Published in final edited form as:
Nature. 2018 March 15; 555(7696): 392–396. doi:10.1038/nature25964.

Epigenetic reprogramming enables the primordial germ cell-to-gonocyte transition

Peter W. S. Hill[1,2], **Harry G. Leitch**[#1,2], **Cristina E. Requena**[#1,2], **Zhiyi Sun**[#3], **Rachel Amouroux**[#1,2], **Monica Roman-Trufero**[1,2], **Malgorzata Borkowska**[1,2], **Jolyon Terragni**[3], **Romualdas Vaisvila**[3], **Sarah Linnett**[1,2], **Hakan Bagci**[1,2,5], **Gopuraja Dharmalingham**[1,2], **Vanja Haberle**[1,2,4], **Boris Lenhard**[1,2], **Yu Zheng**[3], **Sriharsa Pradhan**[3], and **Petra Hajkova**[1,2,†]

[1]MRC London Institute of Medical Sciences (LMS), Du Cane Road, London, UK, W12 0NN

[2]Institute of Clinical Sciences (ICS), Faculty of Medicine, Imperial College London, Du Cane Road, London, UK, W12 0NN

[3]New England Biolabs, Inc., 240 County Road, Ipswich, MA 01938, USA

[#] These authors contributed equally to this work.

Abstract

Gametes are highly specialised cells that can give rise to the next generation through their ability to generate a totipotent zygote. In mouse, germ cells are first specified in the developing embryo as primordial germ cells (PGCs) starting around embryonic day (E) 6.251 (Fig. 1a). Following subsequent migration into the developing gonad, PGCs undergo a wave of extensive epigenetic reprogramming at E10.5/E11.52–11, including genome-wide loss of 5-methylcytosine (5mC)2–5,7–11 (Fig. 1a). The underlying molecular mechanisms of this process have remained enigmatic leading to our inability to recapitulate this step of germline development *in vitro*12–14. Using an integrative approach, we show that this complex reprogramming process involves the coordinated interplay between promoter sequence characteristics, DNA (de)methylation, Polycomb (PRC1) complex and both DNA demethylation-dependent and -independent functions of Tet1 to enable the activation of a critical set of germline reprogramming responsive (GRR) genes involved in gamete generation and meiosis. Our results also unexpectedly reveal a role for Tet1 in safeguarding but

[†]Correspondence: petra.hajkova@lms.mrc.ac.uk, Tel: +44 (0)20 83838264.
[4]Research Institute of Molecular Pathology (IMP), Vienna Biocenter (VBC), Dr Bohr-Gasse 7, 1030 Vienna, Austria
[5]Medical Research Council Laboratory of Molecular Biology, Francis Crick Avenue, Cambridge CB2 0QH, UK

Author Contributions

P.H. and P.W.S.H. conceived the study; P.W.S.H. performed the experiments and analysed the data; H.G.L. carried out mESC experiments with the help of M.B. and M.R.T.; C.E.R. generated *Tet1*-KO *Dnmt*-TKO mESC line with the help of H.B.; R.A. carried out LC/MS experiments and analysed the data with the help of S.L.; J.T. made AbaSeq libraries with support from Y.Z.; computational analysis was carried out by P.W.S.H with the help of Z.S., G.D., V.H., and B.L.; R.V. performed experiments; P.W.S.H and P.H. wrote the manuscript with assistance from S.P. and B.L.

Author information

The authors declare no conflict of interest or competing financial interest.

Data availability

NGS data reported in this paper is tabulated in SI Tables 1-4 and archived on GEO (Accession: GSE76973). All other data are available from P.H. upon request.

Europe PMC Funders Author Manuscripts

not driving DNA demethylation in gonadal PGCs. Collectively, our work uncovers a fundamental biological role for gonadal germline reprogramming and identifies the epigenetic principles of the PGC-to-gonocyte transition that will be instructive towards recapitulating complete gametogenesis *in vitro*.

In order to address the potential role and underlying molecular mechanisms of gonadal germline reprogramming, we first set out to investigate the dynamics of and relationship between 5mC and 5-hydroxymethylcytosine (5hmC), which has previously been implicated in DNA demethylation in PGCs3,6,9–11. We did this quantitatively and at single base resolution using liquid chromatography/mass spectrometry (LC/MS) coupled with whole genome bisulphite sequencing (WGBS, Extended Data Fig. 1a) and AbaSeq15 (Extended Data Fig. 1b-e). WGBS provides information regarding combined levels of 5mC and 5hmC16, while AbaSeq15 enables robust site-specific quantification and accurate comparison of 5hmC levels genome-wide within a given sample and between samples when combined with LC/MS (see Methods, Extended Data Fig. 1b-e).

By LC/MS we observed that global levels of genomic 5mC remain stable between migratory (E9.5) and early gonadal (E10.5) PGCs, followed by a significant reduction between E10.5 and E11.5 and much more limited DNA demethylation between E11.5 and E13.5 (Fig. 1b). With respect to 5hmC, LC/MS analysis surprisingly revealed that global levels in PGCs are lower than those in mouse embryonic stem cells (mESCs) grown in serum-containing culture conditions (Fig. 1b). Furthermore, the global 5hmC levels in PGCs are relatively constant between E9.5 and E13.5, with a slight decrease in females starting at E12.5 (Fig. 1b). Importantly, 5hmC levels are consistently an order of magnitude lower than either total 5mC levels at E10.5 or the amount of 5mC lost between E10.5 and E11.5 (Fig. 1b-c), documenting that DNA demethylation is globally not accompanied by a reciprocal increase in 5hmC levels, as has previously been suggested3,17 (Extended Data Fig. 5a).

Consistent with our LC/MS measurements, WGBS analysis revealed near complete loss of combined 5mC/5hmC between E10.5 and E11.5 at features within uniquely mapped regions of the genome, with limited further DNA demethylation observed between E11.5 and E12.5 (Extended Data Fig. 3a). Loss of DNA methylation was also observed at consensus repeat sequences, although some repetitive elements such as LINE-1A and ERV-IAP retrotransposons retained comparatively high levels of combined 5mC/5hmC in E12.5 PGCs, as previously suggested8 (Extended Data Fig. 3b). Detailed analysis of 5hmC localisation by AbaSeq in E10.5 PGCs revealed that, although global levels are lower (Fig. 1b), 5hmC localisation in PGCs is remarkably similar to that of serum-grown mESCs, even at imprint control regions (ICRs; Extended Data Fig. 2a,b,f). Overall, 5hmC was enriched at putative active enhancers, present in intergenic regions and gene bodies, depleted at promoters, and absent on the vast majority of CpG islands (Extended Data Fig. 2b-f). With respect to transcription (SI Table 7), both 5mC and 5hmC at promoter regions show an inverse relationship with gene expression levels (Extended Data Fig. 2c). Within gene bodies, 5mC and 5hmC are clearly enriched at expressed genes compared to genes without detectable expression, although a non-linear relationship with gene expression is observed

Europe PMC Funders Author Manuscripts

for 5hmC while combined 5mC/5hmC levels show a clear positive correlation (Extended Data Fig. 2c).

Detailed analysis of 5hmC patterns across examined developmental stages uncovered that the majority of 5hmC is lost from uniquely mapped regions of the genome and re-localised to repetitive elements (Fig. 1d, Extended Data Fig. 3a-b). This relocalisation was also clearly evident by immunofluorescence staining (Fig. 1e). Our data thus shows that both 5mC and 5hmC are lost in PGCs throughout the uniquely mapped regions of the genome, although with different kinetics with 5hmC showing a more gradual decrease (Extended Data Fig. 4b). However, this was not consistent with passive dilution of 5hmC through cell divisions3 as demonstrated by poor Pearson and Spearman correlations between stages (Extended Data Fig. 4a, 5a). To the contrary, we conclude that 5hmC is a dynamic mark in PGCs.

We next explored the relationship between 5hmC deposition and DNA demethylation in gonadal PGCs between E10.5 and E12.5 for all initially methylated 2kb windows (i.e. min. 20% methylation at E10.5). DNA demethylation involving a 5hmC intermediate predicts a direct correlation between 5hmC appearance and 5mC loss (Extended Data Fig. 5a-b). To our surprise, we observed no correlation between either the total or relative 5hmC levels at E10.5 or E11.5 and the extent that the combined 5mC/5hmC levels decrease between these stages (Extended Data Fig. 4c-f). However, for all initially methylated 2kb windows, we did observe a negative correlation between the relative 5hmC level and the combined 5mC/5hmC levels at E11.5 (Extended Data Fig. 4g). Thus, 5hmC represents a much higher proportion of combined 5mC/5hmC at regions that are newly hypomethylated at E11.5, regardless of their original DNA methylation levels. Although 5hmC-depleted regions contain slightly more 5mC at E11.5 than regions enriched for 5hmC, sequences depleted of 5hmC in both E10.5 and E11.5 PGCs still undergo considerable DNA demethylation between these two stages (Extended Data Fig. 4h-i), indicating that the presence of detectable 5hmC is not a prerequisite for 5mC loss in gonadal PGCs. Our observations thus implicate involvement of 5hmC in the regulation of the post-DNA demethylation locus-specific 5mC levels in germ cells rather than in the initial wave of global DNA demethylation (Extended Data Fig. 5c).

To expand on this observation, we used a previously published *Tet1*-KO mouse model18 (Extended Data Fig. 6a-c). Initial LC/MS analysis revealed that loss of Tet1 leads to approximately 50% reduction in global 5hmC levels in E10.5 *Tet1*-KO germ cells (Fig. 2c). In agreement with the high level of Tet1 expression at E12.53,9,11 (Extended Data Fig. 6a-c), LC/MS analysis confirmed that Tet1 represents the primary 5mC oxygenase in demethylated PGCs, with approximately 85% decrease in global 5hmC levels observed in E14.5 *Tet1*-KO germ cells (Fig. 2a,c). Importantly, the genome of both *Tet1*-KO and wild type PGCs reached near-complete depletion of 5mC by E13.5 (Fig. 2b,d), highlighting that Tet1-mediated 5mC oxidation is not directly responsible for the bulk of DNA demethylation in gonadal PGCs.

In support of our LC/MS measurements, only a limited number of differentially methylated regions were detected in E14.5 *Tet1*-KO PGCs by reduced representation bisulphite sequencing (RRBS) (Fig. 2e). Intriguingly, these regions initially undergo extensive DNA

demethylation in both *Tet1*-KO and wild type PGCs, followed by a subsequent increase in 5mC levels specifically in *Tet1*-KO PGCs between E12.5 and E14.5 (Fig. 2e). In contrast, 5mC levels remain stable and/or undergo slight further reduction between these stages in wild type germ cells (Fig. 2e). The same DNA demethylation/remethylation kinetics were also observed at the few examples of previously reported9,10 germline gene promoters and ICRs that were found hypermethylated in E14.5 *Tet1*-KO PGCs by RRBS (Extended Data Fig. 6d-e, SI Table 8). Although significant enrichment of 5mC is indeed observed at the *Dazl* promoter by targeted bisulphite sequencing in demethylated PGCs, the extent of hypermethylation observed at the *Peg3* and *IG-DMR* ICRs is in fact very limited (Extended Data Fig. 6f-g). Furthermore, for all three regions, very few clones retained full methylation while a number of clones had heterogeneous methylation patterns consistent with a stochastic failure to remove aberrant residual/*de novo* DNA methylation in *Tet1*-KO PGCs (Extended Data Fig. 6f-g).

We next analysed the observed 5mC and 5hmC dynamics in combination with RNA-Seq datasets derived from E10.5-E14.5 PGCs (Extended Data Fig. 7). Initial clustering analysis of all genes based on their promoter DNA methylation dynamics revealed that, while most promoters become completely demethylated, there is a small subset of transcriptionally silenced promoters that retain high levels of 5mC/5hmC during global DNA demethylation (cluster 2, Extended Data Fig. 7a). These promoters significantly overlap with LINE1 and LTR (p-values = 9.5×10^{-24} and 7.2×10^{-83} respectively, hypergeometric test) containing endogenous retroviruses that are likely to determine this epigenetic status (Extended Data Fig. 3b). Overall, although high levels of promoter 5mC and 5hmC are associated with transcriptional repression in E10.5 pre-reprogramming PGCs, loss of these marks does not generally result in transcriptional activation (Extended Data Fig. 7a).

As the influence of 5mC on transcriptional activity of a gene has been shown in mammals to be highly dependent on promoter CpG content19, we performed clustering analysis specifically at genes with either high-CpG (HCPs), intermediate-CpG (ICPs) or low-CpG (LCPs) promoters19 (Fig. 3a and Extended Data Fig. 7b-c). Interestingly, this yielded a group of HCP genes that became DNA demethylated during the course of germline epigenetic reprogramming, and showed progressive transcriptional activation (cluster 3; Fig. 3a). Differential expression analysis confirmed that these genes show a significant enrichment among all genes up-regulated concurrent with epigenetic reprogramming in PGCs (p-value < 0.001, hypergeometric test), with 45 genes commonly activated in both sexes (Fig. 3a-c). Considering their promoter methylation dynamics and timing of their activation, we termed these 45 genes germline reprogramming responsive (GRR) genes (Fig. 3c). Interestingly, GRR genes shows significant enrichment for factors involved in gamete generation and meiosis, including *Dazl*, *Sycp1-3*, *Mael*, *Hormad1*, and *Rad51c* (Fig. 3c and SI Table 5 and 7).

Considering that GRR genes (n=45) constituted less than 25% of the entire subset of HCP genes that undergo DNA demethylation (n=226; Fig. 3a-c), DNA demethylation is likely an important factor for transcriptional activation of methylated HCPs, with other factors additionally necessary. Indeed, GRR gene promoters showed both exceptionally high CpG density and 5hmC levels compared to other methylated and demethylating HCPs (Extended

Data Fig. 9a-b). We also noted that, unusually for promoters, 5hmC levels transiently increased at GRR gene promoters in PGCs immediately following the major wave of DNA demethylation (Extended Data Fig. 3a, 9b). In addition, and in agreement with their high CpG density and 5hmC levels20,21, GRR gene promoters have been shown to be bound by Tet1 in both mESCs21 and PGCs9 (Fig. 3b).

The observed binding of Tet1 is functionally relevant, as the extent of GRR gene upregulation is considerably lower in *Tet1*-KO PGCs (Fig. 4a, Extended Data Fig. 9c, SI Table 7). Although GRR gene promoters undergo normal DNA demethylation in the absence of Tet1 by E12.5, they show slight hypermethylation later in E14.5 *Tet1*-KO PGCs (Fig. 4b). However, this limited DNA hypermethylation shows only weak correlation with the decreased expression (Extended Data Fig. 9e). Furthermore, lower expression of GRR genes in *Tet1*-KO germ cells is already apparent at E12.5 in the absence of any methylation differences (Fig. 4a-b, Extended Data Fig. 9e), suggesting that Tet1 is potentially acting as a transcriptional regulator outside its role in 5mC removal21,22. In addition to GRR genes, transposable elements (TE) show accumulation of 5hmC during gonadal epigenetic reprogramming (Extended Data Fig. 3b, 8). Alongside reduction in DNA methylation, some TEs show transcriptional activation concurrent with epigenetic reprogramming, especially from evolutionary young retrotransposons (Extended Data Fig. 8, SI Table 9). Interestingly, the lack of Tet1 appears to also reduce the extent of transcriptional activation of normally activated TEs (Extended Data Fig. 8, SI Table 9).

To further mechanistically probe the causal relationship between epigenetic reprogramming and GRR gene activation, we turned to an *in vitro* model. Serum-grown mESCs represented an ideal system, as these cells are not germ line-restricted yet have highly similar epigenetic modifications at GRR gene promoters to what is observed *in vivo* in pre-reprogramming gonadal PGCs (Extended Data Fig. 10a-d). Consistent with what we observed *in vivo*, promoter DNA demethylation represents a dominant epigenetic reprogramming event for GRR gene activation also *in vitro*. *Dnmt*-TKO23 mESCs display increased expression of GRR genes (Fig. 4c). However, even in the complete absence of DNA methylation, this is crucially dependent on the presence of Tet1 as *Tet1*-KO *Dnmt*-TKO mESCs fail to activate GRR genes as a group (Fig. 4c, Extended Data Fig. 9f).

Although these *in vitro* observations clearly supported our *in vivo* data with respect to the roles of 5mC and Tet1, the extent to which GRR genes were up-regulated in *Dnmt*-TKO mESCs (Fig. 4c) or in E10.5 PGCs that have undergone precocious DNA demethylation by conditional deletion of *Dnmt1* (*Dnmt1*-CKO)24 (Extended Data Fig. 9d) was relatively mild. We thus hypothesised that other factors, including potentially other epigenetic barriers, may regulate GRR gene expression. In this context, gonadal epigenetic reprogramming has been previously linked with erasure of epigenetic information at various distinct levels4,25, with removal of Polycomb Repressive Complex 1 (PRC1) previously shown to coordinate the timing of meiosis initiation in DNA demethylated E11.5/E12.5 PGCs26. Remarkably, genes aberrantly up-regulated following PRC1 deletion in PGCs show a significant enrichment for GRR genes (Extended Data Fig. 11a) and promoters of GRR genes in serum-grown mESCs are enriched for Ring1b binding and H2AK119ub (Extended Data Fig. 10a,e,f). In view of this, we simultaneously abolished both DNA methylation and PRC1

activity using highly specific chemical inhibition of PRC1 in the context of *Dnmt*-TKO mESCs to test the role of both DNA methylation and PRC1 in GRR gene regulation, thus mimicking gonadal epigenetic reprogramming. Culture of mESCs with **PRT416527** resulted in significant inhibition of PRC1-mediated H2A ubiquitination after only 6h of culture (Extended Data Fig. 11b). Dual inhibition of 5mC/PRC1 repression strikingly resulted in the activation of 33 out of 45 GRR genes with 25 and 10 genes activated upon the sole inhibition of either 5mC or PRC1 repression, respectively (Fig. 4d, Extended Data Fig. 11c). Combined, these observations show that gonadal epigenetic reprogramming entails a composite erasure of epigenetic systems4,25 to potentiate the expression of GRR genes.

Our study has identified a set of germline reprogramming responsive (GRR) genes crucial for the correct progression of gametogenesis. These genes have unique promoter sequence characteristics, with high levels of both 5mC and 5hmC, and are targets of Tet1 and PRC1. We show that combined loss of DNA methylation and PRC1 repression is uniquely required for GRR gene activation, with this epigenetically poised state further requiring Tet1 to potentiate both full and efficient activation. Tet1 appears to be particularly important in female PGCs9, which initiate meiotic prophase soon after completion of epigenetic reprogramming, thus posing a requirement on the timely high expression of these genes. Importantly, although we observed slight hypermethylation at GRR gene promoters in E14.5 *Tet1*-KO PGCs, our study clearly documents that Tet1 stimulates transcription of GRR genes also via a DNA demethylation-independent mechanism21,22. In this context, previous studies have shown that Tet1 recruits OGT to gene promoters22, thus facilitating deposition of H3K4me3 via SET1/COMPASS28 leading to transcriptional activation. In further support, GRR gene promoters in mESCs are marked by low but detectable H3K4me3, the levels of which are significantly decreased in the absence of Tet1 without changes in DNA methylation (Fig. 4b, Extended Data Fig. 10g). Tet1 may additionally potentiate transcription through regulation of 5mC/5hmC levels at non-promoter *cis*-elements, such as enhancers. Last, but not least, our study shows that Tet1 is not directly involved in initiation of global DNA demethylation during epigenetic reprogramming in gonadal PGCs, but rather we define a critical role for Tet1 in the subsequent removal of aberrant residual and/or *de novo* DNA methylation (Extended Data Fig. 12). This is reminiscent of the role of Tet3-driven 5mC oxidation in protection against *de novo* DNA methylation during zygotic DNA demethylation29, suggesting that global reprogramming events require efficient protection from *de novo* DNA methylation following removal of 5mC to stabilise the newly acquired epigenetic state. Collectively, our study reinforces the idea that gonadal epigenetic reprogramming entails complex erasure of epigenetic information4 and suggests that a central function of this process is to ascertain the timely and efficient activation of GRR genes, thus enabling progression towards gametogenesis (Extended Data Fig. 12).

Methods

Mice

All animal experiments were carried out under and in accordance with a UK Home Office Project Licence in a Home-Office designated facility. Except for direct comparison with *Tet1*-KO PGCs, wild type PGCs were isolated from embryos produced by crossing outbred

MF1 females with mixed background GOF18ΔPE-EGFP5 transgenic males. The sex of embryos from E12.5 onwards was determined by visual inspection of the gonads. For study of *Tet1*-KO PGCs, the *Tet1* knockout mouse strain (B6;129S4-Tet1$^{tm1.1Jae}$ /J)18 was purchased from Jackson Laboratory and bred onto the GOF18ΔPE-EGFP5 transgenic mouse line. Wild type and *Tet1*- KO PGCs were isolated from embryos produced from crosses between *Tet1*-heterozygous GOF18ΔPE-EGFP-homozygous females and males. For genotyping of embryos produced by crossing *Tet1*-heterozygous GOF18ΔPE-EGFP-homozygous males and females, PCR was always carried out twice using two different sets of primers (see below) to confirm exon 4 deletion. The sex of the embryos from E12.5 onwards was determined by visual inspection of gonads and additionally confirmed by PCR for *Sry*. In all cases, the mating is timed in the way that appearance of a vaginal plug at noon is defined as E0.5.

The following genotyping primers were used in this study: TCAGGGAGCTCATGGAGACTA (*Tet1* forward primer 1); AACTGATTCCCTTCGTGCAG (*Tet1* forward primer 2); TTAAAGCATGGGTGGGAGTC (*Tet1* reverse primer); TTGTCTAGAGAGCATGGAGGGCCATGTCAA (*Sry* forward primer); CCACTCCTCTGTGACACTTTAGCCCTCCGA (*Sry* reverse primer).

PGC isolation by flow cytometry

PGC isolation was carried out as previously described4. Briefly, the embryonic trunk (E10.5) or genital ridge (E11.5-E14.5) was digested at 37°C for 3 min using 0.05% Trypsin-EDTA (1x) (Gibco) or TrypLE Express (Thermo). Enzymatic digestion was followed by neutralization with DMEM/F-12 (Gibco) containing 15% foetal bovine serum (Gibco) and manual dissociation by pipetting. Following centrifugation, cells were re-suspended in DMEM/F-12 supplemented with hyaluronidase (300 μg/ml; Sigma), and a single cell suspension was generated by manual pipetting. Following centrifugation, cells were re-suspended in ice-cold PBS supplemented with poly-vinyl alcohol (10 μg/ml) and EGTA (0.4 mg/ml, Sigma). GFP positive cells were isolated using an Aria IIu (BD Bioscience) or Aria III (BD Bioscience) flow cytometer and sorted into ice cold PBS supplemented with poly-vinyl alcohol (10 μg/ml) and EGTA (0.4 mg/ml, Sigma).

Generation of *Tet1*-KO *Dnmt*-TKO mESCs

Tet1-KO *Dnmt*-TKO mESC line was generated by CRISPR/Cas9-mediated genome editing. pX330 (Addgene, #42230) with the sgRNA targeting *Tet1*31 (GGCTGCTGTCAGGGAGCTCA) was co-transfected with a reporter GFP plasmid in 5 x 106 *Dnmt*-TKO mESCs23 using Lipofectamine 3000. The day after, GFP positive cells were sorted by FACS (BD FACS Aria III) in a 96-well plate. Cells were cultured for a week before being frozen down and extracting gDNA. Colonies were screened for mutations using surveyor assay (Surveyor Mutation Detection Kit from IDT, and Taq DNA polymerase from Qiagen). *Tet1*-KO *Dnmt*-TKO mESC selected clone was further analysed by genotype sequencing, which confirmed the presence of a frameshift mutation. Loss of Tet1 was verified by RNA-Seq and western blot. The following primers were used for genotype sequencing and surveyor assay: 5’ TTGTTCTCTCCTCTGACTGC 3’ and 5’ TGATTGATCAAATAGGCCTGC 3’.

mESC cell culture

J1 (wild type), *Dnmt*-TKO23 and *Tet1*-KO *Dnmt*-TKO mESCs were cultured in FCS/LIF medium without feeders on 0.1% gelatin. FCS/LIF medium consists of GMEM (Gibco) supplemented with 10% FCS, 0.1 mM MEM nonessential amino acids, 2 mM l-glutamine, 1 mM sodium pyruvate, 0.1 mM 2-mercaptoethanol and mouse LIF (ESGRO, Millipore). For inhibitor experiments, mESCs were plated at a density of 1.5 $x10^{4}/cm^{2}$ and left overnight. The next morning medium was exchanged for FCS/LIF medium containing either 50μM PRC1 inhibitor PRT4165 (Ismail et al., 2013) or DMSO control and cells pelleted at the indicated time for analysis.

AbaSeq library preparation

Total DNA was isolated from 10,000 sorted PGCs using the QIAamp DNA Micro Kit (Qiagen). AbaSeq libraries for 5hmC profiling were constructed as previously described15. In brief, genomic DNA was glucosylated, then digested by AbaSI enzyme (NEB). Biotinylated P1 adapters were ligated onto the AbaSI digested DNA then fragmented using a Covaris S2 sonicator (Covaris), following the manufacturer's instructions. Sheared P1-ligated DNA was then captured by mixing with Dynabeads MyOne Streptavidin C1 beads (Life Technologies) according to the manufacturer's specifications. End repair and dA-tailing were carried out on the beads by using the NEBNext End Repair Module (NEB) and the NEBNext dA-tailing Module (NEB) at 20°C and 37 °C respectively for 30 min. P2 adapters were ligated to the random sheared ends of the dA-tailed DNA. Finally, the entire DNA was amplified using the Phusion DNA polymerase (NEB) with the addition of 300 nM forward primer (PCR_I) and 300 nM reverse primers (PCR_IIpe) for 16 cycles. The libraries were purified using AMPure XP beads (Beckman-Coulter) and sequenced on the Illumina HiSeq 2000 instrument.

Whole genome bisulphite sequencing (WGBS) library preparation

Total DNA was isolated from 10,000 sorted PGCs using the QIAamp DNA Micro Kit (Qiagen). In some cases, unmethylated λ phage DNA (Promega) was spiked in following DNA isolation to assess bisulphite conversion rate. DNA was fragmented using a Covaris S2 sonicator (Covaris), as per manufacturer's instructions. Libraries were made following the NEBNext Library Prep protocol, with methylated adaptors and the following modifications: following adaptor ligation, bisulphite conversion was carried out using the Imprint Modification Kit (Sigma); and PCR enrichment was carried out for 16 cycles using the NEXTflex Bisulphite-Seq Kit for Illumina Sequencing (Bioo Scientific) master mix and the NEBNext Library Prep universal and index primers (NEB). The libraries were purified by AMPure XP beads (Beckman-Coulter). Libraries were sequenced on the Illumina HiSeq 2000 or 2500 instrument.

Reduced representation bisulphite sequencing (RRBS) library preparation

Total DNA from FACS-sorted PGCs isolated from individual *Tet1*-KO or wild type embryos was isolated using ZR-Duet DNA-RNA MiniPrep kit (Zymo), and DNA from between two to six embryos (equivalent to 1,000 to 8,000 cells) of the same genotype, stage and sex was pooled and concentrated to 26 μL final volume using the Savant SpeedVac Concentrator

(Thermo) and following the manufacturer's instructions. Genomic DNA was digested by 20 units of MspI enzyme (NEB) in NEB buffer 2 at 37°C for 3 hrs, and digested DNA was purified using AMPure XP beads (Beckman-Coulter). Libraries were made following the NEBNext Ultra DNA Library Prep protocol with methylated adaptors and the following modifications: following adaptor ligation, bisulphite conversion was carried out using the Imprint Modification Kit (Sigma); and PCR enrichment was carried out for 18 cycles using the KAPA Uracil$^+$ DNA polymerase master mix (KAPA Biosystems) and the NEBNext Library Prep universal and index primers (NEB). The libraries were purified by AMPure XP beads (Beckman-Coulter). Pooled libraries were sequenced on the Illumina HiSeq 2500 instrument, using the 'dark sequencing' protocol, as previously described32.

RNA-Seq library preparation

For study of *Tet1*-KO PGCs, total RNA from sorted PGCs isolated from individual *Tet1*-KO or wild type embryos was isolated using ZR-Duet DNA-RNA MiniPrep kit (Zymo), and RNA from between two to six embryos (equivalent to 1,000 to 8,000 cells) of the same genotype, stage and sex was pooled and concentrated to 6 μL final volume using the RNA Clean and Concentrator 5 kit (Zymo). For study of wild type PGCs isolated from embryos produced by crossing MF1 females with GOF18ΔPE-EGFP males, total RNA from 600-1,000 sorted E10.5 PGCs was isolated using the Nucleospin RNA XS kit (Macherey-Nagel). cDNA synthesis and amplification (15 cycles) was performed with the SMARTer Ultra Low Input RNA kit (Clontech) using between 100 pg and 3ng total RNA and following the manufacturer's instructions. The amplified cDNA was fragmented by a Covaris S2 sonicator (Covaris) and following the manufacturer's instructions. Sheared cDNA was converted to sequencing libraries using the NEBNext DNA Library Prep kit (NEB), following the manufacturer's instructions and using 15 cycles of amplification. For study of mESCs, total RNA was isolated using ZR-Duet DNA-RNA MiniPrep kit (Zymo). cDNA synthesis and library prep was performed starting with 500 ng total RNA following manufacturer's instructions using the NEBNext Ultra Library Prep Kit (NEB) and the NEBNext Poly(A) mRNA Magnetic Isolation Module (NEB). All libraries were purified by AMPure XP beads (Beckman-Coulter) and sequenced on the Illumina HiSeq 2500 instrument.

Bioinformatics

Whole genome bisulphite sequencing (WGBS) and Tet-assisted bisulphite sequencing (TAB-Seq) alignment and downstream analysis—Raw reads were first trimmed using Trim Galore (version 0.3.1) with the --paired --trim1 options. Alignments were carried out to the mouse genome (mm9, NCBI build 37) with Bismark (version 0.13.0) with the -n 1 parameter; where appropriate, the λ phage genome was added as an extra chromosome. Aligned reads were deduplicated with deduplicate_bismark. Where appropriate, the bisulphite conversion rate was computed using reads aligned to the λ phage genome and using the to-mr script (parameters: -m bismark) and bsrate script (paramters: -N) of Methpipe (version 3.3.1). CpG methylation calls were extracted from the deduplicated mapping output using the Bismark methylation extractor. The number of methylated and unmethylated cytosines in a CpG context was extracted using bismark2bedGraph and coverage2cytosine. Symmetric CpGs were merged with custom R script. For all downstream

analysis, only symmetric CpGs with minimum 8× coverage were used. All WGBS analysis was carried out on data from merged biological replicates. For assessing DNA modification levels at specific repetitive elements, Bismark (version 0.14.4) was used to map all reads from each data set against consensus sequences constructed from Repbase with the -n 1 parameter set. CpG methylation calls were extracted from the mapping output using the Bismark methylation extractor (version 0.14.4).

The mapBed function of BEDtools (version 2.24.0) was used to compute the combined 5mC/5hmC level for the following genomic features: 1) all 2kb windows (containing a minimum 4 symmetric CpGs); 2) gene promoters (defined as Ensembl 67 gene start sites -1kB/+500bp); 3) gene bodies (defined as the region contained within Ensembl 67 gene start and gene end sites); 4) putative active enhancers in day 6 PGCLCs33; 5) imprint control regions; 6) CpG islands (UCSC); 7) intergenic regions. For metagene plots, a genomic feature was divided into equally sized bins using BEDtools (version 2.24.0), including: 1) gene bodies (defined as the region contained within Ensembl 67 gene start and gene end sites) +/- 0.5 * gene body length (100 bins); 2) putative active enhancers in day 6 PGCLCs 33 +/- 1 * putative active enhancer length (90 bins); and 3) CpG islands (UCSC) +/- 1 * CpG island length (90 bins). In all cases, the combined 5mC/5hmC level was expressed as the mean of individual CpG sites.

For k-means clustering of the combined mean 5mC/5hmC levels, high CpG (HCP), intermediate CpG (ICP) and low CpG (LCP) promoters, as defined using the same parameters as previously published19,34. Briefly, LCPs contain no 500-bp window with a CpG ratio > 0.45; HCPs contain at least one 500-bp window with a CpG ratio >0.65 and GC content > 55%; ICPs do not meet the previous criteria.

For determining locus-specific methylation levels in wild type mESCs grown in serum-containing media, raw WGBS reads were downloaded from GSE4851930 and processed as above. TAB-Seq reads for E14 mESCs were downloaded from GSE3617335 and processed as above, with the exception that only symmetric CpGs with minimum 12× coverage were used.

AbaSeq alignment and downstream analysis—For the uniquely mappable part of the genome, AbaSeq reads were processed as previously described15. In brief, raw sequencing reads were trimmed for adaptor sequences and low quality bases using Trim Galore. The trimmed reads were mapped to the mouse genome (mm9, NCBI build 37) using Bowtie (version 0.12.8) with parameters -n 1 -l 25 --best --strata -m 1. Calling of 5hmC was based on the recognition sequence and cleavage pattern of the AbaSI enzyme (5′-$CN_{11–13}\downarrow N_{9–10}$G-3′/3′-$GN_{9–10}\downarrow N_{11–13}$C-5′) using custom Perl scripts. For assessing relative enrichment of 5hmC at repetitive elements and non-repetitive elements, AbaSeq alignments were divided into two groups: unique (single best alignment) and ambiguous (map to multiple locations with equal alignment score). Both groups were then mapped to the repetitive elements defined by the RepeatMasker track of mm9 (UCSC Genome Browser) separately. For comparison with 5hmC levels in mESCs, AbaSeq reads were downloaded from GSE4289815 and aligned in the same way.

For quantification of relative 5hmC levels at symmetric CpGs in the uniquely mapped part of the genome, the number of counts per symmetric CpG for a given sample were normalised to the combined number of uniquely mapped and ambiguously mapped reads for a given library, and then further multiplied by a stage-specific normalisation factor based on the mean 5hmC level for each stage as computed by LC/MS (E14 ESC = 1.64; E10.5 = 1.0; E11.5 = 1.13; E12.5F = 0.76; E12.5M = 1.0). All symmetric CpGs falling within genomic intervals blacklisted by the mouse (mm9) ENCODE project were excluded from all further downstream analysis. Unless stated otherwise, all AbaSeq analysis was carried out on data from merged biological replicates.

The mapBed function of BEDtools (version 2.24.0) was used to compute the 5hmC level for the same genomic features as was carried out with WGBS datasets (see above). In all cases, the 5hmC level was expressed as the mean of individual CpG sites.

To identify 5hmC enriched or depleted regions in E10.5 and E11.5 PGCs, the mm9 genome was first divided into 2kb windows (minimum 4 symmetric CpGs) and the mean 5hmC level for each window was computed using BEDtools (version 2.24.0). To determine the significance of 5hmC enrichment in each 2kB window, upper-tail (to determine 5hmC enriched regions) or lower-tail (to determine 5hmC depleted regions) Poisson probability p-values were computed using ppois(x, λ), where x is the observed 5hmC mean value for each 2kb window and λ is the mean of 5hmC mean values for all 2kb windows at E10.5. Benjamini-Hochberg correction was then applied to correct for multiple testing, giving a final adjusted upper-tail and lower-tail p-value for each 2kb window. Windows with adjusted upper-tail p-value < 0.05 were considered relatively enriched for 5hmC while windows with adjusted lower-tail p-value < 0.05 were considered relatively depleted for 5hmC.

For assessing relative enrichment of 5hmC at specific repetitive elements, Bowtie was used to map all reads from each data set against consensus sequences constructed from Repbase with parameters -n 1 -M 1 --strata –best. The number of reads mapped to each sequence within a given sample was first normalised to the library size of that particular sample, and then normalised to both a stage-specific normalisation factor based on the mean 5hmC level for each stage as computed by LC/MS (E10.5 = 1.0; E11.5 = 1.13; E12.5F = 0.76; E12.5M = 1.0) and the mean proportion of reads mapped to a given sequence in E10.5 PGCs.

Reduced representation bisulphite sequencing (RRBS) alignment and downstream analysis—Raw RRBS reads were first trimmed using Trim Galore (version 0.3.1) with --rrbs parameter. Alignments were carried out to the mouse genome (mm9, NCBI build 37) with Bismark (version 0.13.0) with the -n 1 parameter. CpG methylation calls were extracted from the mapping output using the Bismark methylation extractor (version 0.13.0). The number of methylated and unmethylated cytosines in a CpG context was extracted using bismark2bedGraph.

RnBeads (version 1.0.0) and RnBeads.mm9 (version 0.99.0) were used to identify differentially methylated regions between two test groups for the following genomic features, with filtering.missing.value.quantile set to 0.95 and filtering.missing.coverage.threshold set to 8: 1) all 2kb windows (containing a minimum 4

symmetric CpGs); 2) gene promoters (defined as Ensembl 67 gene start sites -1kB/+500bp); and 3) imprint control regions (mm9 genome). The following was extracted from the output of RnBeads: 1) the mean methylation level for each group (i.e. stage, sex and/or genotype) for each commonly covered test region; 2) the difference in methylation means between two groups for each commonly covered test region; and 3) the p-value representing the significance of the difference in methylation means between two groups for each commonly covered test region. Differentially methylated regions were identified as regions with a p-value < 0.05 and a difference in methylation means between two groups greater than 10%.

For assessing DNA modification levels at specific repetitive elements, Bismark (version 0.14.4) was used to map all reads from each data set against consensus sequences constructed from Repbase with the -n 1 parameter set. CpG methylation calls were extracted from the mapping output using the Bismark methylation extractor (version 0.14.4). The number of methylated and unmethylated cytosines in a CpG context were extract using bismark2bedGraph and coverage2cytosine. Differentially methylated consensus repeats were identified as regions with a p-value < 0.05 (as computed by two-sided Student's t-test) and a difference in methylation means between two groups greater than 10%.

hMeDIP alignment and downstream analysis—Raw hMeDIP-Seq and input reads for E14 mESCs were downloaded from GSE2850036 and aligned to the mouse genome (mm9, NCBI build 37) with Bowtie (version 0.12.8) with parameters -n 2 -l 25 -m 1. BEDtools multicov was used to identify the number of hMeDIP and input reads overlapping each 2kB window (containing a minimum 4 symmetric CpGs). Final 5hmC levels for each 2kB window were determined by first normalising the number of overlapping hMeDIP reads (normalised to library size) by the number of overlapping input reads (normalised to library size) and then by dividing this value by the number of symmetric CpGs contained within the 2kB window.

ChIP-Seq alignment and downstream analysis—For putative active enhancer calling, raw ChIP-Seq reads for H3K4me3, H3K27me3 and H3K27Ac in day 6 PGC-like cells (PGCLCs) were downloaded from GSE6020433 and raw ChIP-Seq reads for H3K4me3, H3K27me3, H3K4me1 and H3K27Ac in wild type mESCs were downloaded from GSE4851930. Reads were aligned to the mouse genome (mm9, NCBI build 37) with Bowtie (version 0.12.8 or version 1.0.0) with parameters -n 2 -l 25 -m 1and -C where appropriate. Subsequent ChIP-Seq analysis was carried out on data from merged biological replicates. To identify putative active enhancers, we first generated an 8-state chromatin model using ChromHMM. Putative active enhancers were defined as all regions not overlapping any potential promoter regions (Ensembl 67 gene start sites -1kB/+500bp) and contained within the ($H3K27Ac^{+}/H3K4me3^{-}/H3K27me3^{-}$) chromatin state in day 6 PGCLCs or ($H3K4me1^{+}/H3K27Ac^{+}/H3K4me3^{-}/H3K27me3^{-}$) in wild type mESCs.

For analysis of epigenetic modifications and modifiers around transcription start sites (Ensembl 67): raw ChIP-Seq reads for: Tet1 binding in wild type serum-grown mESCs was downloaded from GSE2484321; H2AK119Ub1 levels in wild type serum-grown mESCs were downloaded from GSE3452037; Ring1b binding in wild type serum-grown mESCs were downloaded from ERP00557538; and for H3K4me3 in wild type and *Tet1*-KO serum-

grown mESCs were downloaded from GSE4851930. Reads were aligned to the mouse genome (mm9, NCBI build 37) with Bowtie (version 0.12.8 or version 1.0.0) with parameters -n 2 -l 25 -m 1. Subsequent ChIP-Seq analysis was carried out on data from merged biological replicates. For computing ChIP-Seq signal around transcription start sites (TSS), the genomic interval around the Ensembl 67 gene start sites +/- 5kB (or 2kB) was divided into 100 (or 40) equally sized bins using BEDtools makewindows. BEDtools multicov was then used to compute the number of test and control reads overlapping each bin. The total number of test and control reads per bin for each sample were normalised to the appropriate library size, and fold enrichment for each bin was determined by dividing the number of normalised ChIP-Seq test sample reads by the number of normalised ChIP-Seq control sample reads. For computing ChIP-Seq signal at gene promoters, the genomic interval around the Ensembl 67 gene start sites +500bp/-1kB was

RNA-Seq alignment and downstream analysis—For study of *Tet1*-KO and *Tet1*-WT PGCs, Illumina and Smart-seq adapters from the sequencing reads were first trimmed using Trimmomatic. For other RNA-Seq libraries, fastq files generated from output of next generation sequencing were used directly for alignment. RNA-Seq reads were aligned to the mouse genome (mm9, NCBI build 37) with Bowtie (version 0.12.8) and Tophat (version 2.0.2) with options -N 2 --b2-very-sensitive --b2-L 25. Annotations from Ensembl Gene version 67 were used as gene model with Tophat. Read counts per annotated gene were computed using HTSeq (version 0.5.3p9) and expression level of each gene was quantified by computing the number of fragments detected per kilobase per million of reads (FPKM) using custom R script. Genes were assigned to an expression level bin based on the mean FPKM values of the two biological replicates. Differential expression analysis was performed using DESeq2 (version 1.6.3), and genes with an adjusted p-value < 0.05 were considered differentially expressed. For determining gene expression levels in wild type and *Dnmt1*-conditional knockout and matched wild type E10.5 PGCs, raw RNA-Seq reads were downloaded from GSE7493824 and processed as above.

HCPs methylated and demethylating in PGCs during epigenetic reprogramming (cluster 3, Figure 4A) were ranked based on the significance of activation (α) between gene expression in E10.5 and E14.5 PGCs (Figure 4B). In the case where β represents the directionality of fold change (i.e. if log2(FC) < 0, β = -1; else β = +1) and γ represents the adjusted p-value as computed by DESeq2, $\alpha = \beta \times (1 - \gamma)$. For comparing expression levels of the GRR gene set in 1) wild type, *Dnmt*-TKO, and *Tet1*-KO *Dnmt*-TKO mESCs (Fig. 6A); 2) in wild type + 6h DMSO treatment, *Dnmt*-TKO + 6h DMSO treatment, wild type + 6h PRT4165 treatment, *Dnmt*-TKO + 6h PRT4165 treatment (Fig. 6C); 3) *Tet1*-KO E14.5 PGCs against wild type E14.5 PGCs (Fig. 5B); or 3) *Dnmt1*-CKO E10.5 PGCs against wild type E10.5 PGCs (Extended Data Fig. 9G) pairwise differential expression analysis was initially carried out by DESeq2 for each condition against each other condition. For each pairwise differential expression test, each gene was assigned a statistic α, where if β represents the $\log_2$(FC) and γ represents the adjusted p-value as computed by DESeq2, $\alpha = \beta \times (1 - \gamma)$. The ranked gene list based on α was subsequently used for gene set enrichment analysis (GSEA) for testing general up or down-regulation of the combined GRR gene sets and GSEA hallmark gene sets, and GSEA FWER-adjusted p-values were subsequently used. For

overlap between germline reprogramming responsive genes and genes repressed by PRC1 in PGCs (Figure 6B), the list of genes called up-regulated in E11.5 and/or E12.5 PRC1-KO PGCs was downloaded from 26.

For classification of GRR genes (Extended Data Fig. 10, SI Table 5), pairwise differential expression analysis was first carried out. 5mC-reprogramming dependent GRR genes were defined as genes: 1) up-regulated in *Dnmt*-TKO vs WT, *Dnmt*-TKO+PRC1 inhibitor vs WT, and *Dnmt*-TKO+PRC1 inhibitor vs WT+PRC1 inhibitor; and 2) not up-regulated in WT +PRC1 inhibitor vs WT. PRC1-reprogramming dependent GRR genes were defined as genes: 1) up-regulated in WT+PRC1 inhibitor vs WT, *Dnmt*-TKO+PRC1 inhibitor vs WT, and *Dnmt*-TKO+PRC1 inhibitor vs *Dnmt*-TKO; and 2) not up-regulated in *Dnmt*-TKO vs WT. 5mC/PRC1-reprogramming dependent GRR genes were defined as genes either: 1) up-regulated in WT+PRC1 inhibitor vs WT, *Dnmt*-TKO vs WT, *Dnmt*-TKO+PRC1 inhibitor vs WT, *Dnmt*-TKO+PRC1 inhibitor vs *Dnmt*-TKO, and *Dnmt*-TKO+PRC1 inhibitor vs WT+ PRC1 inhibitor; or 2) up-regulated in *Dnmt*-TKO+PRC1 inhibitor vs WT, *Dnmt*-TKO +PRC1 inhibitor vs *Dnmt*-TKO, and *Dnmt*-TKO+PRC1 inhibitor vs WT+ PRC1 inhibitor, and not up-regulated in WT+PRC1 inhibitor vs WT and *Dnmt*-TKO vs WT. 5mC/PRC1 reprogramming independent or insufficient GRR genes were defined as genes not-upregulated in *Dnmt*-TKO vs WT, *Dnmt*-TKO+PRC1 inhibitor vs WT, and *Dnmt*-TKO +PRC1 inhibitor vs WT+PRC1 inhibitor, and WT+PRC1 inhibitor vs WT. Genes that do fall into one of these five classes were described as low confidence classification (l.c.c.) genes.

Tet1 and 5mC/5hmC detection by immunofluorescence

The embryonic trunk (E10.5) or genital ridge (E12.5/E13.5) was first fixed in 2% PFA (in PBS) for 30 min at 4°C. Following fixation, tissue was washed in PBS three times for 10 min and then incubated in 15% sucrose in PBS overnight. After rinsing with 1% BSA in PBS the following day, the tissue was embedded in OCT Embedding Matrix (Thermo Scientific Raymond Lamb) and frozen using liquid nitrogen. Samples were subsequently stored at -80°C. A Leica CM 1950 cryostat was used to cut 10 μm sections from the frozen embedded tissue. Sections were settled on poly-lysine slides (Thermo Scientific) and post-fixed with 2% PFA in PBS for 3 minutes.

For detection of Tet1, sections were washed three times for 5 min with PBS. After incubating for 30 min at room temperature in 1% BSA/PBS containing 0.1% Triton X-100, the sections were incubated with primary antibodies listed at 4°C overnight in the same buffer. Sections were subsequently washed three times in 1% BSA/PBS containing 0.1% Triton X-100 for 5 min and incubated with secondary antibodies in the same buffer for 1 hour in the dark at room temperature. Secondary antibody incubation was followed by three 5 min washes with PBS. DNA was then stained with DAPI (100 ng/ml). After a final wash in PBS for 10 min, the sections were mounted with Vectashield (Vector Laboratories).

For detection of 5hmC/5mC, sections were washed three times for 5 min with PBS. Post-fixed sections were first permeabilized for 30 min with 0.5% Triton X-100 (in 1% BSA/PBS), and subsequently treated with RNase A (10 mg/ml; Roche) in 1% BSA/PBS for 1 hour at 37°C. Following three 5 min washes with PBS, sections were incubated with 4N HCl for 10-20 min at 37°C to denature genomic DNA, followed by three 10 min washes with

PBS. After incubating for 30 min at room temperature in 1% BSA/PBS containing 0.1% Triton X-100, the sections were incubated with primary antibodies listed at 4°C overnight in the same buffer. Sections were subsequently washed three times in 1% BSA/PBS containing 0.1% Triton X-100 for 5 min and incubated secondary antibodies in the same buffer for 1 hour in the dark at room temperature. Secondary antibody incubation was followed by three 5 min washes with PBS. DNA was then stained with propidium iodide (PI) (0.25 mg/ml). After a final wash in PBS for 10 min, the sections were mounted with Vectashield (Vector Laboratories).

The following primary antibodies were used in this study: anti-SSEA1 (gifted by Dr P. Beverly via Dr G. Durcova Hills); anti-MVH (Abcam 27591 or Abcam 13840); anti-5hmC (Active motif 39791), anti-5mC (Diagenode C15200081-100); anti-Tet1 (GeneTex GTX125888); anti-GFP (Abcam 5450). The following secondary antibodies were used in this study: Alexa Fluor 647 Goat anti-Mouse IgM (Invitrogen A21238); Alexa Fluor 488 Goat anti-Rabbit IgG (Invitrogen A11008); Alexa Fluor 405 Goat anti-Mouse IgG 1:300 (Invitrogen A31553); Alexa Fluor 488 Goat anti-Mouse IgG 1:300 (Invitrogen A11001); Alexa Fluor 405 Goat anti-Rabbit IgG 1:300 (Invitrogen A31556); Alexa Fluor 568 Donkey anti-Rabbit IgG (Invitrogen A10042); Alexa Fluor 488 Donkey anti-Goat IgG (Invitrogen A11055).

Locus-specific bisulphite sequencing

Bisulphite treatment of genomic DNA was carried out using the Imprint DNA modification kit (Sigma). The following primers were used for the semi-nested amplification of the *Dazl* promoter: F1: GATTTTTGTTATTTTTTAGTTTTTTTAGGAT; F2: TTTATTTAAGTTATTATTTTAAAAATGGTATT; R: AGAAACAAGCTAGGCCAGCTGAGAGAATTCT. The following primers were used for the semi-nested amplification of the IG-DMR ICR: F1: GTGTTAAGGTATATTATGTTAGTGTTAGG; F2: ATATTATGTTAGTGTTAGGAAGGATTGTG; R: TACAACCCTTCCCTCACTCCAAAAATT. The following primers were used for the nested amplification of the *Peg3* ICR: F1: TTTTTAGATTTTGTTTGGGGGTTTTTAATA; F2: TTGATAATAGTAGTTTGATTGGTAGGGTGT; R1: AATCCCTATCACCTAAATAACATCCCTACA; R2: ATCTACAACCTTATCAATTACCCTTAAAAA. Methylation levels were assessed by QUMA, using default settings with duplicate bisulphite sequences excluded.

Mass spectrometry

Genomic DNA from between 100 and 2,000 FACS-sorted PGCs was extracted using ZR-Duet DNA/RNA Miniprep kit (Zymo Reasearch) following manufacturer instructions and eluted in LC/MS grade water. DNA was digested to nucleosides using a digestion enzyme mix provided by NEB. A dilution-series made with known amounts of synthetic nucleosides and the digested DNA were spiked with a similar amount of isotope-labelled nucleosides (provided by Dr T. Carell (LMU, Germany)) and separated on an Agilent RRHD Eclipse Plus C18 2.1 × 100 mm 1.8u column by using the UHPLC 1290 system (Agilent) and an Agilent 6490 triple quadrupole mass spectrometer. To calculate the quantity of individual

Europe PMC Funders Author Manuscripts

nucleosides, standard curves representing the ratio of unlabelled over isotope-labelled nucleosides were generated and used to convert the peak-area values to corresponding quantity. Threshold for quantification is a signal-to-noise (calculated with a peak-to-peak method) above 10.

Western blot

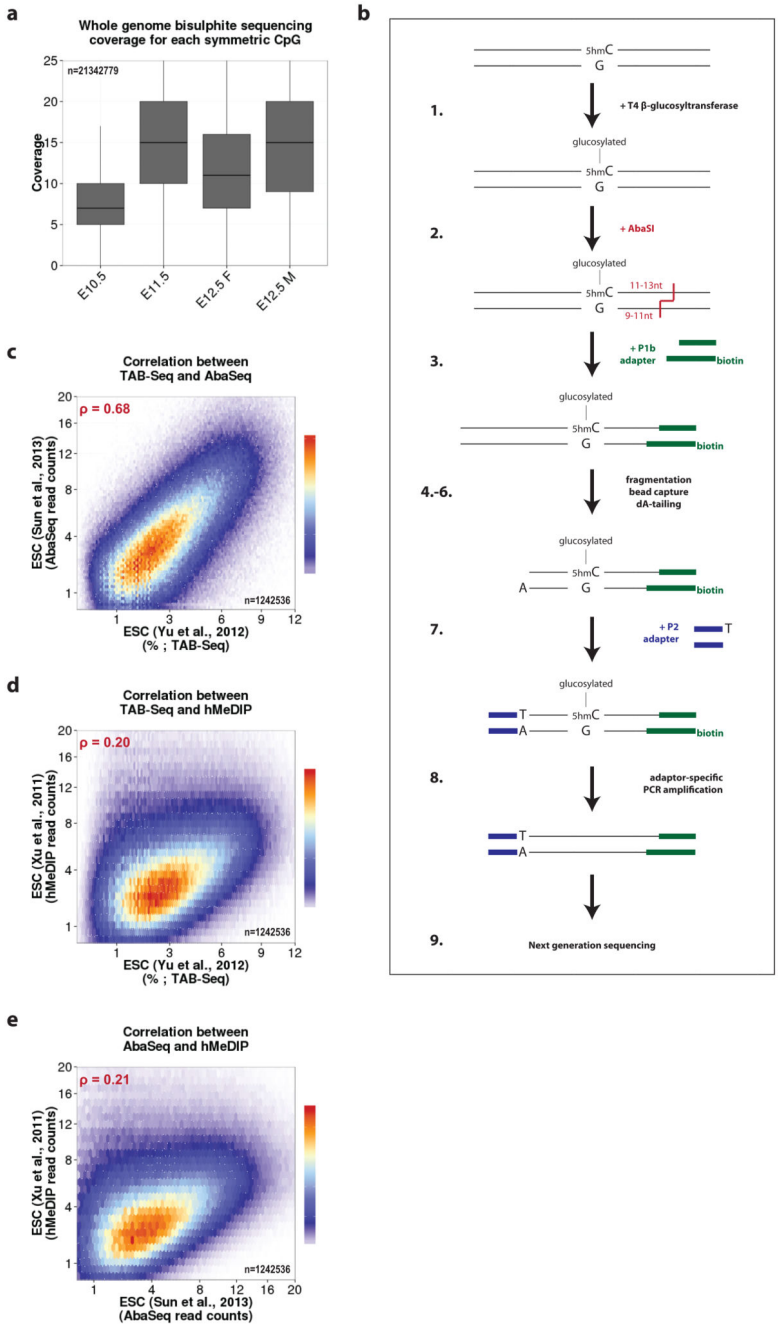
mESCs were lysed by sonication in RIPA buffer (150 mM sodium chloride, 1.0% Triton X-100, 0.5% sodium deoxychlorate, 0.1% sodium dodecylsulfate, 50 mM Tris pH 8.0) and protease-inhibitor cocktail (Roche, 11 697 498 001). Cell debris were removed by centrifugation at 14000 g 5 min 4°C. Protein was quantified using the BCA protein assay (Thermo, 23227). 2 μg (for H2A and H2Aub) or 20 μg (for Tet1) of each protein extract was loaded onto an 15% or 8% SDS polyacrylamide gel and transferred to a PVDF membrane after electrophoresis. Membranes were blocked with 5% BSA for 1 hour and then incubated overnight at 4°C with primary antibodies at the following dilutions: anti-H2A antibody (Abcam, 18255) 1:2000; anti-ubiquityl H2A antibody (Cell Signalling 8240) 1:2000; anti-Tet1 antibody [N1] (GeneTex GTX125888) 1:1000; anti-Lamin B antibody (C20) (Santa Cruz Biotechnologies, sc-6216) 1:10000. Donkey anti-rabbit IgG-HRP (Santa Cruz Biotechnologies, sc-2077) or donkey anti-goat IgG-HRP (Santa Cruz Biotechnologies, sc-2056) secondary antibody were incubated for 1h at room temperature. Blots were developed by using Luminata Crescendo Western HRP substrate (EMD Milipore).

Statistics and Reproducibility

All statistical tests are clearly described in the figure legends and/or in the Methods section, and exact p-values or adjusted p-values are given where possible. For WGBS data (Figs. 3a-b, and Extended Data Figs. 1a, 2c-e, 3a-b, 4, 7, 8), data is derived from cells from either n=1 (E10.5 PGC sample) or n=2 (all other samples) biological replicates, with each replicate from pooled embryos (E10.5: n=39 embryos/4 litters; E11.5: n=8 embryos/1 litter; E12.5M/F: n=4 embryos/1 litter). For AbaSeq data (Figs. 1d, 3a-b, and Extended Data Figs. 1c-e, 2a-f, 3a-b, 4, 7, 8, 9b), data is derived from cells from n=2 biological replicates, with each replicate from pooled embryos (E10.5: n=40 embryos/4 litters; E11.5: n=8 embryos/1 litter; E12.5M/F: n=4 embryos/1 litter). For RNA-Seq of mESCs, samples are derived from n=2 biological replicates corresponding to n=2 independently cultured samples from n=1 cell line. For PGC LC/MS, RNA-Seq and RRBS data, please see SI Table 6 for complete details regarding the number of embryos/litters from which samples were derived. Western blots (Extended Data Figs. 9f, 11b) were performed three times with similar results, and representative blots are shown. All immunostainings (Figs. 1e, 2a-b, Extended Data Fig. 6b) were performed twice with similar results and representative images are shown. Traditional bisulphite sequencing (Extended Data Figs. 6f-g) was carried out twice and a representative methylation profile is shown. For analysis of previously published WGBS (Extended Data Fig. 10a-b), TAB-Seq (Extended Data Fig. 1c-e), AbaSeq (Extended Data Figs. 1c-e, 2b, 10a, 10c) and ChIP-Seq (Fig. 3b, Extended Data Figs. 10a, 10c-g) datasets from mESCs (see Methods for accession numbers), other than H2Aub ChIP-Seq dataset (where n=1), biological replicates were analysed both combined (shown) and separately (not shown) to ensure reproducibility of analysis.

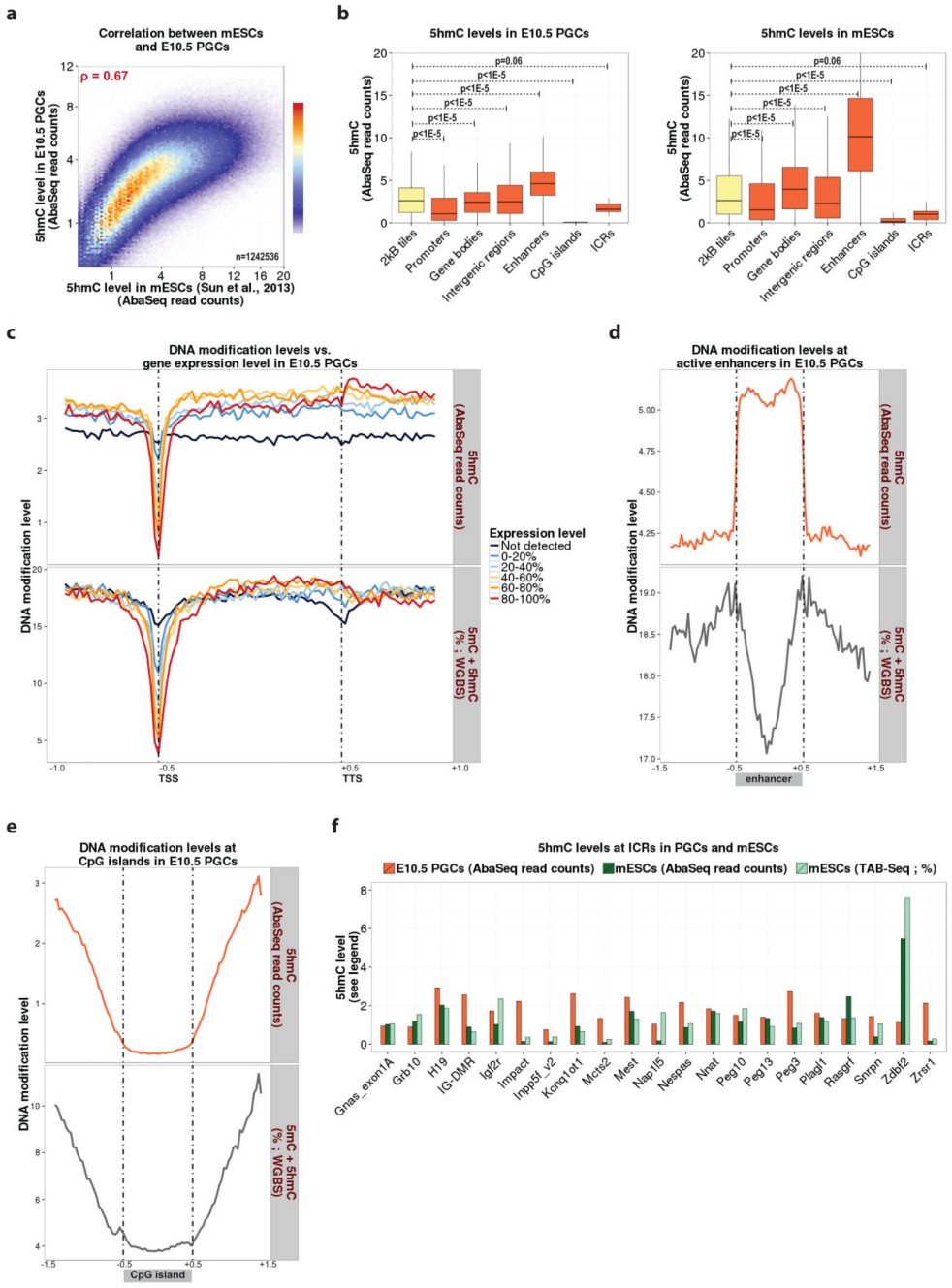
Europe PMC Funders Author Manuscripts

Extended Data



Extended Data Fig. 1. Characterisation of WGBS datasets and validation of AbaSeq method.
a) Distribution of WGBS coverage for each symmetric CpG. For boxplots, the upper and lower hinges correspond to the first and third quartiles, the middle line corresponds to the median, and the maxima and minima respectively correspond to the highest or lowest value within 1.5× the inter-quartile range. **b)** Overview of AbaSeq method15. **c-e)** Density heatmap showing correlation between 5hmC levels at all 2kB windows (minimum 4

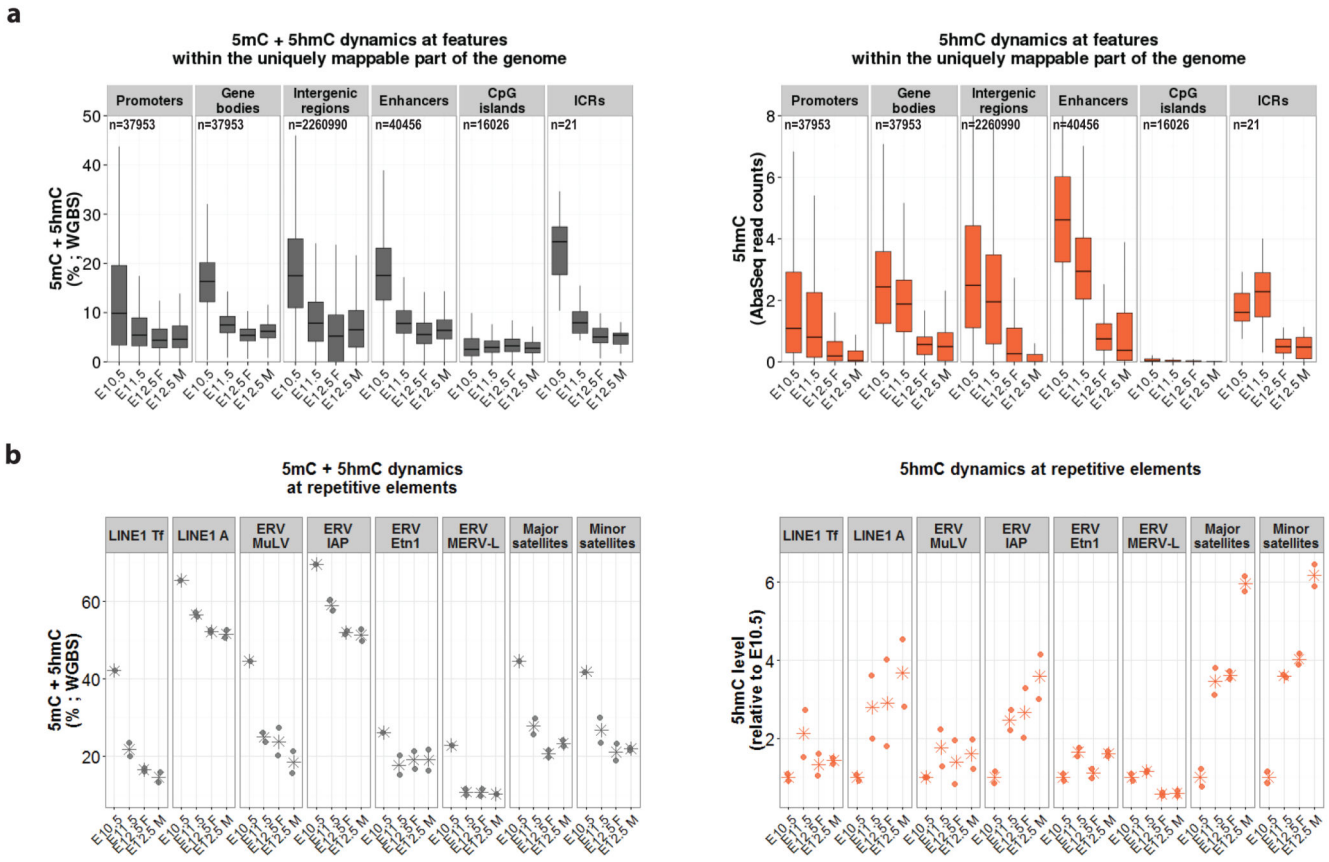
symmetric CpGs) in E14 mESCs as computed by: (c) TAB-Seq35 (x-axis) and AbaSeq15 (y-axis); (d) TAB-Seq35 (x-axis) and hMeDIP36 (y-axis); or (e) AbaSeq15 (x-axis) and hMeDIP36 (y-axis). For (c-e), the Pearson correlation coefficient (ρ) is shown. Specific details regarding sample sizes and how samples were collected are found in Statistics and Reproducibility section.



Extended Data Fig. 2. Further analysis of 5hmC levels in E10.5 PGCs.

Europe PMC Funders Author Manuscripts

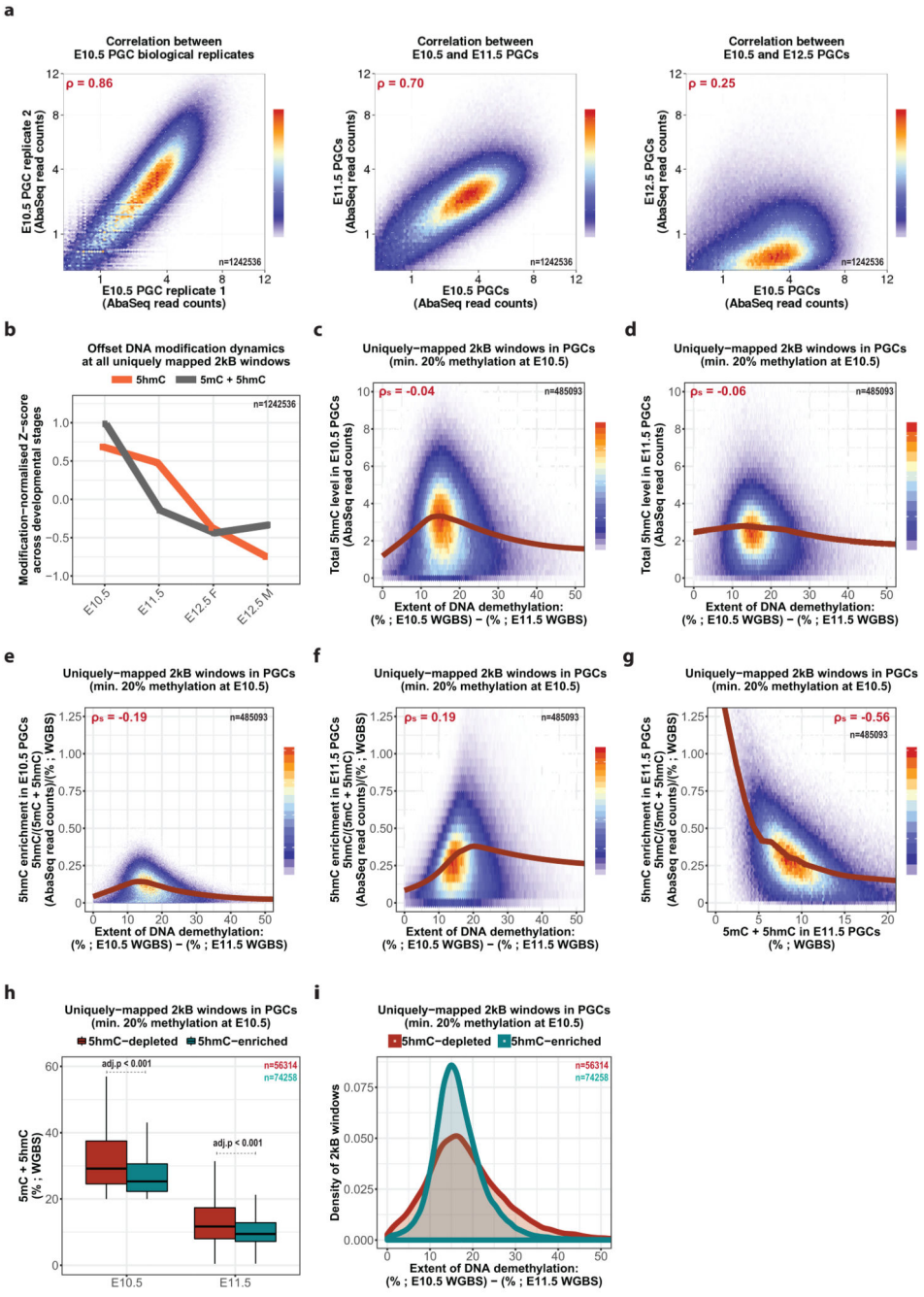
a) Density heatmap showing 5hmC levels per 2kB window (with minimum 4 CpGs) for E10.5 PGCs (y-axis) and E14 mESCs15 (y-axis). Pearson correlation coefficient (ρ) is shown. **b)** 5hmC levels (AbaSeq) at various regulatory elements in E10.5 PGCs (left) or E14 mESCs15. p-values based on ANOVA and Dunnett post hoc test. For boxplots, the upper and lower hinges correspond to the first and third quartiles, the middle line corresponds to the median, and the maxima and minima respectively correspond to the highest or lowest value within 1.5× the inter-quartile range. **c)** Metagene plot showing 5hmC levels (top panel, AbaSeq) and combined 5mC/5hmC levels (bottom panel, WGBS) in E10.5 PGCs across genes expressed at different levels in E10.5 PGCs. **d-e)** Metagene plot showing 5hmC levels (top panel, AbaSeq) and combined 5mC/5hmC levels (bottom panel, WGBS) in E10.5 PGCs across either CpG islands (d) or across putative active enhancers (e). **f)** Bar chart showing 5hmC levels at ICRs in E14 mESCs as determined by TAB-Seq35 (%; light green) or AbaSeq15 (read counts; dark green), or in E10.5 PGCs as determined by AbaSeq (read counts; orange). Specific details regarding sample sizes and how samples were collected are found in Statistics and Reproducibility section.



Extended Data Fig. 3. Further analysis of 5mC and 5hmC dynamics in PGCs.

a) Combined 5mC/5hmC (WGBS; left) or 5hmC (AbaSeq; right) levels at various features within the uniquely mapped part of the genome in PGCs between E10.5 and E12.5. The upper and lower hinges correspond to the first and third quartiles, the middle line corresponds to the median, and the maxima and minima respectively correspond to the

highest or lowest value within 1.5× the inter-quartile range. **b)** The combined 5mC/5hmC (WGBS; left) or 5hmC (AbaSeq; right) levels at various consensus repetitive elements in PGCs between E10.5 and E12.5. Asterisks refer to mean values. For specific details regarding sample sizes and how samples were collected, see the Statistics and Reproducibility section.



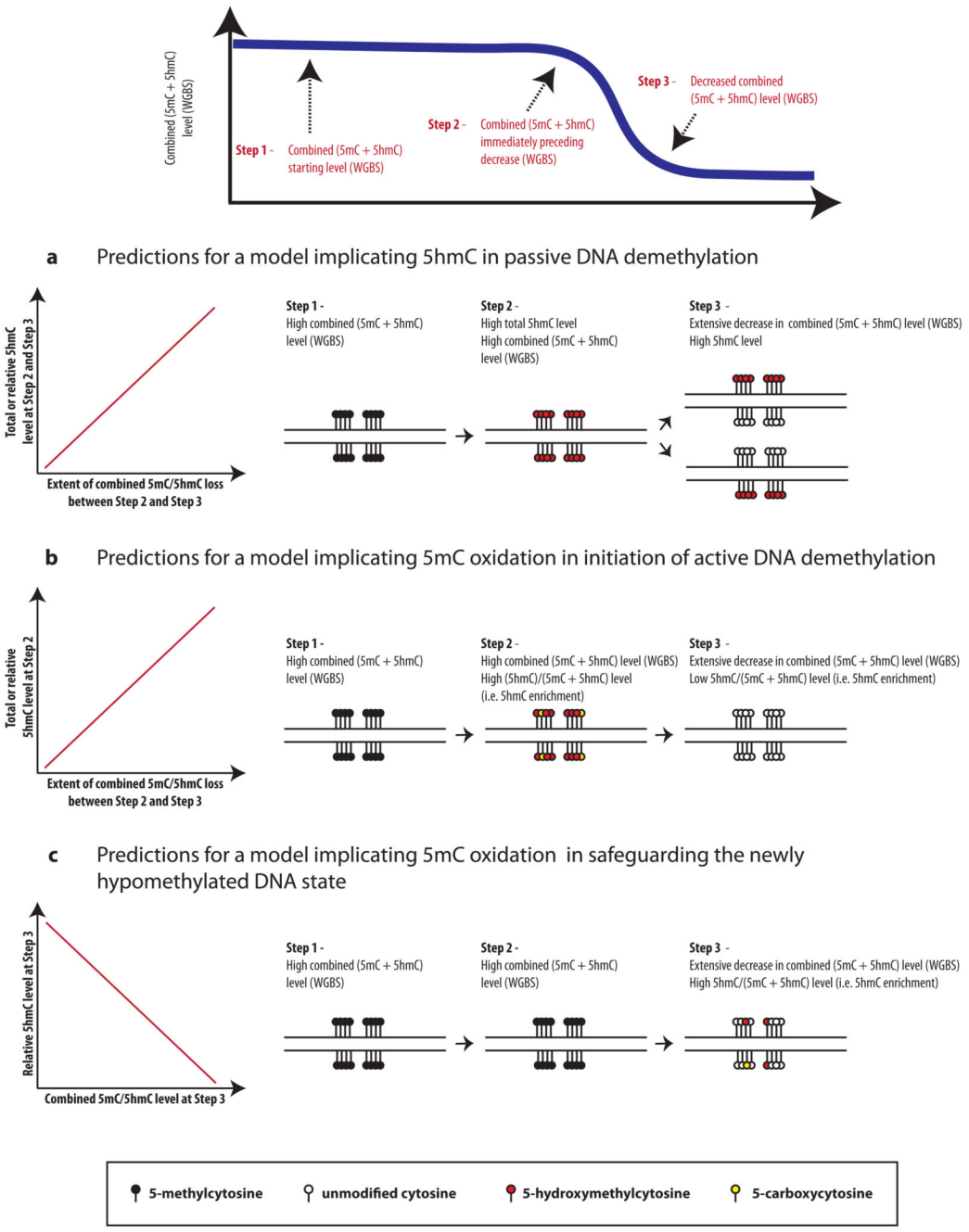
Extended Data Fig. 4. 5hmC is targeted to newly hypo-methylated regions following DNA demethylation in mouse gonadal PGCs (see also Extended Data Fig. 5).

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

a) Density heatmap showing Pearson correlation (ρ) between 5hmC levels for E10.5 biological replicates (left), for E10.5 and E11.5 PGCs (middle), and for E10.5 and E12.5 PGCs (right). **b)** Mean Z-scores depicting 5hmC (orange, AbaSeq) and combined 5mC/5hmC (grey, WGBS) levels for each stage normalised to the average level of either 5hmC (orange, AbaSeq) or combined 5mC/5hmC (grey, WGBS) across stages. Standard error of the mean is shown by too small to see. **c-f)** Density heatmap showing the correlation between the total (c,d; y-axis: AbaSeq read counts) or relative (e,f; y-axis: ratio of (AbaSeq read counts)/(%; WGBS)) 5hmC levels in E10.5 (c,e) or E11.5 (d,f) PGCs and the change in combined 5mC/5hmC levels in PGCs between these two stages (x-axis: %; WGBS) for all 2kB windows with a minimum 20% combined 5mC/5hmC in E10.5 PGCs. **g)** Density heatmap showing the correlation between the relative 5hmC levels in E11.5 PGCs (y-axis: ratio of (AbaSeq read counts)/(%; WGBS)) and the combined 5mC/5hmC level in E11.5 PGCs (x-axis: %; WGBS) for all 2kB windows with a minimum 20% combined 5mC/5hmC in E10.5 PGCs. **h)** Density plot showing the decrease in combined 5mC/5hmC levels in PGCs between E10.5 and E11.5 for 2kB windows with a minimum 20% total DNA modification in E10.5 PGCs that are either 1) enriched for total 5hmC levels at either E10.5 or E11.5 (green, upper-tail adj. Poisson p-value < 0.05), or 2) depleted of total 5hmC at both E10.5 and E11.5 (red, lower-tail adj. Poisson p-value < 0.05). **i)** Combined 5mC/5hmC levels in E10.5 and E11.5 PGCs for 2kB windows with a minimum 20% combined 5mC/5hmC in E10.5 PGCs that are either 1) enriched for total 5hmC levels at either E10.5 or E11.5 (green, upper-tail adj. Poisson p-value < 0.05), or 2) depleted of total 5hmC at both E10.5 and E11.5 (red, lower-tail adj. Poisson p-value < 0.05). For all boxplots, the upper and lower hinges correspond to the first and third quartiles, the middle line corresponds to the median, and the maxima and minima respectively correspond to the highest or lowest value within 1.5× the inter-quartile range. p-values are based on a two-sided Wilcoxon test. Note that for density heatmaps: 1) the Spearman correlation (ρ_S) is shown; and 2) the red line represents the smoothed mean as determined by a generalized additive model. Specific details regarding sample sizes and how samples were collected are found in Statistics and Reproducibility section.

Europe PMC Funders Author Manuscripts

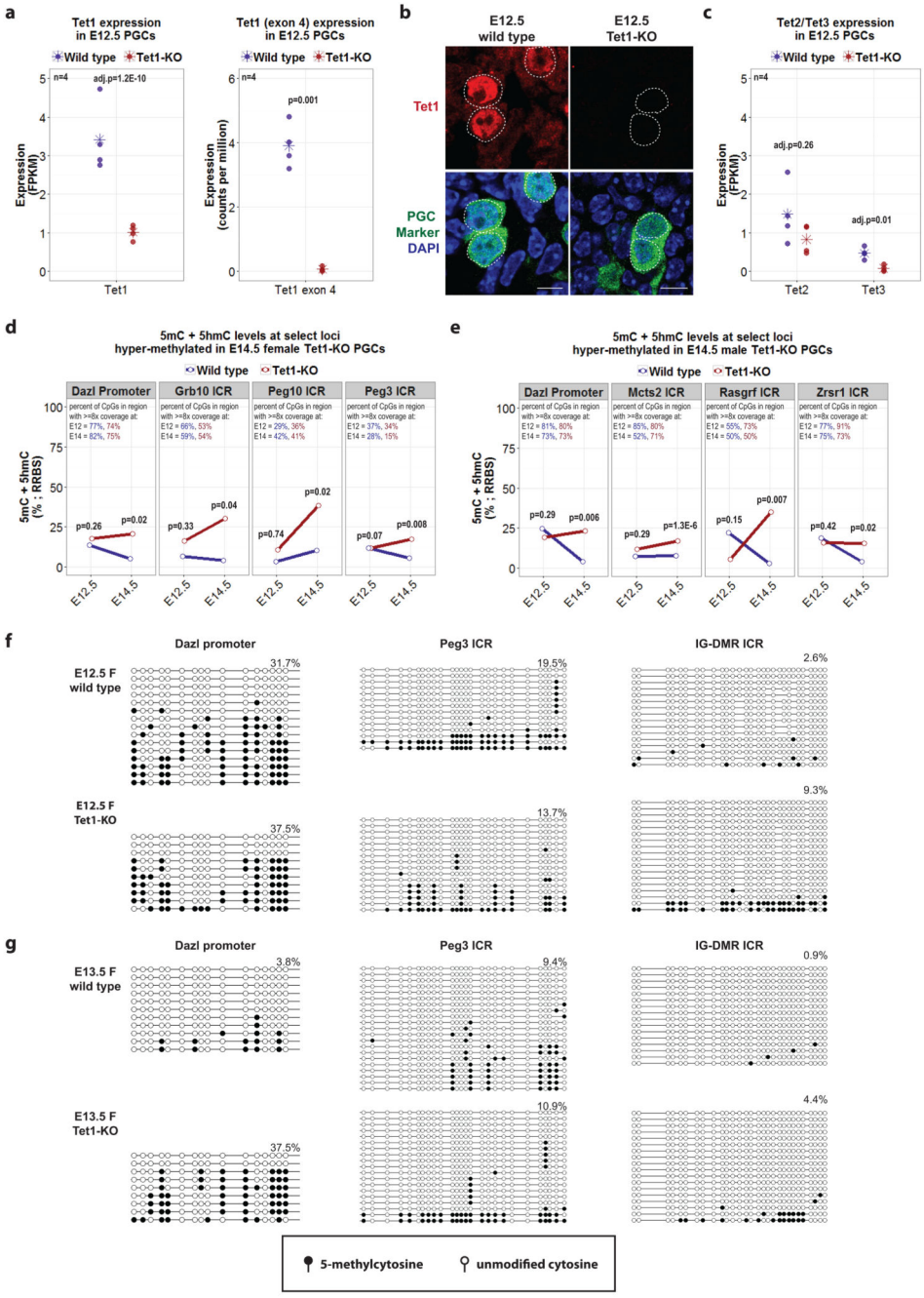


Extended Data Fig. 5. Suggested models implicating 5mC oxidation in DNA demethylation of gonadal PGCs.

a) A model of oxidation followed by passive dilution predicts a positive correlation between the extent to which the combined 5mC/5hmC levels decrease between two stages (i.e. %; WGBS) and the total level of 5hmC at both the stage immediately preceding and following this decrease. **b)** A model implicating 5mC oxidation in triggering DNA demethylation via an active mechanism predicts a positive correlation between the extent to which the combined 5mC/5hmC levels decrease between two stages (i.e. %; WGBS) and the relative

5hmC levels in the stage immediately preceding this decrease, as further oxidation of 5hmC to 5fC is the rate limiting step in the full oxidation of 5mC to 5caC39. **c)** A model implicating oxidation of 5mC in safeguarding DNA hypomethylation following the major wave of DNA demethylation predicts that regions where the majority of DNA demethylation has been lost between two stages (i.e. those that are newly hypomethylated) will have high relative levels of 5hmC in the stage immediately following the major wave of DNA demethylation to remove residual methylation and/or aberrant *de novo* methylation. Thus, a limited correlation between the extent to which the combined 5mC/5hmC levels decrease between two stages (i.e. %, WGBS) and the relative 5hmC levels in the stage immediately following this decrease may also be seen.

Europe PMC Funders Author Manuscripts

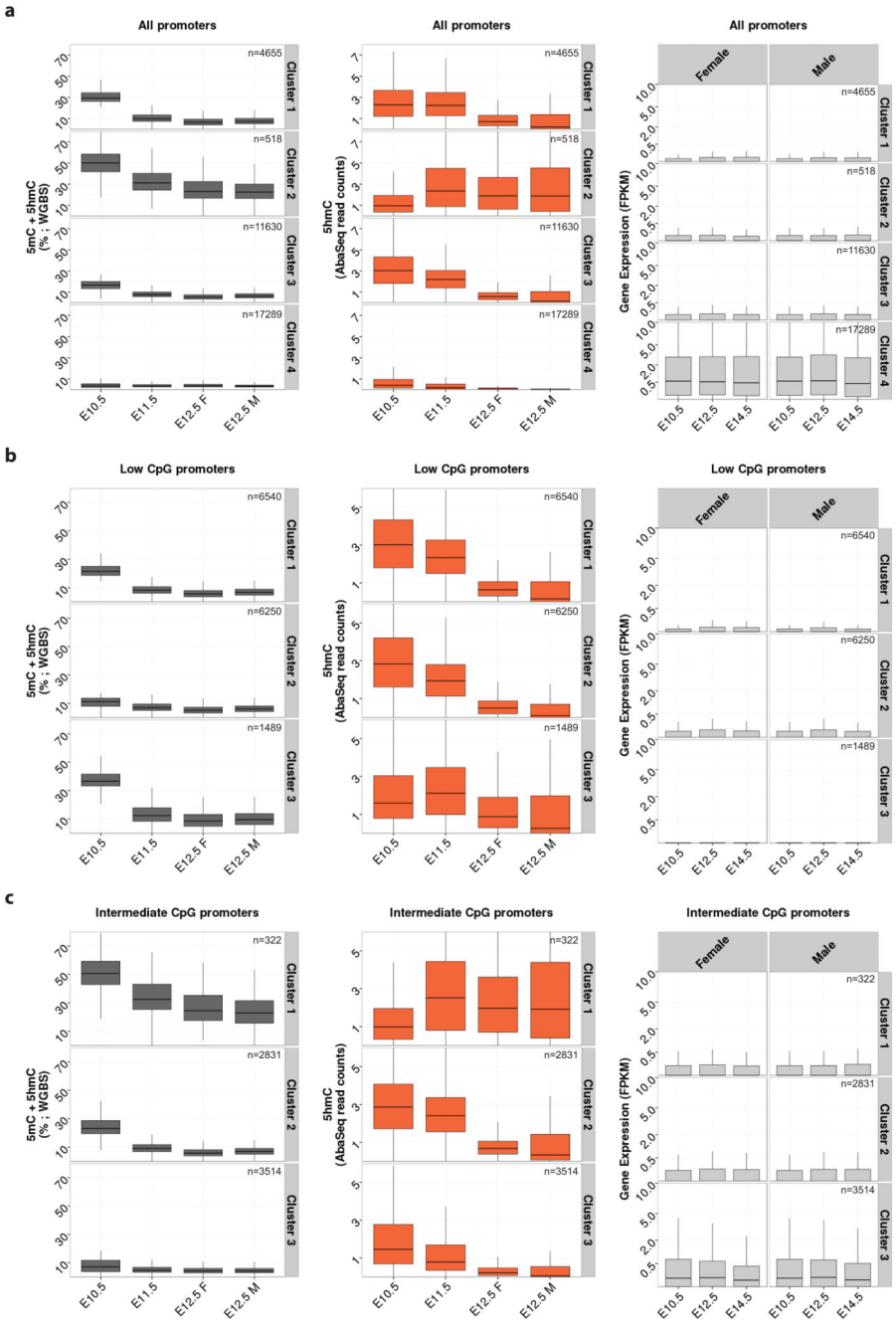


Extended Data Fig. 6. Tet1-3 expression and locus-specific DNA methylation in *Tet1*-KO PGCs during epigenetic reprogramming.

a) Expression of *Tet1* total transcript (left) or deleted exon 4 (right) in E12.5 *Tet1*-KO and wild type PGCs. Adjusted p-values (left) computed by DESeq2 and p-values (right) computed by Student's t-test. Asterisks refer to mean values. **b)** Representative immunostaining against the N-terminus of Tet1 protein in E12.5 wild type and *Tet1*-KO PGCs. Scale bar represents 10 μm. **c)** Expression of *Tet2* and *Tet3* in E12.5 *Tet1*-KO and wild type PGCs. Adjusted p-values computed by DESeq2. Asterisks refer to mean values. **d-**

e) Mean combined 5hmC/5mC levels levels (RRBS) in female (d) or male (e) E12.5 and E14.5 *Tet1*-KO and wild type PGCs for ICRs and germline gene promoters called hypermethylated in E14.5 *Tet1*-KO PGCs. The mean DNA modification level and p-values were computed by RnBeads software (see Methods for details). **f-g)** Locus-specific bisulphite sequencing of the *Dazl* promoter (left), the *Peg3* ICR (middle) and the *IG-DMR* ICR (right) in E12.5 (f) and E13.5 (g) female *Tet1*-KO and wild type PGCs. Specific details regarding sample sizes and how samples were collected are found in Statistics and Reproducibility section.

Europe PMC Funders Author Manuscripts

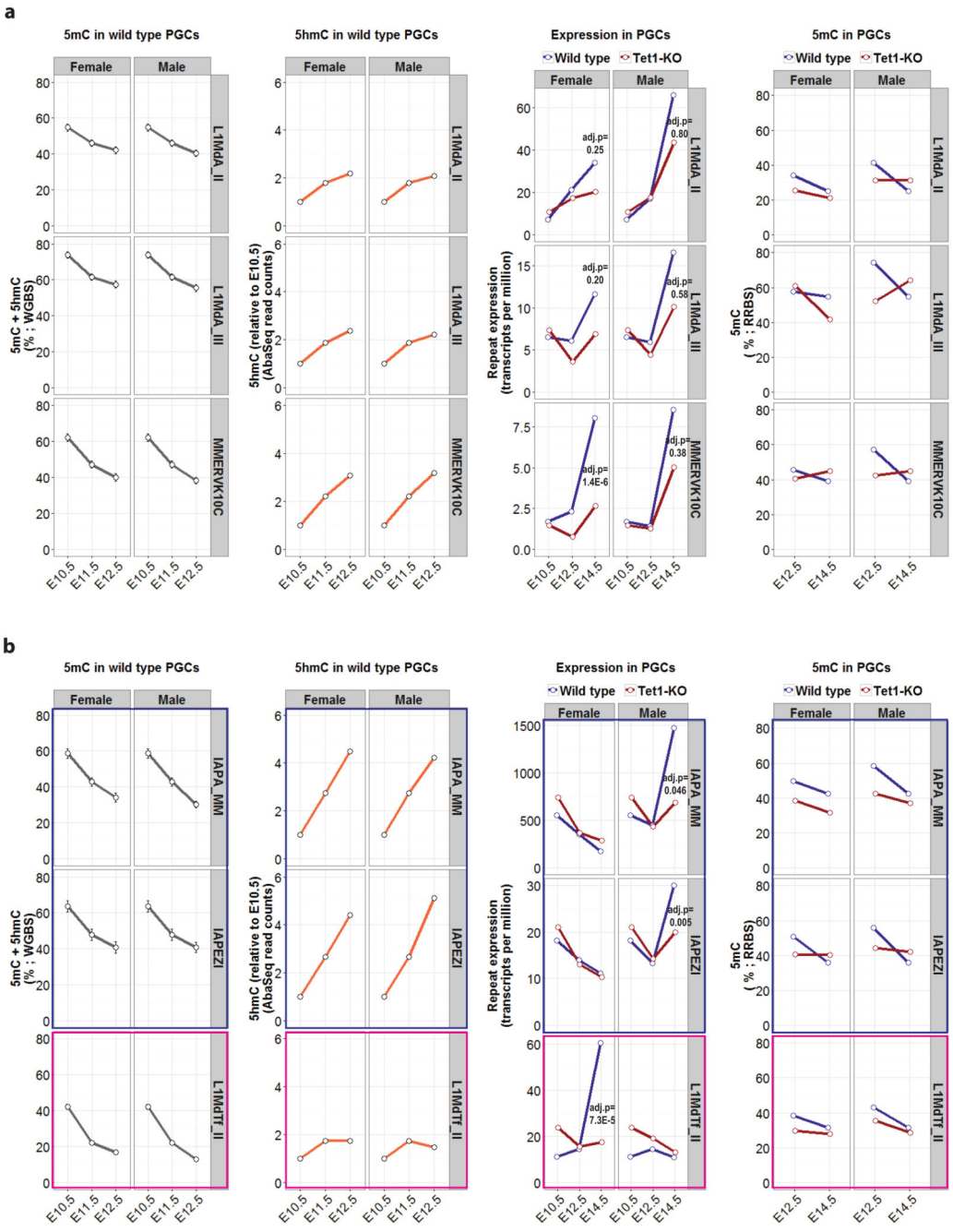


Extended Data Fig. 7. Promoter DNA methylation clustering analysis during germline reprogramming.

a) The combined promoter 5mC/5hmC levels (WGBS, right), promoter 5hmC levels (AbaSeq, centre), or gene expression levels (RNA-Seq, right) in consecutive stages of PGC development for all genes grouped by K-means clustering of the combined 5mC/5hmC dynamics at their promoter regions. **b-c)** Boxplots depicting the combined promoter 5mC/5hmC levels (WGBS, right), promoter 5hmC levels (AbaSeq, centre), or gene expression levels (RNA-Seq, right) in consecutive stages of PGC development for three clusters of

Europe PMC Funders Author Manuscripts

genes with either low CpG promoters (LCPs; b) or intermediate CpG promoters (ICPs; c) grouped by K-means clustering of the combined 5mC/5hmC dynamics at their promoter regions. For all boxplots, the upper and lower hinges correspond to the first and third quartiles, the middle line corresponds to the median, and the maxima and minima respectively correspond to the highest or lowest value within 1.5× the inter-quartile range. Specific details regarding sample sizes and how samples were collected are found in Statistics and Reproducibility section.

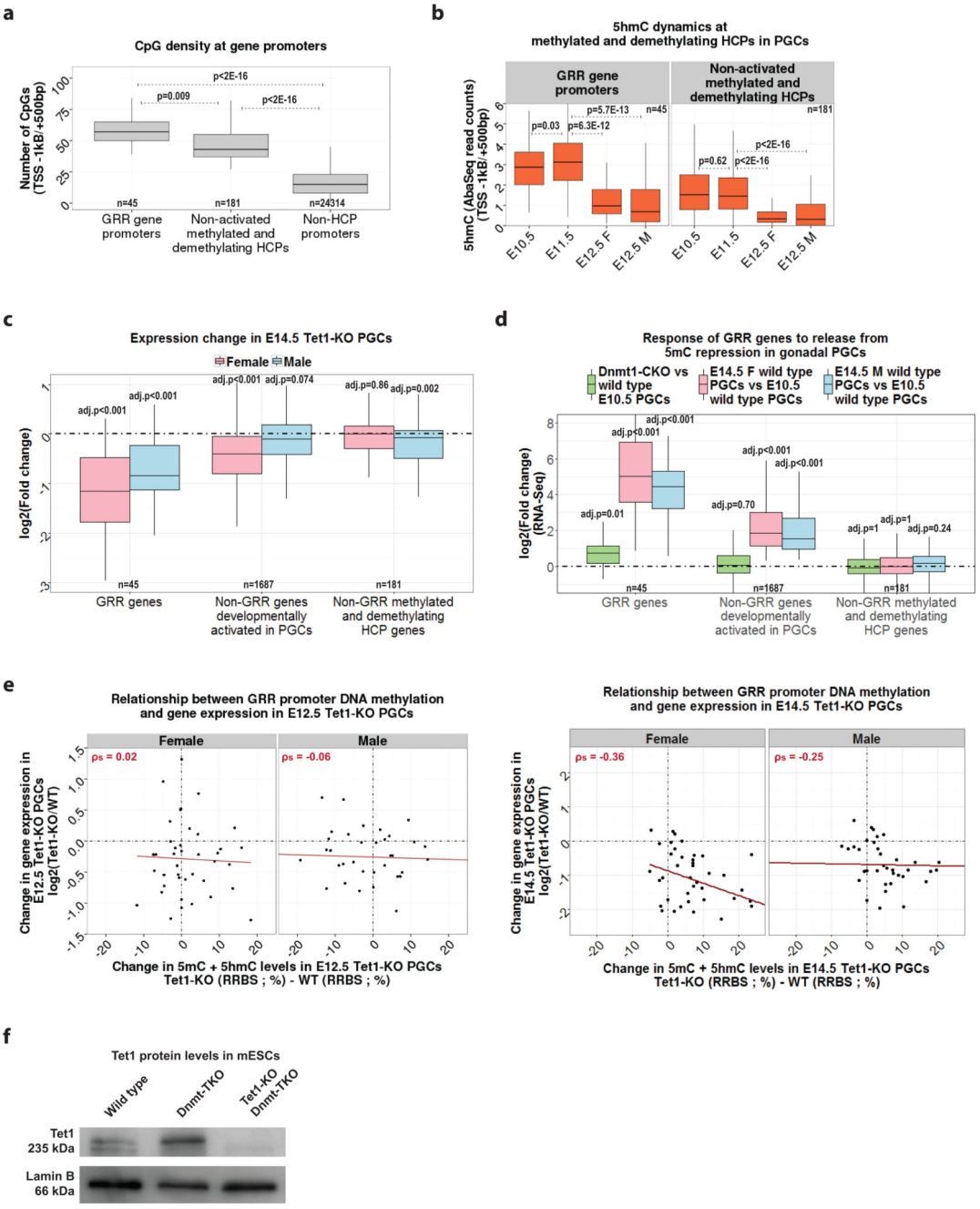


Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

Extended Data Fig. 8. DNA modification and expression dynamics in wild type and *Tet1*-KO PGCs at retrotransposons normally activated concurrent with epigenetic reprogramming.

a-b) Combined 5mC/5hmC dynamics in wild type PGCs (%; WGBS; far left), relative 5hmC dynamics (AbaSeq read counts normalised to E10.5; centre left) in wild type PGCs, the expression dynamics in either wild type or *Tet1*-KO PGCs (transcripts per million (TPM); RNA-Seq; centre right), and combined 5mC/5hmC dynamics in wild type and *Tet1*-KO PGCs (%; RRBS; far right) for representative repetitive elements significantly up-regulated (adj. p-value < 0.05; Sleuth) in a sex-independent manner (a), in a male-specific manner (b, blue box), or in a female-specific manner (b, pink box) between E10.5 and E14.5 in wild type PGCs. Mean values are shown in all cases. Adjusted p-values for differential repeat expression analysis between E14.5 wild type and *Tet1*-KO PGCs are based on Sleuth software. Specific details regarding sample sizes and how samples were collected are found in Statistics and Reproducibility section.



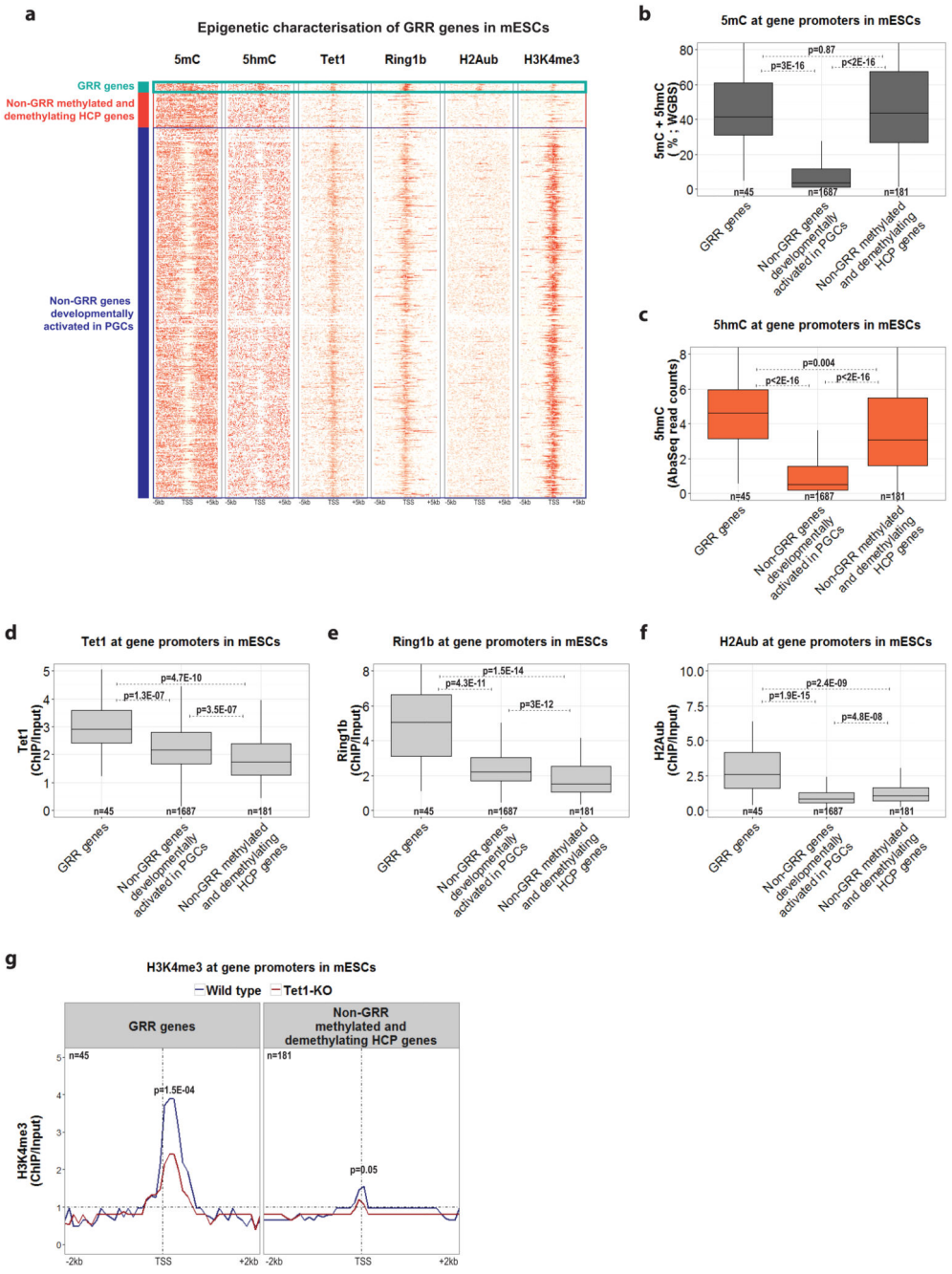
Extended Data Fig. 9. Characterisation of GRR gene regulation by Tet1 and 5mC in PGCs and mESCs.

a) CpG density at GRR gene promoters and other relevant promoters; p-values are based on a two-sided Wilcoxon test. **b)** Mean 5hmC dynamics at GRR gene promoters and non-activated methylated and demethylating HCPs in PGCs; p-values are based on a two-sided paired Wilcoxon test. **c)** Log2-fold change between *Tet1*-KO and wild type E14.5 male (blue) or female (pink) PGCs for GRR genes and other relevant gene sets. FWER-adjusted p-values are based on GSEA software (see Methods for details). **d)** Log2-fold change

Europe PMC Funders Author Manuscripts

between *Dnmt1*-CKO24 and wild type mESCs (green), or between E14.5 female (pink) or male (blue) wild type PGCs and E10.5 wild type PGCs, for GRR genes and other relevant gene sets. FWER-adjusted p-values are based on GSEA software (see Methods for details). **e)** Correlation between the difference in combined 5mC/5hmC levels (x-axis; *Tet1*-KO (RRBS; %) – WT (RRBS; %)) at GRR promoters and the change in GRR gene expression (y-axis; log2(*Tet1*-KO/WT)) in E12.5 (right) and E14.5 (left) *Tet1*-KO PGCs. Spearman correlation is shown. **f)** Representative western blot showing Tet1 and Lamin B protein expression in wild type, *Dnmt*-TKO, and *Tet1*-KO *Dnmt*-TKO mESCs. For all boxplots, the upper and lower hinges correspond to the first and third quartiles, the middle line corresponds to the median, and the maxima and minima respectively correspond to the highest or lowest value within 1.5× the inter-quartile range. For all figures, specific details regarding sample sizes and how samples were collected are found in Statistics and Reproducibility section. Specific details regarding sample sizes and how samples were collected are found in Statistics and Reproducibility section.

Europe PMC Funders Author Manuscripts



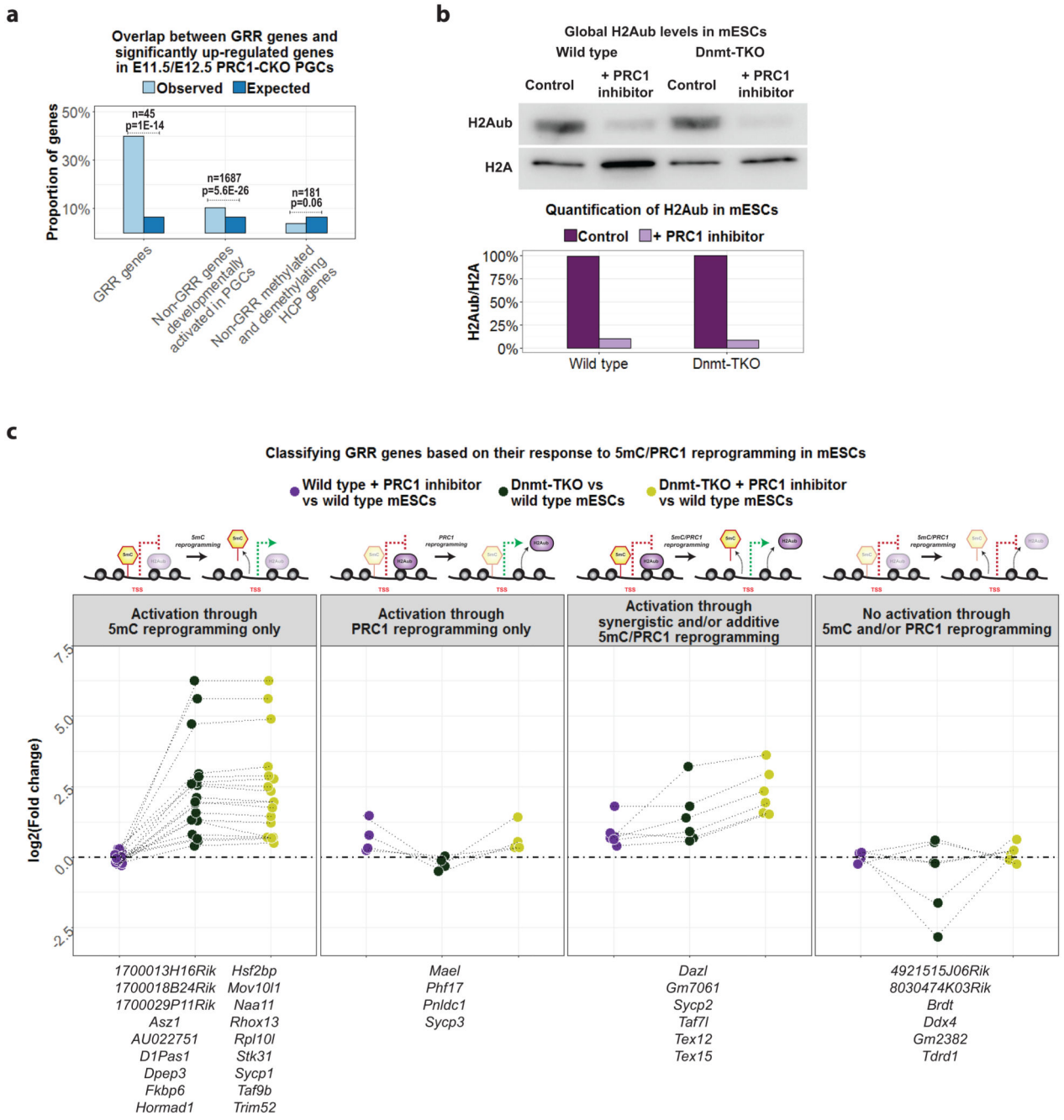
Extended Data Fig. 10. Epigenetic characterisation of GRR gene promoters in mESCs.

a) Genomic sequences centred on TSSs of GRR genes, non-GRR genes activated in both male and female PGCs between E10.5 and E14.5, and non-GRR methylated and demethylating HCP genes in wild type mESCs grown in serum-containing media. Each horizontal line represents one gene; the intensity of red indicates the relative enrichment for the feature shown at the top of each column. The TSS and sequences 5 kb upstream and downstream of the TSS are shown. **b-f)** Boxplots depicting the levels of: (b) combined 5mC/5hmC levels (WGBS)30; (c) 5hmC (AbaSeq)15 ; (d) Tet1 (ChIP-Seq)21; (e) Ring1b (ChIP-

Seq)38 and (f) H2Aub levels (ChIP-Seq)37 at promoters of either GRR genes and other relevant gene sets in wild type mESCs grown in serum-containing media. For all boxplots, the upper and lower hinges correspond to the first and third quartiles, the middle line corresponds to the median, and the maxima and minima respectively correspond to the highest or lowest value within 1.5× the inter-quartile range. p-values are based on two-sided Wilcoxon test. **g)** Metagene plot depicting median H3K4me3 levels (ChIP-Seq)30 around TSS of GRR genes (left) and non-GRR HCP genes that are also initially methlylated and subsequently demethylated during PGC reprogramming (right) in wild type and *Tet1*-KO mESCs grown in serum containing media. p-values are based on paired two-sided Wilcoxon test for region of TSS -1kB/+500bp. Specific details regarding sample sizes and how samples were collected are found in Statistics and Reproducibility section.

Europe PMC Funders Author Manuscripts

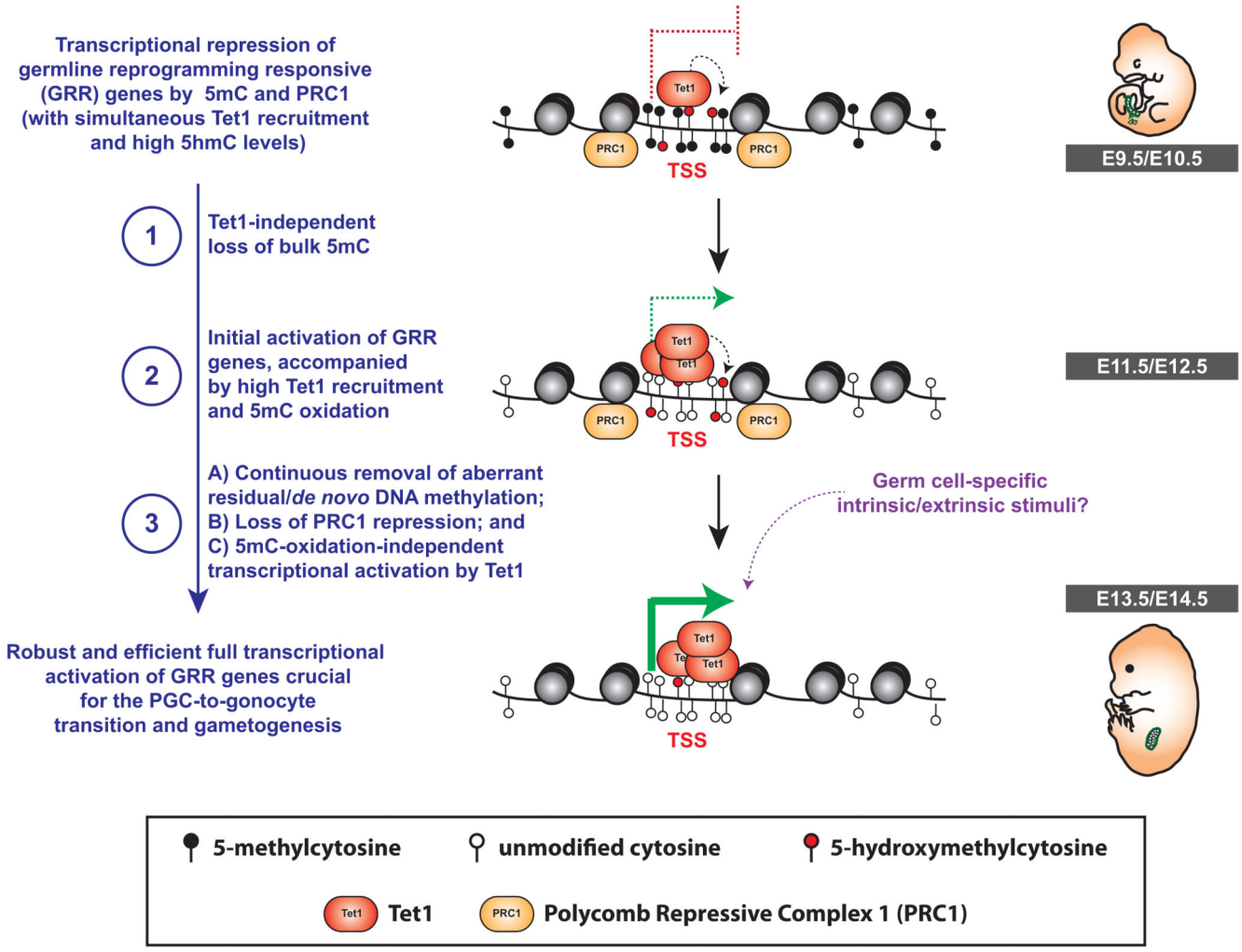
Europe PMC Funders Author Manuscripts



Extended Data Fig. 11. Characterisation of GRR gene regulation by PRC1 and 5mC in PGCs and mESCs.

a) Overlap between GRR genes and genes significantly up-regulated in E11.5 and/or E12.5 PRC1 conditional knockout PGCs compared with wild-type26. p-values based on hypergeometric test. **b)** Representative western blot showing H2Aub and H2A levels in wild type or *Dnmt*-TKO mESCs + 6h DMSO, and wild type or *Dnmt*-TKO mESCs + 6h PRT4165 (PRC1 inhibitor). **c)** Classification of GRR genes depending on their dependency for 5mC and/or PRC1 reprogramming in mESCs (see Methods for details). Specific details

regarding sample sizes and how samples were collected are found in Statistics and Reproducibility section.



Extended Data Fig. 12. Model.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

Work in Hajkova lab is supported by MRC funding (MC_US_A652_5PY70), by FP7 EpiGeneSys network, and by an ERC grant (ERC-CoG- 648879 – dynamicmodifications) to P.H. Y.Z. and S.P.'s lab is supported by grant 1R44GM096723-01A1. P.H. is a member of EMBO Young Investigator Programme. P.W.S.H. is a recipient of an MRC PhD Studentship and MRC-targeted Doctoral Prize Fellowship from Imperial College London. We are grateful to the members of Hajkova lab for discussions and revisions of the manuscript. We would like to thank James Elliot and Thomas Adejumo for help with FACS; Laurence Game for help with NGS; Felix Krueger for providing consensus repetitive element sequences; and Megan Woodberry, Andrew Cameron, and Justyna Glegola for mouse husbandry.

Europe PMC Funders Author Manuscripts

References

1. Lesch B, Page D. Genetics of germ cell development. Nat Rev Genet. 2012; 13:781–794. [PubMed: 23044825]
2. Guibert S, Forné T, Weber M. Global profiling of DNA methylation erasure in mouse primordial germ cells. Genome Res. 2012; 22:633–641. [PubMed: 22357612]
3. Hackett J, et al. Germline DNA demethylation dynamics and imprint erasure through 5-hydroxymethylcytosine. Science. 2013; 339:448–452. [PubMed: 23223451]
4. Hajkova P, et al. Chromatin dynamics during epigenetic reprogramming in the mouse germ line. Nature. 2008; 452:877–881. [PubMed: 18354397]
5. Hajkova P, et al. Epigenetic reprogramming in mouse primordial germ cells. Mech Dev. 2002; 117:15–23. [PubMed: 12204247]
6. Hill P, Amouroux R, Hajkova P. DNA demethylation, Tet proteins and 5-hydroxymethylcytosine in epigenetic reprogramming: an emerging complex story. Genomics. 2014; 104:324–333. [PubMed: 25173569]
7. Lee J, et al. Erasing genomic imprinting memory in mouse clone embryos produced from day 11.5 primordial germ cells. Development. 2002; 129:1807–1817. [PubMed: 11934847]
8. Seisenberger S, et al. The dynamics of genome-wide DNA methylation reprogramming in mouse primordial germ cells. Mol Cell. 2012; 48:849–862. [PubMed: 23219530]
9. Yamaguchi S, et al. Tet1 controls meiosis by regulating meiotic gene expression. Nature. 2012; 492:443–447. [PubMed: 23151479]
10. Yamaguchi S, Shen L, Liu Y, Sendler D, Zhang Y. Role of Tet1 in erasure of genomic imprinting. Nature. 2013; 504:460–464. [PubMed: 24291790]
11. Hajkova P, et al. Genome-wide reprogramming in the mouse germ line entails the base excision repair pathway. Science. 2010; 329:78–82. [PubMed: 20595612]
12. Hayashi K, et al. Offspring from oocytes derived from in vitro primordial germ cell-like cells in mice. Science. 2012; 338:971–975. [PubMed: 23042295]
13. Hayashi K, Ohta H, Kurimoto K, Aramaki S, Saitou M. Reconstitution of the mouse germ cell specification pathway in culture by pluripotent stem cells. Cell. 2011; 146:519–532. [PubMed: 21820164]
14. Hikabe O, et al. Reconstitution in vitro of the entire cycle of the mouse female germ line. Nature. 2016; 539:299–303. [PubMed: 27750280]
15. Sun Z, et al. High-resolution enzymatic mapping of genomic 5-hydroxymethylcytosine in mouse embryonic stem cells. Cell Rep. 2013; 3:567–576. [PubMed: 23352666]
16. Huang Y, et al. The behaviour of 5-hydroxymethylcytosine in bisulfite sequencing. PLoS One. 2010; 5:e8888. [PubMed: 20126651]
17. Yamaguchi S, et al. Dynamics of 5-methylcytosine and 5-hydroxymethylcytosine during germ cell reprogramming. Cell Res. 2013; 23:329–339. [PubMed: 23399596]
18. Dawlaty M, et al. Tet1 is dispensable for maintaining pluripotency and its loss is compatible with embryonic and postnatal development. Cell Stem Cell. 2011; 9:166–175. [PubMed: 21816367]
19. Weber M, et al. Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. Nat Genet. 2007; 39:457–466. [PubMed: 17334365]
20. Tahiliani M, et al. Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. Science. 2009; 324:930–935. [PubMed: 19372391]
21. Williams K, et al. TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity. Nature. 2011; 473:343–348. [PubMed: 21490601]
22. Vella P, et al. Tet proteins connect the O-linked N-acetylglucosamine transferase Ogt to chromatin in embryonic stem cells. Mol Cell. 2013; 49:645–656. [PubMed: 23352454]
23. Tsumura A, et al. Maintenance of self-renewal ability of mouse embryonic stem cells in the absence of DNA methyltransferases Dnmt1, Dnmt3a and Dnmt3b. Genes Cells. 2006; 11:805–814. [PubMed: 16824199]
24. Hargan-Calvopina J, et al. Stage-Specific Demethylation in Primordial Germ Cells Safeguards against Precocious Differentiation. Dev Cell. 2016; 39:75–86. [PubMed: 27618282]

25. Mansour A, et al. The H3K27 demethylase Utx regulates somatic and germ cell epigenetic reprogramming. Nature. 2012; 488:409–413. [PubMed: 22801502]
26. Yokobayashi S, et al. PRC1 coordinates timing of sexual differentiation of female primordial germ cells. Nature. 2013; 495:236–240. [PubMed: 23486062]
27. Ismail I, McDonald D, Strickfaden H, Xu Z, Hendzel M. A small molecule inhibitor of polycomb repressive complex 1 inhibits ubiquitin signaling at DNA double-strand breaks. J Biol Chem. 2013; 288:26944–26954. [PubMed: 23902761]
28. Deplus R, et al. TET2 and TET3 regulate GlcNAcylation and H3K4 methylation through OGT and SET1/COMPASS. EMBO J. 2013; 32:645–655. [PubMed: 23353889]
29. Amouroux R, et al. De novo DNA methylation drives 5hmC accumulation in mouse zygotes. Nat Cell Biol. 2016; 18:225–233. DOI: 10.1038/ncb3296 [PubMed: 26751286]
30. Hon G, et al. 5mC Oxidation by Tet2 Modulates Enhancer Activity and Timing of Transcriptome Reprogramming during Differentiation. Mol Cell. 2014; 56:286–297. [PubMed: 25263596]
31. Yang H, et al. One-step generation of mice carrying reporter and conditional alleles by CRISPR/Cas-mediated genome engineering. Cell. 2013; 154:1370–1379. [PubMed: 23992847]
32. Boyle P, et al. Gel-free multiplexed reduced representation bisulfite sequencing for large-scale DNA methylation profiling. Genome Biol. 2012; 13:R92. [PubMed: 23034176]
33. Kurimoto K, et al. Quantitative Dynamics of Chromatin Remodeling during Germ Cell Specification from Mouse Embryonic Stem Cells. Cell Stem Cell. 2015; 16:517–532. [PubMed: 25800778]
34. Borgel J, et al. Targets and dynamics of promoter DNA methylation during early mouse development. Nat Genet. 2010; 42:1093–1100. [PubMed: 21057502]
35. Yu M, et al. Base-resolution analysis of 5-hydroxymethylcytosine in the mammalian genome. Cell. 2012; 149:1368–1380. [PubMed: 22608086]
36. Xu Y, et al. Genome-wide regulation of 5hmC, 5mC, and gene expression by Tet1 hydroxylase in mouse embryonic stem cells. Mol Cell. 2011; 42:451–464. [PubMed: 21514197]
37. Brookes E, et al. Polycomb associates genome-wide with a specific RNA polymerase II variant, and regulates metabolic genes in ESCs. Cell Stem Cell. 2012; 10:157–170. [PubMed: 22305566]
38. Cooper S, et al. Targeting polycomb to pericentric heterochromatin in embryonic stem cells reveals a role for H2AK119u1 in PRC2 recruitment. Cell Rep. 2014; 7:1456–1470. [PubMed: 24857660]
39. Hashimoto H, et al. Structure of a Naegleria Tet-like dioxygenase in complex with 5-methylcytosine DNA. Nature. 2014; 506:391–395. [PubMed: 24390346]

Europe PMC Funders Author Manuscripts

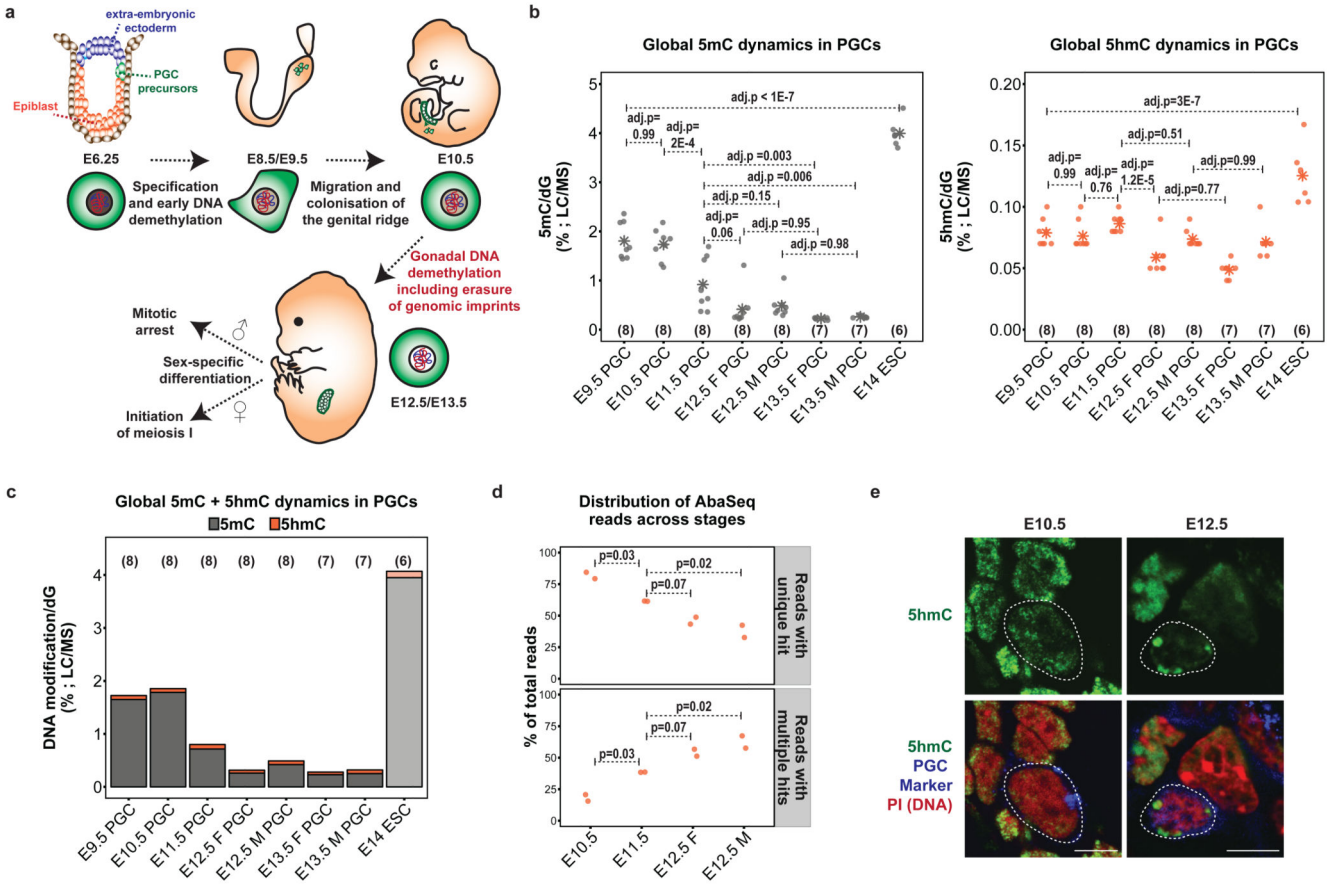


Fig. 1. 5mC and 5hmC dynamics during epigenetic reprogramming.

a) Key events during mouse PGC development. **b-c)** Individual 5mC (b, left) and 5hmC (b, right) and combined 5mC/5hmC (c) levels in mESCs and E9.5 to E13.5 PGCs (LC/MS). Asterisks in (b) refer to mean values. Adjusted p-values are based on ANOVA and Tukey posthoc test. Bar chart in (c) depicts median value of biological replicates depicted in (b). **d)** Re-distribution of 5hmC from the uniquely mapped part of the genome to repetitive elements between E10.5 and E12.5. p-values based on combined ANOVA and Tukey post hoc test. **e)** Representative 5hmC immunostaining in E10.5 and E12.5 PGCs. Scale bar represents 10 μm. Details regarding sample sizes and how samples were collected can be found in Statistics and Reproducibility section.

Europe PMC Funders Author Manuscripts

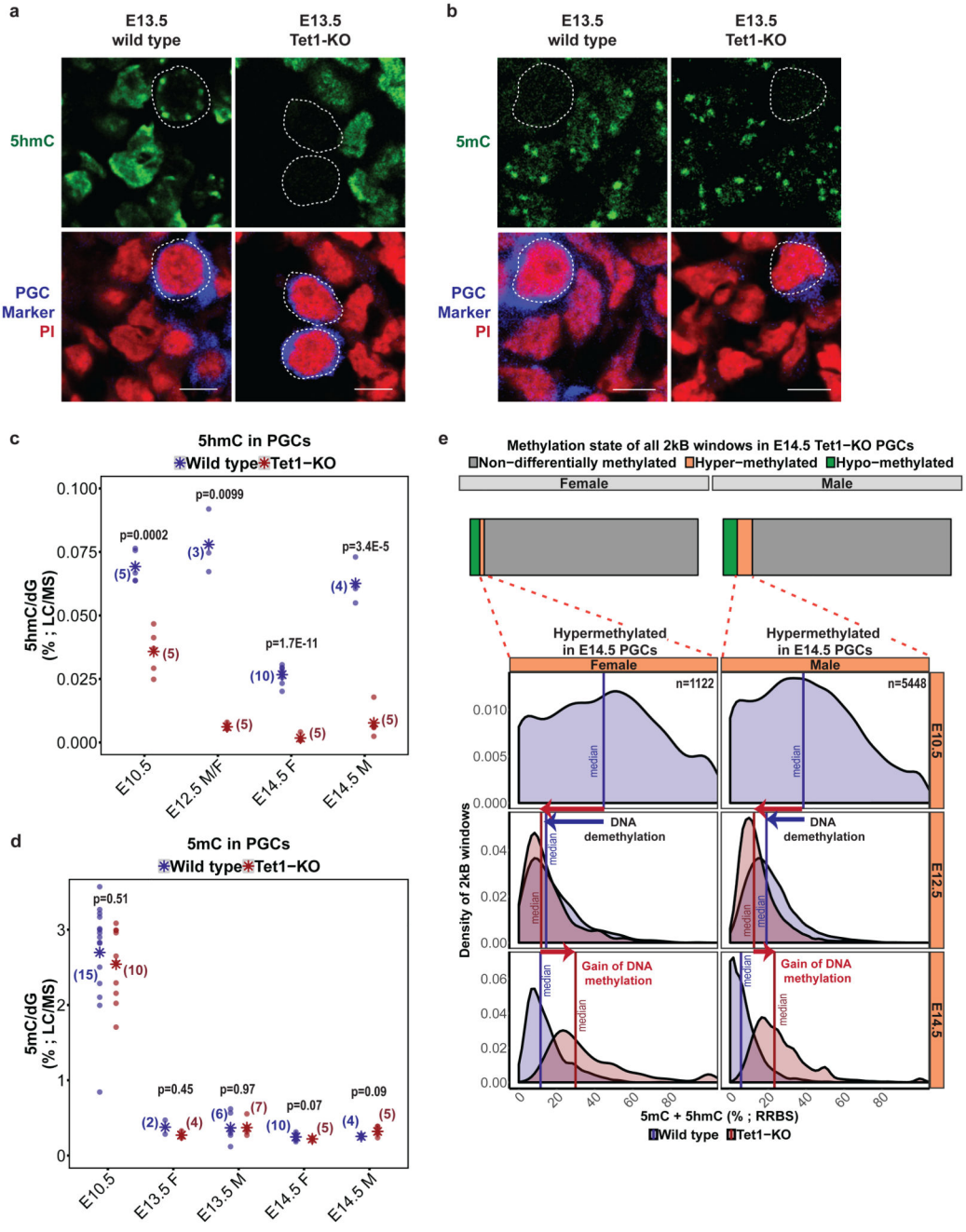


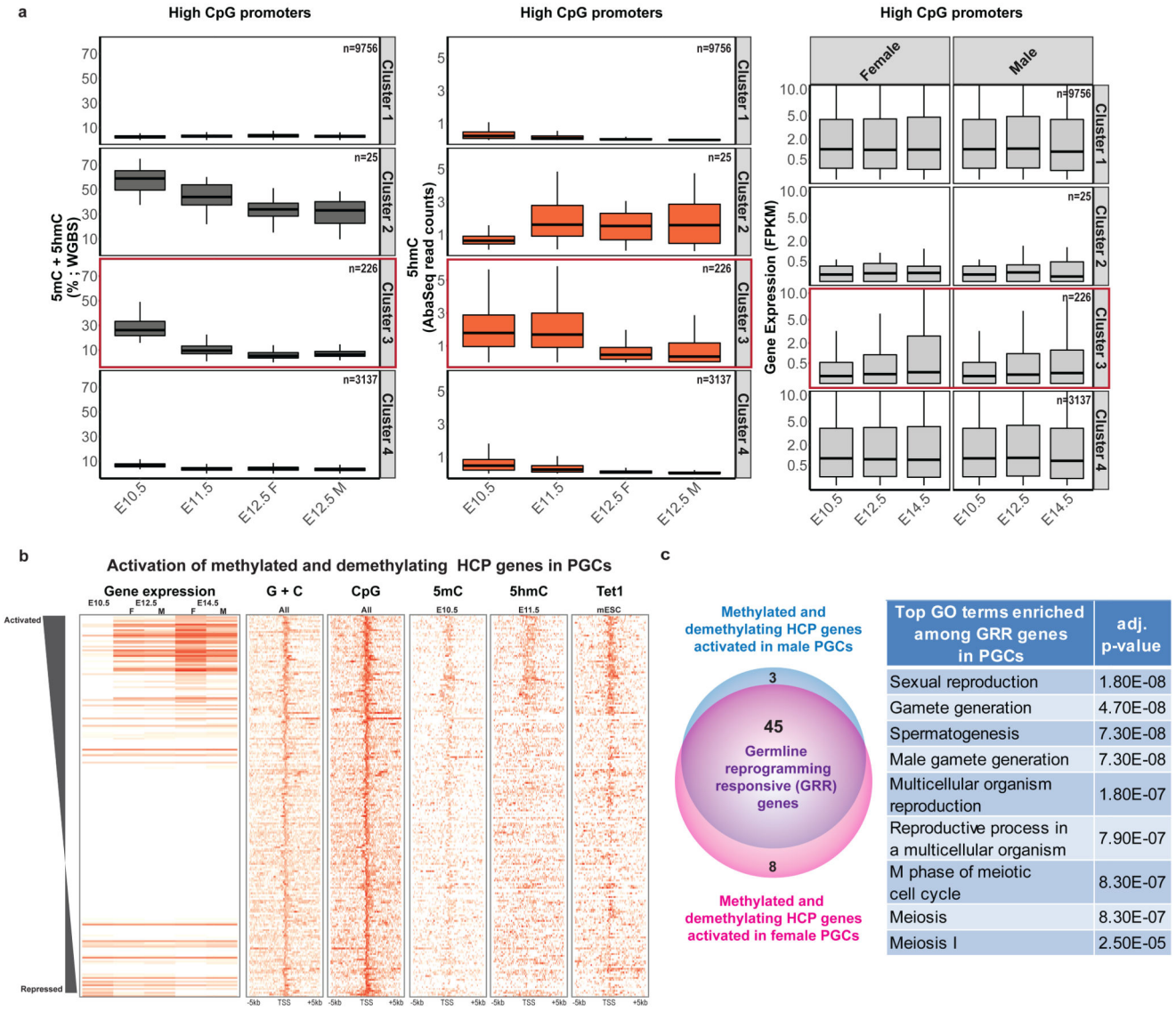
Fig. 2. Tet1 safeguards but does not drive DNA demethylation.

a-b) Representative immunostaining against 5hmC (a) or 5mC (b) in E13.5 wild type and *Tet1*-KO PGCs. Scale bar represents 10 μm. **c-d)** Global 5hmC (c) and 5mC (d) levels (LC/MS) in wild type and *Tet1*-KO PGCs. Sample numbers are indicated on graphs. Asterisks refer to mean values. p-values are based on two-sided Student's t-test**. e) Top Figure:** Proportion of differentially methylated regions in E14.5 *Tet1*-KO PGCs (p<0.05, >10% methylation difference; p-value derived from RnBeads software). **Bottom Figure:** Combined 5mC/5hmC levels (RRBS) in E12.5 (middle) and E14.5 (bottom) *Tet1*-KO (red)

Europe PMC Funders Author Manuscripts

and wild type (blue) PGCs for all E14.5 hypermethylated 2kB windows. DNA modification levels from E10.5 wild type PGCs are also shown (top panel). Median combined 5mC/5hmC levels are denoted by vertical lines. Details regarding sample sizes and how samples were collected can be found in Statistics and Reproducibility section.

Europe PMC Funders Author Manuscripts



Top GO terms enriched among GRR genes in PGCs	adj. p-value
Sexual reproduction	1.80E-08
Gamete generation	4.70E-08
Spermatogenesis	7.30E-08
Male gamete generation	7.30E-08
Multicellular organism reproduction	1.80E-07
Reproductive process in a multicellular organism	7.90E-07
M phase of meiotic cell cycle	8.30E-07
Meiosis	8.30E-07
Meiosis I	2.50E-05

Fig. 3. Germline reprogramming responsive (GRR) genes.

a) Combined promoter 5mC/5hmC levels (right), promoter 5hmC levels (centre), or gene expression levels (right) in consecutive stages of PGC development for HCP gene clusters (see Methods). The upper and lower hinges correspond to the first and third quartiles, the middle line corresponds to the median, and the maxima and minima respectively correspond to the highest or lowest value within 1.5× the inter-quartile range. **b)** Genomic sequences centred on TSSs of methylated and demethylating HCPs (cluster 3, Fig. 3A) ranked based on the significance of up-regulation between E10.5 and E14.5 in wild type PGCs. Each horizontal line represents one gene; the intensity of red indicates the relative enrichment for the feature shown at the top of each column. The TSS +/-5kb is shown. **c)** Gene ontology (GO) terms associated with germline reprogramming responsive (GRR) genes; adj. p-value is based on DAVID software. Details regarding sample sizes and how samples were collected can be found in Statistics and Reproducibility section.

Europe PMC Funders Author Manuscripts

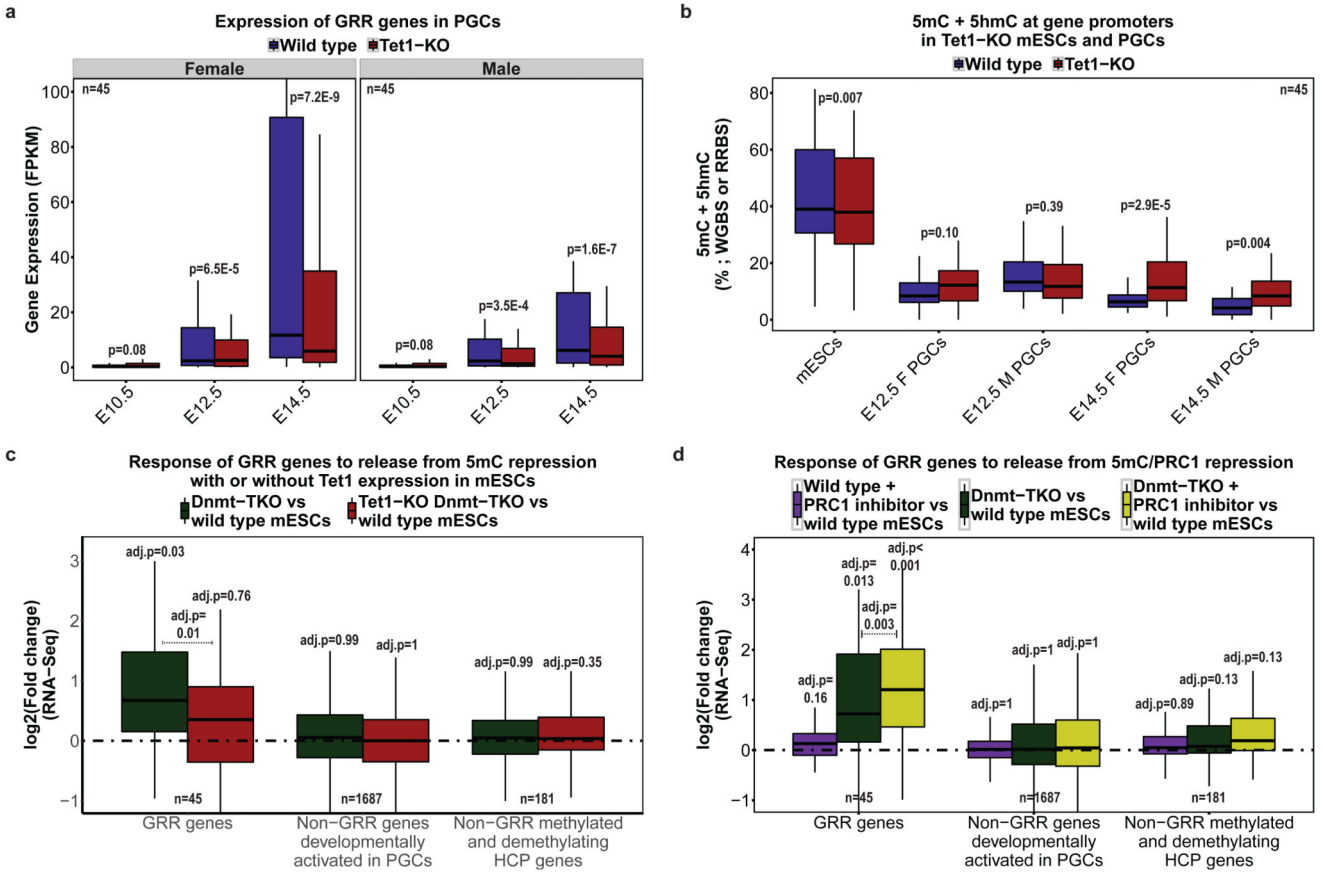


Fig. 4. Epigenetic principles of GRR gene activation.

a) GRR gene expression dynamics in *Tet1*-KO PGCs; p-values are based on a two-sided paired Wilcoxon test. **b)** Combined 5mC/5hmC levels (RRBS) at GRR genes in E12.5 or E14.5 *Tet1*-KO (red) and wild type (blue) PGCs. For comparison, combined 5mC/5hmC levels in mESCs30 (% ; WGBS) are shown. p-values are based on paired two-sided Wilcoxon test. **c-d)** Log2-fold change between *Dnmt*-TKO (green) or *Tet1*-KO *Dnmt*-TKO and wild type mESCs (in c) or between wild type + 6h PRT4165 treatment (purple), *Dnmt*-TKO + 6h DMSO treatment (green) or *Dnmt*-TKO + 6h PRT4165 treatment (yellow) and wild type + 6h DMSO treatment mESCs (in d) for GRR genes and other relevant genes sets. FWER-adjusted p-values are based on GSEA software (see Methods for details). Specific details regarding sample sizes and how samples were collected are found in Statistics and Reproducibility section. For all boxplots, the upper and lower hinges correspond to the first and third quartiles, the middle line corresponds to the median, and the maxima and minima respectively correspond to the highest or lowest value within 1.5× the inter-quartile range.

Europe PMC Funders Author Manuscripts