Loss of SETD1B results in the redistribution of genomic H3K4me3 in the oocyte

Courtney W. Hanna[*,1,2,^], Jiahao Huang[*,1], Christian Belton[1], Susanne Reinhardt[3], Andreas Dahl[3], Simon Andrews[4], A. Francis Stewart[5,6], Andrea Kranz[5,^], Gavin Kelsey[1,2,^]

*These authors contributed equally to this work.

[1]Epigenetics Programme, Babraham Institute, Cambridge, UK; [2]Centre for Trophoblast Research, University of Cambridge, Cambridge, UK; [3]Dresden Concept Genome Center, Center for Molecular and Cellular Bioengineering, Biotechnology Center, Technische Universität Dresden, Germany; [4]Bioinformatics Group, Babraham Institute, Cambridge, UK; [5]Genomics, Center for Molecular and Cellular Bioengineering, Biotechnology Center, Technische Universität Dresden, Germany; [6]Max-Planck-Institute for Cell Biology and Genetics, Dresden, Germany.

^Corresponding authors: Courtney Hanna (courtney.hanna@babraham.ac.uk); Andrea Kranz (andrea.kranz@tu-dresden.de); Gavin Kelsey (gavin.kelsey@babraham.ac.uk)

Summary

Histone 3 lysine 4 trimethylation (H3K4me3) is an epigenetic mark found at gene promoters and CpG islands. H3K4me3 is essential for mammalian development, yet mechanisms underlying its genomic targeting are poorly understood. H3K4me3 methyltransferases SETD1B and MLL2 are essential for oogenesis. We investigated changes in H3K4me3 in *Setd1b* conditional knockout (cKO) GV oocytes using ultra-low input ChIP-seq, in conjunction with DNA methylation and gene expression analysis. *Setd1b* cKO oocytes showed a redistribution of H3K4me3, with a marked loss at active gene promoters associated with downregulated gene expression. Remarkably, many regions gained H3K4me3 in *Setd1b* cKOs, in particular those that were DNA hypomethylated, transcriptionally inactive and CpG-rich - hallmarks of MLL2 targets. Thus, loss of SETD1B appears to enable enhanced MLL2 activity. Our work reveals two distinct, complementary mechanisms of genomic targeting of H3K4me3 in oogenesis, with SETD1B linked to gene expression in the oogenic program and MLL2 to CpG content.

Keywords

Oocyte, oogenesis, histone modification, epigenetics, H3K4me3, DNA methylation, histone methyltransferase, SETD1B, MLL2.

Introduction

Histone 3 lysine 4 trimethylation (H3K4me3) is a hallmark of active gene promoters and is amongst the most conserved eukaryotic epigenetic modifications (Bernstein et al., 2005; Santos-Rosa et al., 2002; X. Zhang et al., 2009). Despite high concordance with promoter activity, the function of H3K4me3 in gene regulation remains enigmatic (Howe et al., 2017). Loss of H3K4me3 has surprisingly little impact on global transcription in several contexts (Hödl and Basler, 2012; Margaritis et al., 2012). In mammalian cells, H3K4me3 is not only correlated with gene transcription, but is strongly associated with CpG islands irrespective of transcriptional activity (Clouaire et al., 2012). Some studies indicate a role for H3K4me3 in transcriptional induction (Vermeulen et al., 2010) as well as the maintenance of transcriptional activity (Weiner et al., 2012). However, H3K4me3 has been paradoxically associated with gene repression in yeast (Briggs et al., 2001; Margaritis et al., 2012) and global transcriptional silencing in oocytes (Andreu-Vieyra et al., 2010). In contrast to doubts regarding the regulation of transcription, H3K4me3 serves to prevent the imposition of repression either by Polycomb-Group action (Klymenko and Müller, 2004) or by inhibiting binding of *de novo* DNA methyltransferases at gene promoters and CpG islands (Ooi et al., 2007; Vermeulen et al., 2007). Indeed, before zygotic genome activation in zebrafish development, H3K4 methylation appears to prevent DNA methylation rather than serving a direct role in transcription (Murphy et al., 2018).

Because of the importance of H3K4me3 in development, fertility and disease, there is a need to understand the mechanisms that govern the placement and maintenance of H3K4me3 in chromatin. Several mechanisms that do not rely on positioning by sequence-specific DNA-binding proteins (i.e., transcription factors) have been identified. In yeast, the sole H3K4 methyltransferase, Set1, associates with both serine 5 and serine 2 phosphorylated forms of elongating RNA Polymerase II (Dehé et al., 2006) and the size of the H3K4me3 promoter peak closely correlates with mRNA production, indicating that H3K4me3 deposition is a consequence of transcription (Choudhury et al., 2019; Soares et al., 2017). The Set1 complex is highly conserved in evolution (Krogan et al., 2002; Roguev et al., 2001; Roguev et al., 2003; Ruthenburg et al., 2007) and includes a subunit involved in

chromatin targeting, Spp1 (in mammals, CXXC1 or CFP1), which contains a PHD finger that binds to H3K4me3 (Brown et al., 2017; Shi et al., 2007; Vermeulen et al., 2007; Vermeulen et al., 2010), and which potentially directs Set1 complex binding to active promoters. CXXC1, which is a subunit of the mammalian SETD1A and SETD1B complexes, also includes the Cxxc zinc finger that binds the other hallmark of mammalian promoters - unmethylated CpG dinucleotides (Clouaire et al., 2012; Skalnik, 2010; Voo et al., 2000).

In mammals, the mechanisms that target H3K4me3 in chromatin have been additionally challenging to disentangle because mammals have six partially redundant H3K4 methyltransferase enzymes (SETD1A, SETD1B, MLL1, MLL2, MLL3, and MLL4) (Bledau et al., 2014; Denissov et al., 2014; Glaser et al., 2006; Ruthenburg et al., 2007). Each of these H3K4 methyltransferases form a complex with a common set of four scaffold proteins: WDR5, RbBP5, ASH2L, and DPY30 (Ernst and Vakoc, 2012; Lee et al., 2004; Miller et al., 2001; Roguev et al., 2001; Ruthenburg et al., 2007). Structures of these highly conserved quintets have recently been solved (Worden et al., 2020). MLL1 and MLL2 appear to include the same combination of a PHD finger that binds H3K4me3 and a Cxxc zinc finger that binds unmethylated CpG dinucleotides, as found in CXXC1. Hence, at least four of the six complexes potentially can bind to CpG island promoters without reliance on transcription factor guidance. Despite this insight, we still have a poor understanding of the differential recruitment of the six H3K4 methyltransferases across genomic regions and in differing cell types. Furthermore, we are far from understanding the distinct functional specificities of the six mammalian H3K4 methyltransferases (Ashokkumar et al., 2020; Bledau et al., 2014; Glaser et al., 2006; Glaser et al., 2009; Jude et al., 2007).

Murine oogenesis presents a unique window into the epigenetic mechanisms that reset the genome for launching the developmental program, which critically rely on H3K4 and DNA methylation. In the oocyte, our previous work established that SETD1A, MLL1 and MLL3 are not required and only SETD1B, MLL2, and at least one more as yet unidentified H3K4 methyltransferase, are required (Andreu-Vieyra et al., 2010; Bledau et al., 2014; Brici et al., 2017). CXXC1 is essential for oogenesis, presumably to support SETD1B function, given the similarities in phenotypes of oocyte-specific *Cxxc1*

and *Setd1b* knockouts (Brici et al., 2017; Sha et al., 2018; Yu et al., 2017). We recently demonstrated the power of ultra-low input native ChIP-seq (ULI-nChIP-seq) in *Mll2* conditional knockout (cKO) oocytes to elucidate the changes in the H3K4me3 landscape (Hanna et al., 2018). Specifically, we identified that MLL2 was responsible for deposition of H3K4me3 at transcriptionally silent, unmethylated genomic regions, dependent on the underlying CpG density, thus giving rise to the characteristic broad domains of H3K4me3 in oocytes (Dahl et al., 2016; B. Zhang et al., 2016), but had relatively little impact on the oocyte transcriptome. Conditional deletion of *Setd1b* in oocytes disrupted the oocyte transcriptome without changing H3K4me3 abundance as evaluated by immunofluorescence (Brici et al., 2017). Having established that MLL2 deposits the majority of H3K4me3 across the oocyte genome shortly before global transcriptional silencing, here we address the role of SETD1B in establishing the oocyte H3K4me3 landscape using ULI-nChIP-seq.

Results

Loss of SETD1B in oogenesis results in gains and losses of H3K4me3

In order to evaluate the role of SETD1B in the H3K4me3 landscape in oocytes, we isolated fully-grown germinal vesicle (GV) oocytes from the ovaries of 21-day old mice after oocyte-specific ablation of *Setd1b* driven by *Gdf9*-Cre (Brici et al., 2017; Lan et al., 2004). Pools of approximately 200 oocytes were collected and processed for ULI-nChIP-seq (Hanna et al., 2018). The genome-wide distribution of H3K4me3, using 5kb running windows, differed only modestly between *Setd1b* cKO (N=3) and *Setd1b* WT (N=3) GV oocytes (Figure 1A, 1B, Supplementary Figure 1A, 1B), in concordance with the observations made by immunofluorescence (Brici et al., 2017). However, the biological replicates for *Setd1b* cKO separated from WT in a hierarchical cluster (Supplementary Figure 1A), and we identified 3.6% of the genome that exhibited significant gains or losses of H3K4me3 (Figure 1A, 1B). The majority of differentially enriched windows were not at gene promoters (Supplementary Figure 1C). However given the potential role in gene regulation, we focused first on gene promoters, stratifying them into active, weak and inactive promoters, as previously described by histone

modifications and gene expression in GV oocytes (Hanna et al., 2018). Promoter-associated 5kb windows that lost H3K4me3 in *Setd1b* cKO oocytes were significantly enriched at active gene promoters, defined by either H3K27ac or high gene expression ($p<0.0001$, Chi-square) (Figure 1C). Conversely, promoter-associated 5kb windows that gained H3K4me3 in the *Setd1b* cKO oocytes were significantly enriched at inactive promoters, defined either by the presence of H3K27me3 or undetectable gene expression ($p<0.0001$, Chi-square) (Figure 1C). This trend was recapitulated when we looked genome-wide at enrichment for H3K27ac and H3K27me3 (Figure 1D, Supplementary Figure 1C). Together these findings suggest that the loss of SETD1B in oocytes results in a redistribution of H3K4me3, away from active to repressed regions of the genome.

Loss of promoter H3K4me3 is linked to downregulated gene expression of oocyte transcriptional regulators in *Setd1b* cKO oocytes

We investigated whether the changes in promoter H3K4me3 in the *Setd1b* cKO oocytes were associated with changes in gene expression. Using single-cell RNA-seq, we identified 1519 differentially expressed genes (DEGs) between *Setd1b* cKO (N=4) and *Setd1b* WT (N=3) GV oocytes, with 594 genes downregulated and 925 genes upregulated in the *Setd1b* cKO (Figure 2A, Supplementary Figure 2A). Using published RNA-seq data from *Setd1b* cKO MII oocytes (Brici et al., 2017), we saw that *Setd1b* cKO DEGs showed consistent changes in expression, despite the distinctive transcriptomes between GV and MII oocytes (Supplementary Figure 2B, 2C). Importantly, we saw no difference in expression of any of the other H3K4 methyltransferases in *Setd1b* cKO GV oocytes (Supplementary Figure 2D).

We then assessed DEG promoters and found that H3K4me3 was significantly lower in *Setd1b* cKO compared to *Setd1b* WT GV oocytes at promoters of downregulated DEGs ($p<0.0001$, two-tailed t-test), while there was no difference at upregulated DEGs ($p=1.0$, two-tailed t-test) (Figure 2B). Gene ontology of down- and up-regulated DEGs further showed a significant enrichment for transcriptional regulators and the egg and ovary among down-regulated DEGs (Figure 2C). Thus, the downregulated expression of oocyte transcriptional regulators in *Setd1b* cKO oocytes may be a direct consequence of

loss of SETD1B-dependent promoter H3K4me3, while upregulated genes may be attributable to indirect effects.

SETD1B-deposited H3K4me3 is required for transcriptional upregulation of a subset of oogenesis genes

Active promoters display H3K4me3 in primary non-growing oocytes (Hanna et al., 2018). Because the *Gdf9*-Cre recombinase used to generate *Setd1b* cKO oocytes is active from postnatal day 3 (Lan et al., 2004), it is likely that only *de novo* and renewed H3K4me3 deposited by SETD1B will be affected and the established, early stage, H3K4me3 will persist. Therefore, we focused on gene expression and H3K4me3 changes at transcripts that are upregulated during oogenesis (Gahurova et al., 2017). Promoter H3K4me3 of upregulated transcripts significantly increased across oogenesis ($p<0.0001$, ANOVA) and decreased in the *Setd1b* cKO compared to WT ($p<0.0001$, *posthoc* Tukey test), while the same trend was not observed across promoters of all annotated gene (ENSEMBL release 95) (Figure 3A, Supplementary Table 1). We then determined whether the transcripts that are upregulated across oogenesis are also specifically enriched among those DEGs that lose expression in the *Setd1b* cKO. Indeed, there was a significant enrichment of genes upregulated during oogenesis among downregulated DEGs ($p<0.0001$, Chi-square) but not upregulated DEGs ($p=0.1$, Chi-square) in *Setd1b* cKO (Figure 3B, Supplementary Figure 3A).

As CXXC1 and SETD1B function together in the SETD1B complex (Lee et al., 2007), and have similar phenotypes when ablated in oocytes (Brici et al., 2017; Sha et al., 2018; Yu et al., 2017), we compared gene expression changes in *Setd1b* cKO with publicly available data for *Cxxc1* cKO GV oocytes (Yu et al., 2017). Similar to the *Setd1b* cKO DEGs, we observed a significant enrichment for genes upregulated during oogenesis among downregulated DEGs in the *Cxxc1* cKOs ($p<0.0001$, Chi-square), but not upregulated DEGs ($p=0.2$, Chi-square) (Figure 3B, Supplementary Figure 3B). Furthermore, we see a more than 10-fold enrichment for *Cxxc1* cKO downregulated DEGs among downregulated DEGs identified in the *Setd1b* cKO oocytes ($p<0.0001$), with less than a 2-fold enrichment of *Cxxc1* cKO downregulated DEGs among up- or downregulated DEGs in *Mll2* cKO oocytes

(p<0.0001 and p=0.04, respectively) (Figure 3C, Supplementary Figure 3C). These data support the proposition that CXXC1 facilitates SETD1B targeting to actively transcribed promoters, and that the corresponding SETD1B-deposited H3K4me3 contributes to upregulation of gene expression during oogenesis, at least for a subset of genes.

Increased CpG-dependent deposition of H3K4me3 after loss of SETD1B

A surprising observation was the gain of H3K4me3 at H3K27me3-marked regions in the *Setd1b* cKO (Figure 1C, 1D, Supplementary Figure 1C). MLL2 is responsible for transcription-independent H3K4me3 in the oocyte (Hanna et al., 2018), so this finding suggests increased MLL2 action after loss of SETD1B. To explore this in more detail, we compared *Mll2* and *Setd1b* cKO H3K4me3 datasets (Figure 4A). We identified regions with MLL2-dependent and independent H3K4me3 using the H3K4me3 peak calls from *Mll2* cKO and WT oocytes (Hanna et al., 2018). However, we did not see a global increase of H3K4me3 in *Setd1b* cKO oocytes across MLL2-dependent regions (Figure 4B), demonstrating that only a subset of MLL2 targets are affected, such as the *Hox* clusters (Figure 4C). Because MLL2 is highly dependent on underlying CpG content (Hanna et al., 2018), we then examined sequence composition in the regions that gain H3K4me3 in the *Setd1b* cKO and found a highly significant enrichment for CpG content compared to a random set of regions (Figure 4D). This concordance lends support to the conclusiuon that the increased H3K4me3 in the absence of SETD1B is due to increased MLL2 activity.

Regions that gain H3K4me3 in *Setd1b* KO oocytes become DNA hypomethylated

H3K4me3 and DNA methylation are mutually exclusive in the oocyte, despite both having atypical genomic patterning compared to somatic cells. In the oocyte, DNA methylation is almost exclusively restricted to transcribed gene bodies (Kobayashi et al., 2012; Smallwood and Kelsey, 2012), whereas H3K4me3 forms broad domains across the remaining unmethylated fraction of the genome (Dahl et al., 2016; B. Zhang et al., 2016), through the activity of MLL2 (Hanna et al., 2018). To evaluate the relationship between DNA methylation and the regions that gain H3K4me3 in *Setd1b* cKO oocytes (N=8,569), we assessed genome-wide patterns of DNA methylation using post-bisulphite adaptor

tagging (PBAT). DNA methylation in *Setd1b* cKO and WT GV oocytes was distinct (Figure 5A), with the vast majority (94%) of differentially methylated regions (DMRs) losing DNA methylation in the *Setd1b* cKOs (Figure 5B). This finding is not explained by a difference in mRNA expression of the de novo *Dnmts*, which are unchanged in *Setd1b* cKO GVs (Supplementary Figure 2D). When we examined DMRs that gained and lost H3K4me3 in *Setd1b* cKO oocytes, we observed a significant overlap between regions that gained H3K4me3 and DMRs that were DNA-hypomethylated in the *Setd1b* cKOs, whereas regions that lost H3K4me3 in the *Setd1b* cKO showed no apparent reciprocal gain in DNA methylation (Figure 5C, 5D, 5E). Therefore, these data again suggest that loss of SETD1B is accompanied by increased MLL2 action at certain sites because these sites also show decreased DNA methylation.

Discussion

Previously we reported that MLL2 was responsible for the broad, transcription-independent, H3K4me3 domains that are uniquely and abundantly found on the oocyte genome. However the H3K4me3 peaks associated with active promoters were largely unaffected (Hanna 2018). Because SETD1B is required for oogenesis (Brici 2017), but MLL1, MLL3 and SETD1A are not, we expected that SETD1B would be responsible for H3K4me3 on active promoters. This expectation was partly fulfilled but, unexpectedly, we observed a considerable reorganisation of genomic H3K4me3 involving both regional losses and gains. Overall, loss of H3K4me3 occurred at active promoters and concorded with reduced expression. Importantly, these genes are enriched for oocyte transcriptional regulators that are normally upregulated during oogenesis. Conversely, the regions that presented increased H3K4me3 correlated with hypomethylated DMRs, H3K27me3-marked promoters, and CpG-rich sequences, which are hallmarks of MLL2 targets (Hanna et al., 2018). These findings support a model whereby SETD1B deposits H3K4me3 on active gene promoters and supports transcriptional activity, while MLL2 deposits H3K4me3 on unmethylated CpG-rich regions (Figure 6). Furthermore, our data suggests that the loss of SETD1B increases either the stability or activity of MLL2, perhaps through increased availability of shared H3K4 methyltransferase cofactors.

In addition to a strong association between H3K4me3 deposited by MLL2 and high CpG content in the oocyte, the association between MLL2 and CpG islands, including transcriptionally inactive promoters, has been observed in other contexts, such as embryonic stem cells (Denissov et al., 2014; Hu et al., 2013). Despite the relatively low affinity of CxxC domains for CpG dinucleotides (Allen et al., 2006; Xu et al., 2018), MLL2 appears to uniformly occupy CpG islands (Denissov et al., 2014), supporting the proposition that MLL2 has a general ability to associate with CpG islands in the naïve epigenome (Glaser et al., 2009). Conversely, SETD1 complexes preferentially bind actively transcribed, H3K4me3-marked CpG island promoters (Brown et al., 2017; Mahadevan and Skalnik, 2016), in part due to the PHD finger of CXXC1 (Brown et al., 2017). Our data from oocytes supports the proposition that SETD1B is targeted to transcriptionally active promoters, likely through its CXXC1 subunit. Although MLL1, and presumably MLL2, contain a PHD finger that binds to H3K4me3 (Chang et al., 2010), the role of this interaction in their genomic distribution remains to be determined. Our observations in *Setd1b* cKO oocytes are consistent with our published findings that MLL2 activity in the oocyte is directed by binding to unmethylated CpGs rather than H3K4me3-marked regions (Hanna et al., 2018).

As the predominant promoter-associated H3K4me3 methyltransferase in oocytes, loss of SETD1B was expected to provoke loss of gene expression. However, in concordance with observations in *Setd1b* cKO MII oocytes (Brici et al., 2017), we find the counterintuitive trend that more mRNAs are up- than down-regulated in the *Setd1b* cKO GV oocyte transcriptome. Furthermore, promoters of these up-regulated genes show no gain in H3K4me3 in *Setd1b* cKO GV oocytes. This paradox may be explained by the observation that the downregulated genes are enriched for negative transcriptional regulators (Brici et al., 2017). In contrast, we found that downregulated genes are linked to reduced H3K4me3, suggesting that the failure to upregulate gene expression in the *Setd1b* cKO oocytes is a direct consequence of the loss of SETD1B-mediated H3K4me3, which in turn leads to substantial consequences for the transcriptional programme of the oocyte.

A surprising observation was that almost as many sites that lost H3K4me3, gained H3K4me3 in the *Setd1b* cKO GV oocytes. This finding explains why no difference in the total amount of H3K4me3 was observed in *Setd1b* cKO oocytes by immunofluorescence (Brici et al., 2017). The gain of H3K4me3 across hypomethylated, H3K27me3-marked, CpG-rich regions indicates that this is a consequence of increased activity of MLL2, but does not exclude the possibility that changes seen at a subset of promoters and/or repetitive elements may be due to loss of transcriptional repressors in *Setd1b* cKO oocytes (Brici et al., 2017). Our observations are consistent with trends seen in *Cxxc1* cKO oocytes, which also showed a gain of H3K4me3 at H3K27me3-marked promoters (Sha et al., 2021). We speculate that in the absence of SETD1B there may be improved stability and/or abundance of MLL2-containing H3K4 methyltransferase complexes due to an increased availability of the core cofactors WDR5, RbBP5, ASH2L and DPY30. While this dynamic has not been widely discussed in the literature, it is consistent with trends observed in *Mll2* KO embryonic stem cells, where a reciprocal gain of H3K4me3 was observed at highly enriched promoters while low enriched bivalent promoters lost H3K4me3 (Denissov et al., 2014). It would be difficult to test whether abundance of the core cofactors is rate-limiting in oocytes, but could warrant future study in other cell contexts.

The redistribution of H3K4me3 in *Setd1b* cKO oocytes apparently impacts the patterning of DNA methylation. The generalised loss of methylation we observe could suggest a delay in the normal *de novo* methylation process in *Setd1b* cKO oocytes, although we did not detect reduced transcript abundance of the key *Dnmts*. Alternatively, and given that *de novo* methylation and non-promoter H3K4me3 deposition occur in parallel but mutually exclusively in growing oocytes (Hanna et al., 2018), increased action or availability of MLL2 complexes could favour accumulation of H3K4me3 over domains that would otherwise be DNA methylated (Figure 6).

In this study, we used ultra-low input sequencing methods to reveal molecular mechanisms underlying the targeting of H3K4me3 in oogenesis. Our findings help to further our understanding of how the H3K4 methyltransferases are functioning *in vivo* and co-ordinately contribute to the gene regulatory landscape in oocytes.

Data accessibility

The sequencing data generated for this study has been deposited into the Gene Expression Omnibus database (GSE167987).

Acknowledgements

We would like to thank Heba Saadeh and Felix Krueger for their contributions in mapping and processing the post-bisulphite adaptor tagging data. We would like to thank the Sequencing and Bioinformatics facilities at the Babraham Institute, and the Animal facility of the Centre for Regenerative Therapies. We thank the Biomedical Services of the Max Planck Institute of Molecular Cell Biology and Genetics, Dresden, for excellent service.

This work was supported by funding from the UK Biotechnology and Biological Sciences Research Council (BBS/E/B/000C0423 to G.K.) and Medical Research Council (MR/K011332/1, MR/S000437/1 to G.K.), the Deutsche Forschungsgemeinschaft (KR2154/6-1 to A.K., STE903/12-1 and STE903/13-1 to A.F.S.), and a Next Generation Fellowship from the Centre for Trophoblast Research (to C.H.). J.H. was supported by 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.

Author contributions

A.F.S., A.K. and G.K. conceptualised the study. S.R., A.D., A.K., J.H. and C.H. collected the data. C.H., J.H., C.B., and S.A. performed data analysis. C.H., C.B., and S.A. generated manuscript figures. C.H. drafted the manuscript, with input from J.H., C.B., S.A., A.K., A.F.S. and G.K.

Declaration of Interests

The authors have no conflicts of interest to declare.

Figure legends

Figure 1. (A) The scatterplot shows average normalised enrichment for H3K4me3 for 5kb running windows (N=544,879) between *Setd1b* cKO (N=3) and WT (N=3) d21 GV oocytes. Differentially enriched windows were identified using LIMMA statistic (p<0.05, corrected for multiple comparisons), and those that show a loss in *Setd1b* cKO are shown in blue and a gain in red. **(B)** The screenshot shows the normalised enrichment for H3K4me3 for 1kb running windows with a 500bp step between *Setd1b* cKO and WT GV oocytes. Significant differentially enriched 5kb windows are shown in the *Setd1b* gain/loss track, with red and blue bars showing windows that gain and lose H3K4me3 in *Setd1b* cKO, respectively. **(C)** The barplot shows the overlap between sets of promoters and differentially enriched windows that gain or lose H3K4me3 compared to a random set of 5kb running windows (Chi-Square statistic, p<0.0001 and p<0.0001, respectively). **(D)** The scatterplot shows enrichment for H3K27ac and H3K27me3 for 5kb running windows (normalised RPKM) in d25 GV oocytes. The windows that gain and lose H3K4me3 in the *Setd1b* cKO are highlighted in red and blue, respectively.

Figure 2. (A) The scatterplot shows the abundance of oocyte transcripts (N=33,437) in *Setd1b* cKO and WT GV oocytes as measured by single-cell RNA-seq. Differentially expressed genes (DEGs) were identified using DESeq2 (p<0.05 corrected for multiple comparisons) and a >2-fold change in expression. **(B)** The heatmap shows H3K4me3 enrichment (normalised RPKM) in *Setd1b* cKO and WT oocytes across downregulated (upper panel) and upregulated (lower panel) DEGs, including 5kb upstream and downstream. **(C)** The barplot shows the –log(corrected p-value) for significant gene ontology categories for down- and upregulated DEGs in *Setd1b* cKO GV oocytes.

Figure 3. (A) The heatmap shows H3K4me3 enrichment (normalised RPKM) in d5 non-growing oocytes (NGOs), d10 growing oocytes (GOs), d15 GV, d25 GV, *Setd1b* cKO GV and WT GV oocytes across transcripts that are upregulated during oogenesis (left panel) and all annotated genes (right panel), including 5kb upstream and downstream of the gene. Promoter H3K4me3 enrichment was statistically compared using one-way ANOVA (p<0.0001) for the overlapping 5kb windows. **(B)** The barplot shows the fold enrichment for genes that are upregulated in oogenesis among the *Setd1b* cKO down- and upregulated DEGs (p<0.0001 and p=0.1, respectively), *Cxxc1* cKO down- and upregulated DEGs (p<0.0001 and p=0.2, respectively) and *Mll2* cKO down- and upregulated DEGs (p=0.05 and p<0.0001, respectively), relative to all oocyte transcripts. **(C)** The barplot shows the fold enrichment for *Cxxc1* cKO downregulated DEGs among the *Setd1b* cKO down- and upregulated DEGs (p<0.0001 and p<0.0001, respectively), and *Mll2* cKO down- and upregulated DEGs (p<0.0001 and p=0.04, respectively), relative to all oocyte transcripts.

Figure 4. (A) Principal component analysis shows H3K4me3 ChIP-seq replicates for *Setd1b* cKO (N=3), *Setd1b* WT (N=3), *Mll2* cKO (N=4), and *Mll2* WT (N=3) GV oocytes. 5kb running windows were quantitated using normalised RPKM. **(B)** The beanplot shows the relative enrichment of H3K4me3 in *Setd1b* cKO (cKO/WT) oocytes at regions with MLL2-independent and -dependent H3K4me3 (t-test, p=n.s.). Enrichment is quantitated as normalised RPKM for 5kb running windows falling within H3K4me3 peaks in the *Mll2* WT oocytes that are either present or absent in the *Mll2* cKO oocytes, defined as MLL2-independent and dependent, respectively. **(C)** Screenshot shows the average normalised enrichment for H3K4me3 for 1kb running windows with a 500bp step between *Setd1b* cKO (N=3), *Setd1b* WT (N=3), *Mll2* cKO (N=4), and *Mll2* WT (N=3) GV oocytes. Significant differentially

enriched 5kb windows in the *Setd1b* cKO are shown in the *Setd1b* cKO gain/loss track, with red and blue bars showing windows that gain and lose H3K4me3 in *Setd1b* cKO, respectively. **(D)** The barplot shows the relative enrichment for dimer content in regions that gain H3K4me3 in the *Setd1b* cKO compared to a random set of 5kb probes.

Figure 5. (A) Hierarchical cluster shows biological replicates for *Setd1b* cKO (N=3) and WT (N=3) GV oocyte DNA methylation patterns, quantitated as 100-CpG windows, with at least 10 informative CpGs (see Methods). **(B)** The scatterplot shows the DNA methylation in *Setd1b* cKO and WT GV oocytes across 100-CpG windows, with at least 10 informative CpGs in each replicate (N=424,105). Differentially methylated regions (DMRs) were identified using logistic regression ($p<0.05$ corrected for multiple comparisons) and a >20% difference in methylation. **(C)** The screenshot shows H3K4me3 enrichment of 1kb running windows with a 500bp step and DNA methylation of 100-CpG windows with at least 10 informative CpGs per replicate for *Setd1b* cKO and WT oocytes. Regions that gain or lose DNA methylation (*Setd1b* DMRs) or H3K4me3 in *Setd1b* cKO are shown in the labelled annotation tracks as red and blue bars, respectively. **(D)** The pie charts show the overlap between 5kb windows that gain (left) and lose (right) H3K4me3 in the *Setd1b* cKO with 100-CpG windows that are hypermethylated (HyperDMR) or hypomethylated (HypoDMR) in the *Setd1b* cKO (Chi-Square statistic, $p<0.0001$). **(E)** The boxplot shows the average difference in DNA methylation (*Setd1b* cKO – WT) of 100-CpG windows that overlap regions that gain and lose H3K4me3 in the *Setd1b* cKO (t-test, $p<0.0001$).

Figure 6. Our study supports the model that in oogenesis, SETD1B and CXXC1 work together to target H3K4me3 to actively transcribed gene promoters, while MLL2 targets transcriptionally inactive regions based on underlying CpG composition (upper panel). When SETD1B is ablated, H3K4me3 is lost at a subset of active promoters, resulting in a downregulation of transcription (lower panel). Loss of SETD1B increases the activity of MLL2 in depositing H3K4me3 at CpG-rich regions, many of which should otherwise be DNA methylated. This gain of H3K4me3 appears to delay the acquisition of DNA methylation or the increased abundance of MLL2 is able to outcompete DNMTs for occupancy at these loci, causing a compensatory loss of DNA methylation.

STAR Methods

Sample collection

Experiments were performed in accordance with German animal welfare legislation, and were approved by the relevant authorities, the Landesdirektion Dresden.

Ovaries were collected from 21-day-old females and digested with 2mg/mL collagenase (Sigma) and 0.02% trypsin solution, agitating at 37°C for 30 min. GV oocytes were collected in KSOM medium (Sigma) and washed with PBS from females with an oocyte conditional deletion of *Setd1b* using Gdf9-cre (*Setd1b* cKO) and females carrying floxed *Setd1b* (*Setd1b* WT). Four GV oocytes for each genotype were collected for single cell RNA-seq (scRNA-seq). Approximately 100 GV oocytes were collected for three biological replicates of each genotype for post-bisulphite adaptor tagging (PBAT). Approximately 200 GV oocytes were collected for four biological replicates of each genotype for H3K4me3 ULI-nChIP-seq. All molecular experiments were performed as previously described (Hanna et al., 2018), but are described in brief below.

Preparation of scRNA-seq libraries

Cells were lysed and RNA was reverse transcribed and amplified according to the SMARTer Ultra Low RNA Kit for Illumina Sequencing (version 1, Clontech). Libraries were prepared from cDNA using the NEBNext Ultra DNA library preparation for Illumina sequencing with indexed adaptors (New England Biolabs). Libraries were multiplexed for 75-bp single-read sequencing on an Illumina HiSeq 2000.

Preparation of PBAT libraries

Cells were lysed with 0.5% SDS in EB buffer and bisulphite treated with the Imprint DNA Modification kit (Sigma). The resulting DNA was purified using columns and reagents from the EZ DNA Methylation Direct kit (Zymo Research). First-strand synthesis was performed with Klenow Exo– enzyme (New England Biolabs) using a customized biotin-conjugated adaptor containing standard Illumina adaptor sequences and 9 bp of random sequences (9N), as previously described (Hanna et al., 2018). Following exonuclease I (New England Biolabs) treatment and binding to Dynabeads M-280 Streptavidin beads (Thermo Fisher Scientific), second-strand synthesis was performed with Klenow Exo– enzyme (New

England Biolabs) using a customized adaptor containing standard Illumina adaptor sequences and 9 bp of random sequences. Ten PCR cycles with Phusion High-Fidelity DNA polymerase (New England Biolabs) were used for library amplification with indexed adaptors. Libraries were multiplexed for 100-bp paired-end sequencing on an Illumina HiSeq 2500.

Preparation of ULI-nChIP-seq libraries

Samples were thawed on ice and permeabilized using nuclei EZ lysis buffer (Sigma) and 0.1% Triton-X-100 /0.1% deoxycholate. Micrococcal nuclease digestion was completed with 200 U of micrococcal nuclease (New England Biolabs) at 21 °C for 7.5 min. Chromatin samples were precleared in complete immunoprecipitation buffer with Protein A/G beads rotating for 2 hours at 4 °C. For each sample, 125ng of anti-H3K4me3 antibody (Diagenode, C15410003) was bound to Protein A/G beads in complete immunoprecipitation buffer for 3 hours rotating at 4 °C. Chromatin was then added to the antibody-bound beads and samples were rotated overnight at 4 °C. Chromatin-bound beads were washed with two low-salt washes and one high-salt wash and DNA was then eluted from the beads at 65 °C for 1.5 hours. Eluted DNA was purified with solid-phase reversible immobilization (SPRI) purification with Sera-Mag carboxylate-modified Magnetic SpeedBeads (Fisher Scientific) at a 1.8:1 ratio. Library preparation was completed using the MicroPlex Library Preparation kit v2 (Diagenode) with indexed adaptors, as per the manufacturers guidelines. Libraries were multiplexed for 75 bp Single End sequencing on an Illumina NextSeq500.

Library mapping and trimming

Fastq sequence files were quality and adaptor trimmed with trim galore v0.4.2 using default parameters. For PBAT libraries, the -clip option was used. Mapping of ChIP-seq data was performed with Bowtie v2.2.9 against the mouse GRCm38 genome assembly. The resulting hits were filtered to remove mappings with a MAPQ scores < 20. Mapping of RNA-seq data was performed with Hisat v2.0.5 against the mouse GRCm38 genome assembly, as guided by known splice sites taken from Ensembl v68. Hits were again filtered to remove mappings with MAPQ scores <20.

Mapping and methylation calling of bisulphite-seq data was performed using Bismark v0.16.3 in PBAT mode against the mouse GRCm38 genome assembly. Trimmed reads were first aligned to the genome in paired-end mode to be able to detect and discard overlapping parts of the reads while writing out unmapped singleton reads; in a second step remaining singleton reads were aligned in single-end mode. Alignments were carried out with Bismark (Krueger and Andrews, 2011) with the following set of parameters: a) paired-end mode: --pbat; b) single-end mode for Read 1: --pbat; c) single-end mode for Read 2: defaults. Reads were then deduplicated with deduplicate_bismark selecting a random alignment for positions that were covered more than once. Following methylation extraction, CpG context files from PE and SE runs were used to generate a single coverage file (the "Dirty Harry" procedure).

Publicly available datasets

Publicly available datasets were downloaded from Gene Expression Omnibus (GEO), including: H3K27me3 and H3K27ac ChIP-seq in GV oocytes, and H3K4me3 ChIP-seq from d5 non-growing oocytes, d10 growing oocytes, d15 GV oocytes and d25 GV oocytes (GSE93941) (Hanna et al., 2018), RNA-seq for GV oocytes (GSE70116) (Veselovska et al., 2015), RNA-seq from size-selected oocytes (GSE86297) (Gahurova et al., 2017), RNA-seq from *Setd1b* cKO and WT MII oocytes (GSE85360) (Brici et al., 2017) and RNA-seq from *Cxxc1* cKO and WT GV oocytes (GSE85019) (Yu et al., 2017). All raw data files were mapped and trimmed using the pipelines described above.

ChIP-seq analysis

5kb running windows were used for quantitative analysis of H3K4me3 ChIP-seq data using RPKM, with mapping artefacts excluded from analysis (RPKM>4 in any input sample). Poor quality H3K4me3 ChIP-seq libraries were excluded from analysis, which included one replicate of *Setd1b* WT and one replicate of *Setd1b* cKO. Poor enrichment was defined as a cumulative distribution plot reflecting a signal-to-noise similar to input and presenting as an outlier in hierarchical clustering. For all subsequent analyses, valid 5kb windows were quantitated using enrichment normalisation of RPKM in SeqMonk, as previously described (Hanna et al., 2018). LIMMA statistic ($p<0.05$ corrected for multiple

comparisons) was used to identify differentially enriched H3K4me3 peaks between *Setd1b* cKO and WT GV oocytes, using an average normalised RPKM for replicates within each group. Overlapping TSSs and promoters (TSS +/-500bp) were classified as previously described (Hanna et al., 2018). In brief, active promoters were defined as those marked with H3K27ac or transcripts with >1 FPKM in GV oocytes, weak promoters were defined those transcripts with an FPKM between 0.1 and 1, and inactive promoters were defined as those transcripts with undetectable gene expression (FPKM<0.1) or marked with H3K27me3. Regions with MLL2-dependent and -independent H3K4me3 were defined using published differential H3K4me3 peak calls between the *Mll2* cKO and WT GV oocytes (GSE93941) (Hanna et al., 2018).

RNA-seq analysis

RNA-seq datasets were quantitated using the RNA-seq quantitation pipeline in Seqmonk, over previously defined oocyte transcripts (Veselovska et al., 2015). Transcript isoforms were merged for all analyses. Differentially expressed genes (DEGs) were identified in three comparisons using DESeq2 ($p<0.05$ corrected for multiple comparisons) and a >2-fold change in expression: (1) *Setd1b* cKO and WT GV oocytes, (2) 10-30um oocytes (non-growing oocytes) and GV oocytes, and (3) *Cxxc1* cKO and WT GV oocytes. Fold enrichment for genes upregulated in oogenesis and DEGs among the *Setd1b* and *Cxxc1* cKOs was calculated relative to all oocyte transcripts and increase in observed over expected overlap between DEGs from different conditions was statistically compared using Chi-square. Gene ontology analysis was done using DAVID (https://david.ncifcrf.gov/) for up- and down-regulated DEGs, using the default settings with the addition of the UP_TISSUE category.

Sequence composition analysis

Dimer composition of candidate 5kb windows (gained H3K4me3 in *Setd1b* cKO and a random set) was assessed using compter (https://www.bioinformatics.babraham.ac.uk/compter/). The mouse genome was used as the background and values were expressed as log2 observed/expected. Enrichment values for each dimer in each condition were summarised by taking the mean, and the difference in mean log ratio (*Setd1b* cKO - Random) was calculated and plotted.

DNA methylation analysis

DNA methylation datasets were analysed using 100-CpG running windows, with a minimum coverage of at least 10-CpGs, using the bisulphite-seq quantitation pipeline in SeqMonk. Differentially methylated regions (DMRs) were identified between *Setd1b* cKO and WT GV oocytes using logistic regression ($p<0.05$ corrected for multiple comparisons) and a >20% difference in methylation.

Supplementary Item Legends

Supplementary Table 1. H3K4me3 enrichment for promoters of genes that upregulate during oogenesis and all genes in d5 non-growing oocytes (NGO), d10 growing oocytes (GOs), d15 germinal vesicle (GV) oocytes, d25 GV oocytes, Setd1b WT GV oocytes and Setd1b cKO GV oocytes. Adjusted p-values are provided for statistical comparisons using one-way ANOVA and posthoc Tukey test for multiple comparison of the means.

Supplementary Figure 1. (A) Hierarchical clustering for H3K4me3 ChIP-seq replicates in *Setd1b* WT (N=3) and *Setd1b* cKO (N=3) GV oocytes, and 10% input replicates (N=4). Clustering was done using 5kb running windows and enrichment normalised RPKM. **(B)** The screenshot shows the normalised enrichment for H3K4me3 for 1kb running windows with a 500bp step between *Setd1b* cKO and WT GV oocyte replicates. Significant differentially enriched 5kb windows are shown in the *Setd1b* gain/loss track, with red and blue bars showing windows that gain and lose H3K4me3 in *Setd1b* cKO, respectively. **(C)** Top: pie charts show the proportion of 5kb windows that gain (left) or lose (middle) H3K4me3 in the *Setd1b* cKO GV oocytes that overlap transcription start sites (TSSs) or genic regions, compared to a random set (right) of 5kb windows ($p<0.0001$ and $p<0.0001$, respectively, Chi-square). Bottom: pie charts show the proportion of the inter-genic 5kb windows that gain (left) or lose (middle) H3K4me3 in the *Setd1b* cKO GV oocytes that overlap combinatorial peaks for histone modifications in GV oocytes (Hanna et al., 2018), compared to the inter-genic random 5kb windows ($p<0.0001$ and $p<0.0001$, respectively, Chi-square).

Supplementary Figure 2. (A) The heatmap shows DEGs in the *Setd1b* cKO and WT replicates, quantitated as per gene-normalised $\log_2$RPKM. **(B)** Principal component analysis shows scRNA-seq replicates for *Setd1b* cKO GV (N=3), *Setd1b* WT GV (N=3), *Setd1b* cKO MII (N=4), and *Setd1b* WT MII (N=4) oocytes. Expression of oocyte transcripts were quantitated using $\log_2$RPM. **(C)** The scatterplot shows the abundance of oocyte transcripts (N=33,437) in *Setd1b* cKO and WT MII oocytes as measured by scRNA-seq. Down- and upregulated differentially expressed genes (DEGs) identified in *Setd1b* cKO GV oocytes are shown in blue and red, respectively. **(D)** The barplot shows the average gene expression for H3K4 methyltransferases (*Setd1a*, *Mll1*, *Mll2*, *Mll3*, *Mll4*), H3K4 methyltransferase complex subunits (*Cxxc1*, *Rbbp5*, *Ash2l*, *Wdr5*, *Dpy30*), and *de novo* DNA methyltransferases (*Dnmt3a*, *Dnmt3b*, *Dnmt3l*) in *Setd1b* cKO and WT GV oocytes. The whiskers show standard deviation between biological replicates, while the asterisks shows those genes that are identified as significant DEGs.

Supplementary Figure 3. (A) The Venn diagrams show the overlap between genes upregulated in oogenesis and up- and downregulated DEGs in *Setd1b* cKO GV oocytes. **(B)** The Venn diagrams show the overlap between genes upregulated in oogenesis and up- and downregulated DEGs in *Cxxc1* cKO GV oocytes. **(C)** The Venn diagrams show the overlap between up- and downregulated DEGs in the *Cxxc1* cKO GV oocytes with up- and downregulated DEGs in *Setd1b* cKO GV oocytes (left panel) and up- and downregulated DEGs in *Mll2* cKO GV oocytes (right panel).

References

Allen, M.D., Grummitt, C.G., Hilcenko, C., Min, S.Y., Tonkin, L.M., Johnson, C.M., Freund, S.M., Bycroft, M., and Warren, A.J. (2006). Solution structure of the nonmethyl-CpG-binding CXXC domain of the leukaemia-associated MLL histone methyltransferase. Embo J. *25,* 4503-4512.

Andreu-Vieyra, C.V., Chen, R., Agno, J.E., Glaser, S., Anastassiadis, K., Stewart, A.F., and Matzuk, M.M. (2010). MLL2 is required in oocytes for bulk histone 3 lysine 4 trimethylation and transcriptional silencing. PLoS Biol. *8,* 10.1371/journal.pbio.1000453.

Ashokkumar, D., Zhang, Q., Much, C., Bledau, A.S., Naumann, R., Alexopoulou, D., Dahl, A., Goveas, N., Fu, J., Anastassiadis, K., Stewart, A.F., and Kranz, A. (2020). MLL4 is required after implantation, whereas MLL3 becomes essential during late gestation. Development *147,* dev186999. doi: 10.1242/dev.186999.

Bernstein, B.E., Kamal, M., Lindblad-Toh, K., Bekiranov, S., Bailey, D.K., Huebert, D.J., McMahon, S., Karlsson, E.K., Kulbokas, E.J.,3rd, Gingeras, T.R., Schreiber, S.L., and Lander, E.S. (2005). Genomic maps and comparative analysis of histone modifications in human and mouse. Cell *120,* 169-181.

Bledau, A.S., Schmidt, K., Neumann, K., Hill, U., Ciotta, G., Gupta, A., Torres, D.C., Fu, J., Kranz, A., Stewart, A.F., and Anastassiadis, K. (2014). The H3K4 methyltransferase Setd1a is first required at the epiblast stage, whereas Setd1b becomes essential after gastrulation. Development *141,* 1022-1035.

Brici, D., Zhang, Q., Reinhardt, S., Dahl, A., Hartmann, H., Schmidt, K., Goveas, N., Huang, J., Gahurova, L., Kelsey, G.*, et al.* (2017). Setd1b, encoding a histone 3 lysine 4 methyltransferase, is a maternal effect gene required for the oogenic gene expression program. Development *144,* 2606-2617.

Briggs, S.D., Bryk, M., Strahl, B.D., Cheung, W.L., Davie, J.K., Dent, S.Y., Winston, F., and Allis, C.D. (2001). Histone H3 lysine 4 methylation is mediated by Set1 and required for cell growth and rDNA silencing in Saccharomyces cerevisiae. Genes Dev. *15,* 3286-3295.

Brown, D.A., Di Cerbo, V., Feldmann, A., Ahn, J., Ito, S., Blackledge, N.P., Nakayama, M., McClellan, M., Dimitrova, E., Turberfield, A.H.*, et al.* (2017). The SET1 Complex Selects Actively Transcribed Target Genes via Multivalent Interaction with CpG Island Chromatin. Cell. Rep. *20,* 2313-2327.

Chang, P.Y., Hom, R.A., Musselman, C.A., Zhu, L., Kuo, A., Gozani, O., Kutateladze, T.G., and Cleary, M.L. (2010). Binding of the MLL PHD3 finger to histone H3K4me3 is required for MLL-dependent gene transcription. J. Mol. Biol. *400,* 137-144.

Choudhury, R., Singh, S., Arumugam, S., Roguev, A., and Stewart, A.F. (2019). The Set1 complex is dimeric and acts with Jhd2 demethylation to convey symmetrical H3K4 trimethylation. Genes Dev. *33,* 550-564.

Clouaire, T., Webb, S., Skene, P., Illingworth, R., Kerr, A., Andrews, R., Lee, J.H., Skalnik, D., and Bird, A. (2012). Cfp1 integrates both CpG content and gene activity for accurate H3K4me3 deposition in embryonic stem cells. Genes Dev. *26,* 1714-1728.

Dahl, J.A., Jung, I., Aanes, H., Greggains, G.D., Manaf, A., Lerdrup, M., Li, G., Kuan, S., Li, B., Lee, A.Y.*, et al.* (2016). Broad histone H3K4me3 domains in mouse oocytes modulate maternal-to-zygotic transition. Nature *537,* 548-552.

Dehé, P.M., Dichtl, B., Schaft, D., Roguev, A., Pamblanco, M., Lebrun, R., Rodríguez-Gil, A., Mkandawire, M., Landsberg, K., Shevchenko, A.*, et al.* (2006). Protein interactions within the Set1 complex and their roles in the regulation of histone 3 lysine 4 methylation. J. Biol. Chem. *281,* 35404-35412.

Denissov, S., Hofemeister, H., Marks, H., Kranz, A., Ciotta, G., Singh, S., Anastassiadis, K., Stunnenberg, H.G., and Stewart, A.F. (2014). Mll2 is required for H3K4 trimethylation on bivalent promoters in embryonic stem cells, whereas Mll1 is redundant. Development *141,* 526-537.

Ernst, P., and Vakoc, C.R. (2012). WRAD: enabler of the SET1-family of H3K4 methyltransferases. Brief Funct. Genomics *11,* 217-226.

Gahurova, L., Tomizawa, S.I., Smallwood, S.A., Stewart-Morgan, K.R., Saadeh, H., Kim, J., Andrews, S.R., Chen, T., and Kelsey, G. (2017). Transcription and chromatin determinants of de novo DNA methylation timing in oocytes. Epigenetics Chromatin *10,* 25-017-0133-5. eCollection 2017.

Glaser, S., Lubitz, S., Loveland, K.L., Ohbo, K., Robb, L., Schwenk, F., Seibler, J., Roellig, D., Kranz, A., Anastassiadis, K., and Stewart, A.F. (2009). The histone 3 lysine 4 methyltransferase, Mll2, is only required briefly in development and spermatogenesis. Epigenetics Chromatin *2,* 5-8935-2-5.

Glaser, S., Schaft, J., Lubitz, S., Vintersten, K., van der Hoeven, F., Tufteland, K.R., Aasland, R., Anastassiadis, K., Ang, S.L., and Stewart, A.F. (2006). Multiple epigenetic maintenance factors implicated by the loss of Mll2 in mouse development. Development *133,* 1423-1432.

Hanna, C.W., Taudt, A., Huang, J., Gahurova, L., Kranz, A., Andrews, S., Dean, W., Stewart, A.F., Colome-Tatche, M., and Kelsey, G. (2018). MLL2 conveys transcription-independent H3K4 trimethylation in oocytes. Nat. Struct. Mol. Biol. *25,* 73-82.

Hödl, M., and Basler, K. (2012). Transcription in the absence of histone H3.2 and H3K4 methylation. Curr. Biol. *22,* 2253-2257.

Howe, F.S., Fischl, H., Murray, S.C., and Mellor, J. (2017). Is H3K4me3 instructive for transcription activation? Bioessays *39,* 1-12.

Hu, D., Garruss, A.S., Gao, X., Morgan, M.A., Cook, M., Smith, E.R., and Shilatifard, A. (2013). The Mll2 branch of the COMPASS family regulates bivalent promoters in mouse embryonic stem cells. Nat. Struct. Mol. Biol. *20,* 1093-1097.

Jude, C.D., Climer, L., Xu, D., Artinger, E., Fisher, J.K., and Ernst, P. (2007). Unique and independent roles for MLL in adult hematopoietic stem cells and progenitors. Cell. Stem Cell. *1,* 324-337.

Klymenko, T., and Müller, J. (2004). The histone methyltransferases Trithorax and Ash1 prevent transcriptional silencing by Polycomb group proteins. EMBO Rep. *5,* 373-377.

Kobayashi, H., Sakurai, T., Imai, M., Takahashi, N., Fukuda, A., Yayoi, O., Sato, S., Nakabayashi, K., Hata, K., Sotomaru, Y., Suzuki, Y., and Kono, T. (2012). Contribution of intragenic DNA methylation in

mouse gametic DNA methylomes to establish oocyte-specific heritable marks. PLoS Genet. *8,* e1002440.

Krogan, N.J., Dover, J., Khorrami, S., Greenblatt, J.F., Schneider, J., Johnston, M., and Shilatifard, A. (2002). COMPASS, a histone H3 (Lysine 4) methyltransferase required for telomeric silencing of gene expression. J. Biol. Chem. *277,* 10753-10755.

Lan, Z.J., Xu, X., and Cooney, A.J. (2004). Differential oocyte-specific expression of Cre recombinase activity in GDF-9-iCre, Zp3cre, and Msx2Cre transgenic mice. Biol. Reprod. *71,* 1469-1474.

Lee, J.H., Hart, S.R., and Skalnik, D.G. (2004). Histone deacetylase activity is required for embryonic stem cell differentiation. Genesis *38,* 32-38.

Lee, J.H., Tate, C.M., You, J.S., and Skalnik, D.G. (2007). Identification and characterization of the human Set1B histone H3-Lys4 methyltransferase complex. J. Biol. Chem. *282,* 13419-13428.

Mahadevan, J., and Skalnik, D.G. (2016). Efficient differentiation of murine embryonic stem cells requires the binding of CXXC finger protein 1 to DNA or methylated histone H3-Lys4. Gene *594,* 1-9.

Margaritis, T., Oreal, V., Brabers, N., Maestroni, L., Vitaliano-Prunier, A., Benschop, J.J., van Hooff, S., van Leenen, D., Dargemont, C., Geli, V., and Holstege, F.C. (2012). Two distinct repressive mechanisms for histone 3 lysine 4 methylation through promoting 3'-end antisense transcription. PLoS Genet. *8,* e1002952.

Miller, T., Krogan, N.J., Dover, J., Erdjument-Bromage, H., Tempst, P., Johnston, M., Greenblatt, J.F., and Shilatifard, A. (2001). COMPASS: a complex of proteins associated with a trithorax-related SET domain protein. Proc. Natl. Acad. Sci. U. S. A. *98,* 12902-12907.

Murphy, P.J., Wu, S.F., James, C.R., Wike, C.L., and Cairns, B.R. (2018). Placeholder Nucleosomes Underlie Germline-to-Embryo DNA Methylation Reprogramming. Cell *172,* 993-1006.e13.

Ooi, S.K., Qiu, C., Bernstein, E., Li, K., Jia, D., Yang, Z., Erdjument-Bromage, H., Tempst, P., Lin, S.P., Allis, C.D., Cheng, X., and Bestor, T.H. (2007). DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. Nature *448,* 714-717.

Roguev, A., Schaft, D., Shevchenko, A., Aasland, R., Shevchenko, A., and Stewart, A.F. (2003). High conservation of the Set1/Rad6 axis of histone 3 lysine 4 methylation in budding and fission yeasts. J. Biol. Chem. *278,* 8487-8493.

Roguev, A., Schaft, D., Shevchenko, A., Pijnappel, W.W., Wilm, M., Aasland, R., and Stewart, A.F. (2001). The Saccharomyces cerevisiae Set1 complex includes an Ash2 homologue and methylates histone 3 lysine 4. Embo J. *20,* 7137-7148.

Ruthenburg, A.J., Allis, C.D., and Wysocka, J. (2007). Methylation of lysine 4 on histone H3: intricacy of writing and reading a single epigenetic mark. Mol. Cell *25,* 15-30.

Santos-Rosa, H., Schneider, R., Bannister, A.J., Sherriff, J., Bernstein, B.E., Emre, N.C., Schreiber, S.L., Mellor, J., and Kouzarides, T. (2002). Active genes are tri-methylated at K4 of histone H3. Nature *419,* 407-411.

Sha, Q.Q., Dai, X.X., Jiang, J.C., Yu, C., Jiang, Y., Liu, J., Ou, X.H., Zhang, S.Y., and Fan, H.Y. (2018). CFP1 coordinates histone H3 lysine-4 trimethylation and meiotic cell cycle progression in mouse oocytes. Nat. Commun. *9,* 3477-018-05930-x.

Sha, Q.Q., Zhu, Y.Z., Xiang, Y., Yu, J.L., Fan, X.Y., Li, Y.C., Wu, Y.W., Shen, L., and Fan, H.Y. (2021). Role of CxxC-finger protein 1 in establishing mouse oocyte epigenetic landscapes. Nucleic Acids Res.

Shi, X., Kachirskaia, I., Walter, K.L., Kuo, J.H., Lake, A., Davrazou, F., Chan, S.M., Martin, D.G., Fingerman, I.M., Briggs, S.D*., et al.* (2007). Proteome-wide analysis in Saccharomyces cerevisiae identifies several PHD fingers as novel direct and selective binding modules of histone H3 methylated at either lysine 4 or lysine 36. J. Biol. Chem. *282,* 2450-2455.

Skalnik, D.G. (2010). The epigenetic regulator Cfp1. Biomol. Concepts *1,* 325-334.

Smallwood, S.A., and Kelsey, G. (2012). Genome-wide analysis of DNA methylation in low cell numbers by reduced representation bisulfite sequencing. Methods Mol. Biol. *925,* 187-197.

Soares, L.M., He, P.C., Chun, Y., Suh, H., Kim, T., and Buratowski, S. (2017). Determinants of Histone H3K4 Methylation Patterns. Mol. Cell *68,* 773-785.e6.

Vermeulen, M., Eberl, H.C., Matarese, F., Marks, H., Denissov, S., Butter, F., Lee, K.K., Olsen, J.V., Hyman, A.A., Stunnenberg, H.G., and Mann, M. (2010). Quantitative interaction proteomics and genome-wide profiling of epigenetic histone marks and their readers. Cell *142,* 967-980.

Vermeulen, M., Mulder, K.W., Denissov, S., Pijnappel, W.W., van Schaik, F.M., Varier, R.A., Baltissen, M.P., Stunnenberg, H.G., Mann, M., and Timmers, H.T. (2007). Selective anchoring of TFIID to nucleosomes by trimethylation of histone H3 lysine 4. Cell *131,* 58-69.

Veselovska, L., Smallwood, S.A., Saadeh, H., Stewart, K.R., Krueger, F., Maupetit-Mehouas, S., Arnaud, P., Tomizawa, S., Andrews, S., and 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,* 209-015-0769-z.

Voo, K.S., Carlone, D.L., Jacobsen, B.M., Flodin, A., and Skalnik, D.G. (2000). Cloning of a mammalian transcriptional activator that binds unmethylated CpG motifs and shares a CXXC domain with DNA methyltransferase, human trithorax, and methyl-CpG binding domain protein 1. Mol. Cell. Biol. *20,* 2108-2121.

Weiner, A., Chen, H.V., Liu, C.L., Rahat, A., Klien, A., Soares, L., Gudipati, M., Pfeffner, J., Regev, A., Buratowski, S*., et al.* (2012). Systematic dissection of roles for chromatin regulators in a yeast stress response. PLoS Biol. *10,* e1001369.

Worden, E.J., Zhang, X., and Wolberger, C. (2020). Structural basis for COMPASS recognition of an H2B-ubiquitinated nucleosome. Elife *9,* 10.7554/eLife.53199.

Xu, C., Liu, K., Lei, M., Yang, A., Li, Y., Hughes, T.R., and Min, J. (2018). DNA Sequence Recognition of Human CXXC Domains and Their Structural Determinants. Structure *26,* 85-95.e3.

Yu, C., Fan, X., Sha, Q.Q., Wang, H.H., Li, B.T., Dai, X.X., Shen, L., Liu, J., Wang, L., Liu, K., Tang, F., and Fan, H.Y. (2017). CFP1 Regulates Histone H3K4 Trimethylation and Developmental Potential in Mouse Oocytes. Cell. Rep. *20,* 1161-1172.

Zhang, B., Zheng, H., Huang, B., Li, W., Xiang, Y., Peng, X., Ming, J., Wu, X., Zhang, Y., Xu, Q.*, et al.* (2016). Allelic reprogramming of the histone modification H3K4me3 in early mammalian development. Nature *537,* 553-557.

Zhang, X., Bernatavichute, Y.V., Cokus, S., Pellegrini, M., and Jacobsen, S.E. (2009). Genome-wide analysis of mono-, di- and trimethylation of histone H3 lysine 4 in Arabidopsis thaliana. Genome Biol. *10,* R62-2009-10-6-r62. Epub 2009 Jun 9.

Figure 1

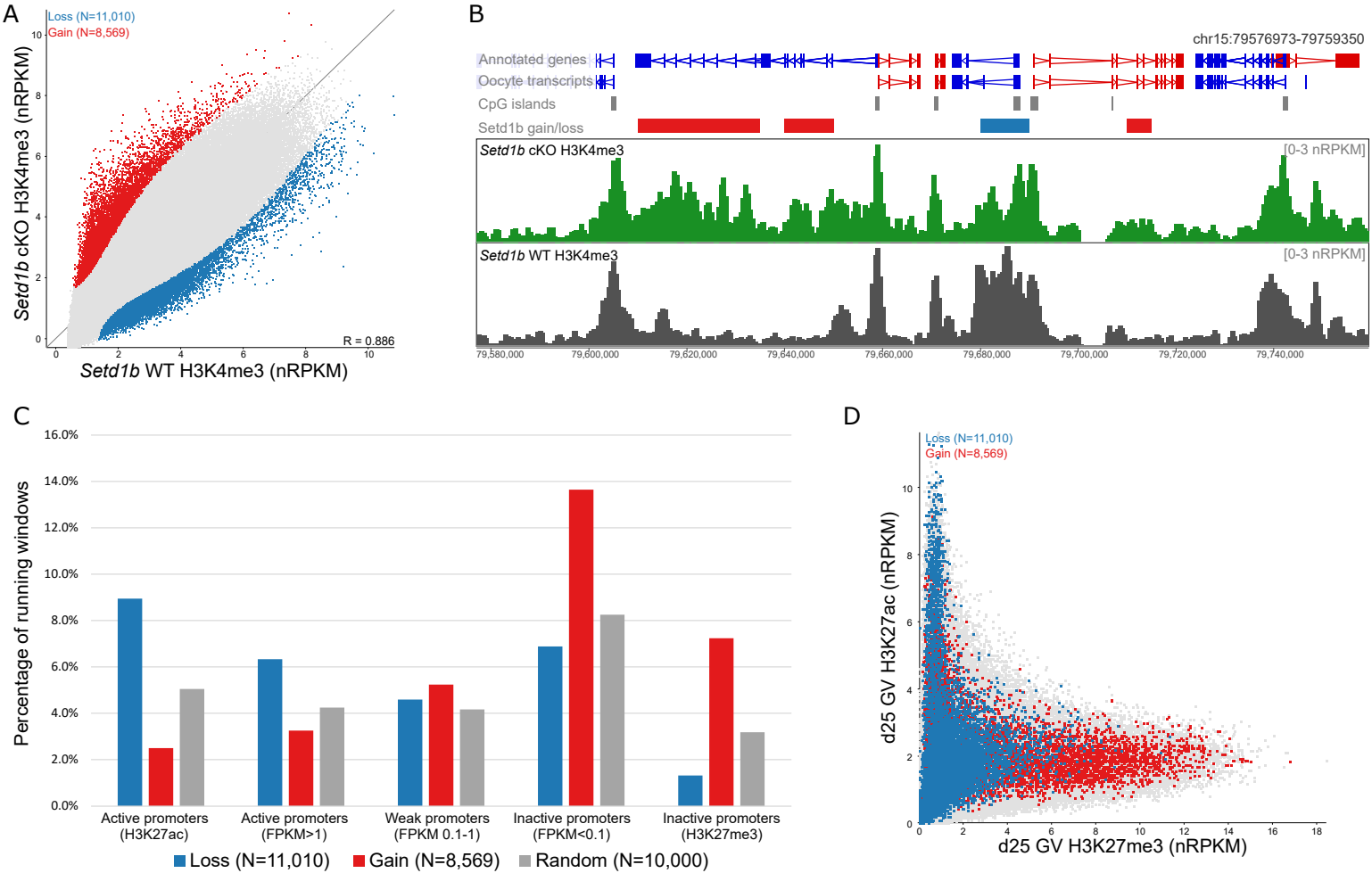

A
Loss (N=11,010)
Gain (N=8,569)
Setd1b cKO H3K4me3 (nRPKM)
Setd1b WT H3K4me3 (nRPKM)
R = 0.886
B
chr15:79576973-79759350
Annotated genes
Oocyte transcripts
CpG islands
Setd1b gain/loss
Setd1b cKO H3K4me3
[0-3 nRPKM]
Setd1b WT H3K4me3
[0-3 nRPKM]
79,580,000
79,600,000
79,620,000
79,640,000
79,660,000
79,680,000
79,700,000
79,720,000
79,740,000
C
Percentage of running windows
16.0%
14.0%
12.0%
10.0%
8.0%
6.0%
4.0%
2.0%
0.0%
Active promoters (H3K27ac)
Active promoters (FPKM>1)
Weak promoters (FPKM 0.1-1)
Inactive promoters (FPKM<0.1)
Inactive promoters (H3K27me3)
Loss (N=11,010)
Gain (N=8,569)
Random (N=10,000)
D
Loss (N=11,010)
Gain (N=8,569)
d25 GV H3K27ac (nRPKM)
d25 GV H3K27me3 (nRPKM)

Figure 2

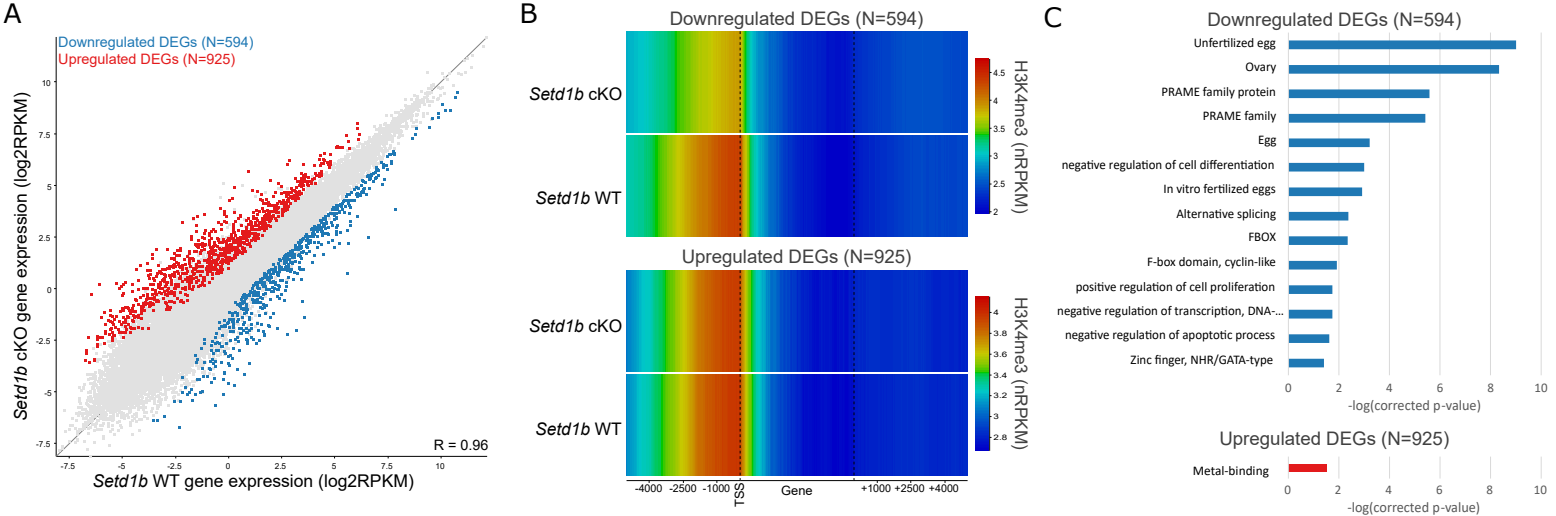
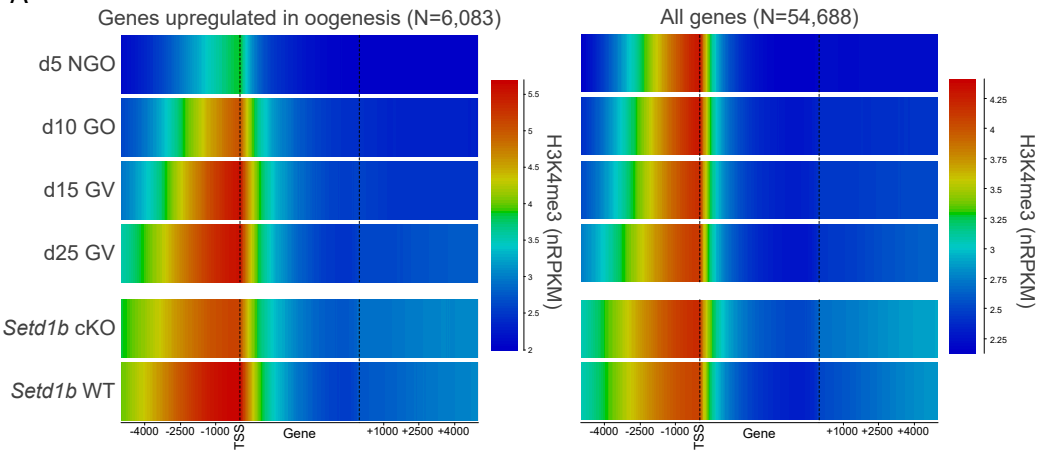
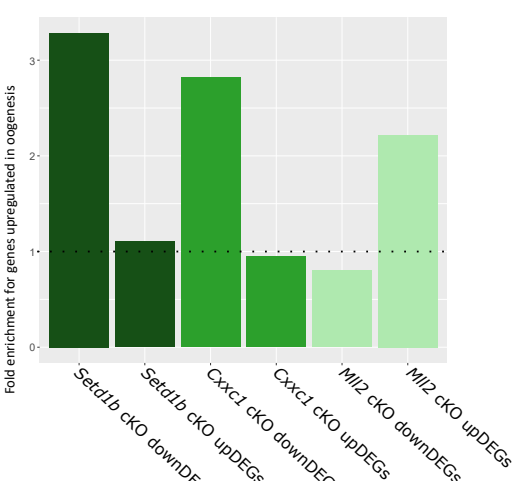

A
Downregulated DEGs (N=594)
Upregulated DEGs (N=925)
Setd1b cKO gene expression (log2RPKM)
Setd1b WT gene expression (log2RPKM)
R = 0.96
B
Downregulated DEGs (N=594)
Setd1b cKO
Setd1b WT
H3K4me3 (nRPKM)
Upregulated DEGs (N=925)
Setd1b cKO
Setd1b WT
H3K4me3 (nRPKM)
-4000 -2500 -1000 TSS Gene +1000 +2500 +4000
C
Downregulated DEGs (N=594)
Unfertilized egg
Ovary
PRAME family protein
PRAME family
Egg
negative regulation of cell differentiation
In vitro fertilized eggs
Alternative splicing
FBOX
F-box domain, cyclin-like
positive regulation of cell proliferation
negative regulation of transcription, DNA-...
negative regulation of apoptotic process
Zinc finger, NHR/GATA-type
-log(corrected p-value)
Upregulated DEGs (N=925)
Metal-binding
-log(corrected p-value)

Figure 3

A


Genes upregulated in oogenesis (N=6,083)
All genes (N=54,688)
d5 NGO
d10 GO
d15 GV
d25 GV
Setd1b cKO
Setd1b WT
H3K4me3 (nRPKM)
-4000 -2500 -1000 TSS Gene +1000 +2500 +4000

B


Fold enrichment for genes upregulated in oogenesis
Setd1b cKO downDEGs
Setd1b cKO upDEGs
Cxxc1 cKO downDEGs
Cxxc1 cKO upDEGs
Mll2 cKO downDEGs
Mll2 cKO upDEGs

C

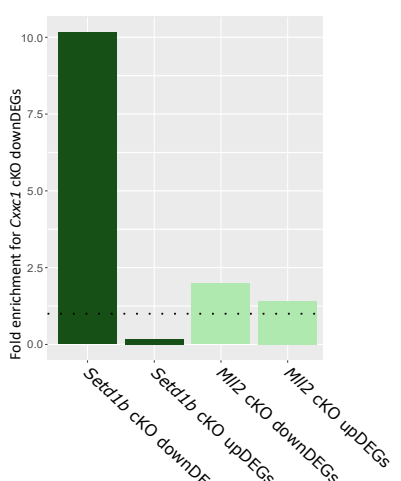

Fold enrichment for Cxxc1 cKO downDEGs
Setd1b cKO downDEGs
Setd1b cKO upDEGs
Mll2 cKO downDEGs
Mll2 cKO upDEGs

Figure 4

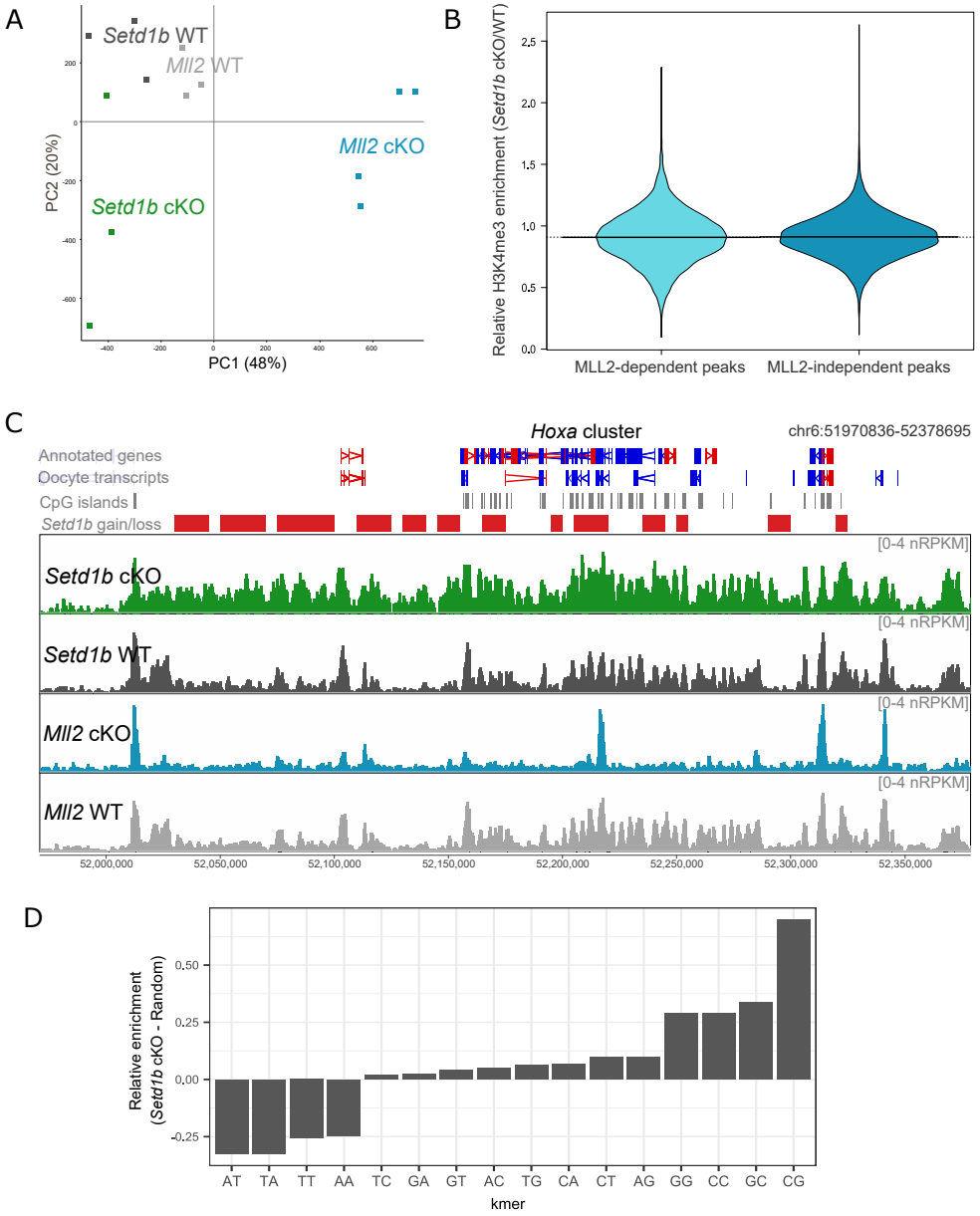

A
Setd1b WT
Mll2 WT
Mll2 cKO
Setd1b cKO
PC2 (20%)
PC1 (48%)
B
Relative H3K4me3 enrichment (Setd1b cKO/WT)
MLL2-dependent peaks
MLL2-independent peaks
C
Hoxa cluster
chr6:51970836-52378695
Annotated genes
Oocyte transcripts
CpG islands
Setd1b gain/loss
Setd1b cKO
Setd1b WT
Mll2 cKO
Mll2 WT
[0-4 nRPKM]
52,000,000
52,050,000
52,100,000
52,150,000
52,200,000
52,250,000
52,300,000
52,350,000
D
Relative enrichment (Setd1b cKO - Random)
AT TA TT AA TC GA GT AC TG CA CT AG GG CC GC CG
kmer

Figure 5

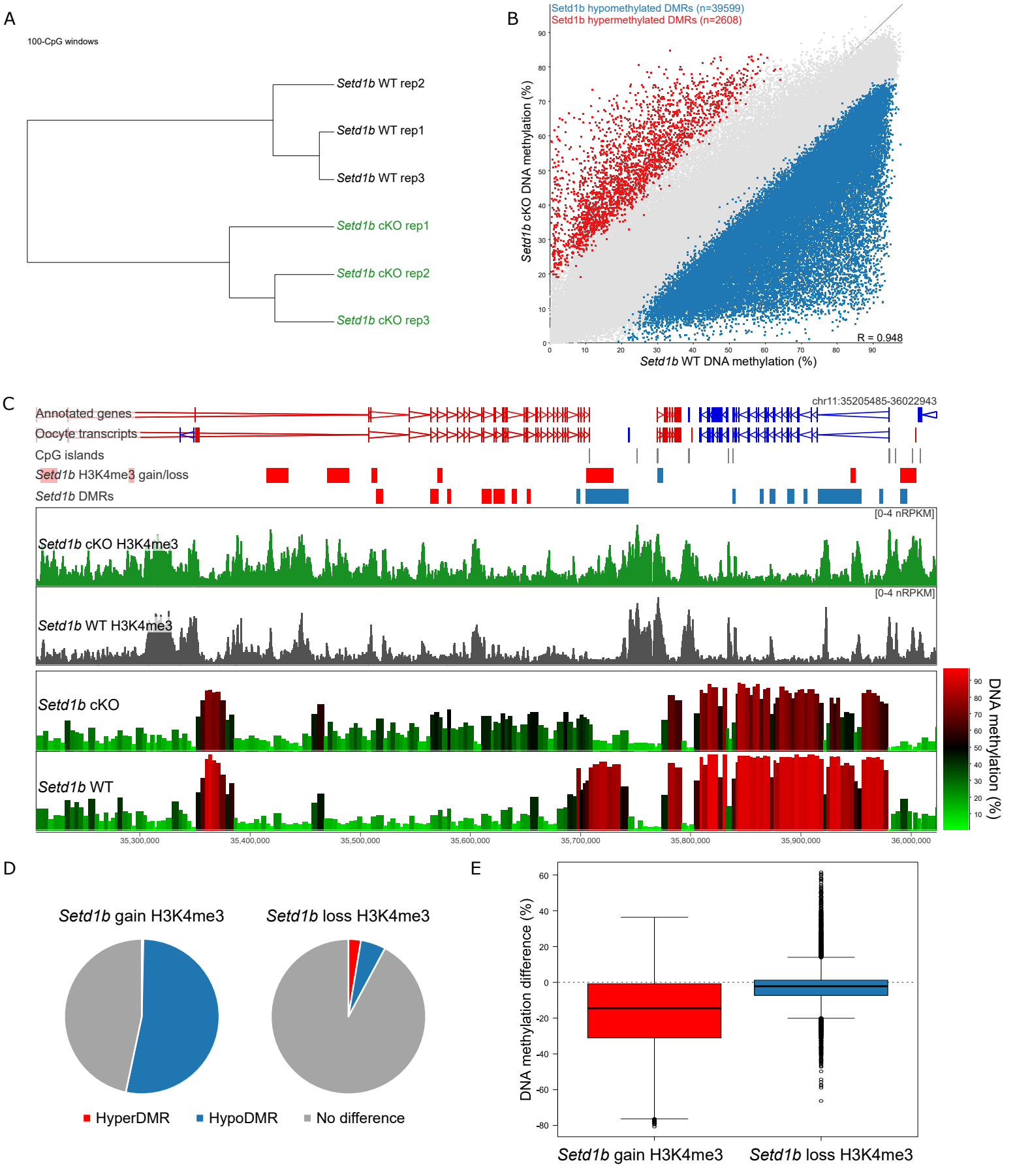
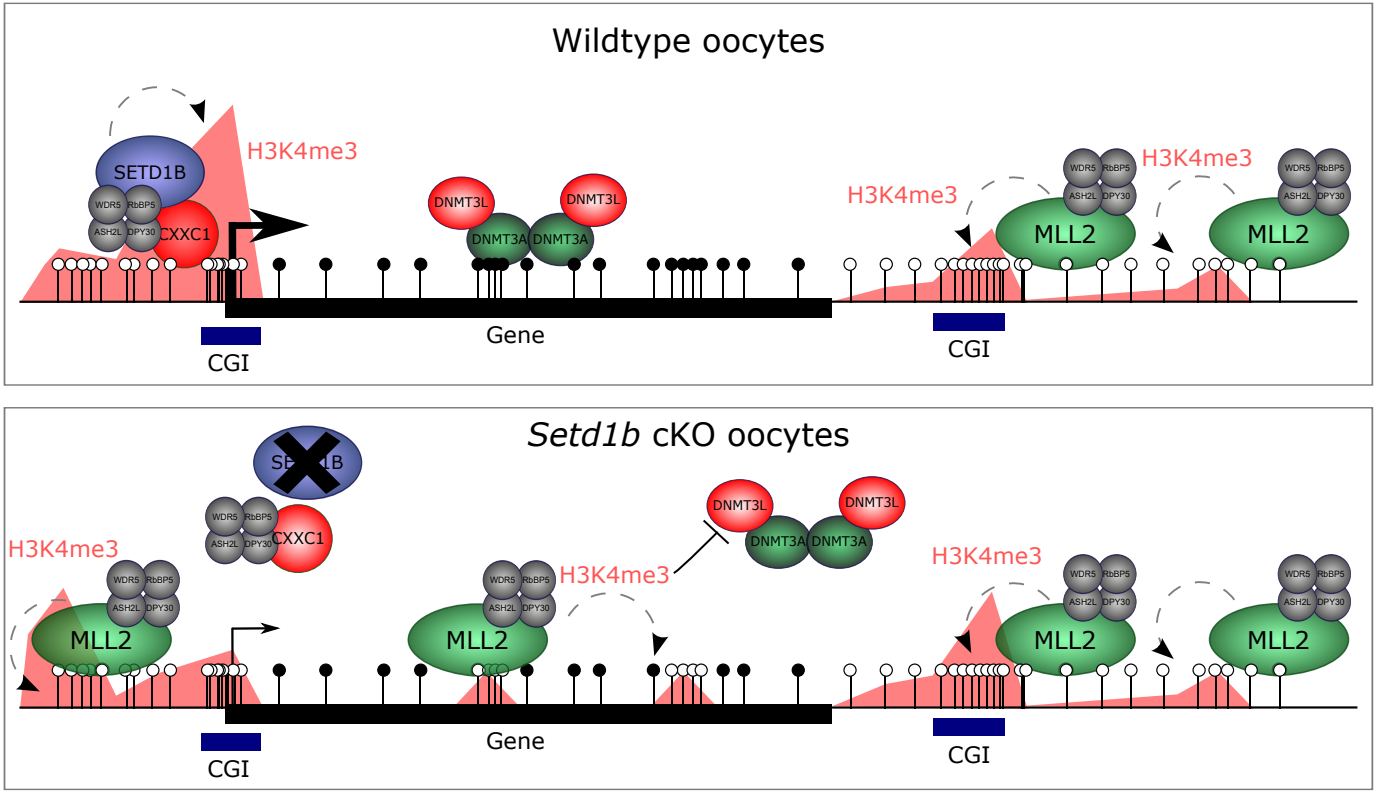
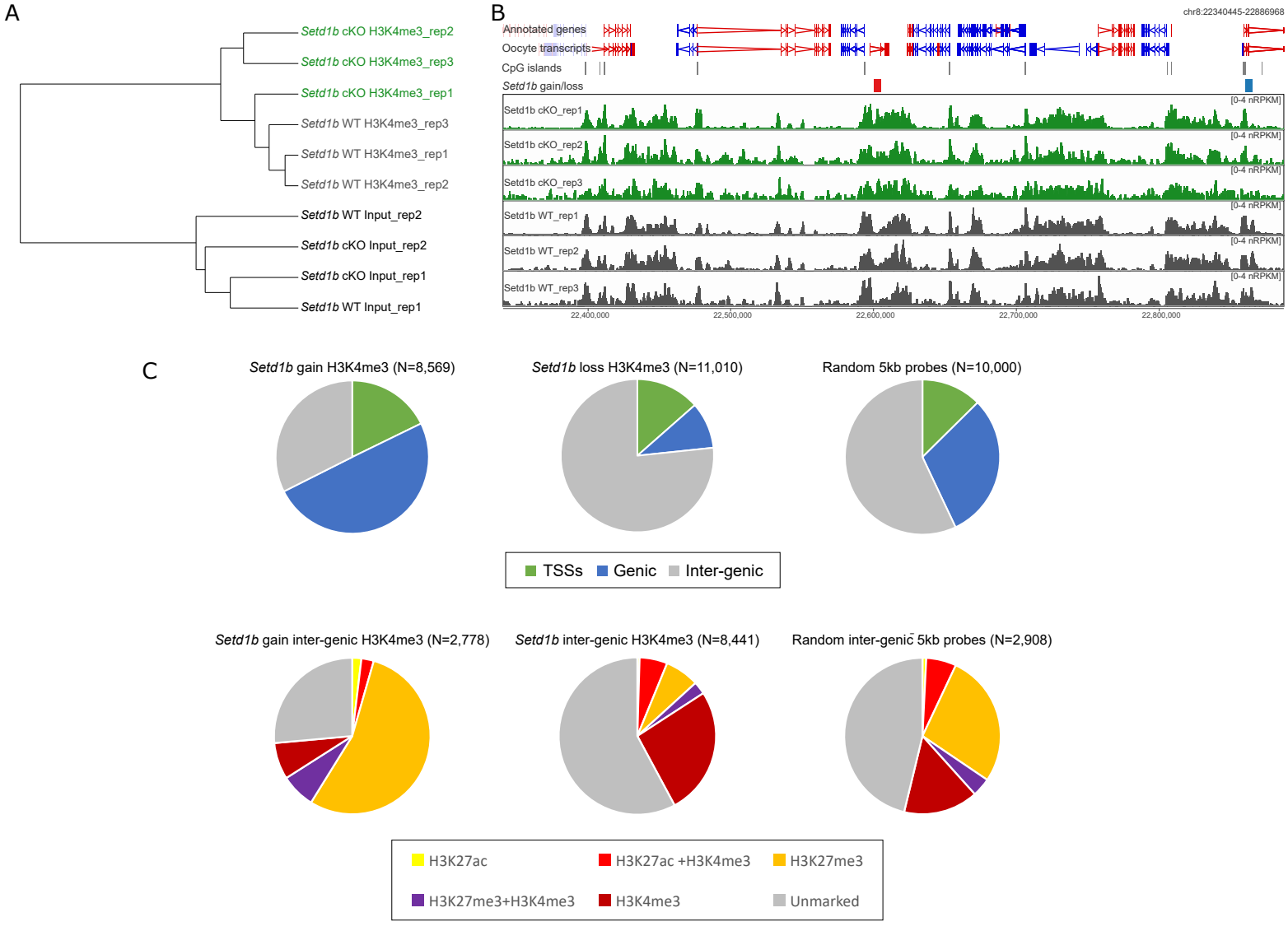

A
100-CpG windows
Setd1b WT rep2
Setd1b WT rep1
Setd1b WT rep3
Setd1b cKO rep1
Setd1b cKO rep2
Setd1b cKO rep3
B
Setd1b hypomethylated DMRs (n=39599)
Setd1b hypermethylated DMRs (n=2608)
Setd1b cKO DNA methylation (%)
Setd1b WT DNA methylation (%)
R = 0.948
C
chr11:35205485-36022943
Annotated genes
Oocyte transcripts
CpG islands
Setd1b H3K4me3 gain/loss
Setd1b DMRs
Setd1b cKO H3K4me3
Setd1b WT H3K4me3
[0-4 nRPKM]
Setd1b cKO
Setd1b WT
DNA methylation (%)
35,300,000
35,400,000
35,500,000
35,600,000
35,700,000
35,800,000
35,900,000
36,000,000
D
Setd1b gain H3K4me3
Setd1b loss H3K4me3
HyperDMR
HypoDMR
No difference
E
DNA methylation difference (%)
Setd1b gain H3K4me3
Setd1b loss H3K4me3

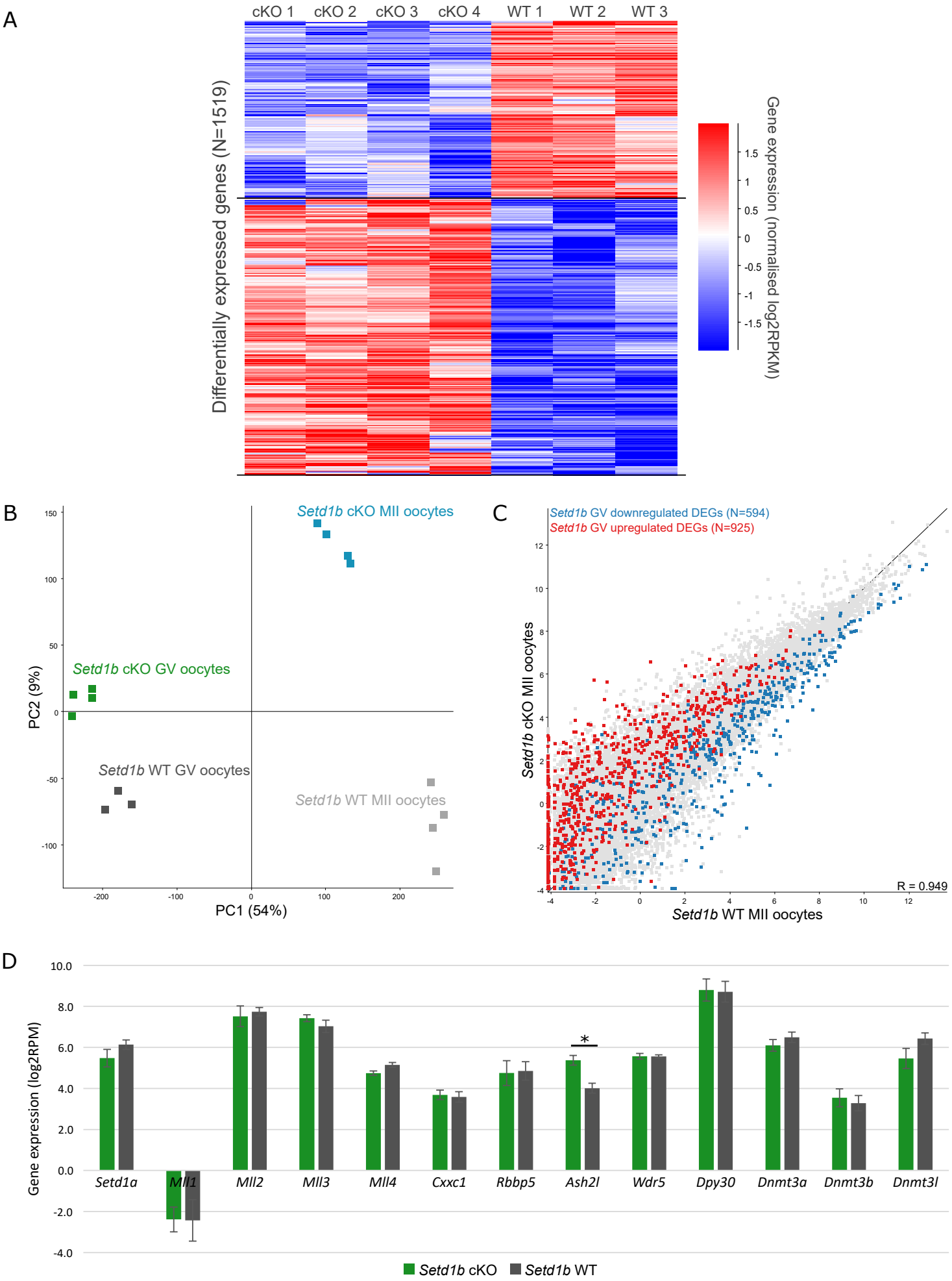
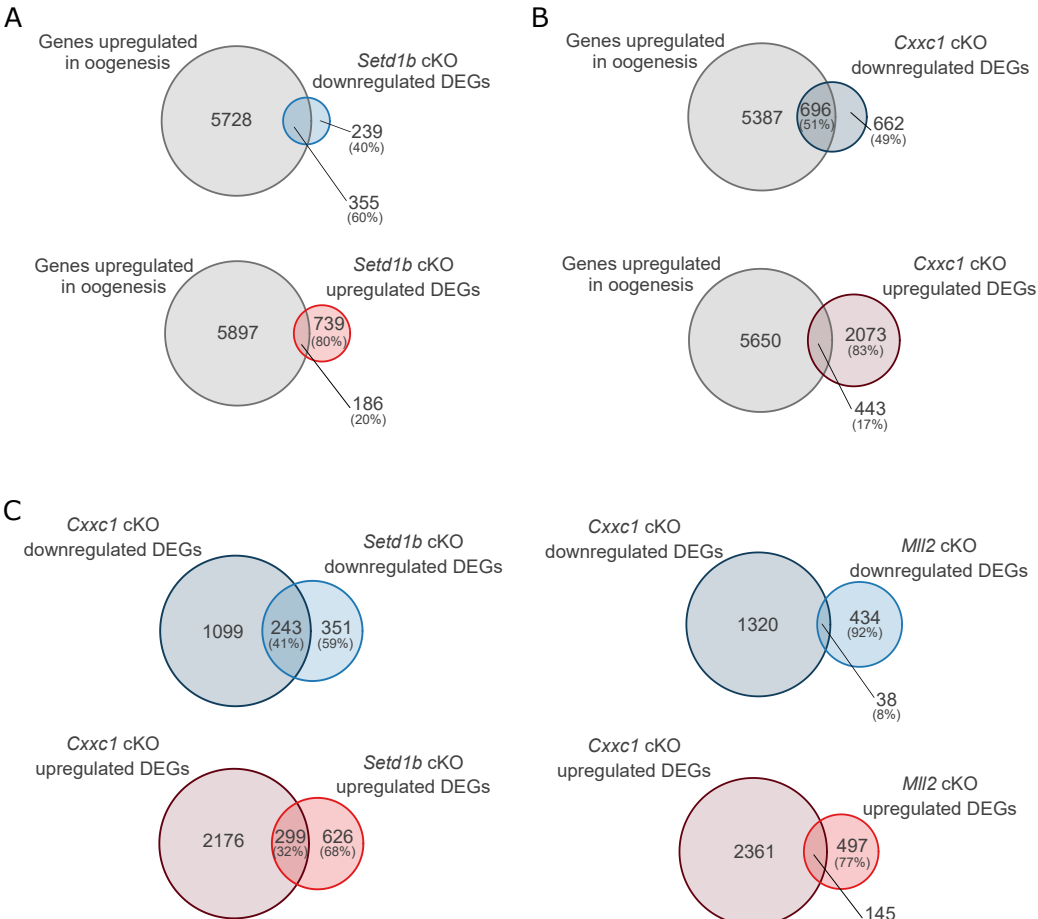
Figure 6


Wildtype oocytes
SETD1B
WDR5
RbBP5
ASH2L
DPY30
CXXC1
H3K4me3
DNMT3L
DNMT3L
DNMT3A
DNMT3A
H3K4me3
H3K4me3
MLL2
MLL2
CGI
Gene
CGI
Setd1b cKO oocytes
SETD1B
CXXC1
H3K4me3
MLL2
DNMT3L
DNMT3L
DNMT3A
DNMT3A
H3K4me3
H3K4me3
MLL2
MLL2
MLL2
CGI
Gene
CGI

Supplementary Figure 1



Supplementary Figure 2



Supplementary Figure 3