Published in final edited form as:
Nat Cell Biol. 2020 April 01; 22(4): 380–388. doi:10.1038/s41556-020-0494-z.

KDM4A regulates the maternal-to-zygotic transition by protecting broad H3K4me3 domains from H3K9me3 invasion in oocytes

Aditya Sankar[1,2,3,*,†], **Mads Lerdrup**[2,†], **Adeel Manaf**[4,†], **Jens Vilstrup Johansen**[2], **Javier Martin Gonzalez**[6], **Rehannah Borup**[1], **Robert Blanshard**[1], **Arne Klungland**[4,5], **Klaus Hansen**[2], **Claus Yding Andersen**[7], **John Arne Dahl**[4,*], **Kristian Helin**[2,3,8,*], **Eva R. Hoffmann**[1,*]

[1]DNRF Centre for Chromosome Stability (CCS), Department of Cellular and Molecular Medicine, Faculty of Health and Medical Sciences, University of Copenhagen, Denmark

[2]Biotech Research Innovation Centre (BRIC), Faculty of Health and Medical Sciences, University of Copenhagen, Denmark

[3]The Novo Nordisk Foundation Center for Stem Cell Biology (DanStem), University of Copenhagen, Denmark

[4]Department of Microbiology, Oslo University Hospital, Rikshospitalet, Norway

*corresponding authors.
†co-first authors

Ethics statement
The human oocyte work was conducted under ethics license no. H-16044731 (The Human MeioMap Project) granted to E.R.H. and ethics license no. H-2-2011-044 (extension license amm. no. 51307) granted to C.Y.A. using appropriate informed consent and GDPR (SUND-2016-60). The study is compliant with all of the relevant ethical regulations regarding research involving human participants. All mouse work was approved by the Danish Animal Ethical Committee ('Dyreforsøgstilsynet') and is compliant with all relevant ethical regulations regarding animal research.

Extended data is available for this paper at https://doi.org/10.1038/s41556-020-0494-z.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author contributions
A.S. conceived the project. A.S., M.L., K.Helin and E.R.H. designed the experiments and supervised the project. A.S., A.M., J.M.G. and R.Blanshard carried out the experiments. A.S., M.L., J.V.J. and R.Borup analysed and interpreted the data, and performed the bioinformatics and statistical analyses. C.Y.A. provided the human oocyte samples. J.A.D. designed all μChIP-Seq experiments and supervised A.M. who performed them. C.Y.A., K.Hansen, A.K. and J.A.D. discussed the data, and provided critical input and supervision. A.S., M.L., J.A.D., K.Helin and E.R.H. wrote the manuscript. All authors read and approved the final manuscript.

Data and code availability
Scripts used for processing and analysis of genome-wide data are available under 10.5281/zenodo.3701229. All sequencing data related to this study have been deposited at the Gene Expression Omnibus (GEO) under the accession number GSE129735. RNA-Seq data for human single oocytes have been submitted to the European Genome Phenome Archive (EGA) under the reference EGAS00001004220. Published H3K4me3 and H3K9me3 μChIP-Seq data re-analysed in this study are available at GEO using the accession numbers GSE72784 and GSM2588560, respectively. A dedicated EaSeq session related to this study is available at http://easeq.net/Sessions/Sankar.eas. Data relating to human single oocyte mRNA-Seq are available under the material transfer agreement and General Data Protection Regulation in accordance with Danish law from the corresponding author along with any other data on reasonable request.

Competing interests
The authors declare no competing interests.

Europe PMC Funders Author Manuscripts

[5]Department of Molecular Medicine, Institute of Basic Medical Sciences, University of Oslo, NO-0317, Oslo, Norway

[6]Transgenic Core Facility, Department of Experimental Medicine, Faculty of Health and Medical Sciences, University of Copenhagen, Denmark

[7]Laboratory of Reproductive Biology, Section 5712, University Hospital of Copenhagen, Denmark

[8]Cell Biology Program and Center for Epigenetics Research, Memorial Sloan Kettering Cancer Center, New York, USA

Europe PMC Funders Author Manuscripts

Abstract

The importance of germline-inherited posttranslational histone modifications on priming early mammalian development is just emerging[1–4]. Histone H3 lysine 9 (H3K9) trimethylation is associated with heterochromatin and gene repression during cell-fate change[5], while histone H3 lysine 4 (H3K4) trimethylation marks active gene promoters[6]. Mature oocytes are transcriptionally quiescent and possess remarkably broad domains of H3K4me3 (bdH3K4me3)[1,2]. It remains unknown as to which factors contribute to the maintenance of the bdH3K4me3 landscape. Lysine-specific demethylase 4A (KDM4A) demethylates H3K9me3 at promoters marked by H3K4me3 in actively transcribing somatic cells[7]. Here, we report that KDM4A-mediated H3K9me3 demethylation at bdH3K4me3 in oocytes is crucial for normal preimplantation development and zygotic genome activation (ZGA) after fertilization. Loss of KDM4A in oocytes causes aberrant H3K9me3 spreading over bdH3K4me3, resulting in insufficient transcriptional activation of genes, endogenous retroviral elements and long terminal repeat initiated chimeric transcripts during ZGA. The catalytic activity of KDM4A is essential for normal epigenetic reprogramming and preimplantation development. Hence, KDM4A plays a crucial role in preserving maternal epigenome integrity required for proper ZGA and transfer of developmental control to the embryo.

Oocytes transmit a host of RNA and protein stores as well as a unique epigenetic landscape distinct from that of the sperm that are critical for development of the early embryo[8]. Only recently, we have begun to decipher these distinct histone landscapes unique to the oocyte, and found that they are characterized by a particularly high abundance (approximately 20-fold higher compared to somatic cells) and broad distribution of H3K4me3 (bdH3K4me3) deposited by KMT2B (also known as MLL2)[1,3]. Critically, bdH3K4me3 mark developmentally-regulated genes that are activated soon after fertilization during the major zygotic genome activation (ZGA)[1]. However, while bdH3K4me3 are already established in mature oocytes that are transcriptionally quiescent, it is unclear how they are maintained, especially with regards to repressive histone marks such as H3K9me3.

Recently, it was shown that the lack of maternal SETD2 (a H3K36me3 methyltransferase) induced global H3K36me3 loss during oogenesis and consequently altered the deposition profile of bdH3K4me3 in mature oocytes post-transcriptional arrest[4]. The impact of SETD2 loss on gamete health is likely multifactorial as evidenced by the severe transcriptional changes upon global H3K36me3 loss, including a drastic reduction in the numbers of mature oocytes[4]. Yet, this study provides a prime example of crosstalk between two histone marks during oogenesis with an impact on maternal epigenome.

Here, we report that the loss of KDM4A alters the global H3K9me3 landscape in mature oocytes at sites occupied by bdH3K4me3, thereby highlighting a key role for KDM4A in preserving maternal epigenome integrity. In the soma, KDM4A and KDM4C exhibit functional redundancy in binding to H3K4me3-positive transcription start sites (TSS) through their double Tudor domains[9] (Extended data 1a), where they continuously catalyze the removal of H3K9me3 through their Jumonji (JmjC) domain. While mice lacking KDM4A or KDM4C individually are fully viable[10,11], their combined loss results in early embryonic lethality[7]. However, KDM4A knockout mice display female-specific infertility[11], due to a major defect in their reproductive tract that causes implantation failure. Additionally, the lack of KDM4A in the oocyte induces a maternal effect lethality in preimplantation embryos through unknown mechanisms[11]. Therefore, we investigated how loss of KDM4A disrupts normal preimplantation development.

We first profiled the expression of all KDM4 family transcripts in single MII oocytes using mRNASeq (Figure 1a). We observed a near-exclusive expression of *Kdm4a* in both mouse and human MII oocytes (Figure 1a and Extended data 1b), in contrast to the combined expression of at least two KDM4 family genes during earlier stages of oogenesis prior to transcriptional arrest (Extended data 1b and Supplementary Table 1). Ablation of KDM4A during oogenesis had no significant impact on ovulation since *Kdm4a*$^{-/-}$ and wild-type females ovulated similar numbers of fertilizable oocytes[11] (Figure 1b). Mature metaphase-II (MII) oocytes from *Kdm4a*$^{-/-}$ females were euploid with no evidence of major chromosomal breakages or aneuploidies (Figure 1c and Extended data 1c).

Previously, we demonstrated that *Kdm4a*$^{-/-}$ male mice are fertile and that *Kdm4a*$^{-/-}$ sperm can establish healthy preimplantation embryos *in vitro*[11]. Therefore, we fertilized *Kdm4a*$^{+/+}$ and *Kdm4a*$^{-/-}$ oocytes *in vitro* with *Kdm4a*$^{-/-}$ sperm to generate *Kdm4a*$^{+/-}$ (control) and *Kdm4a*$^{-/-(MZ)}$ maternal-zygotic mutant embryos (henceforth “MZ mutants”) to track preimplantation development in culture. A majority of MZ mutant embryos suffered developmental delays or arrest soon after the 2-cell stage prior to compaction of the morula (Figure 1d). The MZ mutant embryos also displayed hallmarks of genome instability such as the appearance of DNA bridges and increased micronuclei formation (Figure 1e, f and Extended data 1d). In summary, KDM4A is the major demethylase functional in MII oocytes and is required to maintain the genomic stability of pre-implantation embryos.

H3K9me3 is strongly associated with transcriptionally-silent heterochromatin and is a known barrier to cell fate change during somatic cell nuclear transfer (SCNT)[12] or induced pluripotency-mediated epigenetic reprogramming[5]. Removal of bulk H3K9me3 by ectopic expression of KDM4 demethylases markedly increases the efficiency of such reprogramming[12–15]. As epigenetic reprogramming is also a hallmark of early preimplantation development[8], we investigated the impact of KDM4A loss on this process with a focus on maternal H3K9me3. In zygotes generated from *Kdm4a*$^{-/-}$ oocytes, we observed an increased H3K9me3 abundance in the maternal pronuclei with little effect on pronuclei volume (Figure 2a and Extended data 2a), indicating that KDM4A uses its demethylase activity to ensure correct levels and distribution of H3K9me3 in the oocyte.

As KDM4A-catalyzed H3K9me3 demethylation happens in the context of established H3K4me3-marked chromatin in somatic cells[7,16], we wanted to assess any changes in maternal H3K9me3 in connection to that of H3K4me3 in *Kdm4a*$^{-/-}$ MII oocytes. To achieve this, we generated high quality genome-wide H3K9me3 profiles of MII oocytes from *Kdm4a*$^{+/+}$ and *Kdm4a*$^{-/-}$ females using a micro-scale chromatin immunoprecipitation and sequencing (μChIP-Seq) method[1]. We compared these data with published H3K4me3 data[1] that was used to generate reference region sets for visualization, including 12,133 bdH3K4me3 regions of at least 20kb size and all non-redundant UCSC genes from transcription start (TSS) to termination (TTS) sites (Extended data 3a and Supplementary Table 2). In wild-type oocytes, we observed a strong anti-correlation between H3K9me3 and bdH3K4me3 (Figure 2b, Extended data 3a). However, loss of KDM4A resulted in widespread *de novo* establishment of H3K9me3 at regions normally marked by bdH3K4me3 alone, including transcription start-sites (TSS) and associated upstream promoter regions (Figure 2b, c and Extended data 3a). Furthermore, the difference in H3K9me3 association with gene bodies versus TSS was less pronounced in the mutant, suggesting aberrant H3K9me3 association with gene bodies. This could either be due to more H3K9me3 in gene bodies, a redistribution or both (Extended data 4a-e, also see ChIP-Seq processing section in Methods). Since KDM4A also catalyses H3K36me3 removal[17], we assessed whether the aberrant deposition of H3K9me3 in bdH3K4me3 regions was accompanied by a similar increase of H3K36me3 at those sites. Importantly, we observed largely identical H3K36me3 patterns in *Kdm4a*$^{+/+}$ and *Kdm4a*$^{-/-}$ MII oocytes (Supplementary Fig 3b, c), suggesting that H3K9me3 demethylation is the major function of KDM4A in the oocyte.

Since the gain of H3K9me3 occurred in regions reported to have abundant H3K4me3 in wild-type MII oocytes[1] (Extended data 3d), we investigated whether aberrant H3K9me3 negatively impacts bdH3K4me3 establishment or its maintenance in *Kdm4a*$^{-/-}$ MII oocytes. To this end, we generated genome-wide maps of H3K4me3 in *Kdm4a*$^{+/+}$ and *Kdm4a*$^{-/-}$ oocytes. Remarkably, we found identical patterns of H3K4me3 at TSS and bdH3K4me3 in *Kdm4a*$^{+/+}$ and *Kdm4a*$^{-/-}$ MII oocytes (Figure 2d, e). This resulted in extensive H3K4me3 and H3K9me3 co-occurrence at thousands of loci, a feature that is extremely rare in transcriptionally-active somatic cells[18,19]. Aberrant establishment of H3K9me3 in *Kdm4a*$^{-/-}$ oocytes correlated well with wild-type H3K4me3, and could not be explained by the minor changes in H3K4me3 in *Kdm4a*$^{-/-}$ oocytes (Extended data 3d, e). Hence, we propose that KDM4A functions late in oogenesis to prevent co-occurrence of repressive H3K9me3 and active H3K4me3 marks at an unprecedented scale.

Regions marked by bdH3K4me3 in oocytes significantly overlap with ZGA genes primed for activation after fertilization[1]. Investigating a previously reported list of genes known to be differentially expressed in either the 2-cell embryo or the oocyte[20], we observed aberrant H3K9me3 spreading in *Kdm4a*$^{-/-}$ MII oocytes at a subset of highly expressed 2-cell embryo genes (Extended data 3f, g and Supplementary Table 3). In order to assess the impact of aberrant H3K9me3 on transcriptional activation in the embryo, we performed single embryo polyA mRNA-Seq of control and MZ mutant embryos (Figure 3a and Supplementary Table 4). We observed a clear and distinct separation of stage-specific transcriptomes (Extended data 5a). During the major ZGA, we observed 837 downregulated and 96 upregulated genes in 2-cell embryos from the MZ mutant (Figure 3b and Extended data 5b, c), in line with our

expectation that aberrant H3K9me3 gain in the oocyte would impede transcriptional activation.

By comparing differentially expressed genes to a predefined transcriptomic atlas of early mouse embryogenesis[21] (DBTMEE_v2), KDM4A loss resulted in downregulation of genes linked to major zygotic (ZGA) and mid-preimplantation (MGA) gene activation with higher abundance of maternally-inherited RNAs (Figure 3c). In addition, 4- and 8-cell MZ mutant embryos had increased levels of transcripts associated with minor ZGA, normally found in zygotes (Figure 3c). This suggests that onset of developmental failure in KDM4A MZ mutants originates from defective ZGA in the 2-cell embryo. Gene ontology analysis in 2-cell embryo transcriptomes revealed that a majority of repressed genes were involved in key developmental processes such as ribosome biogenesis (Extended data 5d, e), known markers of ZGA[20,22], transcription factors[23–26], mitotic control[27–29] as well as DNA replication and repair factors[30–32] (Extended data 5f). The latter categories provide a rationale for the DNA damage and arrest observed in MZ mutant embryos.

To understand if gene repression at ZGA is linked to increased H3K9me3, we generated genome-wide H3K9me3 μChIP-Seq profiles from 2-cell embryos. The aberrant H3K9me3 found in knockout oocytes was largely retained in the 2-cell embryos from the MZ mutant. This was particular the case at bdH3K4me3 regions and promoters (Extended data 6a, b). In addition to promoters, aberrant H3K9me3 extended into gene bodies at a majority of downregulated genes during ZGA (Figure 3d-f, Extended data 6c-f). These observations show that H3K9me3 demethylation by KDM4A is a necessary mechanism to ensure proper activation of genes during ZGA.

Oocytes and preimplantation embryos have a high abundance of transcripts, not only from genes but also from specific retrotransposon families suggested to play a role in transcriptional regulation and genome evolution[33–35]. ZGA in human and mouse embryos coincide with expression of several families of long terminal repeats (LTRs) from endogenous retroviruses (ERVs), of which a subset serves as alternative transcription start-sites that generate chimeric transcripts. These have been suggested to contribute towards totipotency in the 2-cell embryo in addition to achieving a proper maternal-zygotic transition[20,36]. Using deep stranded total RNA sequencing, we observed a significant repression of many high-copy number repeat types in 2-cell embryos from the MZ mutant (Figure 4a, Extended data 7a and Supplementary Table 5). The vast majority of these belonged to the LTR class that included MERVL retrotransposons, which have been suggested to regulate transcription of genes at the 2-cell stage[20,36] (Figure 4b). In general, many LTRs are detected in the transcriptomes of both oocytes and embryos[34] to varying degrees throughout preimplantation development (Figure 4c, Extended data 7b). We found several LTR subgroups that are dominantly expressed in the 2-cell embryo to be significantly affected in the MZ mutant, including MERVL and MT2 (Figure 4c, Extended data 7c). Through the CHAF1A complex, active LTRs are silenced during preimplantation development by H3K9me3 acquisition[37]. However, several of these prominent LTRs had already acquired aberrant H3K9me3 in *Kdm4a*$^{-/-}$ MII oocytes, suggesting that they are already primed for repression during ZGA (Figure 4d, Supplementary Table 6).

Recently, several thousand chimeric LTR initiating transcripts (LITs) were identified and annotated in fully grown mouse oocytes at the time of transcriptional arrest[35] (Supplementary Table 7). However, we found that a majority of these LITs were expressed in our low depth full-length cDNA libraries both in MII oocytes as well as 2-cell embryos. Interestingly, over half of these LITs had higher levels in the embryo compared to the MII oocyte (Extended data 8a, Supplementary Table 8), suggesting that such transcripts could be activated during ZGA. Relative to conventional genes, we found a reduced read coverage at the beginning of these LITs coinciding with the initiating LTR sites - in oocytes and embryos lacking KDM4A (Figure 4e, Extended data 8b, c). Similar to that seen for differentially expressed genes, we found that H3K9me3 spreading across a subset of LITs correlated with their downregulation in both MII oocytes and 2-cell embryos lacking KDM4A (Figure 4f, g and Extended data 8d). Taken together, H3K9me3 removal by KDM4A is required for optimal activation of a subset of LTRs and LITs in the oocyte and pre-implantation embryo during ZGA.

Finally, we wanted to understand whether embryonic fate of MZ mutant embryos could be reversed after aberrant H3K9me3 establishment in the MII oocyte. Therefore, we injected human wild-type or catalytic dead (H188A) *Kdm4a* mRNA[12] into *Kdm4a*$^{-/-}$ MII oocytes, followed by *in vitro* fertilization with *Kdm4a*$^{-/-}$ sperm (Figure 5a, Extended data 9a). A significant proportion of *Kdm4a*$^{-/-}$ MII oocytes that received wild-type *Kdm4a* mRNA progressed to form healthy, compacted morulae and blastocysts, while those receiving the catalytic-dead *Kdm4a* mRNA suffered developmental failure (Figure 5a). Based on these findings, we propose that KDM4A-mediated H3K9me3 removal in oocytes is crucial for normal ZGA and epigenetic reprogramming that confers developmental competence (Figure 5b).

Histone modifications inherited in the gametes are faithfully propagated to embryos across cell divisions with the help of co-factors[37,38]. During chromatin replication in somatic cells, H3K9me3-marked histones are preferentially recycled and re-used in daughter cells[39]. In addition to transcriptional repression, an increased maternal load of H3K9me3-marked histones in *Kdm4a*$^{-/-}$ oocytes could persist in daughter cells during early embryonic mitoses. This could potentially create a higher number of late-replicating sites that lengthen the cell cycle and thereby compound embryonic fate. This hypothesis is supported by the finding that KDM4A expression is cell-cycle regulated and in turn regulates S-phase timing via H3K9me3 demethylation[40].

Our findings show that transcriptional quiescence in the oocyte likely makes it is capable of existing in a state of extensive co-occupation by H3K4me3 (an euchromatic mark) and H3K9me3 (a heterochromatic mark), a phenomenon normally not found in actively transcribing cells. KDM4A therefore plays a major role in preventing the formation of a H3K9me3 reprogramming barrier during normal development. Therefore, it is also possible that the variable expression of KDM4A in individual human MII oocytes could lead to the formation of such reprogramming barriers in oocytes with low expression of KDM4A, and potentially contribute to defective ZGA and mitoses observed in a large proportion of human embryos in fertility clinics.

Online Methods

Mouse housing and handling

Generation of *Kdm4a* null mice have been described previously[7,11]. Mouse breeding pairs were housed in conventional cages behind a specific-pathogen-free barrier and experimental mice were housed in groups at the University of Copenhagen in individually ventilated cages. All of the animals were exposed to 12 h of light (6:00-18:00) and had free access to water and a standard mouse chow diet. The mice used in experiments were re-genotyped before use and after sacrifice to confirm identity. All mouse work was approved by the Danish Animal Ethical Committee ("Dyreforsøgstilsynet") and is compliant with all relevant national ethical regulations regarding animal research.

Collection and processing of mouse oocytes and embryos

The procedure can be found at *Protocol Exchange*[41].

In vitro fertilization of mouse oocytes

The protocol can be found at *Protocol Exchange* [42].

In vitro transcription and microinjection of Kdm4a mRNA

The protocol can be found at *Protocol Exchange* [43].

Immunofluorescence of mouse zygotes

The protocol can be found at *Protocol Exchange* [44].

Imaging was completed in a single session using a Leica SP8 confocal microscope with a 63 objective and a system optimized step size of 0.898 μm. The maternal pronuclei volumes and intensities were measured using Volocity 3D for each zygote summing up values derived from 15 image stacks for each zygote.

Metaphase spread of mouse oocytes

The protocol can be found at *Protocol Exchange* [45].

μChIP-Seq for mouse oocytes and 2-cell embryos

This procedure is based on previously published protocols and has been described below in stages beginning at the harvesting of oocytes/embryos and ending with library preparation and sequencing[1,41,42,46,47].

Cross-linking of oocytes and embryos for chromatin immunoprecipitation

The cross-linking was performed as previously described[1]. Briefly, oocytes were collected in 50 μl M2 medium. Next, 50 μl PBS containing 20mM Na-Butyrate with 2% formaldehyde was added to cross-link the oocytes (1% final concentration). For the two-cell embryos, the samples were recovered from KSOM-Embryomax medium (Merk Millipore) and briefly washed in M2 medium and cross-linked similarly to the oocytes. The samples were carefully mixed by gentle vortexing and incubated at room temperature for 8 min; the samples were

vortexed once more during the incubation step. Glycine (12 μl; final concentration of 125 mM) was added to stop the fixation. The oocytes were mixed by gentle vortexing and incubated for 5 min at room temperature; the samples were vortexed once more during the incubation step. Next, the samples were centrifuged at 750*g* for 10 min and the supernatant was discarded. The samples were then washed twice with a 400-μl drop of ice-cold PBS with Na-butyrate. Only 10 μl supernatant was left behind after the last wash and the samples were snap frozen in liquid nitrogen and stored at -80 °C before further processing.

Binding of antibodies to paramagnetic beads

The stock solution of paramagnetic Dynabeads protein A was vortexed thoroughly to ensure the suspension was homogenous before pipetting. A 100-μl volume of Dynabeads stock solution was transferred to a 1,5 ml tube. The tube was placed in a magnetic rack and the beads were captured on the tube wall. The buffer was discarded and the beads were washed two times in 500 μl RIPA buffer (10mM Tris–HCl pH8.0, 140mM NaCl, 1mM EDTA, 0.5mM EGTA, 1% Triton X-100, 0.1% SDS and 0.1% Na-deoxycholate) and resuspended in RIPA buffer to a final volume of 100 μl. For each μChIP reaction, 97 μl RIPA buffer was aliquoted into 200-μl PCR tubes on ice. The washed beads were vortexed thoroughly, 1.5 μl of the bead suspension was added to each of the 200-μl PCR tubes and 1.75 μg anti-H3K4me3 (Millipore, monoclonal no. 04-745, lot number 25911879), anti-H3K9me3 (Abcam, polyclonal no. ab8898, lot number GR153112-1) or anti-H3K36me3 (Abcam, polyclonal no. ab9050, lot number GR233723-2) antibody was added to each tube. The antibody–bead solution was incubated overnight at 40 r.p.m. on a ‘head-over-tail’ tube rotator at 4 °C.

Chromatin preparation

Tubes containing oocytes or embryos were collected from -80 °C and placed on dry-ice in an insulated box. Oocytes or embryos from several tubes were pooled into the tube containing the highest cell number using 120 μl 0.8% SDS lysis buffer supplemented with 20 mM Na-butyrate, 1 mM PMSF and protease inhibitor cocktail for pooling. Each tube containing cross-linked oocytes or embryos contributed 10 μl PBS with Na-butyrate to the pooling. Depending on the number of tubes used for pooling, additional PBS with 20 mM Na-butyrate was added to obtain a final volume of 160 μl for sonication. The samples were incubated on ice for a minimum of 5 min. The lysate was sonicated with a UP100H Ultrasonic Processor (Hielscher) fitted with a 2-mm probe, 27% power and pulse settings of 0.5 cycles. After sonication, 230 μl RIPA dilution buffer (10 mM Tris–HCl pH 8.0, 175 mM NaCl, 1 mM EDTA, 0.625 mM EGTA, 1.25% Triton X-100, 0.125% Na-deoxycholate, 20 mM sodium butyrate, 1 mM PMSF and protease inhibitor cocktail) was added to each sample. The samples were centrifuged at 12,000*g* for 10 min in a swinging-bucket rotor at 4 °C and 380 μl of the supernatant was carefully transferred to a new 0.6 ml tube. RIPA dilution buffer (150 μl) was added to 10 μl of the leftover volume and pellet, and the samples were re-sonicated. The samples were centrifuged again at 12,000*g* in a swing-bucket for 10 min at 4 °C. The supernatant (150 μl) was added to the 0.6 ml tube containing 380 μl chromatin. The samples were mixed well with gentle vortexing and quick-spun in a mini-centrifuge. For input, about 5% (25 μl) of the chromatin was removed from the 530 μl.

Immunoprecipitation and washes

Pre-incubated antibody–bead complexes were washed twice with 130 µl RIPA buffer by vortexing roughly; the tubes were centrifuged in a mini-centrifuge to collect any solution trapped in the lid and the antibody–bead complexes were captured in a magnetic rack cooled on ice. The antibody–bead complexes were resuspended in 50 µl chromatin and the entire sample was transferred to a ChIP-ready chromatin-containing tube, resulting in a total volume of 505 µl. The samples were incubated at 40 r.p.m. on a head-over-tail tube rotator at 4 °C for 36–44 h. The chromatin–antibody–bead complexes were washed four times with 100 µl ice-cold RIPA buffer with SDS and NaCl, the concentrations of which were titrated for each antibody to optimize the signal-to-noise ratio. The tubes were placed on ice-cold magnets for at least 2 min between each wash. The H3K9me3 µChIP samples were washed four times with RIPA buffer containing 0.1% SDS and 140 mM NaCl. The H3K4me3 and H3K36me3 µChIP samples were washed as follows: first wash, RIPA buffer with 0.10% SDS and 140 mM NaCl; second and third washes, RIPA buffer with 0.20% SDS and 300 mM NaCl; and fourth wash, RIPA buffer with 0.10% SDS and 140 mM NaCl (each wash involved rough vortexing at full speed, repeated twice with pauses on ice in between). The samples were resuspended in 100 + 50 µl Tris–EDTA buffer and transferred to a clean 200 µl tube.

DNA purification

The beads were pelleted using a magnet and the Tris–EDTA buffer was removed. ChIP elution buffer (150 µl; 20 mM Tris–HCl pH 7.5, 50 mM NaCl, 5 mM EDTA, 1% SDS and 30 µg RNase A) was added to remove the RNA. The samples were incubated at 37 °C for 1 h at 1,200 r.p.m. on a Thermomixer. Protein digestion and reversal of cross-linking was performed by adding 1 µl proteinase K (20 mg ml^{-1} stock) to each tube followed by incubation at 68 °C for 4 h at 1,250 r.p.m. The eluate was transferred to a 1.5-ml tube and a second elution with 150 µl was performed for 5 min. Each sample was pooled with the first supernatant. The ChIP DNA was extracted using phenol–chloroform–isoamylalcohol, ethanol precipitated with 11 µl acrylamide carrier as described previously[23] and dissolved in 43 µl EB (10 mM Tris–HCl) overnight on ice.

ChIP library preparation and sequencing

The ChIP and input library preparation was carried out using a Qiagen ultra low input library preparation kit (Qiagen, 180492) according to the manufacturer’s instructions. The libraries were quantified using a KAPA library quantification kit (Roche, kk4824). The libraries were pooled at an equimolar ratio and purified using AMPure XP beads at a 1:1 ratio. The final purified library pool was eluted in 25 µl Qiagen EB. The libraries were sequenced on an Illumina NextSeq500.

ChIP-Seq data processing, visualization and normalization

The protocol can be found at *Protocol Exchange*[48].

RNASeq of oocytes and embryos

The wild-type and knockout samples were sampled and processed simultaneously using the developmental stage of wild-type embryos as a reference. Sixteen biological replicates were sampled per developmental stage for each genotype for mRNA-Seq and six biological replicates were sampled for each genotype for the two-cell total RNA-Seq. Following RNA-Seq sampling as described earlier, the samples were lysed in-tube and the cDNA was amplified according to the manufacturer's instructions (Takara Bio; mRNA-Seq, SMART-Seq v4 ultra low input RNA kit, cat. no. 634894; total RNA-Seq, SMART-Seq stranded kit, cat. no. 634442). For both the mRNA and total RNA-Seq, individual cDNA libraries were run and verified on an Agilent 2100 Bioanalyzer instrument using a high sensitivity DNA kit (Agilent, 5067-4626) with no observable dropouts during lysis or cDNA amplification. Human MII oocytes were processed similar to mouse MII oocytes for cDNA amplification and library preparation.

For mRNA-Seq, libraries were prepared with 100 pg input using the Nextera XT DNA library preparation kit (Illumina, FC-131-1024) and the Nextera XT index kit v2 (FC-131-2002). For total RNA-Seq, we followed the manufacturer's instructions for library preparation using built-in Illumina compatible indexes (Takara Bio, 634442). The final libraries were purified and quantified on a Qubit 3.0 fluorimeter using a high sensitivity DNA assay (Thermo Fisher Scientific, Q32854) and verified on the Agilent 2100 Bioanalyzer high sensitivity DNA chip as described above and pooled to 4 nM. The 4-nm library pools were denatured and loaded according to the recommended NextSeq500 guidelines.

Single cell/embryo RNASeq data processing

The protocol can be found at *Protocol Exchange* [49].

ChIP-seq and RNA-seq data integration

Unless otherwise stated, all integration and visualization were performed using EaSeq v.1.05-1.111 (http://easeq.net)[52] and integrated tools using the default settings. The significantly regulated part of the Deseq2 output together with their corresponding genomic coordinates were used for analysis. In addition, we used zygotically activated genes previously identified and provided as Supplementary Table 1 in the publication by Marcfarlan and colleagues[20], which were translated to mm10 coordinates using the liftover tool[3]. Both sets of genes were imported as a 'Regionset' into EaSeq. Regions were sorted according to the $\log_2$ fold difference in expression using the 'Sort' tool and used as viewpoints in heat maps showing the ratios of µChIP-Seq signal. These heat maps were generated using the 'RatioMap' tool and adjusted to show 200% of the visualized gene bodies. For the bee swarm plots, H3K9me3 values were quantified over the entire span of the gene bodies using the 'Quantify' tool and the settings 'Start position = start', 'Start offset = 0', 'End position = end', 'End offset = 0' as well as for ±2 kbp windows near the TSS using the settings 'Start position = start', 'Start offset = -2000', 'End position = start', 'End offset = 2,000'. The $\log_2$ fold difference values were calculated using the 'Calculate' tool with 'Division' and 'Logarithm' checked to the respective quantified values. Bee swarm plots were made using R[50] (https://www.R-project.org) and beeswarm package[51]. For *k*-

means clustering of LITs according to transcriptional differences, the single-cell or embryo polyA-RNA levels over the entire span of each LIT were quantified using the Quantify tool using the settings 'Start position = start', 'Start offset = 0', 'End position = end' and 'End offset = 0', and the Regionset was exported for normalization in Excel. The values for each single-cell/embryo sample were $\log_2$-fold normalized to the mean of all samples and the conditions at that particular LIT. After re-import as a Regionset, these normalized values alone were used for *k*-means clustering using the 'ClusterP' tool and the settings 'No normalization', 'Log Transform' unchecked, '*k*-means clustering' checked and $k = 8$. Pair-wise $\log_2$-fold differences between the means of different conditions, were visualized in bee swarm and box plots made using R[50] (https://www.R-project.org/) and beeswarm packages[51]. Ratiometric heat maps showing pseudocolored differences in H3K9me3 levels at LITs were generated using the ratiomap plot type as described above. The abundance of particular repeat types in the clusters was done by counting the abundance of repeats identified in the transcripts in the original studies within each cluster and comparing this to the number expected if each cluster contained an average number of repeats. Two-tailed Fischer's exact tests and Benjamini–Hochberg corrections for multiple testing were done using R.

Repeat data visualization

The values of the $\log_2$ fold difference and count matrices normalized to million reads (CPM) were generated using edgeR as described above. The mean RPKM values per instance were derived by dividing the mean read counts from all conditions with the number of repeat instances in the mm10 reference genome for each repeat and by the median size of the repeat in kbp. The genomic coverage was calculated as the number of instances multiplied by the median repeat size.

Integration with previously published data on expression from repeat types[34] was done as follows. The mean expression for each repeat type at each stage was calculated based on Supplementary Table 2 in the published work[34], $\log_2$-normalized to the mean signal at all shown stages and visualized in a simple heat map. Repeats were ordered according to the $\log_2$-fold difference between the two-cell and MII samples, and divided into two overall groups—maternal and zygotic—based on this sorting and on whether or not the repeat expression predominantly peaked in the maternal stages. We then assembled our Repenrich results, which corresponded to the groups from Franke et al. [34], and visualized them in bubble plots. In cases where multiple subtypes of repeats existed within the categories used by Franke and colleagues, we chose to visualize all of these on the same *y* coordinate in the bubble plots. The statistical test was based on the comparison of the number of significant repeat types with the expected in the zygotic and maternal groups. It should be noted that a lot of repeat types in the ORR1A group were significantly regulated and we also performed the analysis without this group to ensure that significance was not due to this group alone, and still observed more significantly regulated groups among the zygotic groups than expected.

Europe PMC Funders Author Manuscripts

Visualization and data integration at LITs

Unless otherwise stated, all integration and visualization were performed using EaSeq v.1.05-1.111 (http://easeq.net)[52] and integrated tools using the default settings. The previously published LITs of mouse oocytes[35] were acquired from the Supplementary Data 1 and imported as a Regionset into EaSeq. The total RNA as well as single-cell or embryo polyA-RNA levels at each LIT were quantified and normalized to the FPKM using the Quantify tool and the settings 'Start position = start', 'Start offset = 0', 'End position = end', 'End offset = 0', and used to calculate the $\log_2$ fold difference between the average knockout and average wild-type transcript levels. The order of the LITs was sorted according to this value using the 'Sort' tool, and heat maps were generated using the 'RatioMap' tool and adjusted to show 200% of the visualized LITs. The values used in plots along the entire body of gene bodies or LITs as well as surrounding loci were generated using the 'FillTrack' tool and exported using the beta-tool 'Export values from plots'. For each relative point along the *x* axis, values from each replicate were $\log_2$ normalized to the mean of all conditions and plotted in Microsoft Excel 2016 or as box plots in R.

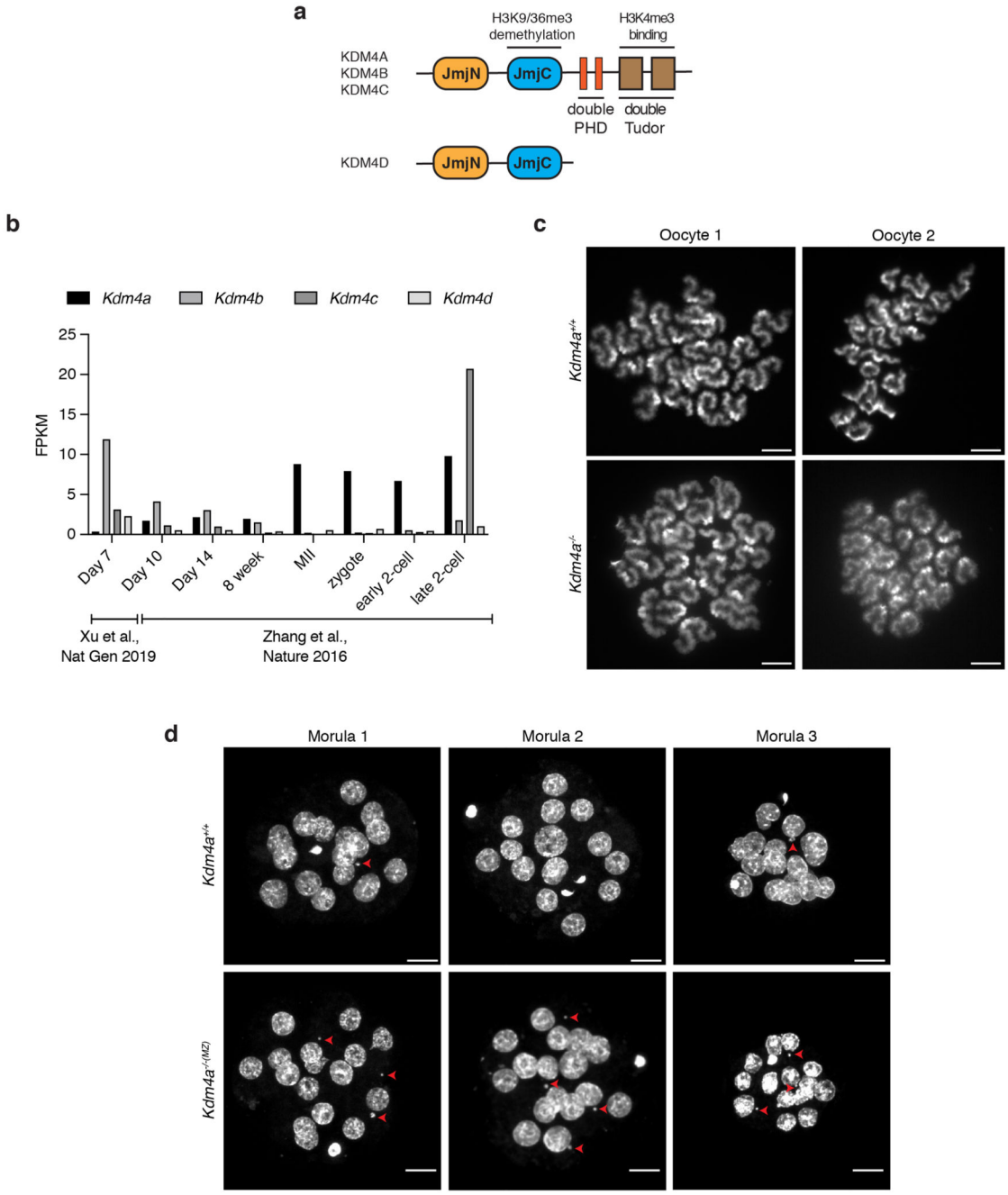
Statistics and reproducibility

Calculations and statistical analyses were performed using Microsoft Excel 2016, GraphPad Prism and R. The nature of statistical testing and the exact p-values for statistics have been provided wherever possible. Due to space constraints in Extended data 5 (panels e, f), the maximal p-value is highlighted and exact gene specific p-values are available in the associated source data. Taking biological and technical variations into account, reproducibility of results was ensured through a combination of using biological replicates and/or independent repetitions which have been explicitly mentioned wherever applicable.

Statistics and reproducibility

Calculations and statistical analyses were performed using Microsoft Excel 2016, GraphPad Prism and R. The nature of the statistical tests and the exact *P* values for the statistical analyses have been provided wherever possible. Due to space constraints in Extended Data Fig. 5e,f, the maximal *P* value is highlighted and the exact gene-specific *P* values are available in the associated Source data. Taking biological and technical variations into account, reproducibility of results was ensured through a combination of biological replicates and/or independent repetitions, which have been explicitly mentioned wherever applicable.

Extended Data



Extended Data Fig. 1. Extended characterization of the role of KDM4A in oocytes and preimplantation embryos

(a) Structure of KDM4A/B/C/D with corresponding catalytic JmjC domain(s) and H3K4me3 binding double Tudor domain(s) shown.

(b) FPKM table of KDM4A/B/C/D expression in P7, P10, P14, 8 weeks and MII oocytes as well as in post-fertilization zygote and 2-cell embryos from published sources.

(c) Representative images of metaphase chromosome spreads from two independent *Kdm4a* $^{+/+}$ and *Kdm4a* $^{-/-}$ MII oocytes with DNA counterstained using DAPI. Spreads were

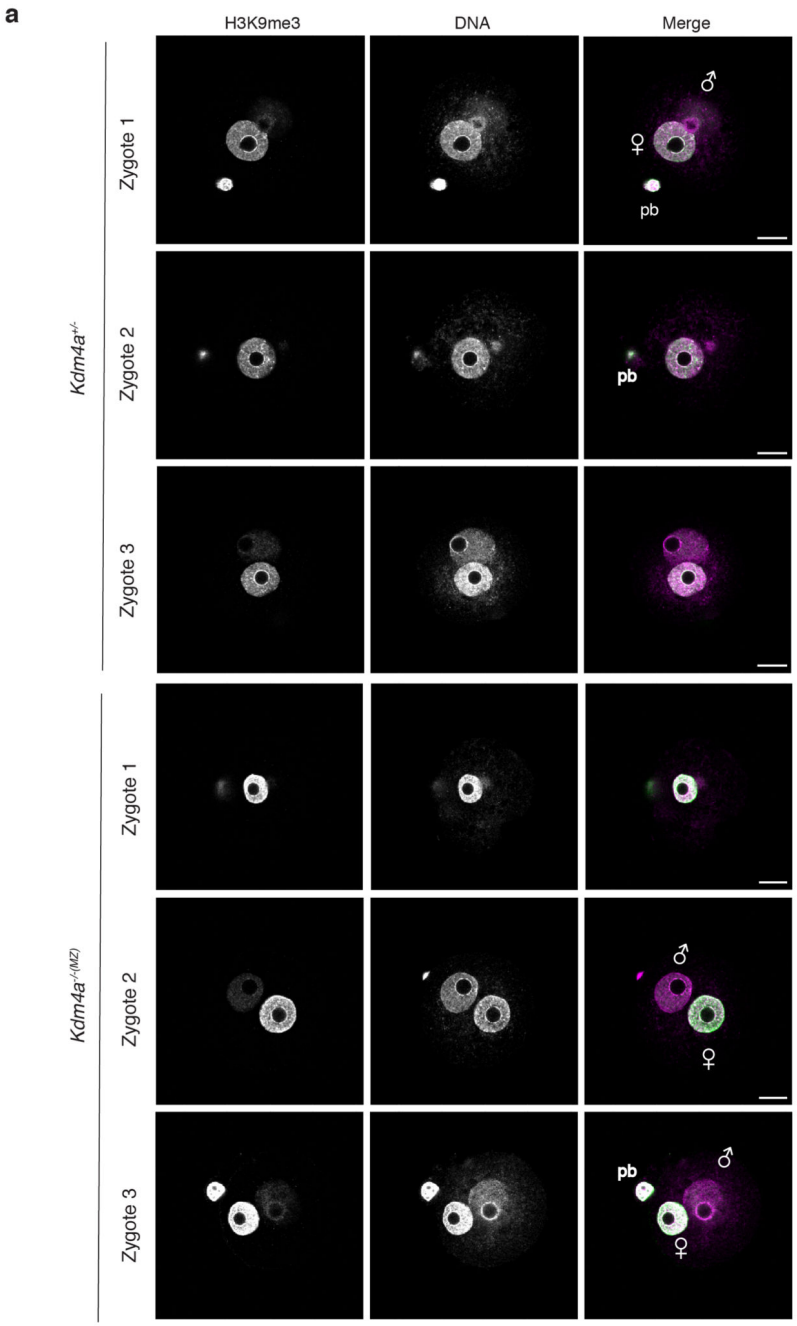
Europe PMC Funders Author Manuscripts

performed on ten MII oocytes derived from three independent donor females (Scale bar: 10μm).

(d) Flattened Z-stack images of three independent *Kdm4a*$^{+/-}$ control and *Kdm4a*$^{-/-(MZ)}$ preimplantation embryos from three independent mice per genotype. All embryos were imaged in a single session. Micronuclei are highlighted with red arrows (Scale bar: 10μm). Statistical source data are provided in source data extended data fig. 1.

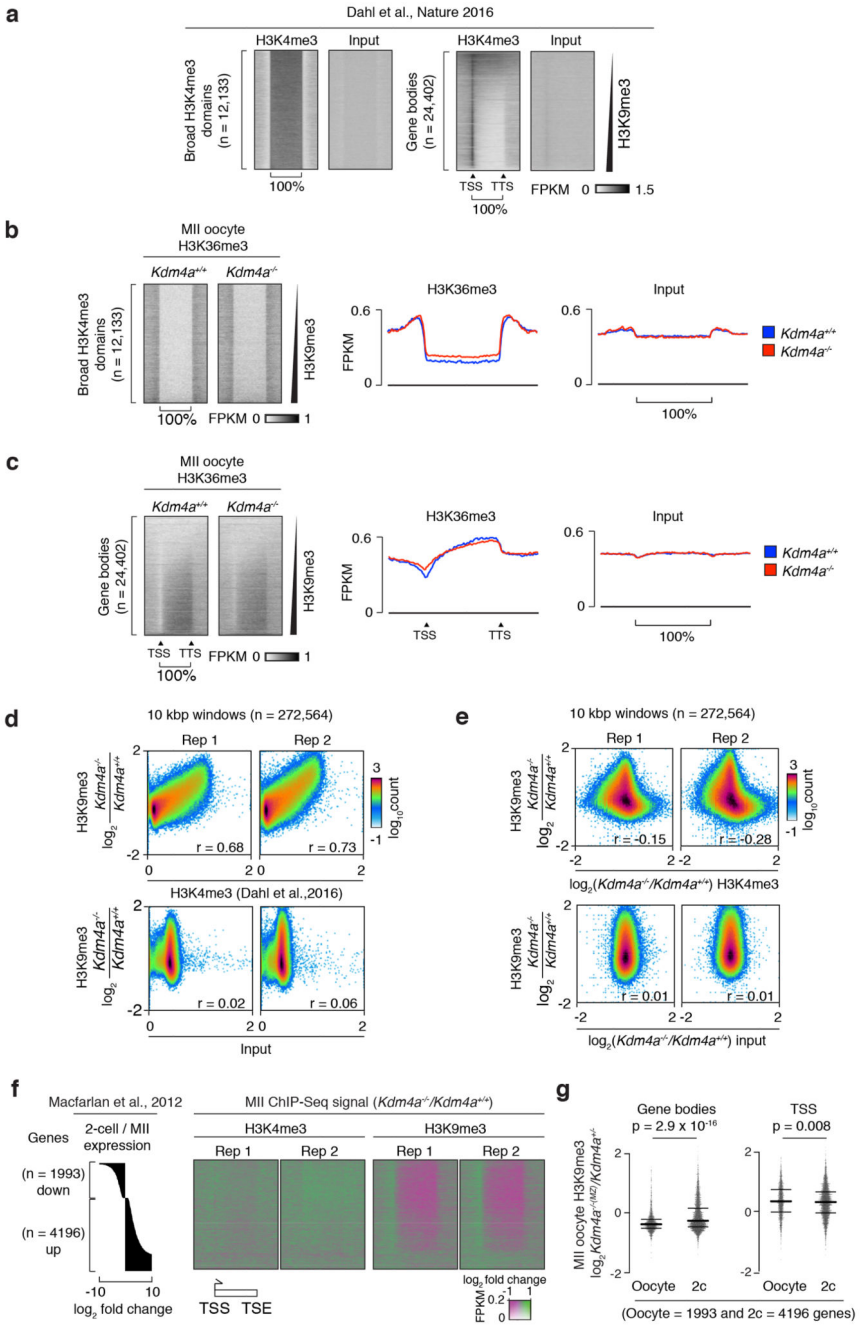
Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts



Extended Data Fig. 2. Increased maternal H3K9me3 in KDM4A MZ mutant zygotes

(a) Immunofluorescence imaging of H3K9me3 in maternal pronuclei of from *Kdm4a*$^{+/-}$ and *Kdm4a*$^{-/-(MZ)}$ zygotes. Representative additional mid-section Z-stack images from the same batch of zygotes as in Figure 2a. DNA is counterstained using DAPI. All images are of a maternal pronuclei mid-section shown for comparison (Scale bar: 10μm).



Extended Data Fig. 3. Comparing changes in H3K36me3 and H3K9me3 to H3K4me3 in KDM4A knockout oocytes and ZGA specific genes

(a) Heatmap of published H3K4me3 data at region-sets: bdH3K4me3 and UCSC genes with non-redundant transcription start-sites. Region-sets were ordered according to enrichment of H3K9me3 (FPKM) in our wildtype sample.

(b) Heat maps (left panel) and line graphs (right panel) of averaged H3K36me3 μChIP-Seq signal in *Kdm4a*$^{+/+}$ (blue) and *Kdm4a*$^{-/-}$ (red) MII oocytes at bdH3K4me3 region set.

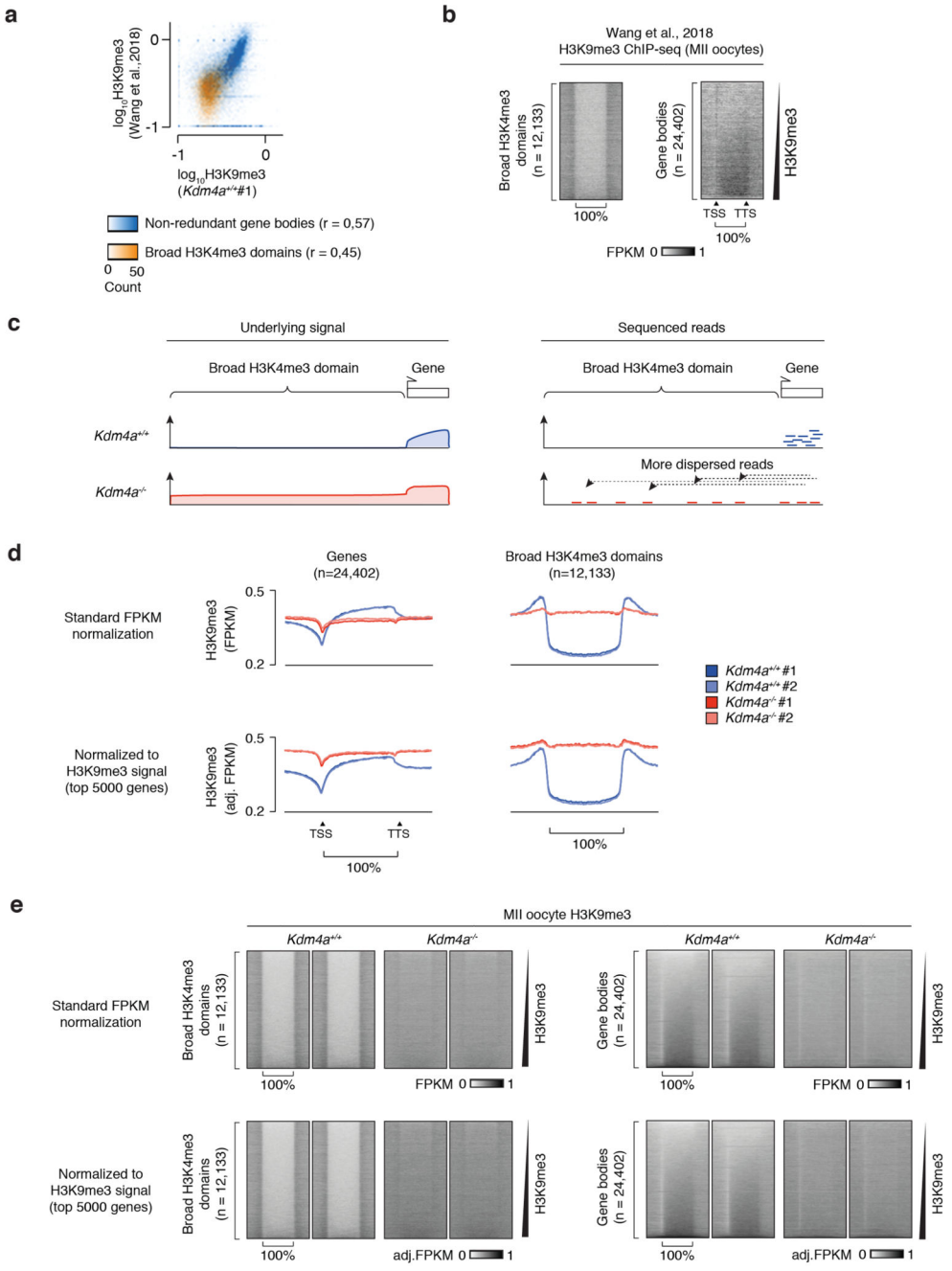
(c) Heat maps (left panel) and line graphs (right panel) of averaged H3K36me3 μChIP-seq signal in *Kdm4a*$^{+/+}$ (blue) and *Kdm4a*$^{-/-}$ (red) MII oocytes at UCSC genes as defined in (a).

(d) 2D-histograms comparing H3K9me3 gain in *Kdm4a*$^{-/-}$ MII oocytes (Y-axis) to that of published H3K4me3 (top) or the corresponding input (bottom) sample. The log10 counts in each bin are color-coded according to the left side color scale, values on X- and Y-axes are in FPKM, and r-values represent Pearson correlation coefficients.

(e) 2D-histograms comparing H3K9me3 changes (Y-axis) to that of H3K4me3 (top) or the corresponding input (bottom) sample in *Kdm4a*$^{-/-}$ MII oocytes. The log10 counts in each bin are color-coded according to the left side color scale, values on X- and Y-axes are in FPKM, and r-values represent Pearson correlation coefficients.

(f) Published genes with specific expression in oocytes or 2-cell embryos (left panel), with associated changes in H3K4me3 and H3K9me3 changes in such genes within *Kdm4a-/-* MII oocytes (middle panel). The μChIP signal is expressed as shown in a ratio with strength in FPKM.

(g) Beeswarm plot (right panel) quantifying H3K9me3 abundance in *Kdm4a*$^{-/-}$ MII oocytes at gene bodies and TSS of genes with differential expression in oocyte or 2-cell embryos according to Macfarlan et al., 2012. n represents genes enriched for oocytes or 2-cell embryos. Significance was calculated using a two-sided Wilcoxon-Mann-Whitney test. Bold and narrow lines indicate the median and interquartile ranges, respectively. Statistical source data are provided in source data extended data fig. 3.



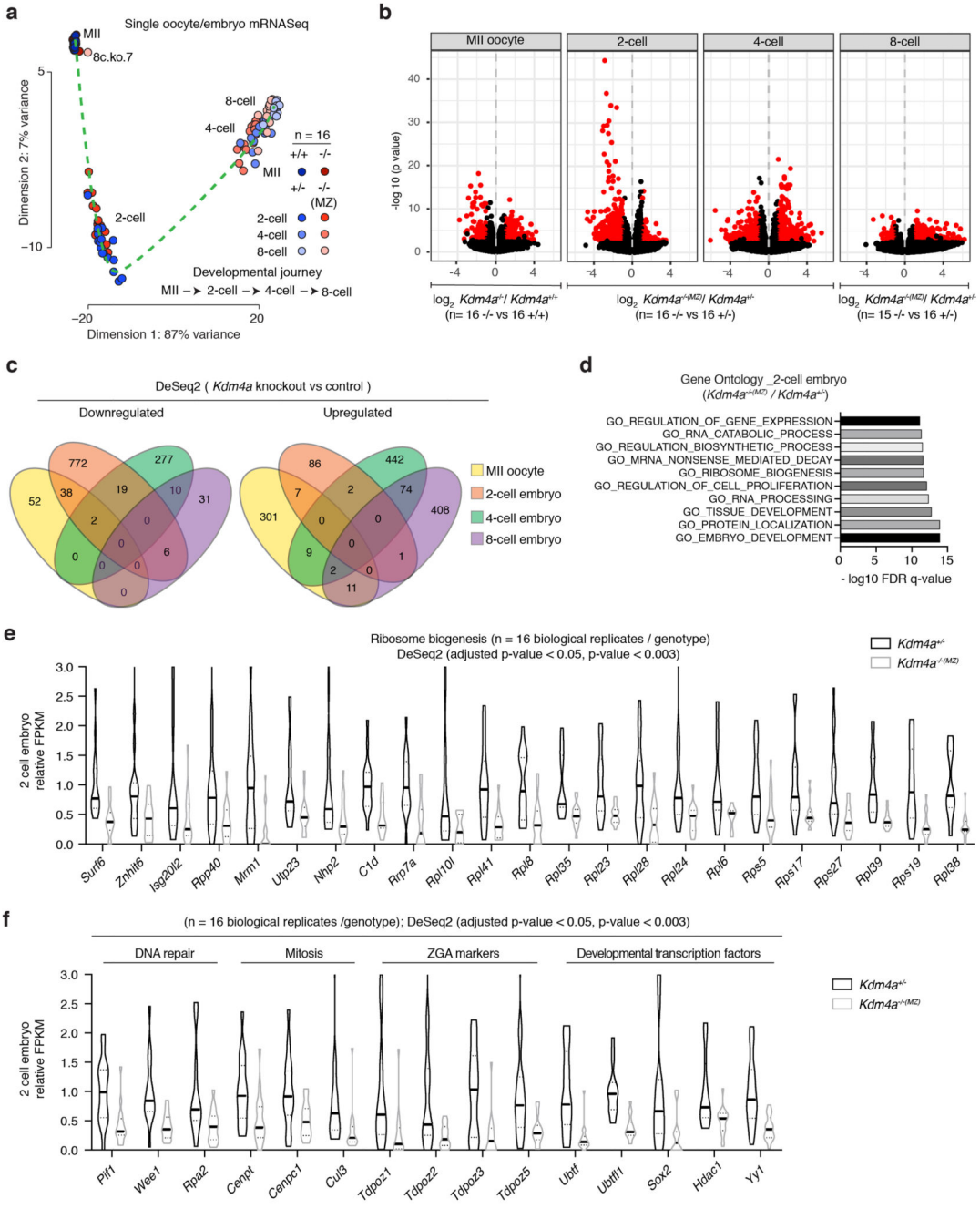
Extended Data Fig. 4. H3K9me3 μChIP-Seq signal using standard FPKM normalization

(a) 2D-histogram comparing H3K9me3 signal in wild type MII oocytes replicate 1 (X-axis) to that of previously published MII oocyte H3K9me3 (Y-axis) quantified within 24,402 non-redundant gene bodies (blue) and 12,133 bdH3K4me3 (orange) of each type of loci. Values are FPKM normalized, and represent average signal at each locus independent of its size. Pearson's correlation coefficients(r) are shown for each subset of loci.

(b) Heat maps of published H3K9me3 μChIP-seq signal in *Kdm4a*$^{+/+}$ MII oocytes at bdH3K4me3 (left) and entire genes (right), ordered and visualized as in Figure 2 b, c.

Europe PMC Funders Author Manuscripts

(c) Illustration of changes in read distribution for H3K9me3 towards covering bdH3K4me3 areas previously devoid of the mark, resulting in signal dispersion and lowering read densities at canonically enriched areas such as gene bodies.
(d) Line graphs of averaged H3K9me3 μChIP-Seq signal in *Kdm4a*$^{+/+}$ (blue) and *Kdm4a*$^{-/-}$ (red) MII oocytes at bdH3K4me3 (left) and surrounding loci as well as non-redundant UCSC genes (right), using two different normalization strategies: standard FPKM (upper row) and FPKM followed by scaling of *Kdm4a*$^{-/-}$ signal to *Kdm4a*$^{+/+}$ samples based on 5000 gene bodies with the highest H3K9me3 signal in each sample.
(e) Heat maps of averaged H3K9me3 μChIP-Seq signal in *Kdm4a*$^{+/+}$ (blue) and *Kdm4a*$^{-/-}$ (red) MII oocytes at bdH3K4me3 (left) and surrounding loci as well as non-redundant UCSC genes (right) ordered and visualized as in Figure 2, using two different normalization strategies as described above in (d). Statistical source data are provided in source data extended data fig. 4.



Extended Data Fig. 5. Differential gene expression analysis of KDM4A knockout oocytes and preimplantation embryos

(a) Principal component analysis (PCA) plot visualizing individual transcriptomes by developmental stage and genotype. n = 16 biological replicates per genotype.

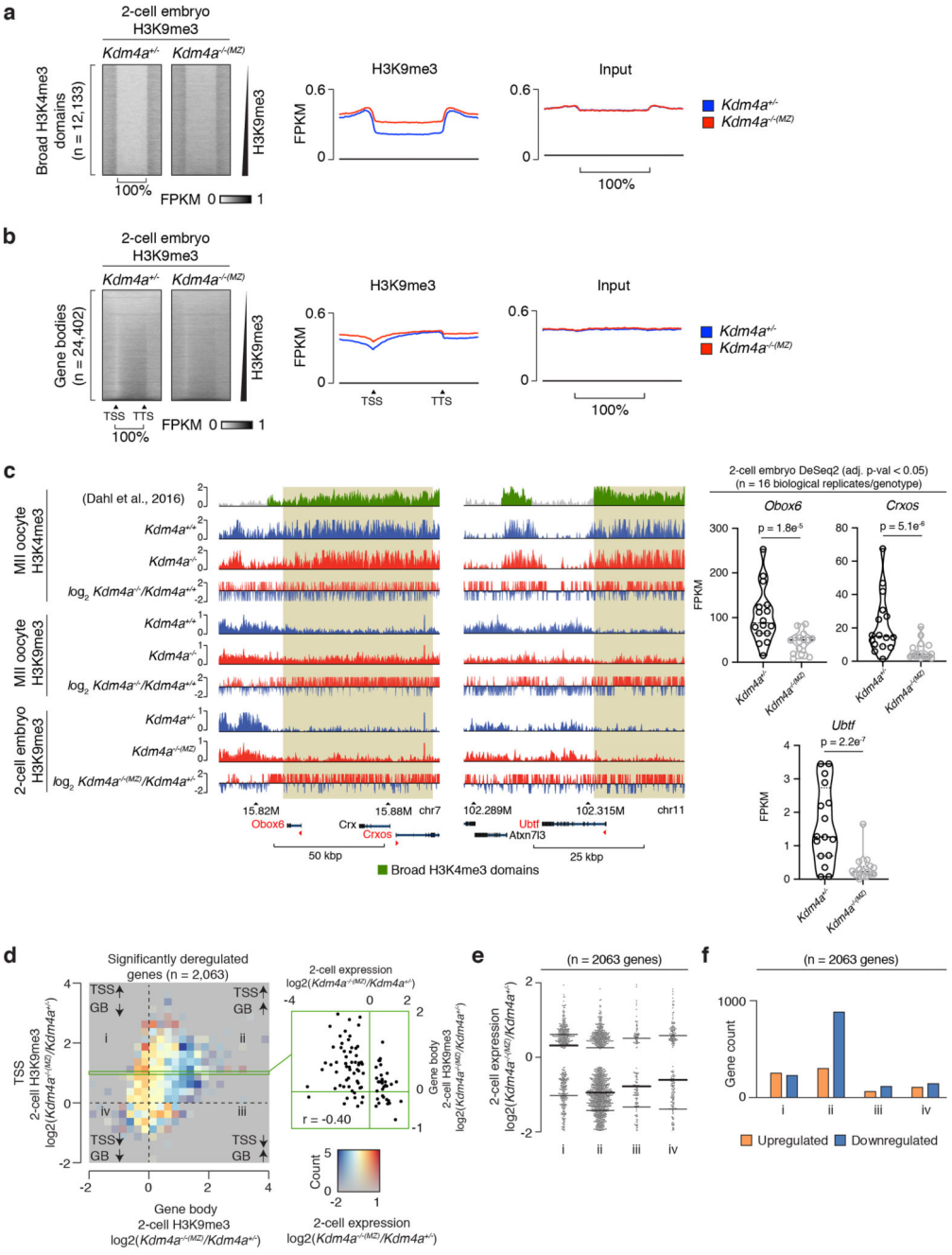
(b) Volcano plots of log2- fold changes and p-values from DeSeq2 analysis on all samples from (a). Significantly deregulated genes are defined as having FDR <0.05 and log2 fold change >=1 are in red. DeSeq2 uses a two-sided Wald test with p-values adjusted using Benjamini-Hochberg procedure, where n represents biological replicates.

(c) Significantly deregulated genes from (b) numbers plotted in a four way Venn diagram.

Europe PMC Funders Author Manuscripts

(d) Overlap of genes downregulated from (b) in MZ 2-cell embryos with MSigDB Gene ontology collection where FDR q-values are the false discovery rate analog of hypergeometric p-value after correction according to Benjamini-Hochberg procedure.
(e) DeSeq2 FPKM values (with adjusted p-value <0.05) plotted as fold change relative to mean expression levels in the control samples in each 2-cell embryo for ribosome biogenesis category genes in (d) that are significantly deregulated, where n represents biological replicates. Violin plot elements show data distribution with median and quartiles. Lowest p-value range for the DeSeq2 two-sided Wald test is shown. Exact gene specific p-values are available in the associated source data.
(f) DeSeq2 FPKM values (with adjusted p-value <0.05) plotted as in (e) for candidate genes involved in ZGA, DNA repair, mitosis and transcription regulation, where n represents biological replicates and violin plots display data distribution with median and quartiles. Lowest p-value range for the DeSeq2 two-sided Wald test is shown. Exact gene specific p-values are shown in the associated source data. Statistical source data are provided in source data extended data fig. 5.

Europe PMC Funders Author Manuscripts



Extended Data Fig. 6. Integration of changes in H3K9me3 to that of gene expression during ZGA in KDM4A MZ mutant embryos

(a) Heat maps (left panel) and line graphs (right panel) of averaged H3K9me3 μChIP-Seq signal in *Kdm4a*$^{+/-}$ (blue) and *Kdm4a*$^{-/-(MZ)}$ (red) 2-cell embryos at bdH3K4me3 region set as in Figure 2b.

(b) Heat maps (left panel) and line graphs (right panel) of averaged H3K9me3 μChIP-seq signal in *Kdm4a*$^{+/-}$ (blue) and *Kdm4a*$^{-/-(MZ)}$ (red) 2-cell embryos at non-redundant UCSC genes region set as in Figure 2c.

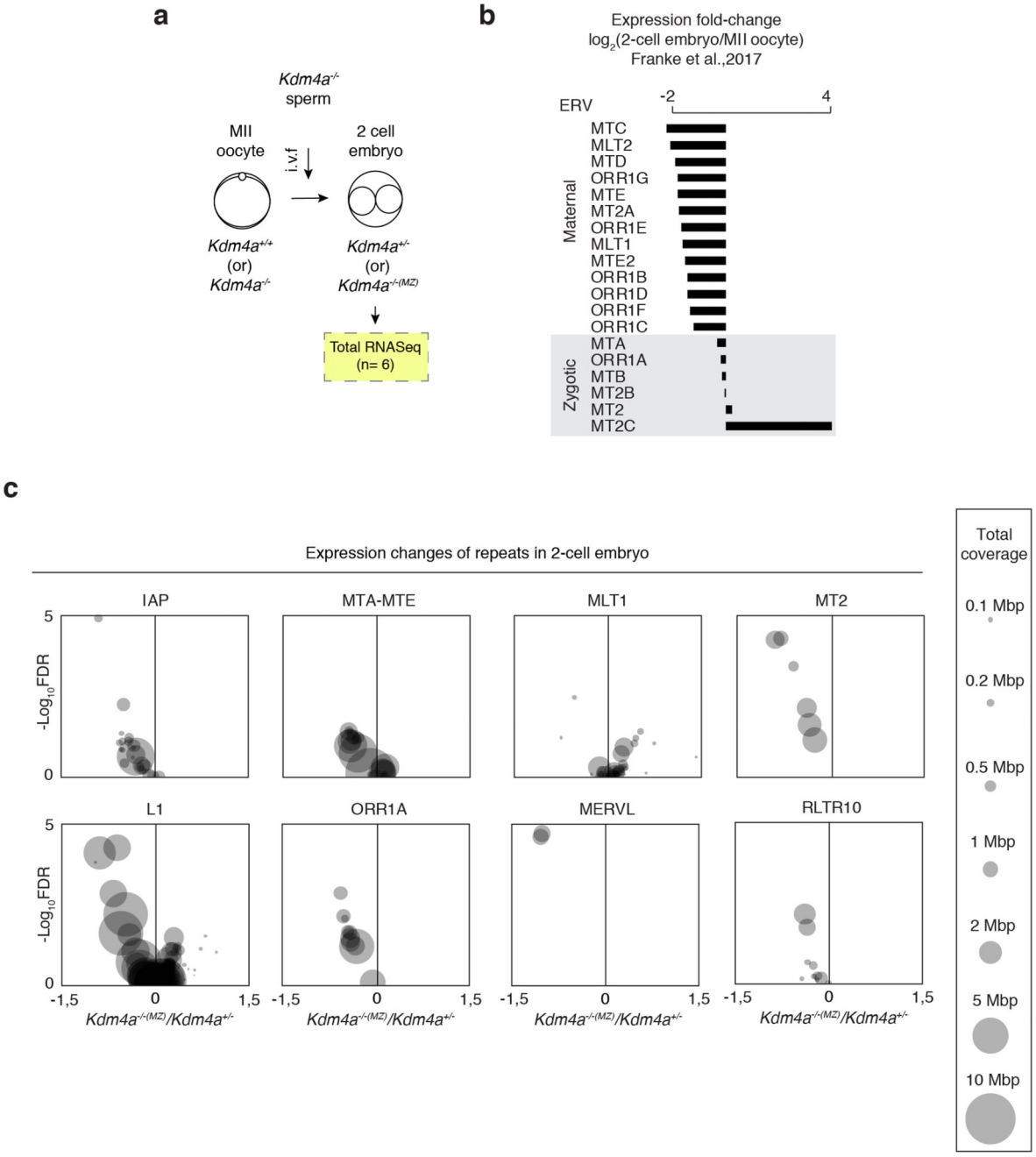
Europe PMC Funders Author Manuscripts

(c) Tracks of μChIP-Seq signal for H3K4me3 and H3K9me3 in oocytes, and of H3K9me3 signal in 2-cell embryos at the *Obox6/Crxos* and *Ubtf loci*. Red arrowheads depict transcript directionality and and bdH3K4me3 are highlighted in green. FPKM values from single-embryo RNA-seq expression analysis of selected genes. Genes correspond to those highlighted in the right part of the panel. p-values derived using DeSeq2 and two-sided Wald test and adjusted p-value < 0.05. n represents number of embryos examined, and violin plots show data distribution with median and quartiles (right panel).

(d) 2-D histogram plotting H3K9me3 abundance at gene bodies (GB) along X-axis, with TSS along Y-axis, for DeSeq2 derived differentially expressed genes in *Kdm4a*$^{-/-(MZ)}$ 2-cell embryos. n represents number of genes, colour indicates expression trend, and intensity indicates μChIP signal strength (left). The green frame indicates the subpopulation used for the scatter plot in the right side of the panel. The r-value indicates the Spearman correlation coefficient within a small subpopulation of genes with similar differences in TSS ($0.95 < \log_2 fd < 1.05$), where changes in gene body H3K9me3 correlates with expression of these genes (green bounding box).

(e) Beeswarm plot quantifying trends in gene expression change for total differentially expressed genes according to each quadrant as defined in panel (d). The lines represent medians and quartiles and n represents the number of total differentially expressed genes.

(f) Bar diagram depicting the number of up (orange) or down-regulated (blue) genes in *Kdm4a*$^{-/-(MZ)}$ 2-cell embryos subdivided within each quadrant as in (d, e). Statistical source data are provided in source data extended data fig. 6.



Extended Data Fig. 7. KDM4A loss results in reduced activation of LTR subgroups

(a) Sampling scheme for generation of stranded total RNASeq samples.

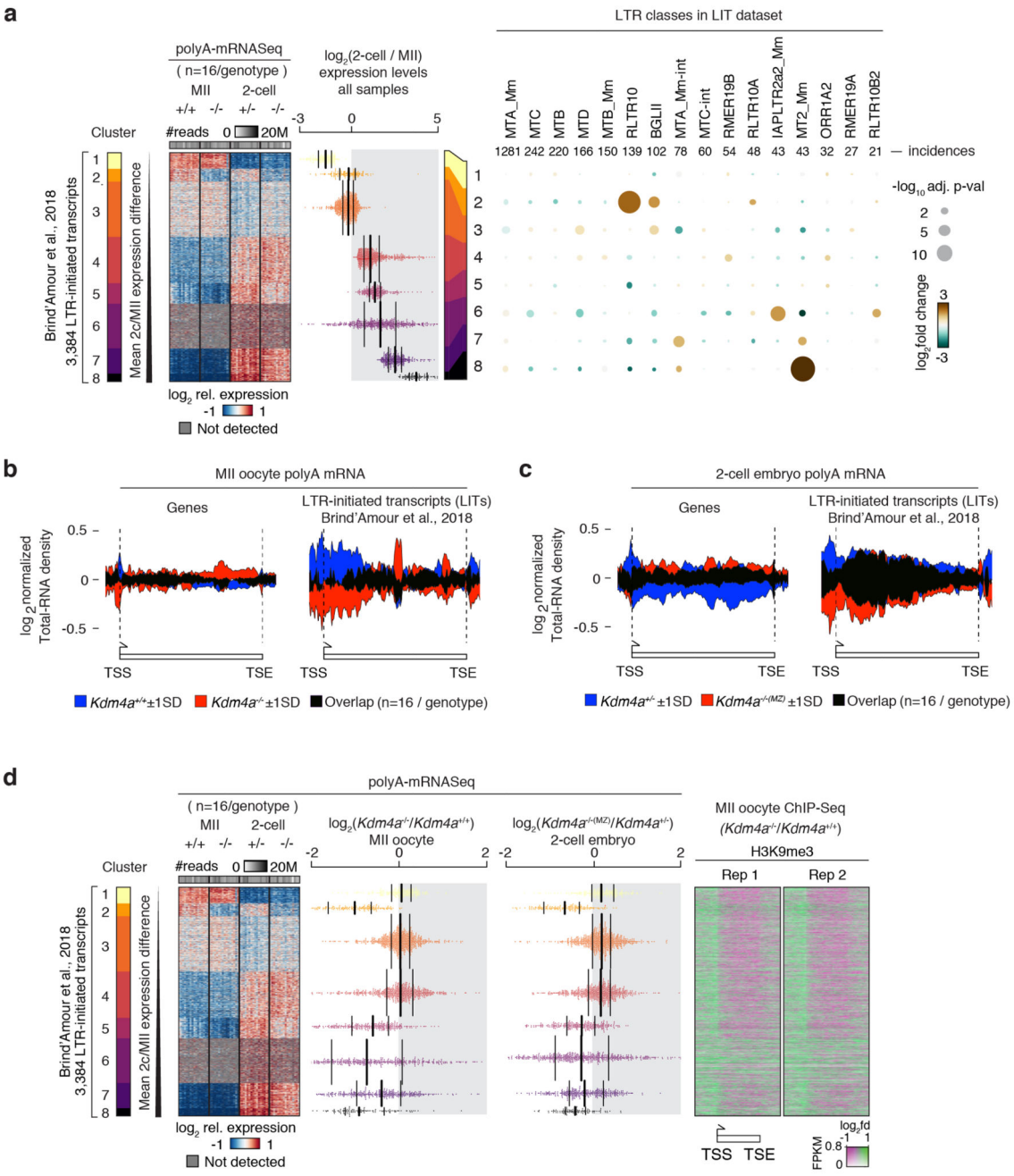
(b) Bar diagram showing changes in expression of selected groups of LTR-containing ERVs expressed in oocytes and 2-cell embryos. Data are from Franke et al., 2017 and developmental categories are from Figure 4c.

(c) Volcano plots showing the relationship between log2 fold difference in expression between *Kdm4a*$^{+/-}$ and *Kdm4a*$^{-/-(MZ)}$ 2-cell embryos (x-axis) as well as –log10 p-values (y-axis) at selected groups of repeats. n represents six biological replicates per genotype. Each

Europe PMC Funders Author Manuscripts

plot shows individual types of repeat-masker entries within the repeat group and the size of each data point is adjusted to its total coverage in the genome. Repeats with atleast six non-zero read count samples and a minimum threshold of 200 reads were fitted to a negative binomial generalized log-linear model and statistic tested using a two-sided gene-wise likelihood ratio test and Benjamini-Hochberg adjusted for multiple testing.

Europe PMC Funders Author Manuscripts



Extended Data Fig. 8. LTR-initiated transcripts are downregulated in the absence of KDM4A

(a) Composed panel showing LITs clustered according to maternal or zygotic enrichment along expression differences in *Kdm4a*$^{-/-}$ MII oocytes and *Kdm4a*$^{-/-(MZ)}$ mutant 2-cell embryos. n represents sixteen biological replicates. Heatmap (left panel) shows the expression in each replicate normalized to the average expression across all replicates per cell-type. Beeswarm plot (middle panel) shows average log2 expression change of LITs in 2-cell embryo against MII oocyte. Bold and narrow lines indicate median and interquartile ranges, respectively. Bubble plot (right panel) shows cluster-wise abundance of certain LTR-

Europe PMC Funders Author Manuscripts

containing ERV types annotated in the LITs relative to the randomly expected abundance (color), with size of the bubble representing –log10 of adjusted p-values. Statistical significance was calculated using Chi-square tests and Benjamini-Hochberg corrected for multiple testing.

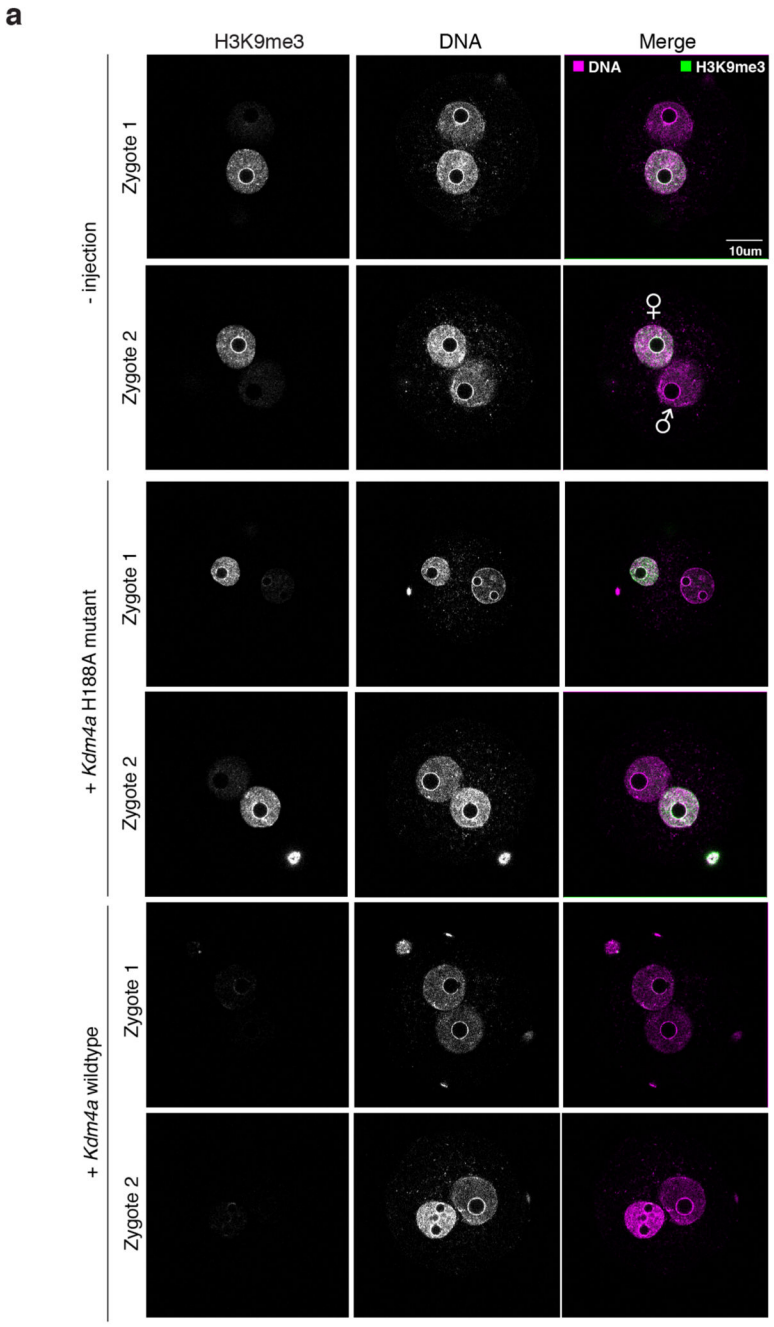
(b) Normalized mRNA-Seq read coverage in *Kdm4a*$^{+/+}$ and *Kdm4a*$^{-/-}$ MII oocytes along regular gene bodies (left panel) and known annotated LTR-initiated transcripts (LITs) (right panel). X-axis reflects the read position along feature and Y-axis values are normalized to the average of all replicates.

(c) Normalized mRNA-Seq read coverage in *Kdm4a*$^{+/-}$ and *Kdm4a*$^{-/-(MZ)}$ 2-cell embryos along regular gene bodies (left panel) and known annotated LTR-initiated transcripts (LITs) (right panel). X-axis reflects the read position along feature and Y-axis values are normalized to the average of all replicates.

(d) Composed panel showing LITs clustered with expression changes as in panel (a). n represents sixteen 16 biological replicates per genotype per cell-type. Beeswarm plots show changes in LIT expression in *Kdm4a*$^{-/-}$ MII oocytes or *Kdm4a*$^{-/-(MZ)}$ 2-cell embryos. Bold and narrow lines indicate median and interquartile ranges, respectively. Heatmap (right panel) show associated changes in H3K9me3 μChIP-seq signals in *Kdm4a*$^{-/-}$ MII oocytes in a manner described previously. Statistical source data are provided in source data extended data fig. 8.

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts



Extended Data Fig. 9. Testing efficacy of *Kdm4a* mRNA on maternal H3K9me3 removal in oocytes

(a) Representative immunofluorescence images against H3K9me3 and DNA (DAPI) of control zygotes microinjected with *Kdm4a* wildtype or H188A mutant mRNA to visualize efficiency of maternal H3K9me3 removal by KDM4A protein. Non-injected mouse zygotes were used as control for H3K9me3 staining. The experiment was performed once on oocytes pooled from three female mice to ensure quality control of freshly prepared mRNA intended for preparing multiple single thaw frozen aliquots for use in rescue experiment (Scale bar: 10μm).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We thank the members of the Dahl, Helin and Hoffmann laboratories for discussions of the data. We thank Y. Antoku of the BRIC Microscopy core facility for technical help, and V. Shukla and J. Hussain for their help with mouse genotyping. The work in the Klungland laboratory is funded by the South-Eastern Norway Regional Health Authority. The work in the Dahl laboratory is funded by the South-Eastern Norway Regional Health Authority (Early Career grant no. 2016058), the Norwegian Cancer Society and the Research Council of Norway (Young Research Talent Grant). The work in the Helin laboratory was supported by the Danish National Research Foundation (grant no. DNRF82), the Independent Research Foundation (grant no. DFF 7016-00067), a centre grant from the NNF to the NNF Center for Stem Cell Biology (grant no. NNF17CC0027852) and the Memorial Sloan Kettering Cancer Center Support Grant NIH (grant no. P30 CA008748). The Hoffmann laboratory was supported by a NNF Young Investigator Grant (grant no. NNF15OC0016662), ERC Consolidator Grant (grant no. 724718-ReCAP), a project grant from the DFF-FSS (grant no. 1111831001) and DNRF Center Grant (grant no. 6110-00344B).

Funding

The work in the Klungland laboratory is funded by the South-Eastern Norway Regional Health Authority. The work in the Dahl laboratory is funded by the South-Eastern Norway Regional Health Authority, Early Career Grant (2016058), the Norwegian Cancer Society and the Research Council of Norway, Young Research Talent Grant. The work in the Helin laboratory was supported by the Danish National Research Foundation (DNRF82), the Independent Research Foundation (DFF 7016-00067), through a center grant from the NNF to the NNF Center for Stem Cell Biology (NNF17CC0027852), and through the Memorial Sloan Kettering Cancer Center Support Grant (NIH P30 CA008748). The Hoffmann lab was supported by a NNF Young Investigator Grant (NNF15OC0016662), ERC Consolidator Grant (724718-ReCAP), a project grant from the DFF-FSS (1111831001) and DNRF Center Grant (6110-00344B).

References

1. Dahl JA, et al. Broad histone H3K4me3 domains in mouse oocytes modulate maternal-to-zygotic transition. Nature. 2016; 537:548–552. [PubMed: 27626377]
2. Zhang B, et al. Allelic reprogramming of the histone modification H3K4me3 in early mammalian development. Nature. 2016; 537:553–557. [PubMed: 27626382]
3. Hanna CW, et al. MLL2 conveys transcription-independent H3K4 trimethylation in oocytes. Nat Struct Mol Biol. 2018; 25:73–82. [PubMed: 29323282]
4. Xu Q, et al. SETD2 regulates the maternal epigenome, genomic imprinting and embryonic development. Nat Genet. 2019; 51:844–856. [PubMed: 31040401]
5. Becker JS, Nicetto D, Zaret KS. H3K9me3-Dependent Heterochromatin: Barrier to Cell Fate Changes. Trends Genet. 2016; 32:29–41. [PubMed: 26675384]
6. Ruthenburg AJ, Allis CD, Wysocka J. Methylation of lysine 4 on histone H3: intricacy of writing and reading a single epigenetic mark. Mol Cell. 2007; 25:15–30. [PubMed: 17218268]
7. Pedersen MT, et al. Continual removal of H3K9 promoter methylation by Jmjd2 demethylases is vital for ESC self-renewal and early development. EMBO J. 2016
8. Eckersley-Maslin MA, Alda-Catalinas C, Reik W. Dynamics of the epigenetic landscape during the maternal-to-zygotic transition. Nat Rev Mol Cell Biol. 2018; 19:436–450. [PubMed: 29686419]
9. Huang Y, Fang J, Bedford MT, Zhang Y, Xu RM. Recognition of histone H3 lysine-4 methylation by the double tudor domain of JMJD2A. Science. 2006; 312:748–751. [PubMed: 16601153]
10. Pedersen MT, et al. The demethylase JMJD2C localizes to H3K4me3-positive transcription start sites and is dispensable for embryonic development. Mol Cell Biol. 2014; 34:1031–1045. [PubMed: 24396064]
11. Sankar A, et al. Maternal expression of the histone demethylase Kdm4a is crucial for pre-implantation development. Development. 2017; 144:3264–3277. [PubMed: 28827393]

12. Chung YG, et al. Histone Demethylase Expression Enhances Human Somatic Cell Nuclear Transfer Efficiency and Promotes Derivation of Pluripotent Stem Cells. Cell Stem Cell. 2015; 17:758–766. [PubMed: 26526725]
13. Matoba S, et al. Embryonic development following somatic cell nuclear transfer impeded by persisting histone methylation. Cell. 2014; 159:884–895. [PubMed: 25417163]
14. Cheloufi S, et al. The histone chaperone CAF-1 safeguards somatic cell identity. Nature. 2015; 528:218–224. [PubMed: 26659182]
15. Chen J, et al. H3K9 methylation is a barrier during somatic cell reprogramming into iPSCs. Nat Genet. 2013; 45:34–42. [PubMed: 23202127]
16. Agger K, et al. Jmjd2/Kdm4 demethylases are required for expression of Il3ra and survival of acute myeloid leukemia cells. Genes Dev. 2016
17. Kooistra SM, Helin K. Molecular mechanisms and potential functions of histone demethylases. Nat Rev Mol Cell Biol. 2012; 13:297–311. [PubMed: 22473470]
18. Rugg-Gunn PJ, Cox BJ, Ralston A, Rossant J. Distinct histone modifications in stem cell lines and tissue lineages from the early mouse embryo. Proc Natl Acad Sci U S A. 2010; 107:10783–10790. [PubMed: 20479220]
19. Matsumura Y, et al. H3K4/H3K9me3 Bivalent Chromatin Domains Targeted by Lineage-Specific DNA Methylation Pauses Adipocyte Differentiation. Mol Cell. 2015; 60:584–596. [PubMed: 26590716]
20. Macfarlan TS, et al. Embryonic stem cell potency fluctuates with endogenous retrovirus activity. Nature. 2012; 487:57–63. [PubMed: 22722858]
21. Park SJ, Shirahige K, Ohsugi M, Nakai K. DBTMEE: a database of transcriptome in mouse early embryos. Nucleic Acids Res. 2015; 43:D771–776. [PubMed: 25336621]
22. Huang CJ, Chen CY, Chen HH, Tsai SF, Choo KB. TDPOZ, a family of bipartite animal and plant proteins that contain the TRAF (TD) and POZ/BTB domains. Gene. 2004; 324:117–127. [PubMed: 14693377]
23. Ma P, Schultz RM. Histone deacetylase 1 (HDAC1) regulates histone acetylation, development, and gene expression in preimplantation mouse embryos. Dev Biol. 2008; 319:110–120. [PubMed: 18501342]
24. Pan H, Schultz RM. Sox2 modulates reprogramming of gene expression in two-cell mouse embryos. Biol Reprod. 2011; 85:409–416. [PubMed: 21543769]
25. Hamdane N, et al. Conditional inactivation of Upstream Binding Factor reveals its epigenetic functions and the existence of a somatic nucleolar precursor body. PLoS Genet. 2014; 10:e1004505. [PubMed: 25121932]
26. Yamada M, et al. Involvement of a novel preimplantation-specific gene encoding the high mobility group box protein Hmgpi in early embryonic development. Hum Mol Genet. 2010; 19:480–493. [PubMed: 19915186]
27. Prendergast L, et al. The CENP-T/-W complex is a binding partner of the histone chaperone FACT. Genes Dev. 2016; 30:1313–1326. [PubMed: 27284163]
28. Singer JD, Gurian-West M, Clurman B, Roberts JM. Cullin-3 targets cyclin E for ubiquitination and controls S phase in mammalian cells. Genes Dev. 1999; 13:2375–2387. [PubMed: 10500095]
29. Moghe S, et al. The CUL3-KLHL18 ligase regulates mitotic entry and ubiquitylates Aurora-A. Biol Open. 2012; 1:82–91. [PubMed: 23213400]
30. Schulz VP, Zakian VA. The saccharomyces PIF1 DNA helicase inhibits telomere elongation and de novo telomere formation. Cell. 1994; 76:145–155. [PubMed: 8287473]
31. Russell P, Nurse P. Negative regulation of mitosis by wee1+, a gene encoding a protein kinase homolog. Cell. 1987; 49:559–567. [PubMed: 3032459]
32. Toledo LI, et al. ATR prohibits replication catastrophe by preventing global exhaustion of RPA. Cell. 2013; 155:1088–1103. [PubMed: 24267891]
33. Peaston AE, et al. Retrotransposons regulate host genes in mouse oocytes and preimplantation embryos. Dev Cell. 2004; 7:597–606. [PubMed: 15469847]
34. Franke V, et al. Long terminal repeats power evolution of genes and gene expression programs in mammalian oocytes and zygotes. Genome Res. 2017; 27:1384–1394. [PubMed: 28522611]

35. Brind'Amour J, et al. LTR retrotransposons transcribed in oocytes drive species-specific and heritable changes in DNA methylation. Nat Commun. 2018; 9
36. Rodriguez-Terrones D, Torres-Padilla ME. Nimble and Ready to Mingle: Transposon Outbursts of Early Development. Trends Genet. 2018; 34:806–820. [PubMed: 30057183]
37. Wang C, et al. Reprogramming of H3K9me3-dependent heterochromatin during mammalian embryo development. Nat Cell Biol. 2018; 20:620–631. [PubMed: 29686265]
38. Gaydos LJ, Wang W, Strome S. Gene repression. H3K27me and PRC2 transmit a memory of repression across generations and during development. Science. 2014; 345:1515–1518. [PubMed: 25237104]
39. Alabert C, et al. Two distinct modes for propagation of histone PTMs across the cell cycle. Genes Dev. 2015; 29:585–590. [PubMed: 25792596]
40. Black JC, et al. Conserved antagonism between JMJD2A/KDM4A and HP1gamma during cell cycle progression. Mol Cell. 2010; 40:736–748. [PubMed: 21145482]
41. Sankar A. Collection and handling of mouse oocytes/embryos. 2020 Jan 27.
42. Sankar A, Gonzalez JM. In vitro fertilization of mouse oocytes. 2020 Jan 27.
43. Sankar A. In vitro transcription and micro-injection of Kdm4a mRNA into mouse oocytes. 2020 Jan 27.
44. Sankar A. Immunofluorescence of mouse zygotes and preimplantation embryos. 2020 Jan 27.
45. Sankar A. Metaphase spread of mouse oocytes. 2020 Jan 27.
46. Dahl JA, Klungland A. Micro chromatin immunoprecipitation (muChIP) from early mammalian embryos. Methods Mol Biol. 2015; 1222:227–245. [PubMed: 25287350]
47. Dahl JA, Collas P. A rapid micro chromatin immunoprecipitation assay (microChIP). Nat Protoc. 2008; 3:1032–1045. [PubMed: 18536650]
48. Sankar A, Lerdrup M. ChIP-Seq data processing, normalization and visualization. 2020 Jan 27.
49. Sankar A, Johansen JV. Single oocyte/embryo RNASeq data processing. 2020 Jan 27.
50. R Core Team. R Foundation for Statistical Computing. R: A language and environment for statistical computingVienna, Austria: 2017. https://www.R-project.org
51. Eklund A. The Bee Swarm Plot, an Alternative to Stripchart. 2016
52. Lerdrup M, Johansen JV, Agrawal-Singh S, Hansen K. An interactive environment for agile analysis and visualization of ChIP-sequencing data. Nat Struct Mol Biol. 2016; 23:349–57. [PubMed: 26926434]

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

Reporting Summary

Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41556-020-0494-z.

Europe PMC Funders Author Manuscripts

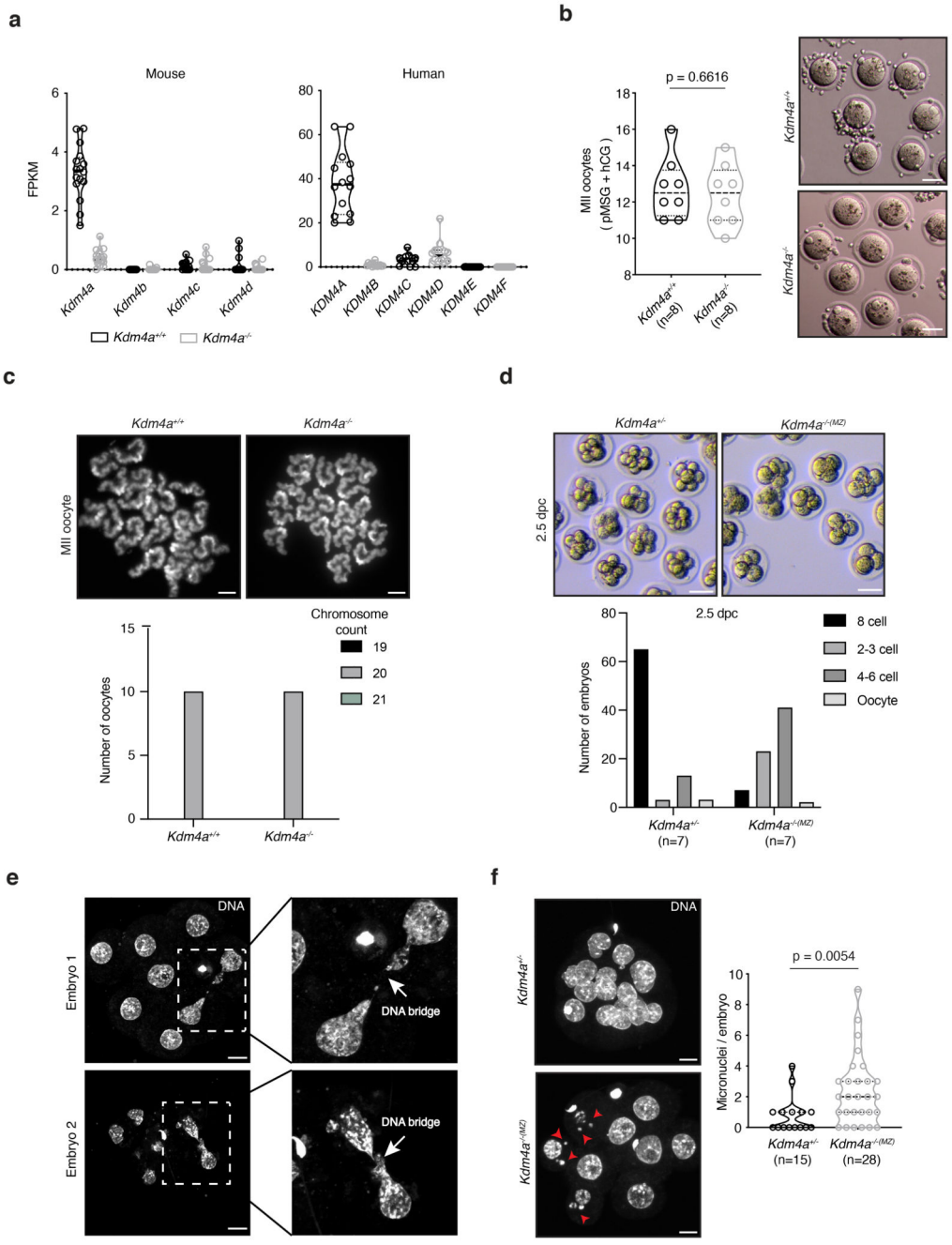


Figure 1. KDM4A plays a critical role in maintaining genomic stability of early preimplantation embryos

(a) Fragments Per Kilobase of transcript per Million mapped reads (FPKM) values for all KDM4 family genes in mouse (n=16 individual eggs from three females) and human (n=14 individual eggs from three women) MII oocyte transcriptomes. Violin plots show distribution with median and quartiles where n represents number of oocytes profiled. (b) Quantification of ovulated MII oocytes from eight-week-old *Kdm4a*$^{+/+}$ and *Kdm4a*$^{-/-}$ females stimulated with ovulation hormones (left panel) with representative images (right panel). n represents the number of female mice examined. Statistical significance was

computed using a two-tailed unpaired t-test with Welch correction (Scale bar: 100 μm). Violin plot elements show data distribution with median and quartiles.

(c) Metaphase spread of chromosomes from *Kdm4a*$^{+/+}$ and *Kdm4a*$^{-/-}$ MII oocytes with DNA counterstained using 4',6-diamidino-2-phenylindole (DAPI) with quantification below (Scale bar: 10μm).

(d) Representative images of 2.5 dpc *Kdm4a*$^{+/-}$ and *Kdm4a*$^{-/-(MZ)}$ preimplantation embryos (top) (Scale bar: 100 μm) along with distribution of cleavage embryos found in each group (bottom panel). n represents the number of female mice examined.

(e) Incidental observations of DNA bridges between dividing blastomeres of a fixed uncompacted 8-cell *Kdm4a*$^{-/-(MZ)}$ embryo (Scale bar: 10μm). Two of 12 mutant embryos showed bridges. No bridges were observed amongst 10 wildtype embryos.

(f) Representative images of micronuclei formation in *Kdm4a*$^{+/-}$ and *Kdm4a*$^{-/-(MZ)}$ preimplantation embryos with DNA counterstained with DAPI (left panel). Quantification of micronuclei, where n represents the number of embryos examined and violin plot elements show data distribution with median and quartiles (right panel). Statistical significance was calculated using a two-tailed unpaired t-test with Welch correction (Scale bar: 10μm). Statistical source data are provided in source data fig. 1.

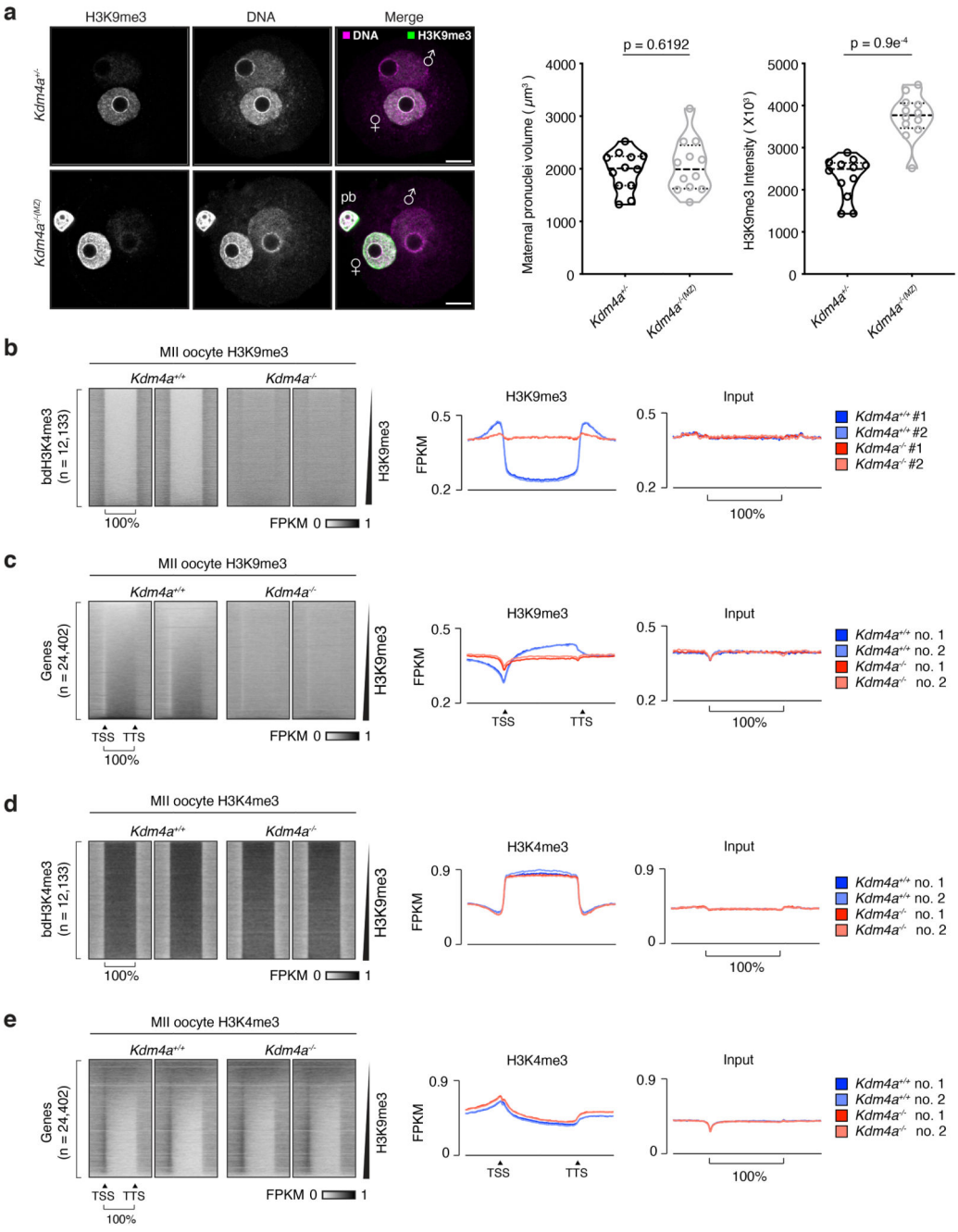


Figure 2. KDM4A preserves maternal epigenome integrity and prevents against widespread co-occurrence of H3K4me3-H3K9me3 in oocytes

(a) Immunofluorescence imaging of H3K9me3 in maternal pronuclei of *Kdm4a*$^{+/-}$ and *Kdm4a*$^{-/-(MZ)}$ zygotes where DNA is counterstained using DAPI. Representative images of a mid-section zygote are shown with both parental pronuclei (left panel). Maternal pronuclei H3K9me3 intensity levels and volume are compared for significance using a two-tailed unpaired t-test with Welch correction (right panel). Violin plots show data distribution with median and quartiles for 12 zygotes assessed for each genotype. (Scale bar: 10μm).

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

(b) Heat maps (left) and line graphs (right) of averaged H3K9me3 ChIP-Seq signal in *Kdm4a*$^{+/+}$ (blue) and *Kdm4a*$^{-/-}$ (red) MII oocytes at previously defined bdH3K4me3 (>= 20kb size) and surrounding loci, all adjusted to fit the same visual space. The bdH3K4me3 set of regions is ordered according to H3K9me3 levels in a wild-type replicate with signal enrichment shown in FPKM.

(c) Heat maps (left panel) and line graphs (right panel) of averaged (bottom) H3K9me3 ChIP-seq signal in *Kdm4a*$^{+/+}$ (blue) and *Kdm4a*$^{-/-}$ (red) MII oocytes at genes including transcription start-sites (TSS), entire gene bodies and surrounding loci. All non-redundant genes annotated in UCSC database were adjusted to fit the same visual space. The genes are ordered according to H3K9me3 levels in a wild-type replicate with signal enrichment shown in FPKM.

(d) Heat maps (left panel) and line graphs (right panel) of averaged H3K4me3 ChIP-Seq signal in *Kdm4a*$^{-/-}$ (red) and *Kdm4a*$^{+/+}$ (blue) MII oocytes at bdH3K4me3 and surrounding loci as defined in panel (b).

(e) Heat maps (left panel) and line graphs (right panel) of averaged H3K4me3 ChIP-seq signal in *Kdm4a*$^{-/-}$ (red) and *Kdm4a*$^{+/+}$ (blue) MII oocytes at non-redundant UCSC genes ordered as in panel (c). The ChIP-seq assays in b-e were performed on two independent pools of oocytes. Statistical source data are provided in source data fig. 2.

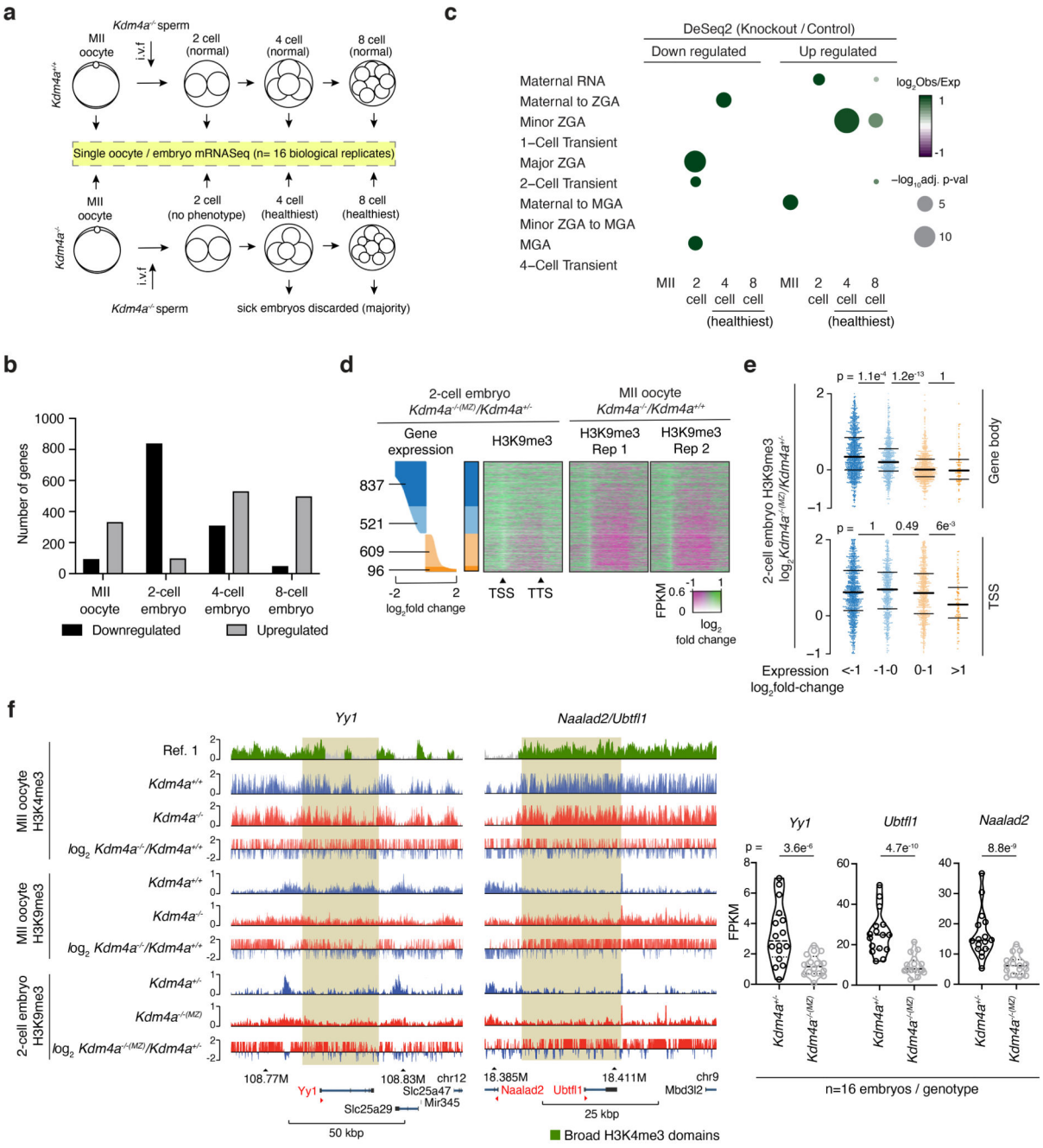


Figure 3. H3K4me3-H3K9me3 co-occurrence in oocytes leads to defective activation of ZGA genes

(a) Sampling scheme for single oocyte/embryo mRNASeq in control and KDM4A knockout samples where “n” represents biological replicate of oocytes or embryos.

(b) Number of significantly deregulated genes in KDM4A knockout oocytes and MZ mutant embryos from (a) according to DeSeq2 algorithm that fits RNA-Seq count data to a negative binomial generalized linear model (GLM) and performs a two-sided Wald test on the coefficients. p-values were adjusted for multiple comparisons using the Benjamini-Hochberg

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

procedure. Data were filtered for log2 expression fold change of 1 and adjusted p-value <0.05.

(c) Bubble plot presenting overlap of all differentially expressed genes in DeSeq2 (adjusted p-value < 0.05) with DBTMEE transcriptome categories. Bubble plot sizes show –log10 adjusted p-values derived from a hypergeometric test for over-representation with p-values corrected for multiple testing (Benjamini-Hochberg procedure). Bubble color shows the log2 ratio of observed gene expression changes compared to randomly expected frequencies.

(d) All significantly deregulated genes (left) (from panel c, adjusted p-value < 0.05) including heatmap of associated H3K9me3 changes across such genes in *Kdm4a*$^{-/-(MZ)}$ 2-cell embryos (middle) and oocyte inherited H3K9me3 changes at such genes (right). The ChIP signal is expressed as shown in a ratio with strength in FPKM.

(e) Bee swarm plot quantifying H3K9me3 read abundance at gene bodies and TSS of significantly deregulated genes (n=2063 genes) shown in (d). Statistical significance was calculated using a Wilcoxon-Mann-Whitney test and Bonferroni-corrected for multiple testing and p-value plotted. n represents the number of differentially regulated genes corresponding to 837, 521, 609, and 96 genes for the groups “<-1”, “-1-0”, “0-1”, and “>1”, respectively.

(f) Tracks of ChIP-Seq signal for H3K4me3 and H3K9me3 in oocytes, H3K9me3 in 2-cell embryos at two genes with transcript directionality (red arrowhead) and bdH3K4me3 (green) shown. Highlighted area shows H3K9me3 change (left panel). FPKM values of individual 2-cell embryos and adjusted p-value (p<0.05) from DeSeq2 shown (two-sided Wald test). n represents the number of embryos (right panel). Violin plots show data distribution with median and quartiles. Statistical source data are provided in source data fig. 3.

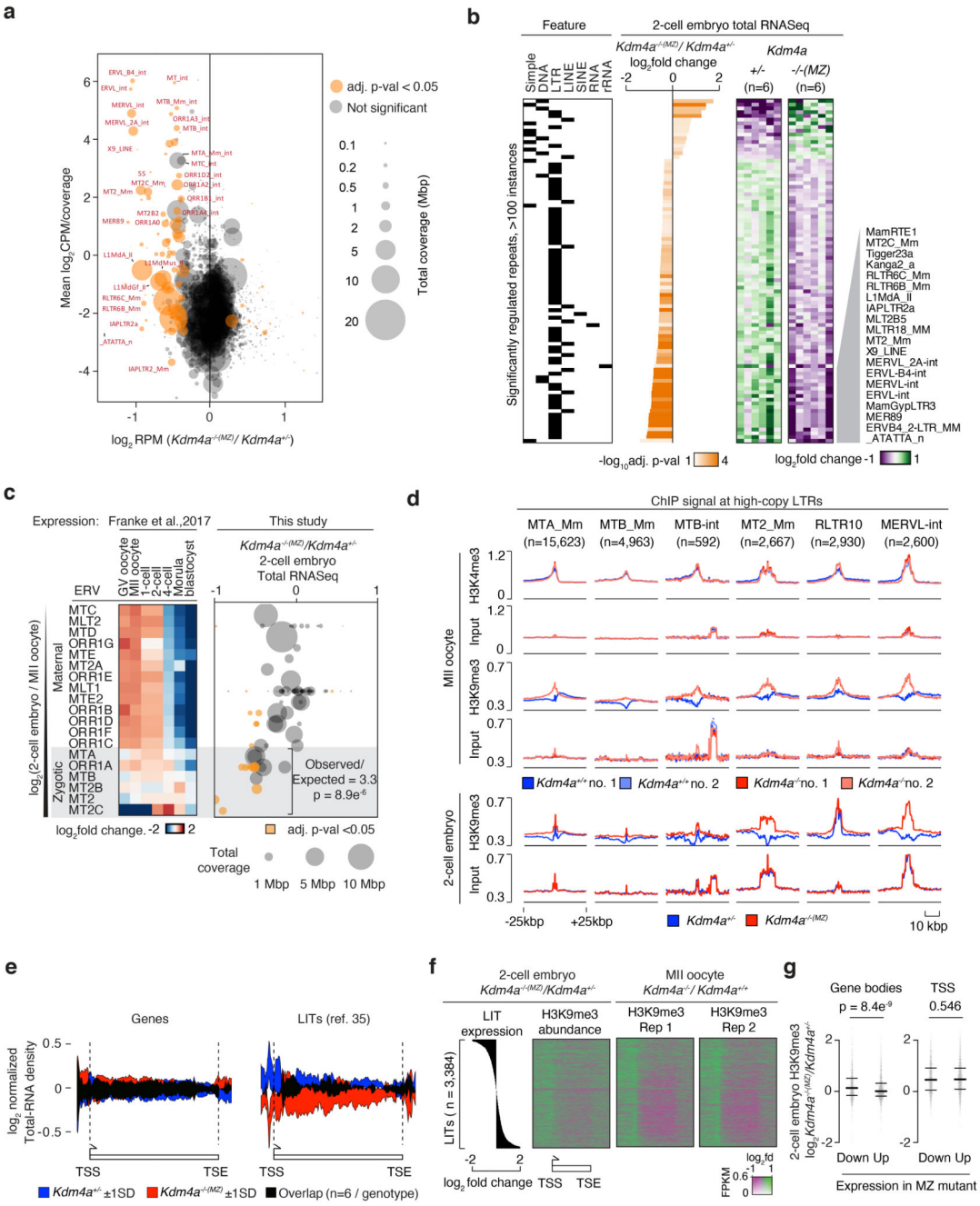


Figure 4. Aberrant H3K9me3 negatively impacts expression of a subset of high-copy LTRs and LTR initiated transcripts

(a) Volcano plot of significantly-deregulated repeats (orange) in *Kdm4a*$^{-/-(MZ)}$ 2-cell embryos. n represents biological replicates. Dot areas represent repeat coverage across the genome. Repeats with ≥6 non-zero read-count samples and ≥200 reads were fitted to a negative binomial generalized log-linear model and coefficients tested using a two-sided gene-wise likelihood ratio test. p-values were adjusted for multiple testing in all analyses throughout Figure 4 (Benjamini-Hochberg).

Europe PMC Funders Author Manuscripts

(b) Classification ('feature') of significantly deregulated (middle panel) repeats in *Kdm4a*$^{-/-(MZ)}$ 2-cell embryos and a heatmap of aggregated read counts normalized to average across all samples (right panel) for those repeats in individual replicates. n represents biological replicates.

(c) Analysis of LTR expression in *Kdm4a*$^{-/-(MZ)}$ 2-cell embryos compared to published expression of LTRs [34] in oocytes and pre-implantation embryos. Transcripts are ranked as predominantly maternal or zygotic based on expression dominance (left panel). Significantly expressed LTR subgroups (orange) as identified in (a) and (b) (right panel). Significance testing (χ-test) of the number of down-regulated zygotic LTR subgroups (n=10) compared to by random based on all significantly down-regulated groups of LTRs within the set (n=17).

(d) Average H3K4me3 and H3K9me3 enrichments in MII oocytes and 2-cell embryos at prominent long terminal repeat (LTR) types of ERVs. n represents the number of repeat instances in each group of LTRs.

(e) Total RNA-Seq normalized read coverage in *Kdm4a*$^{+/-}$ and *Kdm4a*$^{-/-(MZ)}$ 2-cell embryos along genes (left) and 3384 LTR-initiated transcripts (right). X-axis reflects the read position along the feature and Y-axis values are normalized to the average of all replicates. n represents the number of RNA-seq samples.

(f) Differentially expressed LITs (n=3,384) in *Kdm4a*$^{-/-(MZ)}$ 2-cell embryos (left) with associated H3K9me3 changes in *Kdm4a*$^{-/-}$ oocytes and MZ mutant 2-cell embryos along the LIT feature. n represents the number of LITs.

(g) Bee-swarm plot of H3K9me3 abundance of down (n = 1,539) or up-regulated (n = 1,845) LITs. P-values from two-sided Wilcoxon-Mann-Whitney test. Bold and narrow lines illustrate median and quartile ranges, respectively. Statistical source data are provided in source data fig. 4.

Europe PMC Funders Author Manuscripts

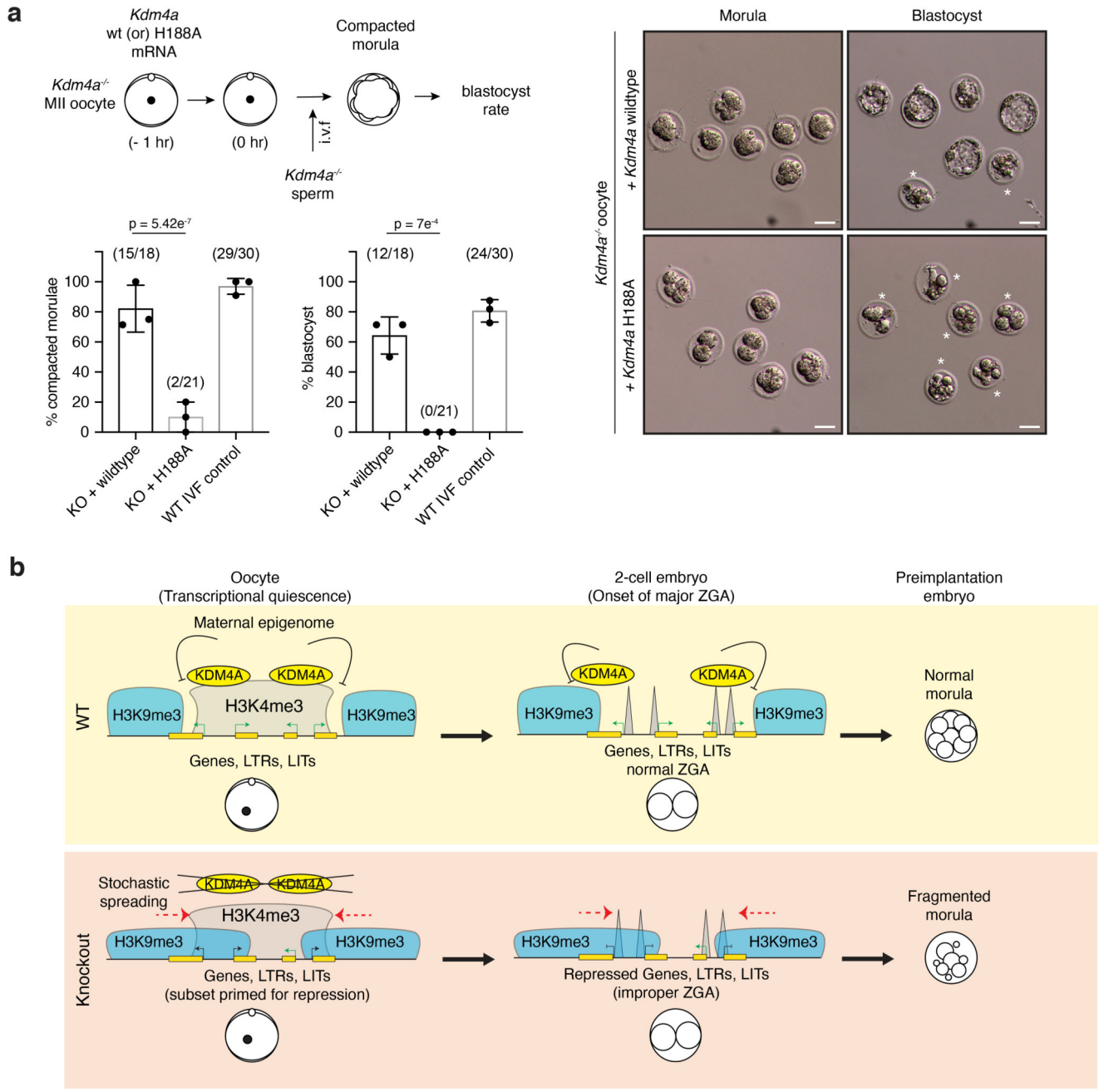


Figure 5. The catalytic activity of KDM4A is required for preimplantation development.

(a) Scheme showing the timeline of *Kdm4a* mRNA injection and fertilization (upper left panel). Percentage of compacted morula and cavitated blastocysts derived from successfully fertilized zygotes is plotted for three rounds of injections (lower left panel). Numbers in parentheses represent the fraction of embryos that reached the compacted morula or blastocyst stage compared to the total fertilized p-values were calculated using a two-tailed Fisher's exact test comparing wildtype and H188A. Error bars represent mean with S.D (n = 3 rounds of injections). Representative images (right panel) show developmental progression

Europe PMC Funders Author Manuscripts

of fertilized oocytes supplemented with wild-type and H188A mutant *Kdm4a* mRNA (Scale bar: 100 μm).

(b) Model depicting the role of KDM4A in the oocyte to preserve the maternal epigenome and its function in ensuring proper zygotic genome activation (ZGA) for normal preimplantation development. Statistical source data are provided in source data fig. 5.

Europe PMC Funders Author Manuscripts