

Dynamic changes in histone modifications precede de novo DNA methylation in oocytes

Kathleen R. Stewart,¹ Lenka Veselovska,¹ Jeesun Kim,² Jiahao Huang,¹ Heba Saadeh,^{1,3} Shin-ichi Tomizawa,⁴ Sébastien A. Smallwood,¹ Taiping Chen,² and Gavin Kelsey^{1,5}

¹Epigenetics Programme, Babraham Institute, Cambridge CB22 3AT, United Kingdom; ²Department of Epigenetics and Molecular Carcinogenesis, The University of Texas M.D. Anderson Cancer Center, Smithville, Texas 77030, USA; ³Bioinformatics Group, Babraham Institute, Cambridge CB22 3AT, United Kingdom; ⁴School of Medicine, Yokohama City University, Yokohama 236-0027, Japan; ⁵Centre for Trophoblast Research, University of Cambridge CB2 3EG, Cambridge, United Kingdom

Erasure and subsequent reinstatement of DNA methylation in the germline, especially at imprinted CpG islands (CGIs), is crucial to embryogenesis in mammals. The mechanisms underlying DNA methylation establishment remain poorly understood, but a number of post-translational modifications of histones are implicated in antagonizing or recruiting the de novo DNA methylation complex. In mouse oogenesis, DNA methylation establishment occurs on a largely unmethylated genome and in nondividing cells, making it a highly informative model for examining how histone modifications can shape the DNA methylome. Using a chromatin immunoprecipitation (ChIP) and genome-wide sequencing (ChIP-seq) protocol optimized for low cell numbers and novel techniques for isolating primary and growing oocytes, profiles were generated for histone modifications implicated in promoting or inhibiting DNA methylation. CGIs destined for DNA methylation show reduced protective H3K4 dimethylation (H3K4me2) and trimethylation (H3K4me3) in both primary and growing oocytes, while permissive H3K36me3 increases specifically at these CGIs in growing oocytes. Methylome profiling of oocytes deficient in H3K4 demethylase KDM1A or KDM1B indicated that removal of H3K4 methylation is necessary for proper methylation establishment at CGIs. This work represents the first systematic study performing ChIP-seq in oocytes and shows that histone remodeling in the mammalian oocyte helps direct de novo DNA methylation events.

[*Keywords:* oocytes; DNA methylation; genomic imprinting; histone modifications; ChIP-seq]

Supplemental material is available for this article.

Received September 4, 2015; revised version accepted November 4, 2015.

5-cytosine methylation—or DNA methylation—is a repressive epigenetic mark that characterizes the majority of CpGs within mammalian genomes. However, CpG-dense regions called CpG islands (CGIs) are generally unmethylated (Deaton and Bird 2011; Jones 2012). In the male and female germlines, a number of CGIs comprise imprinted germline differentially methylated regions (gDMRs), which acquire methylation in the germline and retain this monoallelic methylation throughout development and postnatal life (Sasaki and Matsui 2008). These loci resist the genome-wide demethylation that occurs following fertilization, and failure to establish or maintain methylation at imprinted gDMRs results in embryonic lethality (Kaneda et al. 2004; Smith and Meissner 2013). Despite its importance, the mechanisms responsible for targeting DNA methylation to CGIs, especially *in vivo*,

remain poorly understood. However, it is known that DNA methylation positively and negatively correlates with a number of post-translational modifications of histones, indicating that histone modifications may direct the DNA methyltransferase complex, including to certain CGIs, during *de novo* events (Cheng and Blumenthal 2010; Blackledge and Klose 2011).

Starting from a largely unmethylated genome, ~2000 CGIs gain methylation in mouse oocytes (Kobayashi et al. 2012; Shirane et al. 2013), making oogenesis a biologically important system for epigenetic studies. It is also an especially useful system, as only in oogenesis does epigenetic reprogramming occur in the absence of replication, meaning that *de novo* methylation events can be uncoupled from the requirement for DNA methylation maintenance between cell divisions (Smallwood and Kelsey 2012). This *de novo* methylation is concordant with

Corresponding author: gavin.kelsey@babraham.ac.uk

Article published online ahead of print. Article and publication date are online at <http://www.genesdev.org/cgi/doi/10.1101/gad.271353.115>. Freely available online through the *Genes & Development* Open Access option.

© 2015 Stewart et al. This article, published in *Genes & Development*, is available under a Creative Commons License [Attribution-NonCommercial 4.0 International], as described at <http://creativecommons.org/licenses/by-nc/4.0/>.

oocyte growth, beginning in mice at around postnatal day 10 (P10) and largely finishing by P21, at the germinal vesicle (GV) stage (Obata and Kono 2002; Hiura et al. 2006; Smallwood et al. 2011; Shirane et al. 2013). Transcription is required for methylation at a number of gDMRs (Chotalia et al. 2009; Smith et al. 2011), and the oocyte methylome closely mirrors the oocyte transcriptome (Smallwood et al. 2011; Kobayashi et al. 2012). Of the DNA methyltransferases DNMT1, DNMT3A, and DNMT3B and the cofactor DNMT3L, DNMT3A and DNMT3L are primarily required for methylation in the mouse oocyte, with DNMT1 playing a subsidiary role in ensuring symmetric methylation of CpG sites (Bourc'his et al. 2001; Hata et al. 2002; Kaneda et al. 2004; Smallwood et al. 2011; Shirane et al. 2013).

How transcription might shape the oocyte methylome remains unclear. It may simply render chromatin more accessible to the DNMT3A/DNMT3L complex. However, the parallels between histone modifications laid down by transcription and histone modifications that correlate or anti-correlate with DNA methylation in somatic contexts suggest that these marks may be the mechanism through which transcription patterns the DNA methylome. Trimethylation and dimethylation of H3K4 (K4me3/K4me2) classically mark sites of transcription initiation (Deaton and Bird 2011; Henikoff and Shilatifard 2011). More generally, K4me2 and K4me3 are hallmarks of CGIs (Illingworth et al. 2008). H3K36 trimethylation (K36me3), on the other hand, is associated with elongating eukaryotic chromatin (Kizer et al. 2005; Edmunds et al. 2008; Yoh et al. 2008). Methylated H3K4 excludes DNMT3A and DNMT3L via their ATRX–DNMT3A–DNMT3L (ADD) domains, effectively protecting associated DNA from DNA methylation (Ooi et al. 2007; Noh et al. 2015). Supporting the idea that K4me2/3 must be removed from CGIs prior to de novo methylation, KDM1B (AOF1), a K4me2 demethylase, is required for imprint establishment at several gDMRs in oocytes (Ciccone et al. 2010). Conversely, K36me3 can recruit both DNMT3A (Dhayan et al. 2010) and DNMT3B (Baubec et al. 2015; Morselli et al. 2015). Since the majority of CGIs that acquire methylation in the oocyte is intragenic (Smallwood et al. 2011; Kobayashi et al. 2012), accumulation of K36me3 may target these loci for DNA methylation.

To date, most of the analysis of the relationship between histone modifications and DNA methylation has been done in systems such as mouse embryonic stem (ES) cells or sperm owing to the relative ease of isolation and large number of such cells available compared with oocytes. Specifically, genome-wide chromatin maps of developing mammalian oocytes have never been generated. Here, we developed new methods to isolate primary oocytes and growing oocyte nuclei by fluorescence-activated cell sorting (FACS) followed by chromatin immunoprecipitation (ChIP) combined with sequencing (ChIP-seq) to profile the dynamics of K4me2, K4me3, and K36me3 during oogenesis. We also performed whole-genome bisulfite sequencing on mature oocytes lacking the K4 demethylase KDM1B or its paralog, KDM1A (LSD1), to determine the extent of their roles in DNA methylation establish-

ment. We found that K36me3 specifically increases at CGIs destined for DNA methylation between the primary and growing oocyte stage, in parallel with activation of the upstream promoters that define the oocyte transcriptome. K4me2 and K4me3 are relatively reduced but largely static at these CGIs between embryonic day 18.5 (E18.5) and P10, but KDM1B-deficient and, to a lesser extent, KDM1A-deficient oocytes show loss of methylation at CGIs, including at most imprinted gDMRs. This indicates a step-wise modulation of CGI chromatin in which K36me3 acquisition occurs prior to methylation establishment, and K4me2 removal occurs predominantly after the onset of de novo methylation.

Results

Isolation of primary and growing oocytes for ChIP-seq using novel FACS protocols

Oocytes are challenging to isolate in large numbers. The most common collection method, picking by mouth pipetting from digested ovaries, is time-consuming and returns 100–200 cells per mouse, making it unsuitable for collecting the numbers of cells required for ChIP-seq. Additionally, a massive wave of oocyte apoptosis begins from around the time of birth in mice (Pepling 2006), making the oocyte an increasingly rare cell in the ovary with increasing age. To circumvent these issues, we developed two methods of oocyte isolation using FACS, which allow us to process many mice in parallel in a relatively short time. We concentrated on two time points in oogenesis: The first, E18.5, represents the “ground state” of the oocyte epigenome, as oocytes contain very little DNA methylation (Smallwood et al. 2011; Shirane et al. 2013); at the second, P10, de novo DNA methylation commences (Tomizawa et al. 2012).

At E18.5, female germ cells have arrested in the diplotene stage of meiosis I; are diploid, with 4N DNA content; and can be considered primary, nongrowing oocytes. Oocytes are also most numerous at this time point. Synaptonemal complex protein 3 (SYCP3) is a structural protein that helps connect homologous chromosomes during meiosis I (Handel and Schimenti 2010). At E18.5, oocytes strongly express SYCP3 (Fig. 1A). We therefore developed a FACS strategy based on SYCP3 staining and obtained, on average, ~3000 oocytes per E18.5 fetus using this method. This strategy yielded pure oocyte populations for ChIP-seq (Supplemental Fig. 1A).

E18.5 oocytes are small, measuring 10–20 μm in diameter, and can readily be sorted by flow cytometry. Growing oocytes, however, are ≥ 30 –65 μm in diameter and cannot be sorted intact. ChIP-seq requires only genomic DNA and its associated proteins; therefore, we isolated nuclei from P10 ovary preparations using a protocol adapted from Rosner et al. (2013). Because all P10 oocytes are meiotically arrested, by staining the DNA and first gating on the 4N population, the small oocyte population in P10 nucleus preparations could be more readily visualized (Fig. 1B, left panel). NOBOX, an oocyte-specific transcription factor that characterizes oocytes in primordial follicles

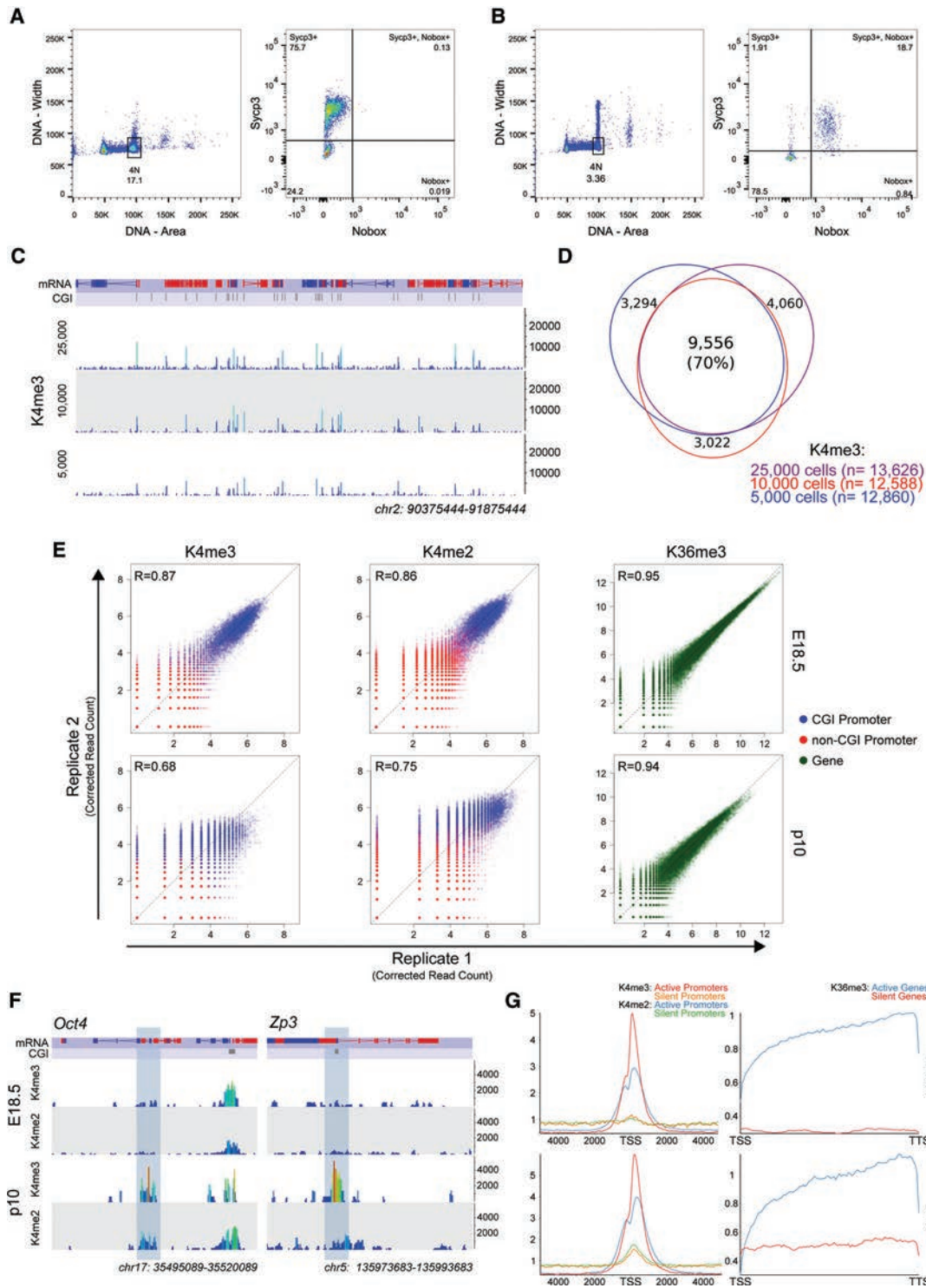


Figure 1. Development of oocyte FACS suitable for CHIP-seq. (A) FACS strategy for E18.5 primary oocyte isolation using SYCP3. (B) FACS strategy for P10 growing oocyte nucleus isolation using DNA content and NOBOX. (C) Genome screenshots showing K4me3 signal (as corrected read count) in E18.5 ovarian somatic cells, starting from 25,000, 10,000, or 5,000 cells. The 5000-cell track represents grouped biological replicates. Shown is a 1-kb window with a 1-kb step. (D) Peaks called in K4me3 ChIP-seqs starting from 25,000, 10,000, or 5,000 cells show high agreement. The 5000-cell peaks were called from grouped biological replicates. (E) Replicate correlation over expected enrichment sites (log-transformed corrected read count) in the E18.5 and P10 oocyte data sets: 4 kb centered on promoters for K4me2 and K4me3 and gene bodies for K36me3. CGI promoters are indicated by blue dots, and non-CGI promoters are indicated by red dots. (F) K4me2 and K4me3 signal (corrected read count) at the non-CGI *Oct4* and CGI *Zp3* promoters reflects the silenced state of these genes at E18.5 (top tracks) and active state at P10 (bottom tracks). Shown is a 100-base-pair (bp) window with a 100-bp step. (G) Trend plots showing K4me2 and K4me3 signal (log-transformed corrected read count) over silent and active promoters (left panels; X-axis is in base pairs) and K36me3 signal across their associated gene bodies (right panels) as determined by RNA sequencing (RNA-seq). (TSS) Transcription start site; (TTS) transcription termination site.

(Suzumori et al. 2002; Rajkovic et al. 2004), is absent at E18.5 but present in most SYCP3⁺ cells at P10 (Fig. 1B, right panel), effectively distinguishing primary and growing oocytes. This FACS strategy yielded pure 4N, NOBOX⁺ oocyte nuclei for ChIP-seq (Supplemental Fig. 1B,C). Using this method, ~200 oocytes per mouse were obtained, and, in our hands, up to 80 mice could be processed in a single FACS sort, representing a significant efficiency improvement over mouth pipetting. Full gating strategies for sorting E18.5 oocytes and P10 oocyte nuclei are shown in Supplemental Figure 1D,E.

For this study, ChIP was carried out using the True MicroChIP kit (Diagenode) with modifications (see the Materials and Methods). E18.5 ChIP-seq was done on 25,000 cells per immunoprecipitation; P10 ChIP-seq used the equivalent of 5000 cells per immunoprecipitation. Before committing precious material, we performed K4me3 ChIP-seq starting with 25,000, 10,000, or 5000 ovarian somatic cells, which showed high agreement in each set when two biological replicates of 5000-cell ChIP-seqs were merged (Fig. 1C,D). For the oocyte ChIP-seq data sets, E18.5 and P10 replicates correlated well over expected enrichment sites (Fig. 1E). Our protocol detected K4me2 and K4me3 signal in oocytes at both CGI promoters, such as that of *Zp3*, and non-CGI promoters, such as *Oct4*, indicating that the resolution of the data is sufficient for downstream analyses (Fig. 1F). Neither of these genes is expressed at E18.5, and, accordingly, K4me2/3 signal at these loci was greatly reduced at this time point, indicating the purity of the isolated oocyte populations. Looking genome-wide, enrichment of these marks over gene bodies and transcription start sites (TSSs) followed expected profiles at active and silent promoters and active and silent genes (Fig. 1G). We therefore used these data sets to investigate distribution and enrichment of these marks with regard to de novo DNA methylation.

H3K4 methylation dynamics at CGIs destined for DNA methylation

To investigate the role of H3K4 methylation at CGIs destined to be methylated or unmethylated in the GV oocyte, we analyzed the ChIP-seq data sets for K4me2 and K4me3 from E18.5 and P10 oocytes. GV methylated (GVmeth) CGIs and GV unmethylated (GVunmeth) CGIs were defined using previously published data (Shirane et al. 2013). We first compared K4me2 and K4me3 at GVmeth and GVunmeth CGIs within the same data set. CpG density is known to positively correlate with enrichment of both K4me2 and K4me3 (Illingworth et al. 2010). Methylated CGIs in both the oocyte and somatic contexts tend to be CpG-poor (Fig. 2A; Thomson et al. 2010; Krebs et al. 2014; Wachter et al. 2014). We therefore put CGIs into subsets by CpG density to compare K4me2 and K4me3 enrichment in GVmeth and GVunmeth CGIs (Fig. 2B). At both E18.5 and P10, there was a relative decrease in both K4me2 and K4me3 in the GVmeth subset, and this decrease became more significant with increasing CpG density (Fig. 2B,C). This trend seems to be a general fea-

ture of these loci rather than oocyte-specific, as a similar trend could be seen in 2i ES cells (Supplemental Fig. 2A; Marks et al. 2012). Controlling for CpG density, GVunmeth CGIs still showed higher H3K4 methylation at both E18.5 and P10, indicating that CpG density alone cannot account for the reduced H3K4 methylation levels at GVmeth CGIs.

We next asked whether there were any broad changes in K4me2 or K4me3 enrichment at GVmeth CGIs between E18.5 and P10. Enrichment for both K4me2 and K4me3 was relatively static between the two time points at these loci, although a wide range of enrichments could be seen among individual loci (Fig. 2D). Enrichment was also relatively unchanged at the maternal imprinted gDMRs (Fig. 2D, black dots). Furthermore, the enrichment trends seen in the GVmeth subset were mirrored in the GVunmeth subset, indicating that GVmeth CGIs as a group are not targeted for modulation of K4me2 or K4me3 between E18.5 and P10.

Apart from CpG density, the other major determinant of H3K4 methylation at CGIs is TSS activity (Henikoff and Shilatifard 2011). To investigate the contribution of CGI localization at E18.5 and P10 to H3K4 enrichment, we divided CGIs into TSS, intragenic, and intergenic categories at each time point using RNA sequencing (RNA-seq) data (Supplemental Table 1). The intergenic category contained both intergenic CGIs and CGIs not associated with an active gene at that time point. Unsurprisingly, at both E18.5 and P10, a smaller proportion of GVmeth CGIs were 5'-most TSSs (11.2% and 6.8%, respectively), while a larger proportion of GVunmeth CGIs were 5'-most TSSs at both time points (~46%) (Supplemental Tables 2, 3). Because the majority of GVmeth CGIs and all maternal imprinted gDMRs are intragenic at P10, we focused on the GVmeth P10 intragenic category. As expected, CGIs that acted as TSSs at E18.5 had, on average, more K4me2 and K4me3 than intragenic or intergenic CGIs at that time point (Fig. 2E, top panels). However, this relative enrichment persisted at P10 (Fig. 2E, bottom panels), when these CGIs were all intragenic, suggesting that the legacy of TSS activity was reflected in K4me2/3 enrichment at these loci. Retention of H3K4 methylation could also be seen at imprinted gDMRs like *Impact* (TSS at E18.5 and intragenic at P10) compared with others such as *Peg13* (intragenic at both E18.5 and P10) (Fig. 2F). Relatively few GVmeth CGIs were TSSs at E18.5 and intragenic at P10 (102 of 1465 P10 intragenic CGIs), indicating that this shift describes only a minority of GVmeth CGIs. This indicates that K4me2 and K4me3 enrichment over GVmeth CGIs at the onset of de novo DNA methylation is influenced by their transcriptional state at E18.5, and the majority of these CGIs are either intragenic or intergenic in the primary oocyte.

Growing oocytes show global H3K36me3 enrichment and targeted H3K4me2/3 loss at GVmeth CGIs

We next investigated K36me3 dynamics between E18.5 and P10 at GVmeth CGIs. A greater proportion of GVmeth CGIs was enriched in K36me3 at P10 compared

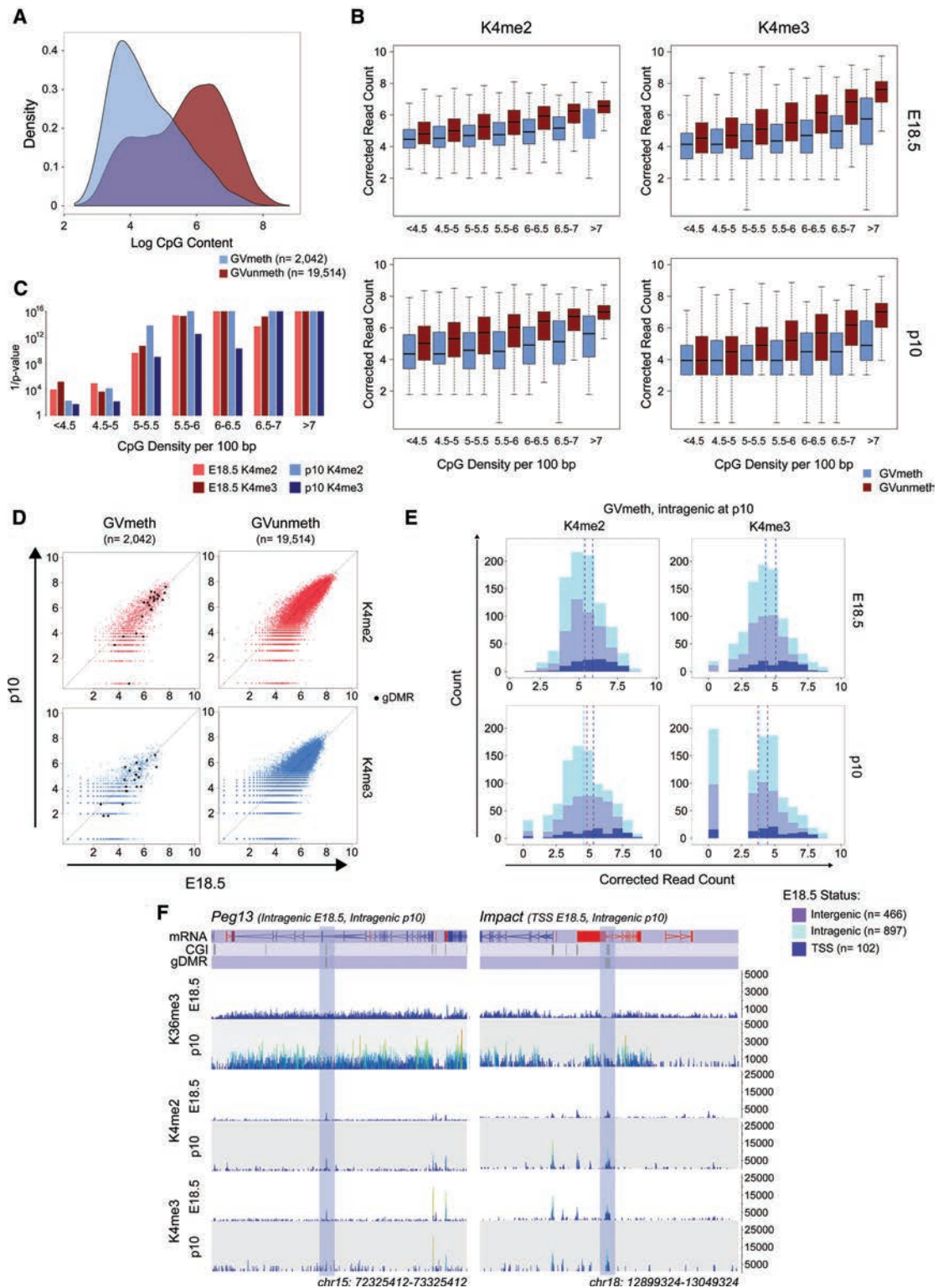


Figure 2. H3K4 methylation dynamics at CGIs prior to de novo DNA methylation. (A) Density plot of CpG densities of GVmeth (blue) and GVunmeth (red) CGIs. (B) E18.5 and P10 K4me2 and K4me3 signal (log-transformed corrected read count) at 4 kb centered on CGIs, separated into GVmeth and GVunmeth CGIs and by CpG density. (Black line) Median; (dashed lines) 1.5 \times interquartile range. (C) Significance of K4me2 or K4me3 enrichment difference between GVmeth and matched GVunmeth CGIs within each CpG density subset at E18.5 and P10; unpaired one-sided Mann-Whitney test. (D) Scatter plots showing E18.5 versus P10 enrichment (log-transformed corrected read count) for K4me2 (top panels) and K4me3 (bottom panels) in GVmeth and GVunmeth CGIs. (Black dots) Imprinted gDMRs. (E) Histograms showing K4me2 and K4me3 enrichment (log-transformed corrected read count) at GVmeth CGIs that are intragenic at P10, separated by annotation at E18.5. (Magenta) Intergenic; (light blue) intragenic; (dark blue) TSS. Dashed lines indicate the mean. (F) Genome screenshots showing ChIP-seq signal (corrected read count) at E18.5 and P10 over the *Peg13* and *Impact* imprinted gDMRs. Shown is a 100-bp window with a 100-bp step.

with GVmeth CGIs (Fig. 3A). CGIs were generally depleted of K36me3 with respect to flanking regions (Fig. 3B). CGIs tended to be K36me3-depleted irrespective of localization or time point, with the exception of the GVmeth subset of intragenic CGIs at P10, for which there is no depletion of K36me3 with respect to flanking regions (Fig. 3B). This indicates that GVmeth CGIs that are intragenic at P10, which includes the majority of CGIs that

gain methylation in the oocyte, are relatively depleted of K36me3 at E18.5 but preferentially enriched for K36me3 at the onset of de novo DNA methylation.

In our data, as has been observed by others, K36me3 levels correlate positively with transcription levels (Supplemental Fig. 3A). To exclude the possibility that the higher P10 K36me3 levels observed at GVmeth CGIs were caused by higher expression of associated genes,

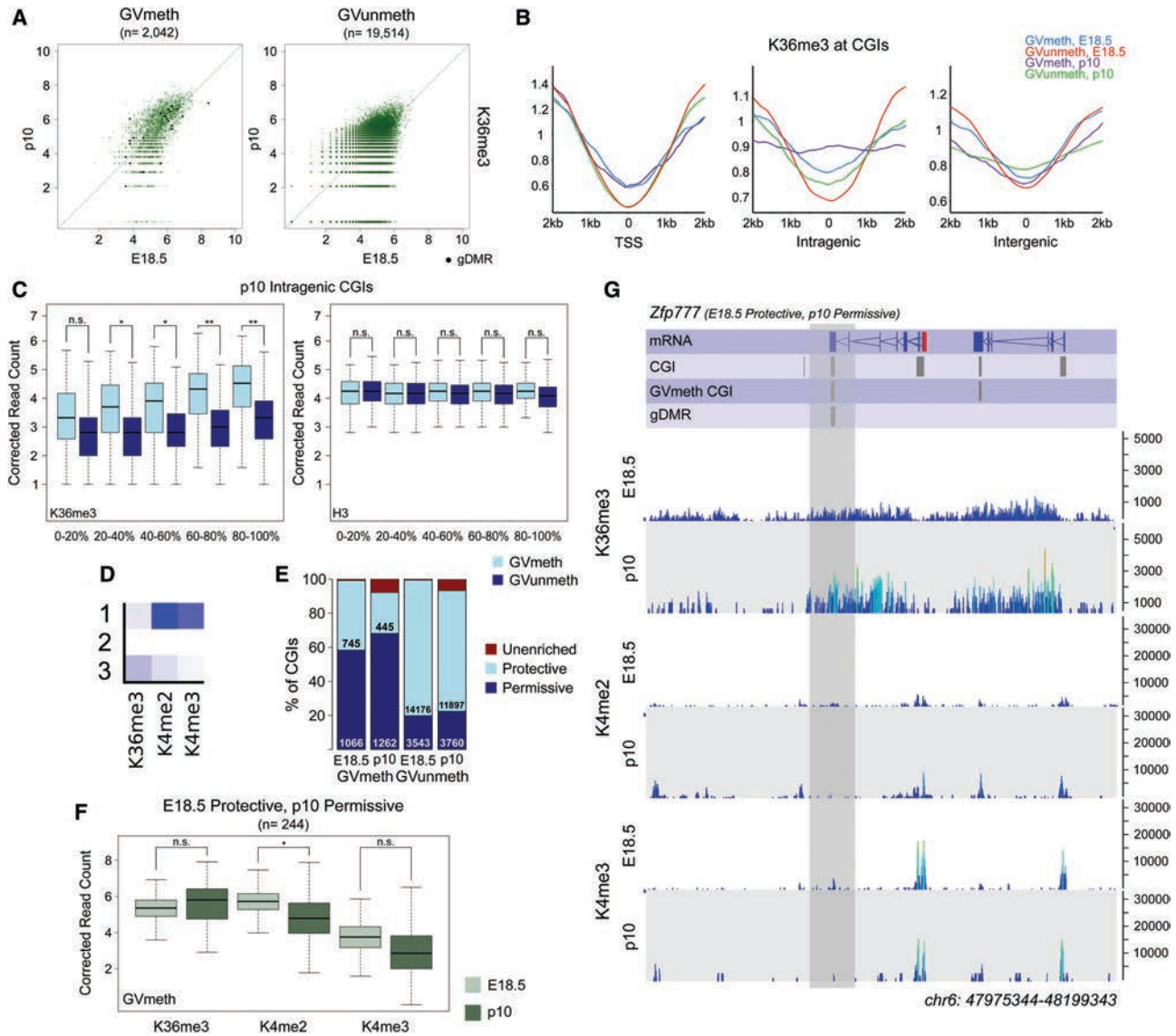


Figure 3. K36me3 is generally enriched, and K4me2/3 is depleted at certain GVmeth CGIs at P10. (A) Scatter plots showing E18.5 versus P10 enrichment (log-transformed corrected read count) for K36me3 in GVmeth and GVunmeth CGIs. (Black dots) Imprinted gDMRs. (B) Trend plots showing K36me3 signal (log-transformed corrected read count) at 4 kb centered on TSS, intragenic, and intergenic GVmeth and GVunmeth CGIs as determined by RNA-seq at E18.5 and P10. (C) Box plots showing K36me3 and H3 (log-transformed corrected read count) at P10 intragenic GVmeth and GVunmeth CGIs, parsed by expression level. (Black line) Median; (dashed lines) 1.5× interquartile range. (*) $P < 1 \times 10^{-10}$; (**) $P < 1 \times 10^{-25}$, Mood's median test. (D) Emission probabilities from a three-state segmentation of K4me2, K4me3, and K36me3 E18.5 and P10 signal over CGIs. (E) Bar chart showing the proportion of protective, permissive, and unenriched CGIs in GVmeth and GVunmeth CGIs at E18.5 and P10. (F) K36me3, K4me2, and K4me3 enrichment (log-transformed corrected read count) at 4 kb centered on GVmeth CGIs that transition from the protective to the permissive state between E18.5 and P10. (Black line) Median; (dashed lines) 1.5× interquartile range. (*) $P < 1 \times 10^{-10}$, Mood's median test. (G) Genome screenshot showing ChIP-seq signal (corrected read count) at E18.5 and P10 over the *Zfp777* imprinted gDMR. Shown is a 100-bp window with a 100-bp step.

intragenic GVmeth and GVunmeth CGIs were segregated into expression quintiles based on the expression of the gene in which they resided. As expected, a positive relationship between K36me3 level at intragenic CGIs and expression could be seen (Fig. 3C, left panel); however, comparing GVmeth and GVunmeth CGIs within the same quintile revealed consistent enrichment in the GVmeth subset. This enrichment was not seen when comparing K36me3 over the gene bodies (excluding the CGIs) of the genes that contain these CGIs (Supplemental Fig. 3B). Therefore, expression level alone does not explain the elevated K36me3 levels at these CGIs. Nucleosome density as measured by H3 ChIP-seq enrichment was the same across all expression quintiles and methylation categories (Fig. 3C, right panel). Unlike H3K4 methylation, K36me3 enrichment at CGIs is largely independent of CpG density (Supplemental Fig. 3C). Taken together, this suggests that some unidentified factor, and not nucleosome density, CpG density, or expression level, accounts for the elevated K36me3 seen at the intragenic GVmeth CGIs in growing oocytes.

To aggregate these data sets and characterize the combined enrichment of K4me2, K4me3, and K36me3 at CGIs at E18.5 and P10, ChIP signal at CGIs was analyzed with unsupervised machine learning (ChromHMM) (Ernst and Kellis 2012). Three states were found: one that is enriched in K4me2 and K4me3 (state 1), one enriched in K36me3 (state 3), and one enriched for none of these marks (state 2) (Fig. 3D). The unenriched state was found to localize to CGIs with little or no ChIP signal, and since state 2 CGIs appear in equal proportion in GVmeth and GVunmeth CGIs at both time points (Fig. 3E), they were not analyzed further. The K4me2/3-enriched state 1 was termed the “protective state” and the K36me3-enriched state 3 was termed the “permissive state” to reflect the relationship between these respective states and the properties of the de novo DNA methylation complex (Smallwood and Kelsey 2012).

Comparing the proportion of CGIs in each state, ~60% of GVmeth CGIs were in the permissive state at E18.5, while the protective state characterized >70% of GVunmeth CGIs at both time points (Fig. 3E). At both E18.5 and P10, the protective state characterized the majority of TSS CGIs, while the permissive state characterized intragenic CGIs (Supplemental Fig. 3D). Twelve percent (244) of GVmeth CGIs transitioned from the protective to the permissive state between E18.5 and P10, while only 5% (1027) of GVunmeth CGIs made the same shift. The majority of the GVmeth CGIs that made this shift were intragenic at both time points (Supplemental Table 4). The GVmeth CGIs that transitioned from protective to permissive were characterized by a modest increase in K36me3 and a significant decrease in K4me2 between E18.5 and P10 (Fig. 3F). These loci seemed to have lower levels of K4me3 at both time points. It is important to note, however, that because ChromHMM segregates the genome into the number of states specified by the user, loci with reasonably different combinations of marks can be assigned the same state based on probability. Therefore, a transition from the protective to the permis-

sive state between E18.5 and P10 could signify either an increase in K36me3 or a decrease in K4me2 or K4me3. Two imprinted gDMRs, *Slc38a4* and *Zfp777*, also transitioned from protective to permissive between E18.5 and P10. At the *Zfp777* locus, this transition appeared to be driven by an increase of K36me3, as K4me2 and K4me3 are low at both time points (Fig. 3G). Taken together, these results suggest that K36me3 enrichment is a general feature of GVmeth CGIs between E18.5 and P10, while a minority of GVmeth CGIs additionally loses H3K4 methylation during this time.

KDM1B is the primary H3K4me2 demethylase required for CGI and imprinted gDMR methylation in the oocyte

We observed enrichment of K4me2 and K4me3 at GVmeth CGIs at both E18.5 and P10; however, before these regions can be methylated, these marks may need to be removed to facilitate recruitment of DNMT3A/3L. To further investigate the role of H3K4 methylation in DNA methylation establishment during oogenesis, we conducted whole-genome bisulfite sequencing using a low-cell post-bisulfite adaptor-tagging (PBAT) method (Miura et al. 2012) on ovulated metaphase II (MII) oocytes deficient in either KDM1A or KDM1B. Unbiased genome-wide sliding window analysis identified 3425 hypermethylated DMRs (hyperDMRs) and 7899 hypomethylated DMRs (hypoDMRs) between controls and the KDM1A knockout; in the KDM1B knockout, the overwhelming majority of all identified DMRs were hypomethylated (45,102 of 45,187) (Supplemental Fig. 4A). DMRs in both knockouts were primarily genic, especially among KDM1B hypoDMRs, of which >90% were intragenic with respect to the oocyte transcriptome (Supplemental Fig. 4B). Assessing DNA methylation at CGIs in the KDM1A knockout, 141 CGIs were hypermethylated and 360 were hypomethylated compared with the control (logistic regression, $P < 0.05$) (Fig. 4A, top). In the KDM1B knockout, 893 CGIs showed methylation levels different from those of the wild-type (logistic regression, $P < 0.05$); the vast majority of these CGIs (885) was hypomethylated (Fig. 4A, bottom). None of the KDM1B hypermethylated CGIs attained full (>75%) methylation; all eight were intermediately (25%–75%) methylated in the knockout and unmethylated (<25%) in the control. Of the KDM1B hypomethylated CGIs, eight were aberrantly hypermethylated and 192 were hypomethylated in the KDM1A data set (Supplemental Table 5; Supplemental Fig. 4C). It therefore appears that the majority of GVmeth CGIs require KDM1B for proper methylation establishment, and at least some CGIs require both KDM1A and KDM1B to gain full DNA methylation.

We next asked how imprinted gDMR methylation is affected in the absence of KDM1A or KDM1B. In the KDM1A knockout, all but two gDMRs, *Gnas1A* and *Cdh15*, retained full methylation ($\geq 75\%$) (Fig. 4B, top). As has been observed previously, several gDMRs retain full methylation in the KDM1B knockout oocyte; in our data, these are *Napl15*, *Zfp787*, *Inpp5f*, and *Cdh15* (Fig.

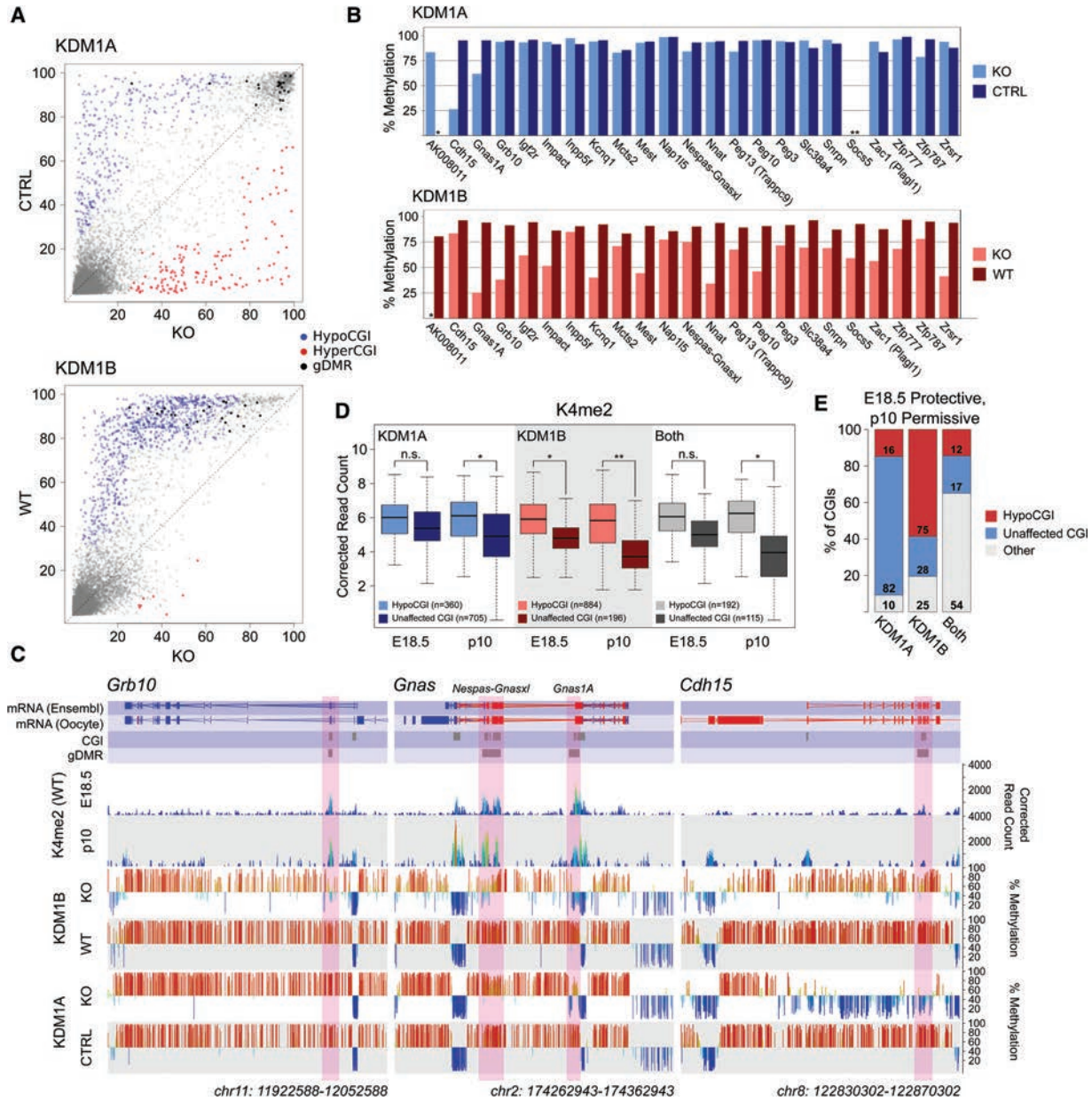


Figure 4. KDM1B, not KDM1A, predominantly affects CGI methylation. (A) Scatter plots showing methylation at CGIs in KDM1A-deficient (top) and KDM1B-deficient (bottom) oocytes, compared with controls. (Blue dots) Significantly hypomethylated CGIs (hypoCGIs); (red dots) significantly hypermethylated CGIs (hyperCGIs); (black dots) imprinted gDMRs; (WT) wild type; (CTRL) *Kdm1a*^{fllox/+}, *ZP3-Cre*-control. (B) Bar plots showing methylation levels at the 23 maternally imprinted DMRs in KDM1A-deficient (top) and KDM1B-deficient (bottom) oocytes, compared with controls. (C) Genome screenshots showing K4me2 ChIP-seq signal (as corrected read coverage) in E18.5 and P10 wild-type oocytes and DNA methylation over the *Grb10*, *Gnas*, and *Cdh15* gDMRs in normal and KDM1A- or KDM1B-deficient MII oocytes. Shown is a 100-bp window with a 100-bp step. (D) Box plots showing E18.5 and P10 K4me2 (log-transformed corrected read count) at 4 kb centered on CGIs that lose methylation in the absence of KDM1A, KDM1B, or both, compared with unaffected CGIs that gain full methylation in the knockouts. (Black line) Median; (dashed lines) 1.5× interquartile range. (*) $P < 1 \times 10^{-10}$; (**) $P < 1 \times 10^{-25}$, Mood's median test. (E) Stacked bar plot depicting the proportion of CGIs transitioning from protective to permissive states between E18.5 and P10 that lose methylation in the absence of KDM1A, KDM1B, or both. (Other) Either gain methylation in the knockout or are intermediately/unmethylated in the control. Shown are CGIs with adequate coverage in both the PBAT and the ChIP-seq data.

4B, bottom). However, no gDMR was completely unmethylated in either the KDM1A or KDM1B knockout (although *Gnas1A*, at 25.3% methylation, nearly met the cutoff in the KDM1B knockout); instead, various levels

of intermediate methylation were observed at all compromised gDMRs in the knockouts. This indicates that, like GVMeth CGIs in general, imprinted gDMRs primarily require KDM1B for methylation establishment, with

KDM1A playing a minor role. The *Grb10* gDMR, for example, shows intermediate methylation in the KDM1B knockout but full methylation in the KDM1A knockout (Fig. 4C, left panel). The *Gnas* locus shows ~74% methylation at the *Nespas-Gnasxl* gDMR but ~25% methylation at the *Gnas1A* gDMR in the KDM1B knockout; *Nespas-Gnasxl* and *Gnas1A* show 84% and 62% methylation, respectively, in the KDM1A knockout (Fig. 4C, middle panel). *Cdh15* is the only uniquely compromised gDMR in the KDM1A knockout; *Gnas1A* exhibits intermediate methylation in the absence of both KDM1A and KDM1B, suggesting that it requires both to gain methylation (Fig. 4C, middle and right panels; Supplemental Table 6). At all three loci, some loss of methylation within the gene body can also be seen in the KDM1B knockout, suggesting that KDM1B activity is important for proper DNA methylation establishment along the entire transcription unit. Strikingly, in the KDM1A knockout, methylation has been lost along nearly the entire *Cdh15* transcription unit, suggesting that loss of methylation at the gDMR is a by-product of down-regulated *Cdh15* transcription in the absence of KDM1A. Indeed, genic hyperDMRs and hypoDMRs tended to be in genes that show increased or decreased expression, respectively, in the knockout, as assessed by RNA-seq on GV oocytes (Supplemental Fig. 4D). In contrast, KDM1B hyperDMRs and hypoDMRs showed no relationship with expression of the associated gene. This suggests that KDM1A and KDM1B play distinct roles within the oocyte: KDM1A may act a transcription factor, and misregulated methylation in the KDM1A knockout is a consequence of misregulated expression, while KDM1B demethylates K4me2 as part of the elongating RNA polymerase II (Pol II) complex (Fang et al. 2010, 2013), and therefore the KDM1B knockout shows severe hypomethylation but normal gene expression.

Since KDM1A and KDM1B have been most extensively characterized as K4me2-me0 demethylases, we compared K4me2 enrichment at hypomethylated and unaffected CGIs and gDMRs in our wild-type E18.5 and P10 ChIP-seq data sets. For comparison, we also looked at K4me2 enrichment at CGIs hypomethylated in the absence of either KDM1A or KDM1B and at CGIs unaffected by the absence of either protein. In all three comparisons, there was increased K4me2 at both time points in the CGIs that fail to gain full methylation when compared with the CGIs that gain methylation normally, but this was most pronounced in the KDM1B knockout (Fig. 4D). Similarly, the *Grb10* and *Gnas* imprinted gDMRs are enriched with K4me2 and show loss of methylation in the KDM1B knockout, while the *Cdh15* imprinted gDMR, whose methylation is unaffected by loss of KDM1B, shows very little K4me2 enrichment at E18.5 or P10 (Fig. 4C). This indicates that KDM1B facilitates DNA methylation establishment by removing K4me2 at the majority of GVmeth CGIs and imprinted gDMRs and that CGIs unaffected by the loss of KDM1B and, to a lesser extent, KDM1A are constitutively in a permissive chromatin state rather than targets of an alternative K4me2 demethylation mechanism.

We then asked whether any CGIs hypomethylated in the absence of KDM1A or KDM1B transition from the protective to the permissive state between E18.5 and P10. Around 15% of E18.5 protective, P10 permissive CGIs are also KDM1A hypomethylated or KDM1A/KDM1B hypomethylated, while ~60% of such CGIs (75) are hypomethylated in the KDM1B knockout (Fig. 4E). However, no tangible change in K4me2 enrichment can be seen at KDM1B hypomethylated CGIs as a whole between E18.5 and P10 (Fig. 4D, middle panel). Instead, this shift appears to be driven by an increase in K36me3 and a decrease of K4me2 specifically at the loci that transition from protective to permissive between E18.5 and P10, compared with KDM1B hypomethylated CGIs that do not change states (809) (Supplemental Fig. 4E). This further supports the idea that H3K4me removal, perhaps facilitated by KDM1B, occurs at specific loci prior to P10, but global H3K4me removal at most GVmeth CGIs occurs sometime after the onset of de novo DNA methylation.

Discussion

Studies of the mammalian oocyte have been hindered by the lack of high-throughput methods for their isolation. Through the FACS approaches developed in this work, molecular and genome-wide studies of the oocyte are much more feasible. Beyond ChIP-seq, with the relevant modifications, this approach could be used to study nucleosome positioning, three-dimensional chromatin conformation, chromosome accessibility, and a number of other aspects of oocyte biology.

Using a low-cell ChIP method and focusing on histone modifications implicated in influencing de novo DNA methylation, our work reveals a step-wise modification of the chromatin at CGIs destined for DNA methylation in the mature oocyte (Fig. 5). Overall, although enrichment in K36me3 and depletion of K4me2/3 are ultimately required as the permissive state for engagement of DNMT3A/3L, the process and timing by which this state is attained may differ among susceptible CGIs. This suggests that there is not a single common mechanism of CGI chromatin modulation downstream from the universal requirement for transcription in CGI methylation in oocytes (Veselovska et al. 2015).

Our work shows that CGIs, even those found within active genes, are usually depleted of K36me3, a mark of elongating chromatin, but K36me3 increases specifically at GVmeth CGIs in parallel with the activation of the oocyte transcriptome. Since K36me3 characterizes the transcription units of active genes, this suggests one of two scenarios: K36me3 is either habitually removed from intragenic CGIs or habitually blocked from being established at these loci. In ES cells, CGIs have been shown to be constitutively depleted of K36me2 by KDM2A, which is recruited to CGIs via its CXXC domain (Blackledge et al. 2010). In vivo, the mammalian H3K36 trimethyltransferase SETD2 appears to work only from a dimethylated template (Yuan et al. 2009), suggesting that depleted K36me3 at CGIs is a consequence of depleted K36me2.

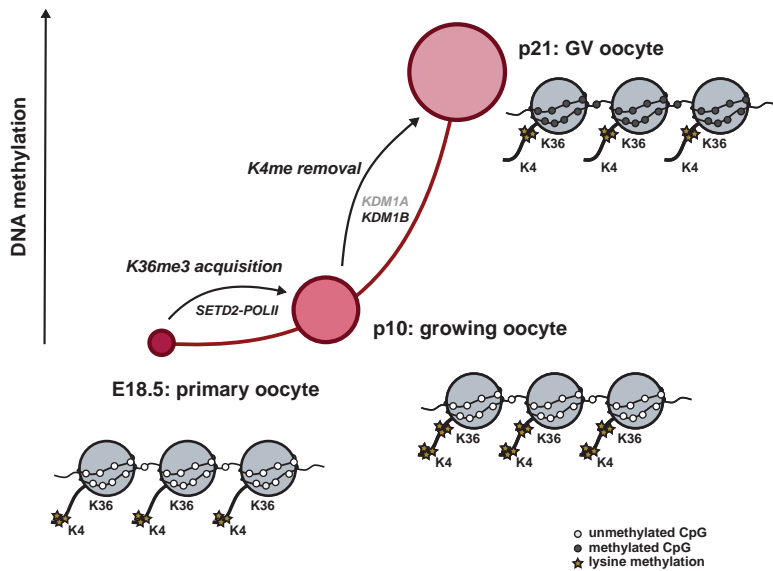


Figure 5. Stepwise modulation of CGI chromatin facilitates DNA methylation acquisition. At E18.5, CGIs destined for DNA methylation are marked with protective K4me2/K4me3 and lack permissive K36me3. In parallel with postnatal activation of the oocyte transcriptome, K36me3 accumulates specifically on CGIs destined for DNA methylation, and, at P10, these CGIs are marked with both K4me2/me3 and K36me3. During de novo DNA methylation between P10 and the GV stage, H3K4 methylation from CGIs must be removed, as evidenced by the methylation defects seen in mature oocytes in the absence of KDM1A and especially KDM1B.

Alternatively, KDM4A, a member of the only family of lysine demethylases that can demethylate K36me3 (Klose et al. 2006), has been shown to be recruited by K4me3 (Huang et al. 2006), suggesting that K4me3-depleted CGIs (and many GVMeth CGIs are constitutively depleted in K4me3) may thus accumulate K36me3. More work will need to be done to determine which scenario explains the increased K36me3 seen at GVMeth CGIs in the growing oocyte.

Between E18.5 and P10, a subset of GVMeth CGIs also loses protective K4me2/3. K4me2 loss can be attributed to the activity of KDM1A and, in particular, KDM1B; however, the demethylases or mechanisms responsible for K4me3 removal remain unknown. The requirement for the KDM1A/KDM1B K4me2 demethylases implies that K4me3 is removed at most oocyte methylated CGIs via conversion to K4me2 rather than through a nucleosome turnover/replacement mechanism. Several K4me3 demethylases have been identified (Mosammaparast and Shi 2010), but which are active at oocyte methylated CGIs remains unknown.

We detected an interesting difference in the effect of KDM1A and KDM1B loss on the oocyte methylome: KDM1A-deficient oocytes had similar numbers of aberrantly hypomethylated and hypermethylated CGIs, while KDM1B-deficient oocytes had overwhelmingly hypomethylated CGIs. In addition, in the case of KDM1A, there was a clear correlation between gene expression changes and DNA methylation changes, a relationship that was not apparent for loci that were differentially methylated in the absence of KDM1B. KDM1A acts as a transcriptional regulator in a number of other contexts (Metzger et al. 2005; Macfarlan et al. 2011; Schenk et al. 2012; Kerenyi et al. 2013). Our work suggests that many of the KDM1A-dependent methylation differences at CGIs are downstream from KDM1A-mediated transcriptional changes and are not reflective of KDM1A K4me2 demethylase activity. Instead, the K4me2 demethylase activity of KDM1B, which is known to associate with the elongat-

ing Pol II complex (Fang et al. 2010, 2013), is directly responsible for the majority of hypomethylation events in KDM1B-null oocytes (Supplemental Fig. 4F).

While K36me3 enrichment at oocyte methylated CGIs was established prior to the onset of de novo methylation at P10, K4me2/3 enrichment remained largely unchanged between E18.5 and P10, which encapsulates the time at which the oocyte transcriptome becomes active from approximately P5. This suggests that K4me2/3 removal is not facilitated by transcriptional activity alone. Among imprinted gDMRs, there are early methylating and late methylating subsets (Obata and Kono 2002; Hiura et al. 2006), and it is possible that this asynchronicity extends generally to CGIs that acquire methylation in oocytes. Therefore, the propensity of a CGI to gain methylation earlier or later in oogenesis may be a consequence of the degree of K4me2/3 that it initially contains as well as the timing of K4me2/3 removal. In somatic contexts, CpG density and transcription factor binding, both of which directly or indirectly enrich for K4me2/3, have been shown to be instrumental in protecting CGIs during de novo methylation (Krebs et al. 2014; Wachter et al. 2014). It may be that, in the oocyte, multiple rounds of transcription and targeting are required to enrich K36me3 and deplete K4me2/3 sufficiently to overcome these protections and allow DNMT3A/3L to access and methylate CGI DNA.

Materials and methods

FACS

All experimental procedures were approved by the Animal Welfare and Ethical Review Body at the Babraham Institute and were performed under licenses issued by the Home Office (UK) in accordance with the Animals (Scientific Procedures) Act 1986. To isolate E18.5 primary oocytes, ovaries from CD1 E18.5 fetuses were digested using 2 mg/mL collagenase (Sigma, C2674) and 0.025% trypsin (Sigma, 93615) in 1× PBS.

The reaction was stopped in 10% final volume of fetal bovine serum (FBS) (Sigma, F7524) and filtered through a 50- μ m mesh. Cells were centrifuged at 1200 rpm for 5 min and incubated in a 1:1000 dilution of Live/Dead fixable dead cell stain (Thermo Scientific, L23101) for 30 min, covered, at 4°C. Cells were washed in FACS buffer (1 \times PBS, 1% FBS) and fixed in 1% formaldehyde (Thermo Scientific, 28906) with 10% FBS for 10 min at room temperature followed by a 5-min quench with glycine (in True MicroChIP kit, Diagenode). The cells were then incubated in a 1:1000 dilution of SYCP3 antibody (Abcam, ab97672) for 30 min, covered, at 4°C followed by incubation with secondary antibody (Alexa fluor 647, donkey anti-mouse IgG). Antibody incubations were done in permeabilization buffer (1 \times PBS, 1% FBS, 0.05% Tween-20, 0.05% Triton-X).

Isolation of P10 oocyte nuclei was largely as described above. Following CD1 ovary digestion, cells were not strained; instead, nuclei were isolated as described by Rosner et al. (2013) except that no washes were done after cell membrane disruption. Nuclei were then fixed using the same method as in E18.5 FACS. Nuclei were incubated in a 1:200 dilution of NOBOX antibody (Abcam, ab41521) for 30 min, covered, at 4°C followed by incubation with secondary antibody (Alexa fluor 568, goat anti-rabbit IgG) and DNA stain (Invitrogen, Hoechst 33342). All sorting was done on a FACSARIAIII (Becton Dickinson).

Primary and growing oocyte RNA-seq

E18.5 oocytes used to extract RNA were collected using a protocol adapted from Pan et al. (2011). The protocol was as above, but the FACS buffer used contained 1% BSA (Sigma, A2058) instead of 1% FBS and, additionally, 100 U/mL Ribolock RNase inhibitor (Thermo Fisher, EO0381) and 5 mM DTT (from SuperScript first strand synthesis system, Thermo Fisher, 11904-018). RNA from sorted E18.5 oocytes was extracted using the RecoverAll total nucleic acid isolation kit (Ambion/Thermo Scientific, AM1975) and purified using the RNA Clean and Concentrator-5 kit (Zymo, R1013) with on-column DNase treatment (RNase-free DNase I; Thermo Scientific, AM2222). Ribosomal RNA was depleted from total RNA using the Ribo-Zero magnetic kit (human/mouse/rat, low input; Epicentre, SCL24G). To generate RNA-seq libraries, RNA was reverse-transcribed using SuperScript III (Life Technologies, 18080044), with the second DNA strand synthesized using dUTPs and DNA polymerase I (New England Biolabs, M0210S). Libraries were constructed using the NEBNext DNA library preparation master mix set for Illumina (New England Biolabs, E6040S), with an additional step of dUTP excision by User enzyme (New England Biolabs, M5505S) before PCR amplification. Growing oocyte RNA-seq was taken from previously published work (Veselovska et al. 2015).

ChIP-seq

Lysis and immunoprecipitation were performed using the True MicroChIP kit (Diagenode, AB-002-0016) with the following modifications. Up to 250,000 cells were sonicated in one lysate and split into 25,000- or 5000-cell equivalents after sonication. Biological replicates were always derived from distinct lysates. Samples were lysed using 25 μ L of buffer tL1 and incubated for 5 min on ice. Seventy-five microliters of 1 \times Hank's buffered salt solution was added, and the lysate was sonicated in 1.5-mL TPX microtubes (Diagenode). Chromatin was sheared using a Bioruptor (Diagenode) with five active cycles (30 sec on, 30 sec off). Sonicate was aliquoted, and an equivalent volume of complete ChIP buffer tC1 was added. For immunoprecipitation, the following antibodies and amounts of antibody were used for

both 25,000- and 5000-cell ChIP: 0.25 μ g of H3K4me3 (Diagenode, pAb-003-050), 0.25 μ g of H3K4me2 (Abcam, ab32356), 5 μ L of H3K36me3 (donated by Robert Klose), and 0.25 μ g of H3 (Abcam, ab1791). Immunoprecipitation and washes were done as described in the True MicroChIP protocol, with elution using SPRI beads (1:1 ratio; Fisher Scientific, 09-981-123). ChIP-seq libraries were generated from immunoprecipitated DNA using the NEBNext Ultra end repair/da-Tailing module (E7742) followed by SPRI purification (1:1 ratio) and 15–20 cycles of amplification using Phusion polymerase (Thermo Scientific, F-530S) before final purification using SPRI beads.

KDM1A/KDM1B MII oocyte collection

Experimental mice were used in accordance with the National Institutes of Health Guide for the Care and Use of Laboratory Animals using institutional care and committee-approved protocols. To disrupt KDM1A in oocytes, mice bearing the *Kdm1a* conditional allele *Lsd1^{fl}* (Kerenyi et al. 2013) were crossed with ZP3-Cre transgenic mice. *Lsd1^{fl/fl}/ZP3-Cre⁺* female mice were referred to as KDM1A knockout; *Lsd1^{fl/fl}/ZP3-Cre⁻* female littermates were used as controls. Generation of KDM1B knockout (*Aof1^{lox/lox}*) mice was described previously (Ciccone et al. 2010), and wild-type littermates were used as controls. To obtain MII oocytes, 6- to 8-wk-old mice were injected intraperitoneally with 5 IU of pregnant mare's serum gonadotrophin (PMSG) (Sigma) and, 48 h later, 5 IU of human chorionic gonadotrophin (hCG) (Sigma). Superovulated mice were euthanized 16 h after hCG injection, and oocytes were collected from the oviducts and released into a hyaluronidase/M2 solution for removal of cumulus cells.

KDM1A/KDM1B PBAT

Genome-wide methylation maps in KDM1A- and KDM1B-deficient MII oocytes and controls were generated using the PBAT method (Miura et al. 2012) as adapted in Peat et al. (2014). For each genotype, triplicate PBAT libraries were generated, each from ~100 MII oocytes. Each sample was spiked with ~5 pg of λ DNA to assess conversion efficiency, which was 98.7%–99.2% for CpGs. Briefly, cells were lysed in EB buffer (Qiagen) with 0.5% SDS (Sigma, 161-0418) and bisulfite-treated using the one-step modification procedure in the Imprint DNA modification kit (Sigma, MOD50-1KT). The resulting DNA was purified using the EZ DNA methylation direct kit (Zymo, D5020). First strand synthesis was performed using Klenow Exo⁻ (New England Biolabs, M0212S) and a custom streptavidin-conjugated adaptor containing standard Illumina adaptor sequences and 9 base pairs (bp) of random sequence (9N). This was followed by exonuclease I treatment (New England Biolabs, M0293S), purification using SPRI beads (1.8 \times ratio; Fisher Scientific, 09-981-123), and binding to biotin beads (Life Technologies, 11205D). Samples were then subjected to second strand synthesis, again using custom primer, followed by 10 cycles of library amplification using Phusion polymerase (Thermo Scientific, F-530S) before final purification using SPRI beads.

Library sequencing and mapping

All ChIP-seq, RNA-seq, and PBAT libraries were sequenced 100-bp paired-end on an Illumina HiSeq1000. Raw reads were trimmed to remove poor-quality calls and adapters using Trim-Galore and mapped to the mouse genome (GRCm38 assembly). ChIP-seq libraries were mapped using Bowtie1 "unique" parameters; RNA-seq libraries were mapped using TopHat version 2.0.9

(option `-g 1`). Paired-end PBAT libraries were mapped using Bismark version 0.12.5 (options `-pbat`, `-unmapped -bowtie2`), any unmapped reads were then mapped as single-end using Bismark version 0.12.5 (read1 options: `-bowtie2`, `-pbat`; read2 options: `-bowtie2`).

RNA-seq analysis

RNA-seq reads were assembled into E18.5 and growing oocyte transcriptomes using Cufflinks (Trapnell et al. 2010; 2012). The E18.5 transcriptome was assembled using Cufflinks version 2.1.1 with default parameters. The strategy for growing oocyte transcriptome assembly has been described (Veselovska et al. 2015). For simplicity, the Ensembl reference gene annotation was used to define silent and active promoters and their associated genes. A promoter or gene was considered active if it contained at least one read and silent if it contained <0.001 reads. Only genes that did not overlap with any other Ensembl gene annotations were analyzed. CGI and maternal gDMR annotations were from previously published data (Illingworth et al. 2010; Tomizawa et al. 2011; Proudhon et al. 2012). CGIs were annotated as (1) a TSS if they overlapped the 5'-most TSS of a gene ± 100 bp, (2) an alternative TSS if they overlapped any other TSS within a gene ± 100 bp, (3) intragenic if they overlapped a gene but were not TSS-associated, and (4) intergenic if they did not overlap an active gene. The true transcriptional activity of alternative TSSs could not be confidently determined, so they were discarded from downstream analyses. CGIs on the Y chromosome were excluded from analysis.

ChIP-seq analysis

Sequencing and quality control details for each ChIP-seq library are in Supplemental Table 7. All data sets were deduplicated prior to analysis; 4 kb centered on each CGI or maternal gDMR was used as the probe length for all relevant analyses. To exclude mapping artifacts, 10-kb nonoverlapping windows were tiled across the genome and quantitated as raw read counts in E18.5 and P10 H3 ChIP-seq data sets. Any window with a value more than three times the median in either the E18.5 or P10 H3 ChIP-seq data set was considered an outlier. Any probe that overlapped by $>5\%$ of its length with one of these outliers was discarded from all downstream analyses.

Peak calling was done essentially as described (Mikkelsen et al. 2010) with the following modifications: For each data set, raw read counts were generated over 1-kb windows with a 100-bp step, and the probabilities were generated for each window from assuming a Poisson distribution with a mean equal to the expected number of ChIP reads given the window size, genome size, and total number of aligned reads. Windows were not normalized to an input control (H3 ChIP-seq at E18.5 and P10 was instead used to filter out coverage outliers, see above). Windows with $P < 1 \times 10^{-6}$ after Benjamini-Hochberg correction for multiple testing were kept; overlapping windows were merged to generate the final peak calls.

For all CGI and maternal gDMR comparisons, biological replicates were first merged. All data were quantitated as follows: Read counts for each time point were corrected to the largest data store, one read was manually added to each probe, and the data were log-transformed. To enable comparison between ChIP data sets of the same histone mark at the two time points, data sets were normalized to one another using a "multiply" correction to match values in each data set's distribution at the 75th percentile. All quantitation was done in Seqmonk version 0.32.0.

ChromHMM analysis

To compare chromatin at CGIs at E18.5 and P10, raw ChIP reads from 4 kb centered on CGIs were used to generate a three-state segmentation of the genome using ChromHMM (Ernst and Kellis 2012) using default parameters. E18.5 and P10 data sets were considered to originate from distinct cell types. To be designated as a particular state, a CGI had to be covered along $\geq 80\%$ of its length by that state.

DNA methylation analysis

Sequencing and quality control details for each PBAT library are in Supplemental Table 8. To filter out mapping artifacts, 25-kb nonoverlapping windows were tiled across the genome as probes and quantitated as read counts corrected to the largest data store. Any window with a value >10 times the median in any PBAT data set was considered an outlier and excluded from downstream analyses. For the CGI and imprinted gDMR analysis, methylation was calculated as the mean over each CGI. To be included in analysis, a CGI or maternal gDMR had to have a least three CpGs covered by three or more reads; probes that did not meet coverage requirements were discarded. Differentially methylated CGIs were defined as those that passed a $P < 0.05$ threshold as calculated by logistic regression after multiple testing corrections, requiring a methylation difference of $\geq 25\%$ between knockout and the appropriate control.

All data associated with this work have been deposited under Gene Expression Omnibus accession numbers GSE73803 and GSE74549.

Acknowledgments

We are grateful to Robert Klose for providing the H3K36me3 antibody. We thank Felix Krueger and Simon Andrews for bioinformatic support; the Babraham Biological Support Unit for mouse husbandry; Rebecca Newman, Rachael Walker, and Arthur Davies for FACS help; Kristina Tabbada for library sequencing; and Courtney Hanna and Wendy Dean for advice and critical reading of the manuscript. This work was supported by grants from the Medical Research Council of the UK (to G.K.), a Rising Star Award (R1108 to T.C.) from the Cancer Prevention and Research Institute of Texas (CPRI), and funding from the People Programme (Marie Curie Actions) of the European Union's Seventh Framework Programme FP7/2007-2013 under Research Executive Agency grant agreement number 290123.

References

- Baubec T, Colombo DF, Wirbelauer C, Schmidt J, Burger L, Krebs AR, Akalin A, Schübeler D. 2015. Genomic profiling of DNA methyltransferases reveals a role for DNMT3B in genic methylation. *Nature* **520**: 243–247.
- Blackledge NP, Klose R. 2011. CpG island chromatin: a platform for gene regulation. *Epigenetics* **6**: 147–152.
- Blackledge NP, Zhou JC, Tolstorukov MY, Farcas AM, Park PJ, Klose RJ. 2010. CpG islands recruit a histone H3 lysine 36 demethylase. *Mol Cell* **38**: 179–190.
- Bourc'his D, Xu GL, Lin CS, Bollman B, Bestor TH. 2001. Dnmt3L and the establishment of maternal genomic imprints. *Science* **294**: 2536–2539.
- Cheng X, Blumenthal RM. 2010. Coordinated chromatin control: structural and functional linkage of DNA and histone methylation. *Biochemistry* **49**: 2999–3008.

- Chotalia M, Smallwood SA, Ruf N, Dawson C, Lucifero D, Frontera M, James K, Dean W, Kelsey G. 2009. Transcription is required for establishment of germline methylation marks at imprinted genes. *Genes Dev* **23**: 105–117.
- Cicccone DN, Su H, Hevi S, Gay F, Lei H, Bajko J, Xu G, Li E, Chen T. 2010. KDM1B is a histone H3K4 demethylase required to establish maternal genomic imprints. *Nature* **461**: 415–418.
- Deaton AM, Bird A. 2011. CpG islands and the regulation of transcription. *Genes Dev* **25**: 1010–1022.
- Dhayalan A, Rajavelu A, Rathert P, Tamas R, Jurkowska RZ, Ragozin S, Jeltsch A. 2010. The Dnmt3a PWWP domain reads histone 3 lysine 36 trimethylation and guides DNA methylation. *J Biol Chem* **285**: 26114–26120.
- Edmunds JW, Mahadevan LC, Clayton AL. 2008. Dynamic histone H3 methylation during gene induction: HYPB/Setd2 mediates all H3K36 trimethylation. *EMBO J* **27**: 406–420.
- Ernst J, Kellis M. 2012. ChromHMM: automating chromatin-state discovery and characterization. *Nat Methods* **9**: 215–216.
- Fang R, Barbera AJ, Xu Y, Rutenberg M, Leonor T, Bi Q, Lan F, Mei P, Yuan GC, Lian C. 2010. Human LSD2/KDM1b/AOF1 regulates gene transcription by modulating intragenic H3K4me2 methylation. *Mol Cell* **39**: 222–233.
- Fang R, Chen F, Dong Z, Hu D, Barbera AJ, Clark EA, Fang J, Yang Y, Mei P, Rutenberg M, et al. 2013. LSD2/KDM1B and its cofactor NPAC/GLYR1 endow a structural and molecular model for regulation of H3K4 demethylation. *Mol Cell* **49**: 558–570.
- Handel MA, Schimenti JC. 2010. Genetics of mammalian meiosis: regulation, dynamics and impact on fertility. *Nat Rev Genet* **11**: 124–136.
- Hata K, Okano M, Lei H, Li E. 2002. Dnmt3L cooperates with the Dnmt3 family of de novo DNA methyltransferases to establish maternal imprints in mice. *Development* **129**: 1983–1993.
- Henikoff S, Shilatifard A. 2011. Histone modification: cause or cog? *Trends Genet* **27**: 389–396.
- Hiura H, Obata Y, Komiya J, Shirai M, Kono T. 2006. Oocyte growth-dependent progression of maternal imprinting in mice. *Genes Cells* **11**: 353–361.
- Huang Y, Fang J, Bedford MT, Zhang Y, Xu R-M. 2006. Recognition of histone H3 lysine-4 methylation by the double tudor domain of JMJD2A. *Science* **312**: 748–751.
- Illingworth R, Kerr A, Desousa D, Jørgensen H, Ellis P, Stalker J, Jackson D, Clee C, Plumb R, Rogers J, et al. 2008. A novel CpG island set identifies tissue-specific methylation at developmental gene loci. *PLoS Biol* **6**: e22.
- Illingworth RS, Gruenewald-Schneider U, Webb S, Kerr ARW, James KD, Turner DJ, Smith C, Harrison DJ, Andrews R, Bird AP. 2010. Orphan CpG islands identify numerous conserved promoters in the mammalian genome. *PLoS Genet* **6**: e1001134.
- Jones PA. 2012. Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nat Rev Genet* **13**: 484–492.
- Kaneda M, Okano M, Hata K, Sado T, Tsujimoto N, Li E, Sasaki H. 2004. Essential role for de novo DNA methyltransferase Dnmt3a in paternal and maternal imprinting. *Nature* **429**: 900–903.
- Kerenyi MA, Shao Z, Hsu Y-J, Guo G, Luc S, O'Brien K, Fujiwara Y, Peng C, Nguyen M, Orkin SH. 2013. Histone demethylase Lsd1 represses hematopoietic stem and progenitor cell signatures during blood cell maturation. *Elife* **2**: e00633.
- Kizer KO, Phatnani HP, Shibata Y, Hall H, Greenleaf AL, Strahl BD. 2005. A novel domain in Set2 mediates RNA polymerase II interaction and couples histone H3 K36 methylation with transcript elongation. *Mol Cell Biol* **25**: 3305–3316.
- Klose RJ, Kallin EM, Zhang Y. 2006. JmjC-domain-containing proteins and histone demethylation. *Nat Rev Genet* **7**: 715–727.
- Kobayashi H, Sakurai T, Imai M, Takahashi N, Fukuda A, Yayoi O, Sato S, Nakabayashi K, Hata K, Sotomaru Y, et al. 2012. Contribution of intragenic DNA methylation in mouse gametic DNA methylomes to establish oocyte-specific heritable marks. *PLoS Genet* **8**: e1002440.
- Krebs AR, Dessus-Babus S, Burger L, Schübeler D. 2014. High-throughput engineering of a mammalian genome reveals building principles of methylation states at CG rich regions. *Elife* **3**: e04094.
- Macfarlan TS, Gifford WD, Agarwal S, Driscoll S, Lettieri K, Wang J, Andrews SE, Franco L, Rosenfeld MG, Ren B, et al. 2011. Endogenous retroviruses and neighboring genes are coordinately repressed by LSD1/KDM1A. *Genes Dev* **25**: 594–607.
- Marks H, Kalkan T, Menafrá R, Denissov S, Jones K, Hofemeister H, Nichols J, Kranz A, Stewart AF, Smith A, et al. 2012. The transcriptional and epigenomic foundations of ground state pluripotency. *Cell* **149**: 590–604.
- Metzger E, Wissmann M, Yin N, Müller JM, Schneider R, Peters AHFM, Günther T, Buettner R, Schüle R. 2005. LSD1 demethylates repressive histone marks to promote androgen-receptor-dependent transcription. *Nature* **437**: 436–439.
- Mikkelsen TS, Xu Z, Zhang X, Wang L, Gimble JM, Lander ES, Rosen ED. 2010. Comparative epigenomic analysis of murine and human adipogenesis. *Cell* **143**: 156–169.
- Miura F, Enomoto Y, Dairiki R, Ito T. 2012. Amplification-free whole-genome bisulfite sequencing by post-bisulfite adaptor tagging. *Nucleic Acids Res* **40**: e136.
- Morselli M, Pastor WA, Montanini B, Nee K, Ferrari R, Fu K, Bonora G, Rubbi L, Clark AT, Ottonello S, et al. 2015. In vivo targeting of de novo DNA methylation by histone modifications in yeast and mouse. *Elife* **4**: e06205.
- Mosammamaparast N, Shi Y. 2010. Reversal of histone methylation: biochemical and molecular mechanisms of histone demethylases. *Annu Rev Biochem* **79**: 155–179.
- Noh K-M, Wang H, Kim HR, Wenderski W, Fang F, Li CH, Dewell S, Hughes SH, Melnick AM, Patel DJ, et al. 2015. Engineering of a histone-recognition domain in Dnmt3a alters the epigenetic landscape and phenotypic features of mouse ESCs. *Mol Cell* **59**: 89–103.
- Obata Y, Kono T. 2002. Maternal primary imprinting is established at a specific time for each gene throughout oocyte growth. *J Biol Chem* **277**: 5285–5289.
- Ooi SKT, Qiu C, Bernstein E, Li K, Jia D, Yang Z, Erdjument-Bromage H, Tempst P, Lin S-P, Allis CD, et al. 2007. DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. *Nature* **448**: 714–717.
- Pan Y, Ouyang Z, Wong WH, Baker JC. 2011. A new FACS approach isolates hESC derived endoderm using transcription factors. *PLoS One* **6**: e17536.
- Peat JR, Dean W, Clark SJ, Krueger F, Smallwood SA, Ficz G, Kim JK, Marioni JC, Hore TA, Reik W. 2014. Genome-wide bisulfite sequencing in zygotes identifies demethylation targets and maps the contribution of TET3 oxidation. *Cell Rep* **9**: 1990–2000.
- Pepling ME. 2006. From primordial germ cell to primordial follicle: mammalian female germ cell development. *Genesis* **44**: 622–632.
- Proudhon C, Duffié R, Ajjan S, Cowley M, Iranzo J, Carbajosa G, Saadeh H, Holland ML, Oakey RJ, Rakyan VK, et al. 2012. Protection against de novo methylation is instrumental in

- maintaining parent-of-origin methylation inherited from the gametes. *Mol Cell* **47**: 909–920.
- Rajkovic A, Pangas SA, Ballow D, Suzumori N, Matzuk MM. 2004. NOBOX deficiency disrupts early folliculogenesis and oocyte-specific gene expression. *Science* **305**: 1157–1159.
- Rosner M, Schipany K, Hengstschlager M. 2013. Merging high-quality biochemical fractionation with a refined flow cytometry approach to monitor nucleocytoplasmic protein expression throughout the unperturbed mammalian cell cycle. *Nat Protoc* **8**: 602–626.
- Sasaki H, Matsui Y. 2008. Epigenetic events in mammalian germ-cell development: reprogramming and beyond. *Nat Rev Genet* **9**: 129–140.
- Schenk T, Chen WC, Göllner S, Howell L, Jin L, Hebestreit K, Klein H-U, Popescu AC, Burnett A, Mills K, et al. 2012. Inhibition of the LSD1 (KDM1A) demethylase reactivates the all-trans-retinoic acid differentiation pathway in acute myeloid leukemia. *Nat Med* **18**: 605–611.
- Shirane K, Toh H, Kobayashi H, Miura F, Chiba H, Ito T, Kono T, Sasaki H. 2013. Mouse oocyte methylomes at base resolution reveal genome-wide accumulation of non-CpG methylation and role of DNA methyltransferases. *PLoS Genet* **9**: e1003439.
- Smallwood SA, Kelsey G. 2012. De novo DNA methylation: a germ cell perspective. *Trend Genet* **28**: 33–42.
- Smallwood SA, Tomizawa S-I, Krueger F, Ruf N, Carli N, Segonds-Pichon A, Sato S, Hata K, Andrews SR, Kelsey G. 2011. Dynamic CpG island methylation landscape in oocytes and preimplantation embryos. *Nat Genet* **43**: 811–814.
- Smith ZD, Meissner A. 2013. DNA methylation: roles in mammalian development. *Nat Rev Genet* **14**: 204–220.
- Smith EY, Futtner CR, Chamberlain SJ, Johnstone KA, Resnick JL. 2011. Transcription is required to establish maternal imprinting at the Prader-Willi syndrome and Angelman syndrome locus. *PLoS Genet* **7**: e1002422.
- Suzumori N, Yan C, Matzuk MM, Rajkovic A. 2002. Nobox is a homeobox-encoding gene preferentially expressed in primordial and growing oocytes. *Mech Dev* **111**: 137–141.
- Thomson JP, Skene PJ, Selfridge J, Clouaire T, Guy J, Webb S, Kerr ARW, Deaton A, Andrews R, James KD, et al. 2010. CpG islands influence chromatin structure via the CpG-binding protein Cfp1. *Nature* **464**: 1082–1086.
- Tomizawa SI, Kobayashi H, Watanabe T, Andrews S, Hata K, Kelsey G, Sasaki H. 2011. Dynamic stage-specific changes in imprinted differentially methylated regions during early mammalian development and prevalence of non-CpG methylation in oocytes. *Development* **138**: 811–820.
- Tomizawa S-I, Nowacka-Wosuzuk J, Kelsey G. 2012. DNA methylation establishment during oocyte growth: mechanisms and significance. *Int J Dev Biol* **56**: 867–875.
- Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L. 2010. Transcript assembly and quantification by RNA-seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol* **28**: 511–515.
- Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Salzberg SL, Rinn JL, Pachter L. 2012. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc* **7**: 562–578.
- Veselovska L, Smallwood SA, Saadeh H, Stewart KR, Krueger F, Maupetit-Méhous S, Arnaud P, Tomizawa S-I, Andrews S, Kelsey G. 2015. Deep sequencing and de novo assembly of the mouse oocyte transcriptome define the contribution of transcription to the DNA methylation landscape. *Genome Biol* **16**: 209.
- Wachter E, Quante T, Merusi C, Arczewska A, Stewart F, Webb S, Bird A, Ferguson-Smith AC. 2014. Synthetic CpG islands reveal DNA sequence determinants of chromatin structure. *Elife* **3**: e03397.
- Yoh SM, Lucas JS, Jones KA. 2008. The Iws1:Spt6:CTD complex controls cotranscriptional mRNA biosynthesis and HYPB/Setd2-mediated histone H3K36 methylation. *Genes Dev* **22**: 3422.
- Yuan W, Xie J, Long C, Erdjument-Bromage H, Ding X, Zheng Y, Tempst P, Chen S, Zhu B, Reinberg D. 2009. Heterogeneous nuclear ribonucleoprotein L is a subunit of human KMT3a/Set2 complex required for H3 Lys-36 trimethylation activity in vivo. *J Biol Chem* **284**: 15701–15707.