *BMC Genomics* **16,** 624 (2015).
3. Mayer, W., Niveleau, A., Walter, J., Fundele, R. & Haaf, T. Embryogenesis - Demethylation of the zygotic paternal genome. *Nature* **403,** 501–502 (2000).
4. Santos, F., Hendrich, B., Reik, W. & Dean, W. Dynamic reprogramming of DNA methylation in the early mouse Embryo. *Developmental Biology* **241,** 172–182 (2002).
5. Smith, Z. D. *et al.* A unique regulatory phase of DNA methylation in the early mammalian embryo. *Nature* **484,** 339–344 (2012).
6. Maenohara, S. *et al.* Role of UHRF1 in de novo DNA methylation in oocytes and maintenance methylation in preimplantation embryos. *PLoS Genet* **13,** e1007042 (2017).
7. Li, Y. *et al.* Stella safeguards the oocyte methylome by preventing de novo methylation mediated by DNMT1. *Nature* **564,** 136–140 (2018).
8. Han, L., Ren, C., Zhang, J., Shu, W. & Wang, Q. Differential roles of Stella in the modulation of DNA methylation during oocyte and zygotic development. *Cell Discov* **5,** 9 (2019).
9. Hajkova, P. *et al.* Genome-wide reprogramming in the mouse germ line entails the base excision repair pathway. *Science* **329,** 78–82 (2010).
10. Gu, T.-P. *et al.* The role of Tet3 DNA dioxygenase in epigenetic reprogramming by oocytes. *Nature* **477,** 606–610 (2011).
11. Guo, F. *et al.* Active and passive demethylation of male and female pronuclear DNA in the mammalian zygote. *Cell Stem Cell* **15,** 447–458 (2014).
12. Peat, J. R. *et al.* Genome-wide bisulfite sequencing in zygotes identifies demethylation targets and maps the contribution of TET3 oxidation. *Cell Rep* **9,** 1990–2000 (2014).
13. Tsukada, Y.-I., Akiyama, T. & Nakayama, K. I. Maternal TET3 is dispensable for embryonic development but is required for neonatal growth. *Sci Rep* **5,** 15876 (2015).
14. Amouroux, R. *et al.* De novo DNA methylation drives 5hmC accumulation in mouse zygotes. *Nat. Cell Biol.* **18,** 225–233 (2016).
15. Kweon, S.-M. *et al.* Erasure of Tet-oxidized 5-methylcytosine by a SRAP nuclease. *Cell Rep* **21,** 482–494 (2017).
16. Wang, L. *et al.* Programming and inheritance of parental DNA methylomes in mammals. *Cell* **15,** 979–991 (2014).
17. Wang, C. *et al.* Reprogramming of H3K9me3-dependent heterochromatin during mammalian embryo development. *Nat. Cell Biol.* **20,** 620–631 (2018).
18. Auclair, G., Guibert, S., Bender, A. & Weber, M. Ontogeny of CpG island methylation and specificity of DNMT3 methyltransferases during embryonic development in the mouse. *Genome Biology* **15,** 545 (2014).

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.26.009977; this version posted March 28, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

19. Smith, Z. D. *et al.* DNA methylation dynamics of the human preimplantation embryo. *Nature* **511,** 611–615 (2014).
20. Eckersley-Maslin, M. A., Alda-Catalinas, C. & Reik, W. Dynamics of the epigenetic landscape during the maternal-to-zygotic transition. *Nat Rev Mol Cell Biol* **19,** 436–450 (2018).
21. Zhu, P. *et al.* Single-cell DNA methylome sequencing of human preimplantation embryos. *Nat Genet* **50,** 12–19 (2018).
22. Okano, M., Bell, D. W., Haber, D. A. & Li, E. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell* **99,** 247–257 (1999).
23. Bourc'his, D., Xu, G. L., Lin, C. S., Bollman, B. & Bestor, T. H. Dnmt3L and the establishment of maternal genomic imprints. *Science* **294,** 2536–2539 (2001).
24. Hata, K., Okano, M., Lei, H. & Li, E. Dnmt3L cooperates with the Dnmt3 family of de novo DNA methyltransferases to establish maternal imprints in mice. *Development* **129,** 1983–1993 (2002).
25. Bourc'his, D. & Bestor, T. H. Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L. *Nature* **431,** 96–99 (2004).
26. Kaneda, M. *et al.* Essential role for de novo DNA methyltransferase Dnmt3a in paternal and maternal imprinting. *Nature* **429,** 900–903 (2004).
27. Hirasawa, R. *et al.* Maternal and zygotic Dnmt1 are necessary and sufficient for the maintenance of DNA methylation imprints during preimplantation development. *Genes & Development* **22,** 1607–1616 (2008).
28. Hammoud, S. S. *et al.* Distinctive chromatin in human sperm packages genes for embryo development. *Nature* **460,** 473–478 (2009).
29. Brykczynska, U. *et al.* Repressive and active histone methylation mark distinct promoters in human and mouse spermatozoa. *Nat. Struct. Mol. Biol.* **17,** 679–687 (2010).
30. Erkek, S. *et al.* Molecular determinants of nucleosome retention at CpG-rich sequences in mouse spermatozoa. *Nat. Struct. Mol. Biol.* **20,** 868–875 (2013).
31. Shirane, K. *et al.* Mouse oocyte methylomes at base resolution reveal genome-wide accumulation of non-CpG methylation and role of DNA methyltransferases. *PLoS Genet* **9,** e1003439 (2013).
32. Qu, J. *et al.* Evolutionary expansion of DNA hypomethylation in the mammalian germline genome. *Genome Research* **28,** 145–158 (2017).
33. Edwards, J. R., Yarychkivska, O., Boulard, M. & Bestor, T. H. DNA methylation and DNA methyltransferases. *Epigenetics Chromatin* **10,** 23 (2017).
34. Siklenka, K. *et al.* Disruption of histone methylation in developing sperm impairs offspring health transgenerationally. *Science* **350,** aab2006 (2015).
35. Yamaguchi, K. *et al.* Re-evaluating the localization of sperm-retained histones revealed the modification-dependent accumulation in specific genome regions. *Cell Rep* **23,** 3920–3932 (2018).
36. Zhang, B. *et al.* Allelic reprogramming of the histone modification H3K4me3 in early mammalian development. *Nature* **537,** 553–557 (2016).
37. Xu, Q. *et al.* SETD2 regulates the maternal epigenome, genomic imprinting and embryonic development. *Nat Genet* **51,** 844–856 (2019).
38. Ooi, S. K. T. *et al.* DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. *Nature* **448,** 714–717 (2007).

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.26.009977; this version posted March 28, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.


39. Smallwood, S. A. *et al.* Dynamic CpG island methylation landscape in oocytes and preimplantation embryos. *Nat Genet* **43,** 811–814 (2011).
40. Zheng, H. *et al.* Resetting epigenetic memory by reprogramming of histone modifications in mammals. *Molecular Cell* **63,** 1066–1079 (2016).
41. Inoue, A., Jiang, L., Lu, F., Suzuki, T. & Zhang, Y. Maternal H3K27me3 controls DNA methylation-independent imprinting. *Nature* **547,** 419–424 (2017).
42. Stewart, K. R. *et al.* Dynamic changes in histone modifications precede de novo DNA methylation in oocytes. *Genes & Development* **29,** 2449–2462 (2015).
43. Brind'Amour, J. *et al.* LTR retrotransposons transcribed in oocytes drive species-specific and heritable changes in DNA methylation. *Nature Communications* **9,** 3331 (2018).
44. Xu, Q. *et al.* SETD2 regulates the maternal epigenome, genomic imprinting and embryonic development. *Nat Genet* **51,** 844–856 (2019).
45. Matsuzaki, H. *et al.* De novo DNA methylation through the 5'-segment of the H19 ICR maintains its imprint during early embryogenesis. *Development* **142,** 3833–3844 (2015).
46. Kobayashi, H. *et al.* High-resolution DNA methylome analysis of primordial germ cells identifies gender-specific reprogramming in mice. *Genome Research* **23,** 616–627 (2013).
47. Richard Albert, J. *et al.* Development and application of an integrated allele-specific pipeline for methylomic and epigenomic analysis (MEA). *BMC Genomics* **19,** 463 (2018).
48. Deaton, A. M. & Bird, A. CpG islands and the regulation of transcription. *Genes & Development* **25,** 1010–1022 (2011).
49. Weber, M. *et al.* Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. *Nat Genet* **39,** 457–466 (2007).
50. Heinz, S. *et al.* Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Molecular Cell* **38,** 576–589 (2010).
51. Bailey, T. L. *et al.* MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Research* **37,** W202–208 (2009).
52. Ernst, J. & Kellis, M. ChromHMM: automating chromatin-state discovery and characterization. *Nat Chem Biol* **9,** 215–216 (2012).
53. Liu, S. *et al.* Setdb1 is required for germline development and silencing of H3K9me3-marked endogenous retroviruses in primordial germ cells. *Genes & Development* **28,** 2041–2055 (2014).
54. Du, J., Johnson, L. M., Jacobsen, S. E. & Patel, D. J. DNA methylation pathways and their crosstalk with histone methylation. *Nat Rev Mol Cell Biol* **16,** 519–532 (2015).
55. Gaysinskaya, V. *et al.* Transient reduction of DNA methylation at the onset of meiosis in male mice. *Epigenetics Chromatin* **11,** 15 (2018).
56. Wu, J. *et al.* The landscape of accessible chromatin in mammalian preimplantation embryos. *Nature* **534,** 652–657 (2016).
57. Matoba, S. *et al.* Loss of H3K27me3 Imprinting in Somatic Cell Nuclear Transfer Embryos Disrupts Post-Implantation Development. *Cell Stem Cell* **23,** 343–354 (2018).

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.26.009977; this version posted March 28, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.


58. Kaneda, M. *et al.* Genetic evidence for Dnmt3a-dependent imprinting during oocyte growth obtained by conditional knockout with Zp3-Cre and complete exclusion of Dnmt3b by chimera formation. *Genes to Cells* **15,** 169–179 (2010).
59. Oswald, J. *et al.* Active demethylation of the paternal genome in the mouse zygote. *Curr Biol* **10,** 475–478 (2000).
60. Wossidlo, M. *et al.* Dynamic link of DNA demethylation, DNA strand breaks and repair in mouse zygotes. *EMBO J* **29,** 1877–1888 (2010).
61. Au Yeung, W. K. *et al.* Histone H3K9 methyltransferase G9a in oocytes is essential for preimplantation development but dispensable for CG methylation protection. *Cell Rep* **27,** 282–293 (2019).
62. Matsuzaki, H. *et al.* Synthetic DNA fragments bearing ICR cis elements become differentially methylated and recapitulate genomic imprinting in transgenic mice. *Epigenetics Chromatin* **11,** 36 (2018).
63. Matsuzaki, H. *et al.* Recapitulation of gametic DNA methylation and its post-fertilization maintenance with reassembled DNA elements at the mouse Igf2/H19 locus. *Epigenetics Chromatin* **13,** 2–19 (2020).
64. Otani, J. *et al.* Structural basis for recognition of H3K4 methylation status by the DNA methyltransferase 3A ATRX-DNMT3-DNMT3L domain. *EMBO reports* **10,** 1235–1241 (2009).
65. Li, X. *et al.* A maternal-zygotic effect gene, Zfp57, maintains both maternal and paternal imprints. *Developmental Cell* **15,** 547–557 (2008).
66. Strogantsev, R. *et al.* Allele-specific binding of ZFP57 in the epigenetic regulation of imprinted and non-imprinted monoallelic expression. *Genome Biology* **16,** 112 (2015).
67. Quenneville, S. *et al.* In embryonic stem cells, ZFP57/KAP1 recognize a methylated hexanucleotide to affect chromatin and DNA methylation of imprinting control regions. *Molecular Cell* **44,** 361–372 (2011).
68. Jacobs, F. M. J. *et al.* An evolutionary arms race between KRAB zinc-finger genes ZNF91/93 and SVA/L1 retrotransposons. *Nature* **516,** 242–245 (2014).
69. Yang, P. *et al.* A placental growth factor is silenced in mouse embryos by the zinc finger protein ZFP568. *Science* **356,** 757–759 (2017).
70. Velasco, G. *et al.* Dnmt3b recruitment through E2F6 transcriptional repressor mediates germ-line gene silencing in murine somatic tissues. *Proceedings of the National Academy of Sciences* **107,** 9281–9286 (2010).
71. Endoh, M. *et al.* PCGF6-PRC1 suppresses premature differentiation of mouse embryonic stem cells by regulating germ cell-related genes. *eLife* **6,** e21064 (2017).
72. Auclair, G. *et al.* EHMT2 directs DNA methylation for efficient gene silencing in mouse embryos. *Genome Research* **26,** 192–202 (2016).
73. Tatsumi, D. *et al.* DNMTs and SETDB1 function as co-repressors in MAX-mediated repression of germ cell-related genes in mouse embryonic stem cells. *PLoS ONE* **13,** e0205969 (2018).
74. Kono, T., Obata, Y., Yoshimzu, T., Nakahara, T. & Carroll, J. Epigenetic modifications during oocyte growth correlates with extended parthenogenetic development in the mouse. *Nat Genet* **13,** 91–94 (1996).

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.26.009977; this version posted March 28, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

75. Kono, T., Sotomaru, Y., Sato, Y. & Nakahara, T. Development of androgenetic mouse embryos produced by in vitro fertilization of enucleated oocytes. *Mol. Reprod. Dev.* **34,** 43–46 (1993).
76. Obata, Y. *et al.* Post-implantation development of mouse androgenetic embryos produced by in-vitro fertilization of enucleated oocytes. *Hum. Reprod.* **15,** 874–880 (2000).
77. de Vries, W. N. *et al.* Expression of Cre recombinase in mouse oocytes: a means to study maternal effect genes. *genesis* **26,** 110–112 (2000).
78. Solter, D. & Knowles, B. B. Immunosurgery of mouse blastocyst. *Proceedings of the National Academy of Sciences* **72,** 5099–5102 (1975).
79. Toh, H. *et al.* Software updates in the Illumina HiSeq platform affect whole-genome bisulfite sequencing. *BMC Genomics* **18,** 31 (2017).
80. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30,** 2114–2120 (2014).
81. Neph, S. *et al.* BEDOPS: high-performance genomic feature operations. *Bioinformatics* **28,** 1919–1920 (2012).
82. Younesy, H., Möller, T., Lorincz, M. C., Karimi, M. M. & Jones, S. J. M. VisRseq: R-based visual framework for analysis of sequencing data. *BMC Bioinformatics* **16 Suppl 11,** S2 (2015).
83. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* **15,** 550 (2014).
84. Pertea, M. *et al.* StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nature Biotechnology* **33,** 290–295 (2015).