Supplementary Online Material

Hackett *et al*.

Materials and Methods

Isolation of PGCs

PGCs were isolated from outbred MF1 females crossed with stud males carrying the distal *Oct4* enhancer-*GFP* transgene, which is expressed exclusively in PGCs from ~E8.0 onwards (*25*). Noon of the day of the vaginal plug was designated as E0.5. We isolated PGCs from dissected embryonic urogenital ridges that were trypsinized, washed in PBS, and filtered, by FACS sorting using a MoFlo 3 laser Cytometer (Beckman Coulter). Typically, we isolated ~300 PGCs per urogenital ridge pair at E10.5, ~1000 at E11.5, ~3500 at E12.5 and ~10,000 at E13.5. The purity of PGCs was independently verified by STELLA and OCT4 staining and was in excess of 99%.

Genomic DNA preparation

Purified and pooled PGCs, embryonic soma or Epiblast stem cells (EpiSC) were treated with RNase cocktail (Roche) for 5 min and subsequently incubated in 0.5% SDS and Proteinase K (100ng/ul) for 2 hours at 55˚C. Genomic DNA was extracted with phenol:chloroform:IAA and then chloroform using phase lock gel, followed by ethanol/NaCl precipitation overnight at -20˚C and two washes with 70% ethanol.

Immunofluorescence and microscopy

A single-cell suspension of PGCs and neighbouring somatic cells, obtained directly from trypsinisation of urogenital ridges, was allowed to attach to poly-L-lysine (Sigma-Aldrich) treated slides. The cells were fixed in 2% paraformaldehyde (PFA, prepared in PBS) for 10 min at room temperature and washed twice with PBS. Fixed cells were permeabilised for 10 min using 0.1% sodium citrate, 0.2% Triton-X buffer, and then washed with PBS, 0.1% Tween. For 5mC, 5hmC, 5fC and 5caC staining, cells were then treated with 3N HCl for 10 mins at room temperature to denature DNA, neutralised with 100mM Tris-HCl for 20 mins, and washed twice (PBS, 0.25% BSA, 0.1% tween). Cells were subsequently blocked for 1hr with PBS, 2.5% BSA, 0.1% Tween and 10% normal goat serum at room temperature followed by antibody staining that was carried out in the same buffer at 4˚C overnight. The slides were subsequently washed three times in PBS, 0.25% BSA, 0.1% Tween-20 (5 min each wash) and incubated with

Alexa fluorophore conjugated secondary antibodies (Molecular Probes) for 1 hr at room temperature in the dark, washed three times in PBS, 0.1% Tween20 for 20 min and once in PBS for 5 m in. Finally, the slides were mounted in Vectashield containing DAPI (Vector Laboratories) and imaged using an Olympus FV1000 confocal microscope.

Antibodies

5-Methylcytosine (5mC), Eurogentec mAb 33D3, 1:150. 5-Hydroxymethylcytosine (5hmC), Active Motif pAb, 1:1000. 5-Formylcytosine (5fC), Active Motif pAb, 1:500. 5-Carboxylcytosine (5caC), Active Motif pAb, 1:500. Stella/PGC7, a kind gift of Prof. Toru Nakano, 1:500. TG1/SSEA1, a kind gift of Prof. Peter Beverley, 1:2. TET1, Millipore pAb, 1:200. TET2, a kind gift of Prof. Guo-Liang Xu, 1:100.

Single cell RNA-seq

Single cell RNA-seq was performed as previously described (*26*), using E6.5-E12.5 PGCs, ESC and E7.5 soma. Briefly, single cells were picked, lysed and RNA was reverse-transcribed with poly-A priming. Following 2nd strand synthesis, cDNA was amplified, sheared and SOLiD P1 & P2 adapters were ligated. Following library amplification two independent PGCs from each developmental timepoint (E6.5-E12.5), as determined by *Blimp1* and *Prdm14* expression, were submitted for SOLiD sequencing.

Glucosyltransferase assay

The method was modified from Quest 5hmC Detection Kit (Zymo Research) to accommodate for 5mC quantification by introducing an additional *HpaII* digestion sample. *MspI* digestion is inhibited by glucosylated 5hmC allowing quantification of 5hmC. *HpaII* is inhibited by both 5mC and glucosylated 5hmC. Therefore subtraction of the *MspI* digest quantification from *HpaII* digest quantification indicates the level of 5mC. The 0% baseline for the assay was set by digesting unglucosylated DNA with *MspI*, while the 100% threshold was set by unmodified and undigested DNA. Four reactions for each sample were set up (glucosylated+*MspI*; glucosylated+*HpaII*; unglucosylated+*MspI*; unglucosylated undigested) each containing equal amounts of DNA. After heat inactivation (10 min at 80˚C) digestion resistant DNA was quantified by quantitative-PCR (qPCR) using a StepOnePlus system (Applied Biosystems) and specific primers that spanned a single CCGG site of interest. DNA loading was controlled for by using a primer set that spans a region lacking a *HpaII/MspI* restriction site (Chr18F, Chr18R). Error bars show standard error of the mean (S.E.M) from triplicate quantifications.

The rate of replication-coupled DNA demethylation curve was constructed using a PGC doubling time of 16hrs (*13*). This gives a function, where the levels of 5hmC and 5mC after time *t* are equal to: $[5hmC+5mC]^0 \cdot 2^{-t/16}$ where $[5hmC+5mC]^0$ represents levels of 5hmC and 5mC at the onset of demethylation, for imprinted genes this is ~50%. It is necessary to evaluate the rate of loss of DNA modification (5mC+5hmC) to account for direct passive loss of 5mC in PGCs (Fig 4B). Significance of fit to observed rates of demethylation was determined by Chi-Squared test.

Bisulfite sequencing

Purified genomic DNA was bisulphite-treated, desulphonated and purified using the EZ DNA methylation kit, as manufacturer's instructions. For locus-specific analysis, a region of interest was amplified using nested primer sets (see primer table) and Hotstar *Taq* DNA Polymerase (Qiagen). A single amplification band was excised from an agarose gel and cloned into pGEM-T Easy vector (Promega) for sequencing. Bisulfite converted reads were analysed using the quantification tool for methylation analysis (QUMA) with default quality control settings. For whole genome bisulfite sequencing, purified PGC genomic DNA was sonicated to ~300bp and end repaired, A-tailed and ligated to methylated TruSeq adapters. Subsequently 400ng of adapter-ligated DNA was bisulfite converted using the EZ DNA methylation kit and amplified for 5 cycles with *PfuTurbo* Cx hotstart polymerase (Clontech) before purification, cluster generation and sequencing. We performed paired-end 50bp sequencing on an Illumina HiSeq and mapped sequence reads to the genome, and quantified methylation state, using the BIFast program.

(h)meDIP with next generation sequencing

Methylated DNA immunoprecipitation (meDIP) and hydroxymethylated DNA immunoprecipitation (hmeDIP) were performed as previously reported, with modifications (*17*). Samples were assayed in biological duplicate and were highly reproducible (Fig. S9). Purified DNA was sonicated to an average size of 150bp (range 100-500bp) with a bioruptor (Diagenode) and equal distribution of DNA fragment size between samples was confirmed with a high sensitivity DNA bioanalyser chip (Agilient). DNA fragments were end-repaired, A-tailed and custom TruSeq adapters (non-methylated) were ligated using the TruSeq DNA sample preparation kit (Illumina). 200ng of purified adapter-ligated DNA was denatured at 100˚C for 5mins, cooled in ice water, and immunoprecipitated in 100ul IP buffer (10mM Na-Phosphate pH7, 140mM NaCl, 0.05% Triton X-100) overnight at 4˚C. For meDIP 1µl of 5mC antibody diluted 1/5 was used for E10.5 PGCs and EpiSC and 1µl diluted 1/20 for E11.5-E13.5 PGCs, to

account for low global 5mC levels. For hmeDIP 1µl of 5hmC antibody was diluted 1/30. DNA-antibody complexes were purified with sheep anti-mouse IgG (meDIP) or Protein-G (hmeDIP) dynabeads (Invitrogen), washed 5 x 10 mins in IP buffer at room temperature and eluted (0.25% SDS, 100ng/ul Proteinase K) for 2 hours at 55˚C. DNA fragments were purified and amplified for 12 c ycles using adapter-specific primers before cluster generation and next generation sequencing. Multiplexed paired-end 36bp sequencing was performed on an Illumina HiSeq. We generated between 33-65 million paired-end reads per replicate sample. Read pairs were mapped to the NCBIm37/mm9 reference genome using BWA 0.5.9 with default settings. The alignment was processed and indexed using SAMTOOLS, duplicate paired-end mappings (which most likely reflect amplification artefacts) were excluded ("samtools rmdup"). Fragments with very low mapping quality (BAM q<10) were also excluded from subsequent analyses. Like ChIP-seq datasets, (h)meDIP-seq profiles provide information on relative distributions as all samples are normalised to total sequencing reads (reads per million). Thus to profile genome-wide loss of methylation, regions of the genome that are both methylated and CpG-dense, such as exons, are most useful as they provide signal over a large dynamic range between the methylated→unmethylated state.

PGCLC induction

We engineered stable ES cell lines carrying doxycycline(DOX)-inducible *Tet1/Tet2* or non-targeting (NT) microRNAs (miR) (see microRNA table), or with constitutive *Tet1*^{CD}*/Tet2* (catalytically active or mutant (*5*)) expression using PiggyBac mediated genomic transposition and antibiotic selection. Stable ES cell lines were maintained in N2B27 medium supplemented with 2i and 1% knockout serum replacement (KSR) (Invitrogen). PGC-like cells (PGCLC) were induced as previously described (*10*) and purified through FACS sorting of distal *Oct4* enhancer GFP-positive cells. Briefly, ground-state ES cells were induced towards epiblast-like cells (EpiLC) on fibronectin in N2B27 with Activin A (20ng/ml), bFGF (12ng/ml) and 1% KSR for two days. PGCLCs were then specified through embryoid body formation of EpiLCs in the presence of cytokines BMP4 (500ng/ml), BMP8a (500ng/ml), EGF (50ng/ml), SCF (100ng/ml) and LIF (1000u/ml). Where relevant, doxycycline (DOX) (650ng/ml) was added to the culture medium after EpiLC induction. GFP-positive cells first appeared after 2 days of PGCLC induction and were FACS sorted at day 6, with an efficiency of derivation of up to 40%. GFP-positive PGCLCs (+ or -DOX) exhibited expression of multiple germ cell-specific markers including STELLA, TG1 and BLIMP1. DNA methylation levels in EpiLCs and PGCLCs were assayed by quantitative bisulfite sequencing in biological duplicate experiments.

Bioinformatics

Quantification of sequence fragments. Sequence data was quantified using PICARD and custom scripts. We counted the number of (h)meDIP fragments overlapping each 50bp window across the genome (quantitation window). We then normalized these raw quantitations by dividing by the number of million mapped and filtered read-pairs (fragments) per library. Thus, (h)meDIP-seq profiles are reads per million per 50bp (normalised tag density).

Distribution relative to genic features. The ensembl gene-set (version 58.37k) was extracted from the ensembl relational database. From the full set, we selected full-length gene bodies (i.e. from the most 5' transcript start to the most 3' transcript end) ranging in length from 5kb to 20kb to exclude artefacts driven by very short or very long genes. We also selected isolated (>1kb from any other exon) internal (>2kb from either end of the transcript) exons of at least 200bp in length as representatives for exonic distribution. We gathered normalized fragment-count data from windows overlapping a given position relative to the feature start or end (or, in the case of gene-body plots) at a given percentage along the length of the gene, and plot the mean normalized count for all windows overlapping each position.

Signal associated with genomic features. CpG island and repeat annotation for the mm9 genome was obtained from the UCSC genome browser download site and mapped to the NCBIm37/mm9 assembly using liftOver. We calculated mean and standard error of normalized tag counts for all 50bp quantitation windows overlapping each target feature class.

Author contributions

(h)meDIP-seq & bisulfite sequencing was performed by J.A.H. Immunofluorescence was performed by J.A.H and R.S. Glu-qPCR was performed by J.J.Z. PGCLC experiments were done by K.M. Bioinformatics and data analysis was performed by T.D and J.A.H. Experimental design and manuscript written by J.A.H and M.A.S. Project supervisor M.A.S.

Supplementary References

25. T. Yoshimizu *et al.*, *Dev Growth Differ* **41**, 675 (1999).
26. K. Q. Lao *et al.*, *J Biomol Tech* **20**, 266 (2009).

Bisulfite sequencing primers

Primer	Sequence (5'-3')	Product (bp)
Dazl_F	TAAAATTTYGAAGGTGGAGTAGAA	670
Dazl_R	CCTTCTCAATATAACTAAAAAAA	670
Dazl_N1F	TAAGGGGGGTTAGGGGATTTA	400
Dazl_N1R	CTAAAAAAAAAAATAAAAAAACCC	400
IAP_All_F	GGYGTTGATAGTTGTGTTTTAAGTGGTAAAT	500
IAP_All_R	ATTCTAATTCTAAAATAAAAAATCTTCCTTA	500
IAP_All_.N1F	GATAGTTGTGTTTTAAGTGGTAAATAAATA	360-411
IAP_All_N1R	ATTCTAATTCTAAAATAAAAAATCTTCCTTA	360-411
IAPchr2_F	TTGTGTAAGGTTTTTATGTTTGT	599
IAPchr2_R	AAAAACTATATCACCTCTCTTC	599
IAPchr2_N1F	TTAAAATAAAAGGAAATGGGGAG	452
IAPchr2_N1R	AACCCCAAATTACTTACAAT	452
IAPchr6_F	GGGTTYGTGTTTTTTTTTTGGT	466
IAPchr6_R	ACTTCATTTCCTTTAATCTCCTA	466
IAPchr6_N1F	GGAATTGTTGATTGGTTTTT	365
IAPchr6_N1R	TAAATTTCTTCCAACACCCTTT	365
IAPchr7_F	YGTATGTGTTAAGGGTAGTTTTT	423
IAPchr7_R	CTTTTTTCTTTTCTTTCTTTCTT	423
IAPchr7_N1F	TTTTTTATTTTATGTGTTTTG	353
IAPchr7_N1R	TTCAACAACCATTTTATAAAT	353
LINE1_F	TTATTTTGATAGTAGAGTT	311
LINE1_R	CRAACCAAACTCCTAACAA	311
Mael_F	TTAGTTYGGTATTTGGTTAATTAG	350
Mael_R	TTTTCCCTCTCAATACTTTAC	350
Mael_N1F	GGGGGTGAAAATAGTTTGAAAT	250
Mael_N1R	CRAAAATATAAAACACCCAAAAACC	250
Sfi1 _F	TTTGTGGGTAGGTAGTATTTTTT	563
Sfi1 _R	CCACTAATTTTCACTATTCTT	563
Sfi1 _N1F	TTTTTATAGAAGAAAGTAGAAGT	401
Sfi1 _N1R	CCRATCCCATACAAAAATTATAAC	401
Srrm2_F	TTGAGTTAATTAGGAAGGTTAGA	611
Srrm2_R	RTCTCCTACTCACAAATAAAA	611
Srrm2_N1F	GGAGAGAGAAAATTAGAATAA	350
Srrm2_N1R	CAAATAAAATCCRAAATCTAAACC	350
Sycp3_F	YGGGAAGATGGATAATTTATAGT	359
Sycp3_R	CATCTAAAAAAACCTAACCCTT	359
Sycp3_N1F	TTTATATAGGTGTGAATGAG	319
Sycp3_N1R	ACCCTTAAACAAAAATAAATAC	319
Vmn2r29_F	ATTTAGAGTTAGATAGAGGGATAGA	446
Vmn2r29_R	CCCRACCAAAATAAAAAAAAACC	446
Vmn2r29_N1F	AGATATAGAGATAGAGAGATAGA	392
Vmn2r29_N1R	AAACCRACCAAACTAATAAATCC	392

Glucosyltransferase-qPCR primers

Primer	Sequence (5'-3')
Dazl _F	GGTCGCCGAGTCACTGAG
Dazl _R	CGCCTATTGGCTGTAGCAC
Igf2r_F	GAGGTGAGGGTTCCACTGAT
Igf2r_R	GCCCAGAAATCTTCACCCTA
Kcnq1ot1_F	TTTAGAATCCGAAGGCCTGA
Kcnq1ot1_R	TTACAGAAGCAGGGGTGGTC
Peg3 _F	ACCAGCAAAGGTGAACAAGG
Peg3 _R	TTGACCCTGGACTGTATCTGG
Peg10_F	GCGGTCACCCAGTGGAC
Peg10 _R	GGGGGCGCTACTTTTATCTC
Sfi1_F	CCAACTTCAAAACGCAAGG
Sfi1 _R	CACCTCTGTCTAGCGTCTGC
Srrm2_F	CAACATCTCGAAGAAGACAGAGG
Srrm2 _R	TTCGAGAACTCTCCTGTCTGG
Vmn2r29 _F	CAGCGGACGCGGACCTGAAC
Vmn2r29 _R	GGCCAGCCGTGGTTCCAGAC
Chr18_F	AACCTCACACACAACAAGCTG
Chr18_R	TGTGATAGGGAGAATGCTTGC

microRNA sequences

miR target	Sequence (5'-3')
Tet1	**TTCTAAAGCAGGCAGGAATGA**GTTTTGGCCACTGACTGACTCATTCCTCTGCTTTAGAA
Tet2	**TAATGAACACAAAGGCAGCCT**GTTTTGGCCACTGACTGACAGGCTGCCTGTGTTCATTA
Non-targeting (NT)	**AAATGTACTGCGCGTGGAGAC**GTTTTGGCCACTGACTGACGTCTCCACGCAGTACATTT

Bold regions represent mature microRNAs.

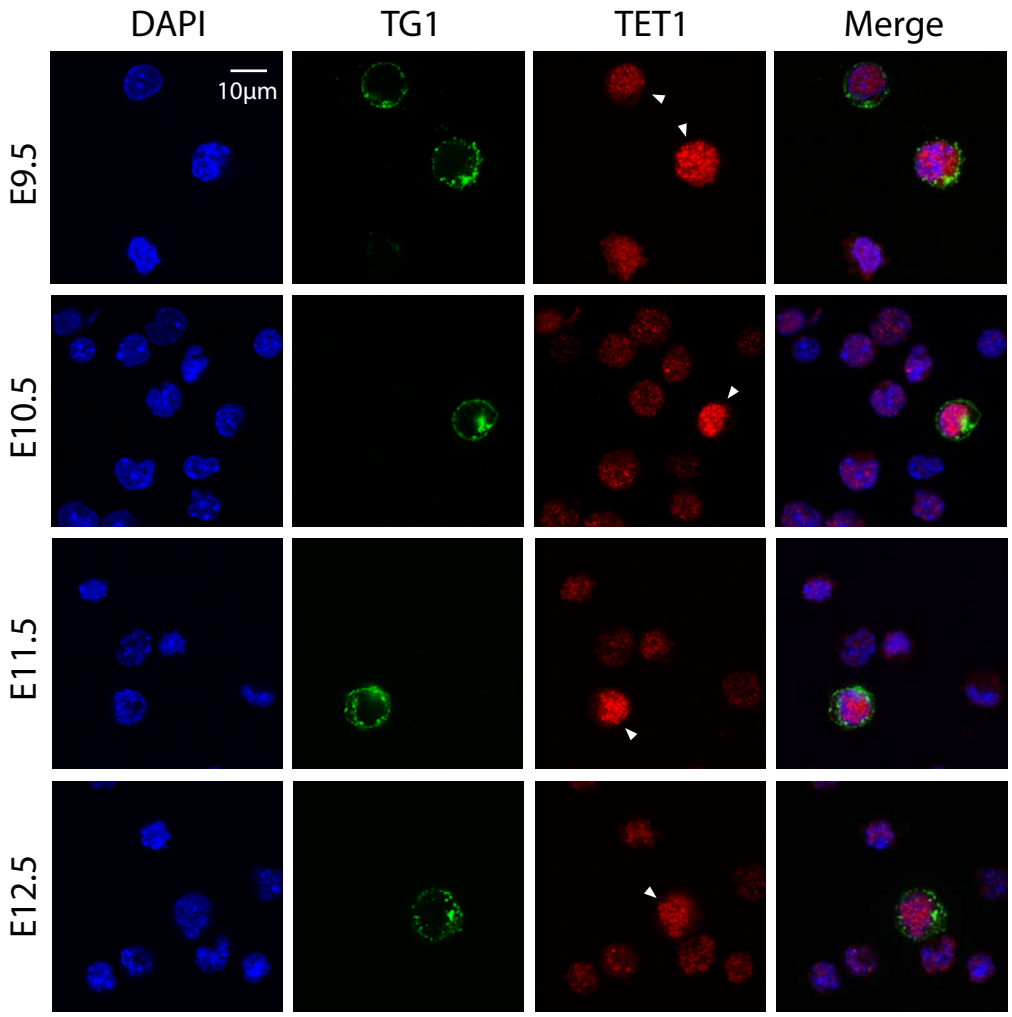


Figure S1. TET1 is upregulated specifically in PGCs between E9.5-E11.5. Immunostaining for TET1 reveals it is highly expressed in PGCs from E9.5 relative to soma, and is predominantly localised to the nucleus. Shown are representative examples that the upregulation of TET1 initiates asynchronously among the PGC population at E9.5 (compare TET1 expression in the two PGCs at E9.5), with almost all PGCs exhibiting strong TET1 staining between E10.5-E11.5 (see also Fig S6). TET1 levels in PGCs thereafter begin to decline in PGCs from ~E12.5. TG1 marks PGCs.

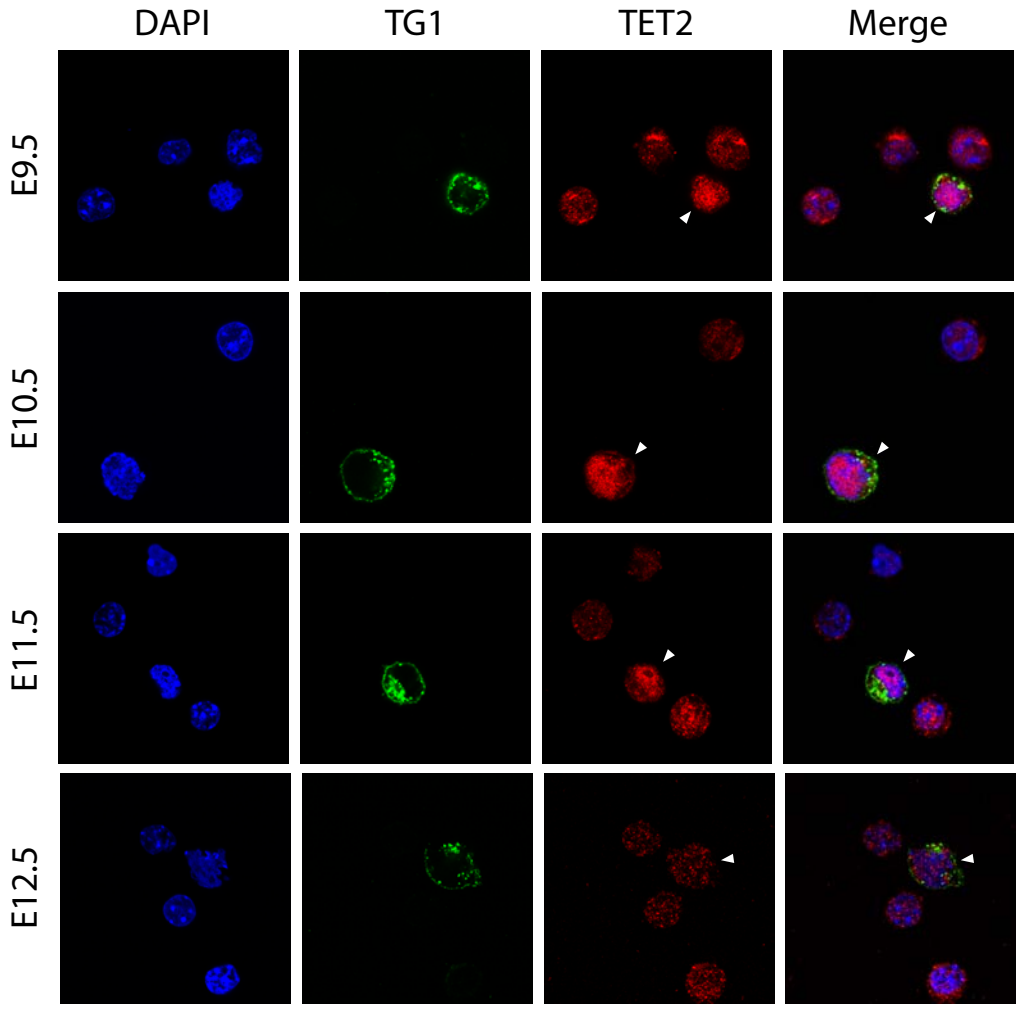


Figure S2. TET2 is upregulated specifically in PGCs between E9.5-E11.5. Immunostaining for TET2 reveals it is upregulated in PGCs at E9.5. TET2 levels thereafter remain high in PGCs and begin to decline from ~E12.5. TG1 marks PGCs.

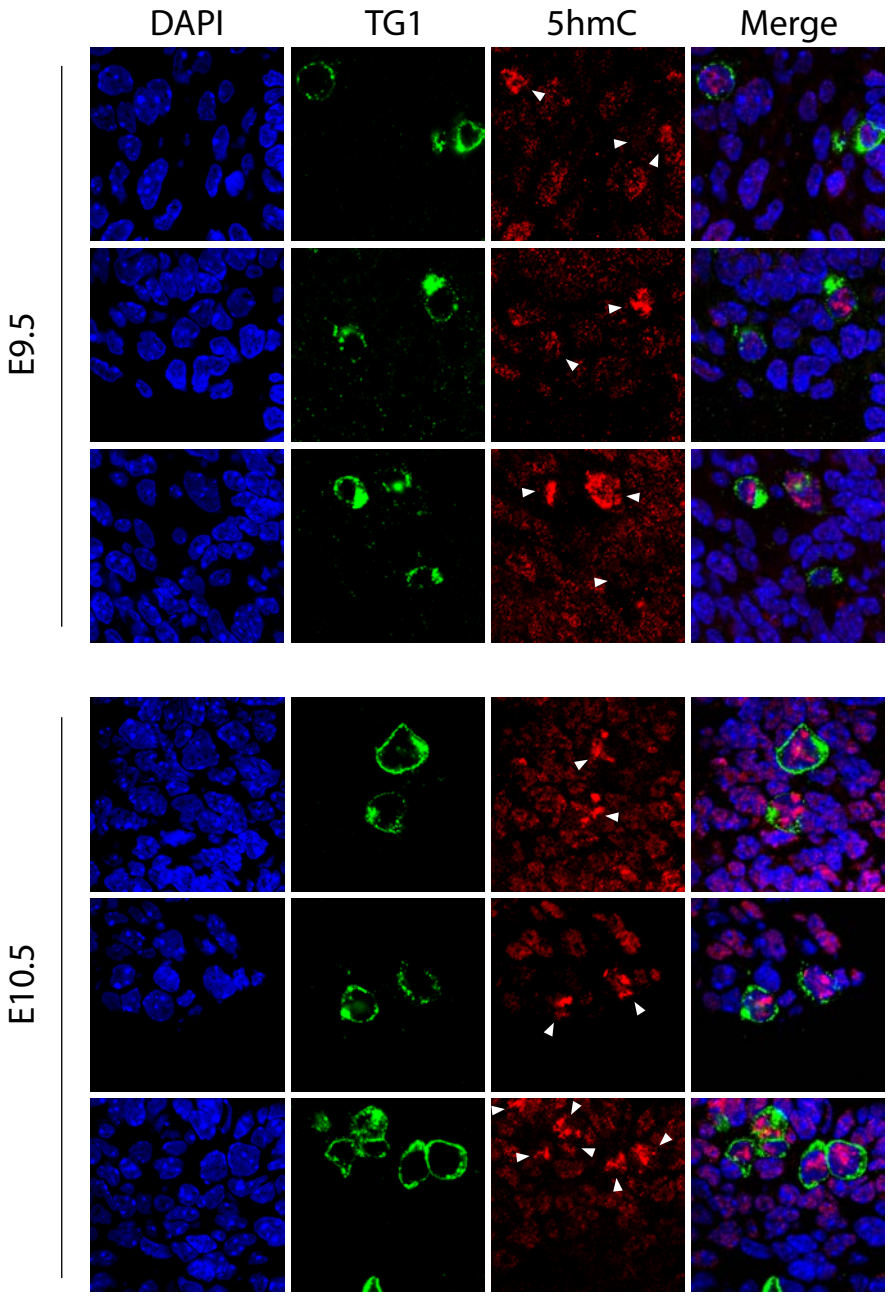


Figure S3. 5hmC conversion initiates asynchronously in PGCs from ~E9.5. Cryosections of urogenital ridges showing that at E9.5 (upper panels), only some PGCs (58%, n=22) (arrowheads) exhibit significantly higher levels of 5hmC than surrounding soma from within the same urogenital ridge. However by E10.5, (lower panels) most PGCs (92%, n=77) are enriched for 5hmC, suggesting that global 5mC to 5hmC conversion starts asynchronously just prior to E9.5 among some PGCs, and most PGCs by E10.5. Note we did not observe 5hmC enrichment (nor significant 5mC erasure) in PGCs at E8.5 (Fig S5). TG1 stains for SSEA1 and marks PGCs.

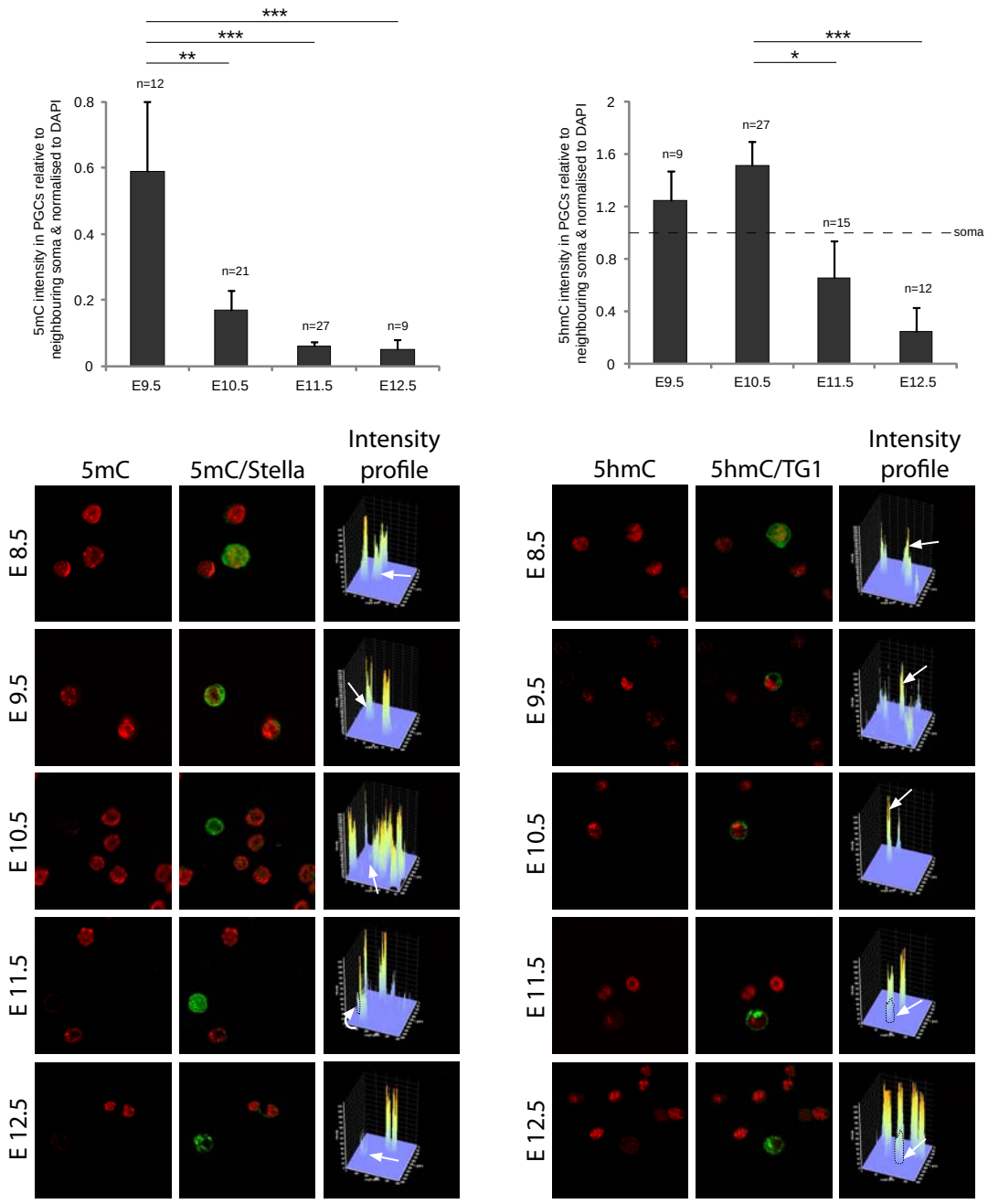


Figure S4. Quantitative intensity profiles for 5mC and 5hmC levels in PGCs. Upper panels: Quantitative IF intensites of 5mC (left) and 5hmC (right) in PGCs between E9.5-E12.5 in single-cell suspension preps. The mean nuclear 5mC/5hmC intensity of PGCs was normalised to DAPI and is shown relative to 5mC/5hmC in neighbouring soma. Lower panels: Representative profiles show 5mC (left) is present in PGCs at E8.5, initiates partial depletiont at E9.5, and with significant erasure from E10.5 onwards. Global 5hmC (right) is at equivalent levels to soma at E8.5 and subsequently becomes enriched in PGCs between E9.5-E10.5 when 5mC is concomitantly depleted. By ~E11.5, 5hmC levels begin to undergo a protracted decline. TG1/Stella mark PGCs. Arrows show PGCs in intensity profiles. *p<0.05; **p<0.01; ***p<0.001; n=number of PGC and soma used to derive intensity.

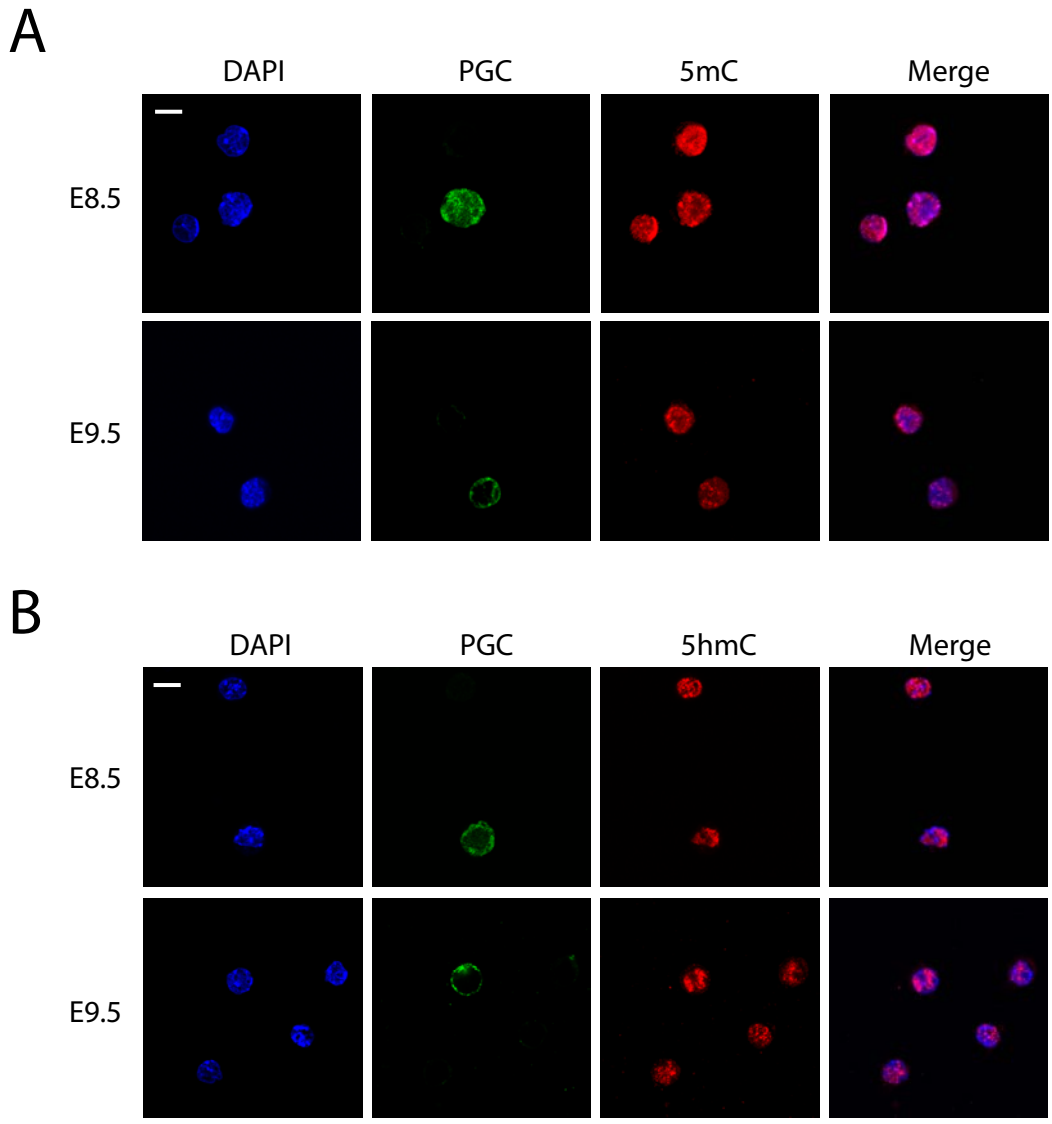


Figure S5. No global changes in 5mC and 5hmC in PGCs at E8.5. (A) PGCs (Oct4-GFP at E8.5; Stella at E9.5) are indistinguishable from somatic neighbours at E8.5. By E9.5 many PGCs have initiated global 5mC erasure, with a clear depletion in all by E10.5 (see Fig. 1B). **(**B) Global 5hmC levels in PGCs (Oct4-GFP at E8.5; TG1 at E9.5) are indistinguishable from somatic neighbours at E8.5 but are enriched in a subset of PGCs by E9.5 (see also Fig 1C). Scale = 10µm.

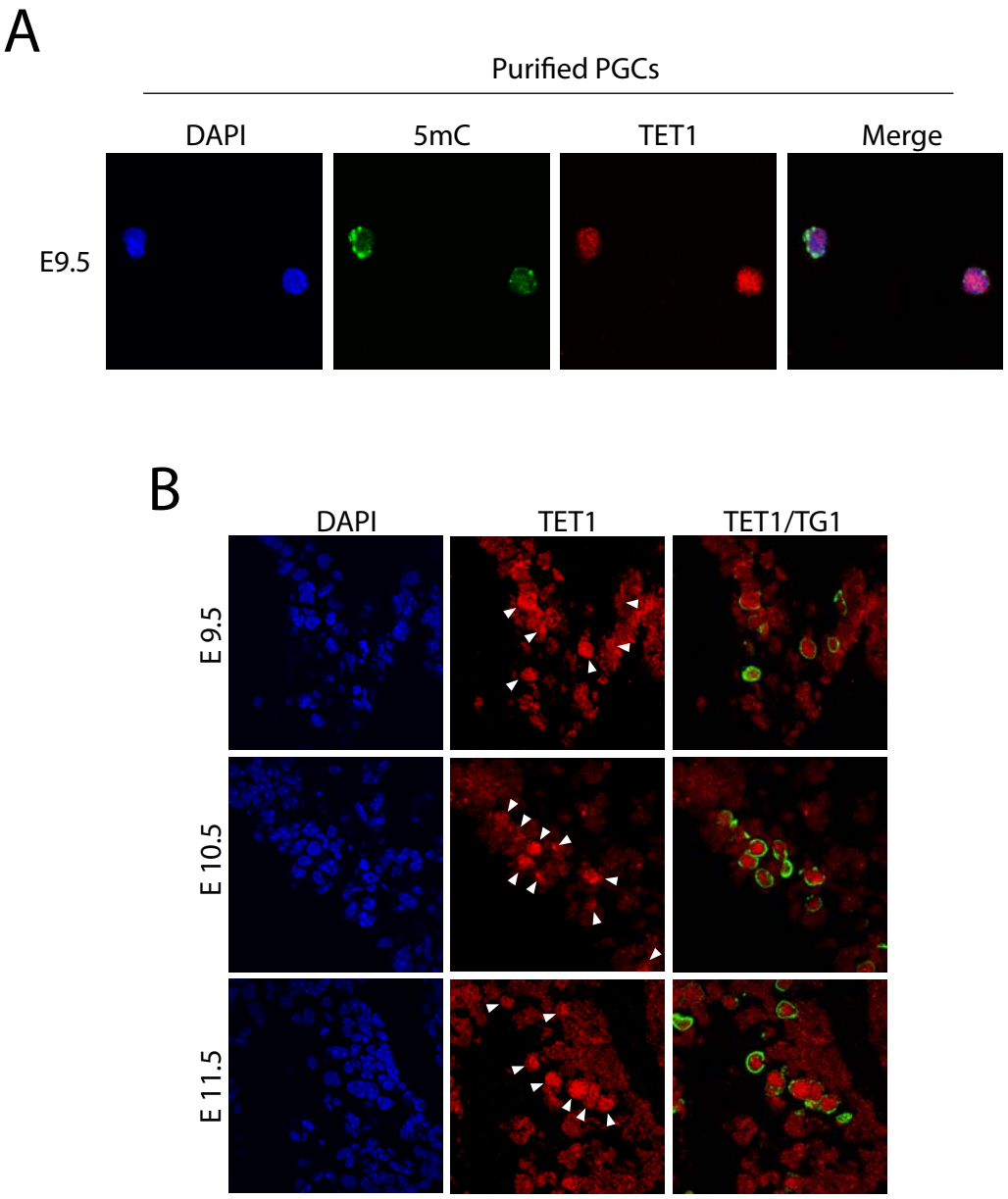


Figure S6. TET1 is upregulated in PGCs at E9.5 coincident with initiation of 5mC erasure. (A) A pure population of FACS-sorted PGCs at E9.5 exhibit asynchronous upregulation of TET1 (red) that is associated with the onset of 5mC erasure (green). See also Fig S1 upper panel for TET1 asychrony. **(**B) Cryosections of genital ridges showing staining for TET1. TET1 is enriched specifically in some PGCs (arrowheads) from E9.5 and most of them by E10.5.

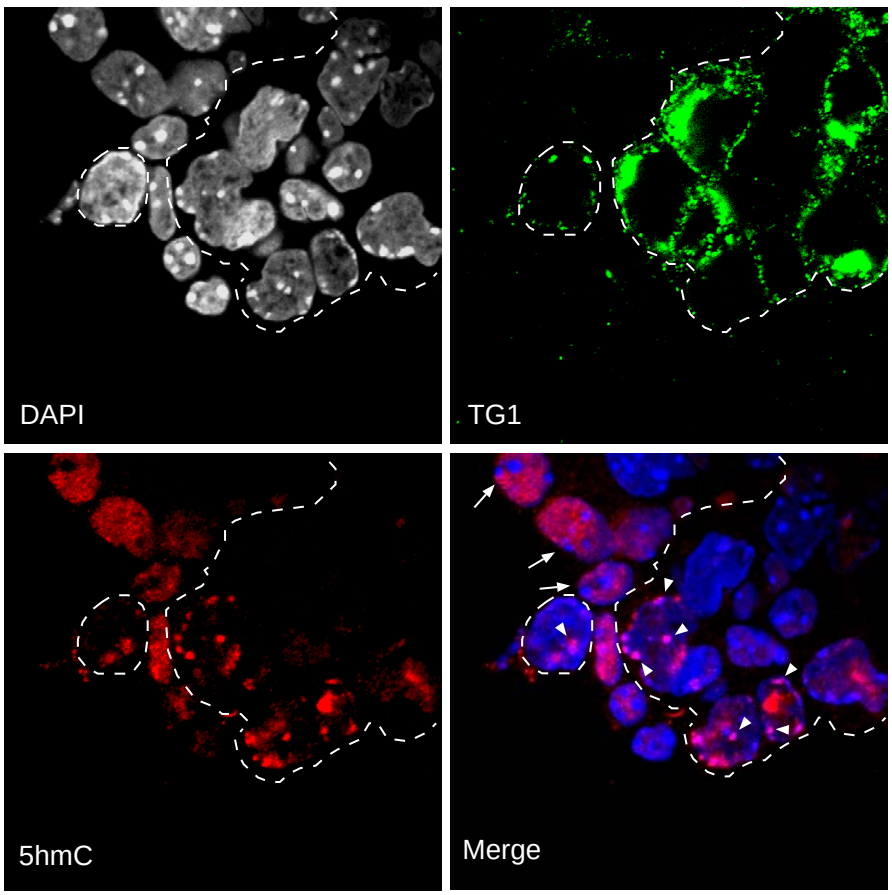


Figure S7. 5hmC co-localises with chromocentres specifically in PGCs. Immunostaining for 5hmC in urogenital ridge cryosections. In PGCs (marked by TG1 and dashed line), most remaining 5hmC at E11.5 strongly overlaps with DAPI-dense chromocentres (arrowheads). In contrast, 5hmC is exclusive of chromocentres in somatic cells (arrows) and ES cells *(6)*. The prevalence of 5hmC at chromocentres, which correspond to 5mC-dense regions of repetitive satellite DNA, suggests a conversion of 5mC to 5hmC at repeats in PGCs, as 5mC staining is also lost from chromocentres in PGCs by this stage (Fig 1B).

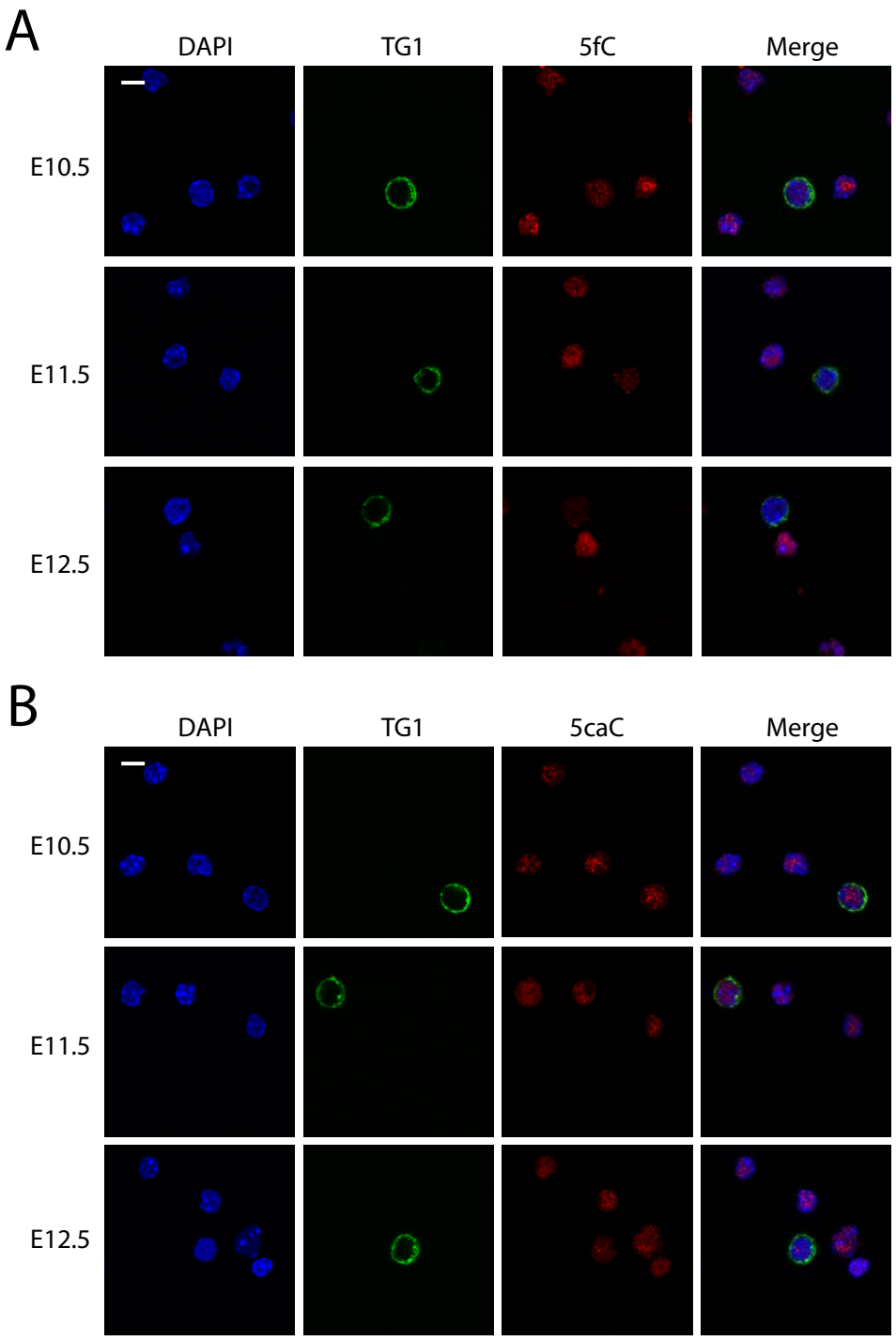


Figure S8. No detectable enrichment of 5fC and 5caC in PGCs. Immunofluorescence of (A) 5-formylcytosine (5fC) and (B) 5-carboxylcytosine (5caC) shows PGCs (TG1-positve) exhibit no global enrichment of these derivatives during the period of demethylation. Additionally, 5fC and 5caC are not observed to overlap chromocentres in PGCs and the overall levels of 5fC and 5caC, as judged by laser intensity required for saturation (and using antibodies with very high avidity), is very low in both PGCs and somatic cells. Scale = 10µm.

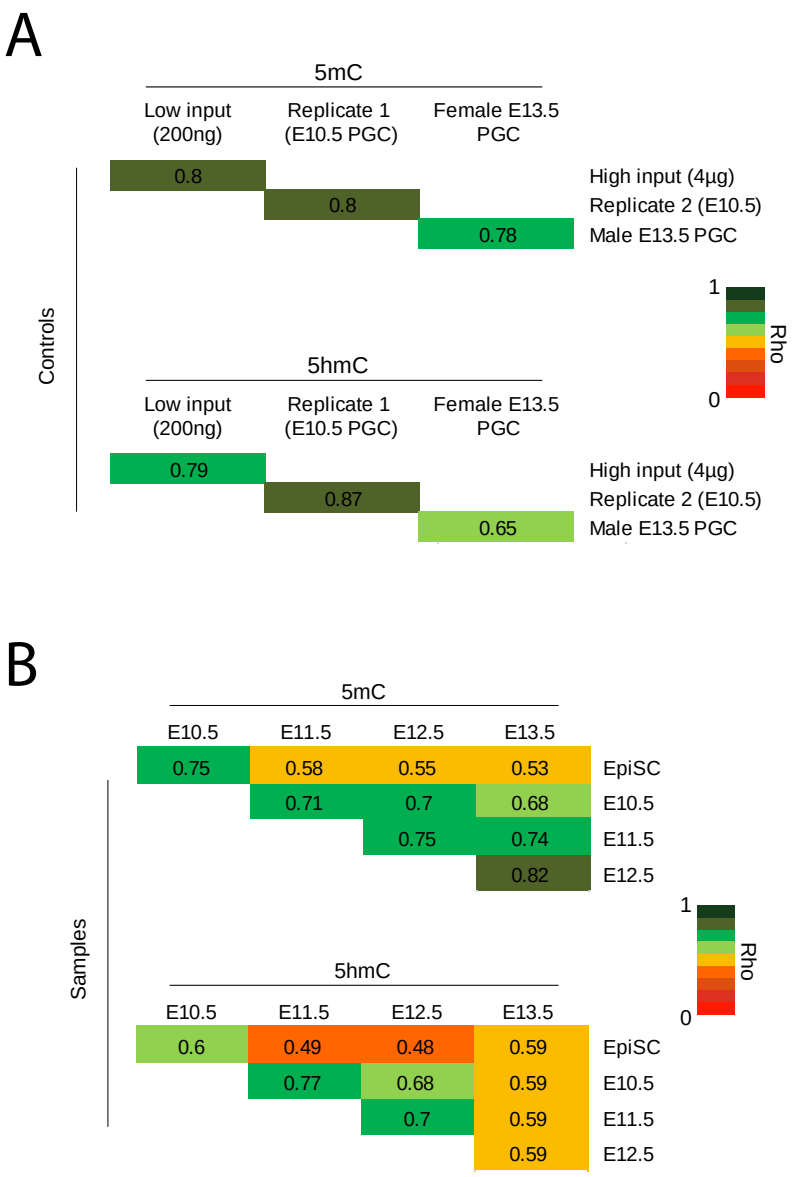


Figure S9. Correlation between control (h)meDIPs and between PGCs. (A) Spearman rank correlation (rho) between control (ESC) low input (200ng) v high input (4µg) samples; replicates 1 & 2 (E10.5 PGCs); and female v male PGCs for meDIP (5mC) and hmeDIP (5hmC). Our optimised low input (h)meDIP protocol, necessary owing to highly limited PGC numbers, generated data highly comparable to that using (h)meDIP with high input (rho = 0.8 and 0.79). Replicates for meDIP and hmeDIP, taken from the most limiting sample (E10.5 PGC) were also highly reproducible (0.8 and 0.87, respectively). (B) Comparison of 5mC correlation between PGCs showed that successive stages of PGC development are increasingly less correlated with EpiSC, as expected owing to ongoing DNA demethylation. Likewise E12.5 and E13.5 PGCs are highly correlated with each other implying that most global 5mC erasure is complete by E12.5. In comparison, the highest degree of 5hmC correlation is between PGCs at E10.5 and E11.5, likely as a result of the global conversion to 5hmC at these stages.

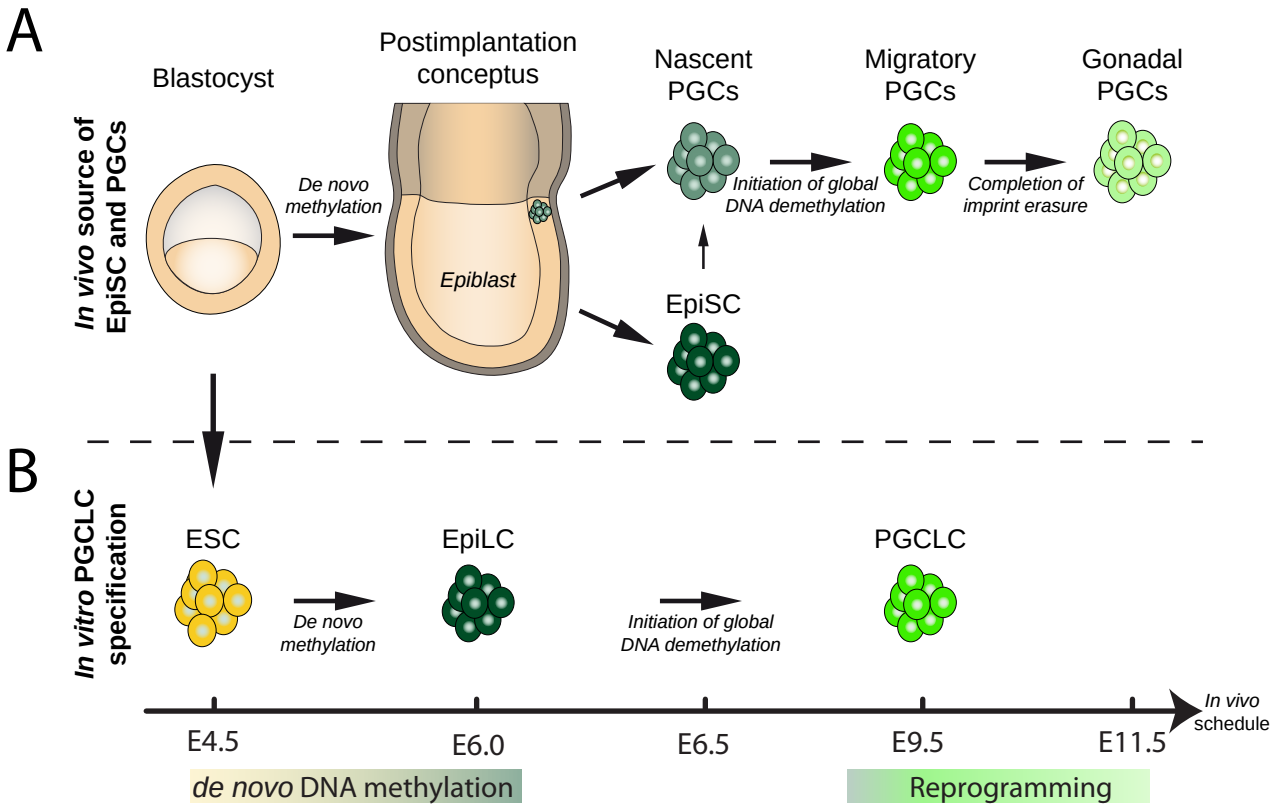


Figure S10. Epigenetic relationship between EpiSCs & PGCs, and *in vitro* derivation of PGCLCs. (A) Both nascent PGCs and epiblast stem cells (EpiSC) originate from post-implantation epiblast cells at ~E6.5, after completion of developmental *de novo* DNA methylation *(17)*. Therefore both EpiSC and nascent PGCs (E6.5-E8.5) likely have established comparable methylomes. The EpiSC 5mC epigenome can therefore be used to infer the starting distribution of 5mC and 5hmC in PGCs prior to reprogramming. This is necessary as early PGCs can't currently be assayed owing to highly limited numbers; ~40-80 per embryo. (B) PGC-like cells (PGCLCs) are derived from pluripotent embryonic stem cells (ESC), which are first induced into highly methylated epiblast-like cells (EpiLC), and then subsequently into PGCLCs *(10)*. PGCLCs undergo similar global DNA demethylation, chromatin reorganisation and gene expression changes to migratory PGCs *in vivo*, with completion of imprint erasure in gonadal PGCs.

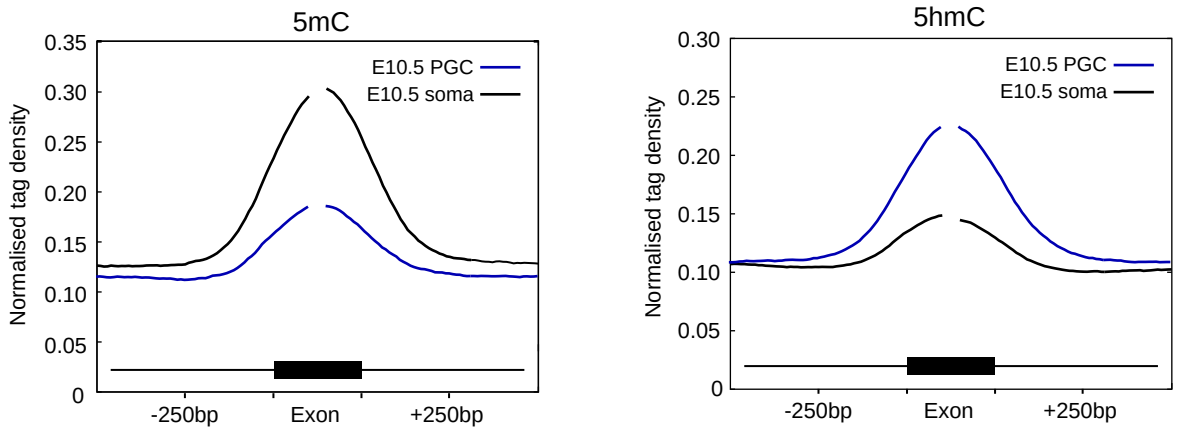


Figure S11. E10.5 PGCs reciprocally lose 5mC and gain 5hmC relative to adjacent E10.5 soma. (h)meDIP profiles showing that by E10.5 PGCs exhibit a significant loss of 5mC signal and reciprocal gain of 5hmC signal at exonic sequences relative to neighbouring somatic cells. This suggests that 5mC is converted to 5hmC as part of a pathway towards DNA demethylation to unmodified cytosine.

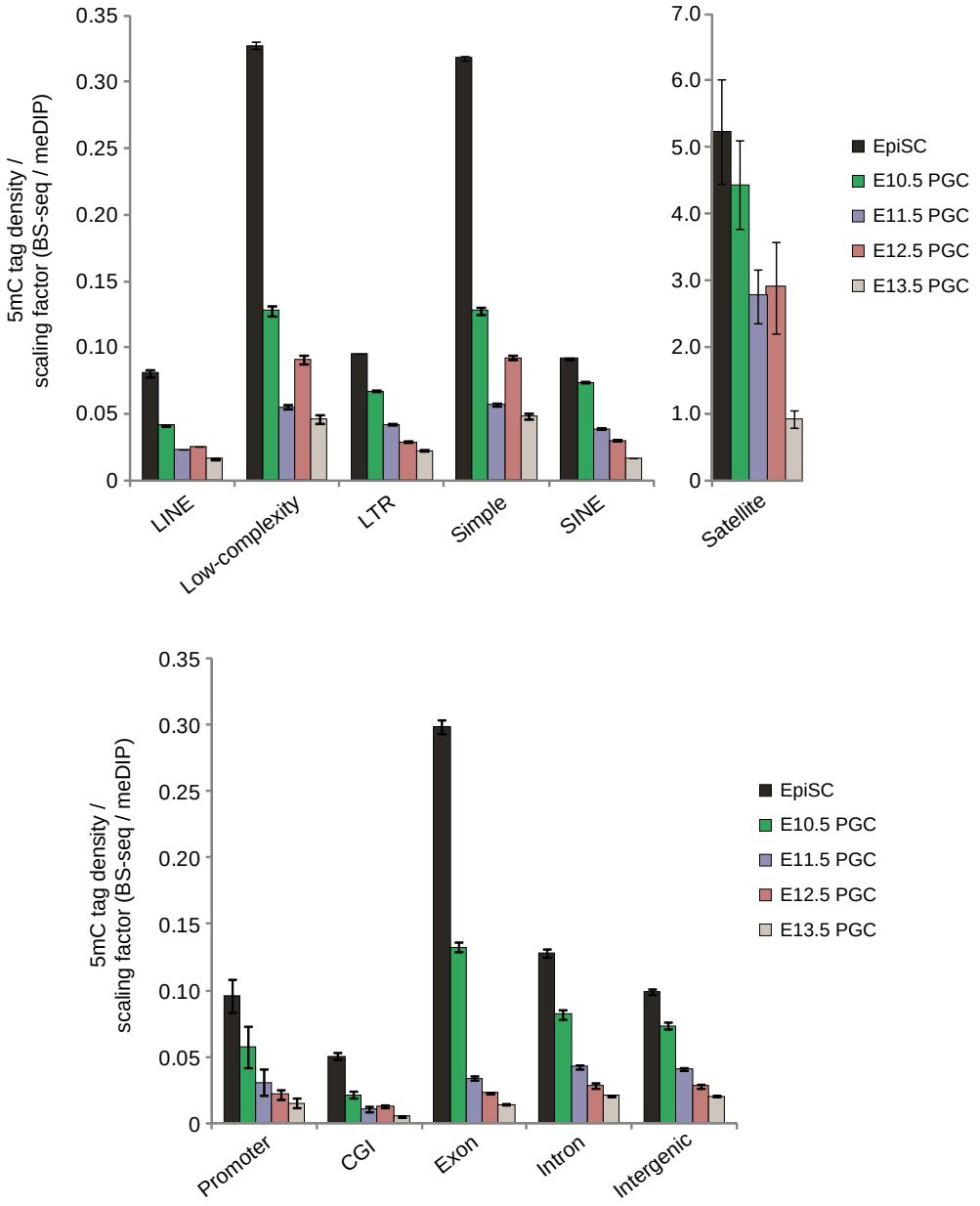


Figure S12. Global DNA demethylation at repeat and genomic landmarks in PGCs. (A) Levels of scaled DNA methylation in PGCs at repeat and retrotransposon sequences. All sequences underwent comparable levels of DNA demethylation. Notably intracisternal A particle (IAP) elements, which form part of LTR-type elements, were at least partially resistant by both meDIP and bisulfite sequencing (Fig. S22), as previously reported *(13)* **(**B) DNA methylation levels according to genomic landmark. Exons are among the most methylated single-copy landmarks in the genome whereas CpG islands (CGI) are among the most hypomethylated, with the exception of some promoter and intergenic CGIs. Because meDIP generates a distribution of 5mC we scaled data between PGCs (Fig. S24) to infer relative demethylation rates.

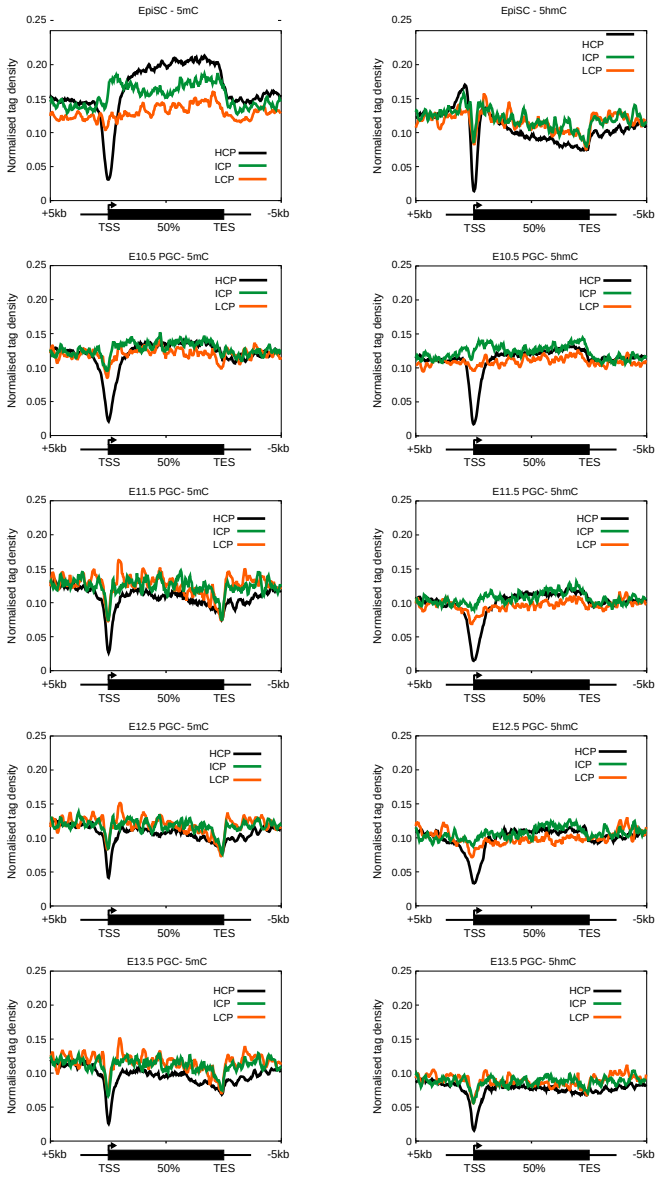


Figure S13. Metagene profiles of 5mC and 5hmC according to promoter CpG-density. Profiles of 5mC and 5hmC at genes with high CpG-density promoters (HCP), intermediate CpG-density promoters (ICP) and low CpG-density promoters (LCP). HCPs are rarely methylated and thus exhibit a discrete trough in EpiSC whereas LCPs are usually methylated but the CpG-density is to low to be efficiently immunoprecipitated by (h)meDIP. ICPs often attract developmental CpG methylation and have a relatively high CpG density and thus exhibit a peak of promoter 5mC that is paralleled by a peak of 5hmC in E10.5-E11.5 PGCs that is not present in EpiSC, indicating conversion to 5hmC. Notably 5mC and 5hmC signal at all promoter classes is progressively flattened in PGCs owing to loss of regions of relative enrichment by global 5mC/5hmC erasure. The preservation of a slight trough at promoters at E13.5 reflects the presence of partially methylated repeats within introns of the metagene, which enriches genic regions relative to comprehensively demethylated promoter regions.

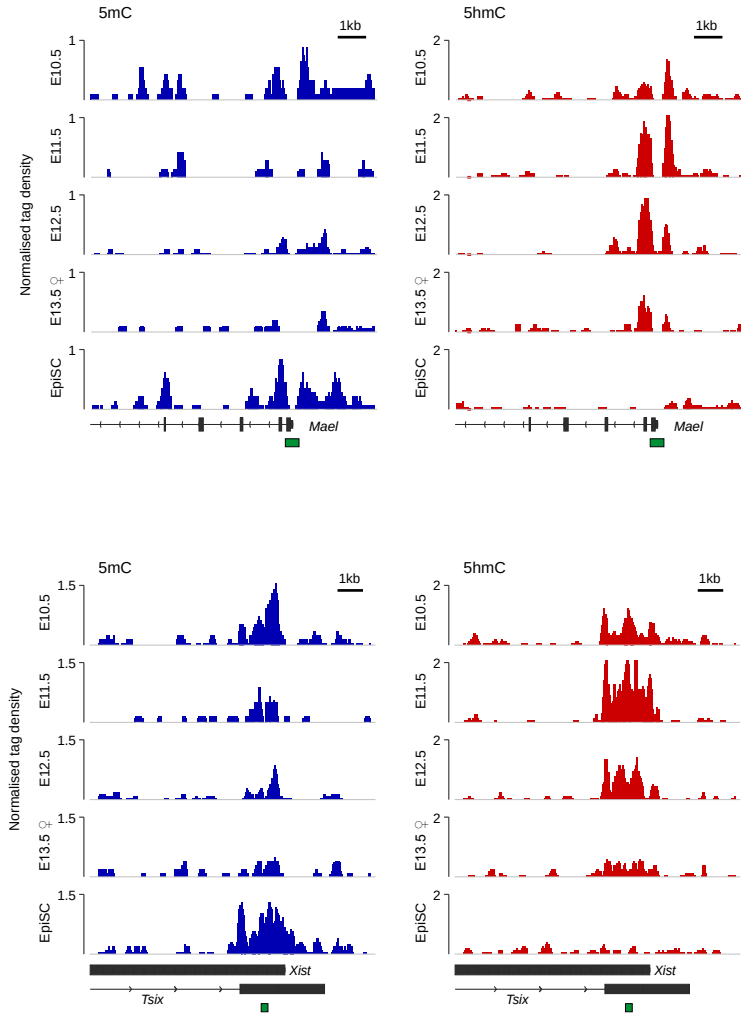


Figure S14. 5mC erasure at germline & *Xist* promoters is associated with 5hmC enrichment. Example profiles of 5mC (blue) and 5hmC (red) at *Mael* and the *Xist* ncRNA in E10.5-E13.5 PGCs, and epiblast stem cells (EpiSC), which indicate the likely 5mC and 5hmC levels prior to the onset of PGC reprogramming. Loss of promoter 5mC signal in PGCs is accompanied by an increase in 5hmC, with 5hmC signal subsequently exhibiting a protracted decline over several cell cycles (E11.5-E13.5). Known methylation-dependent genes such as *Mael*, *Dazl* and others (Fig. S15) are transcriptionally de-repressed in PGCs coincident with promoter demethylation, suggesting this mechanism is likely responsible for their activation for downstream meiotic functions in germ cells. Green boxes indicate CpG islands.

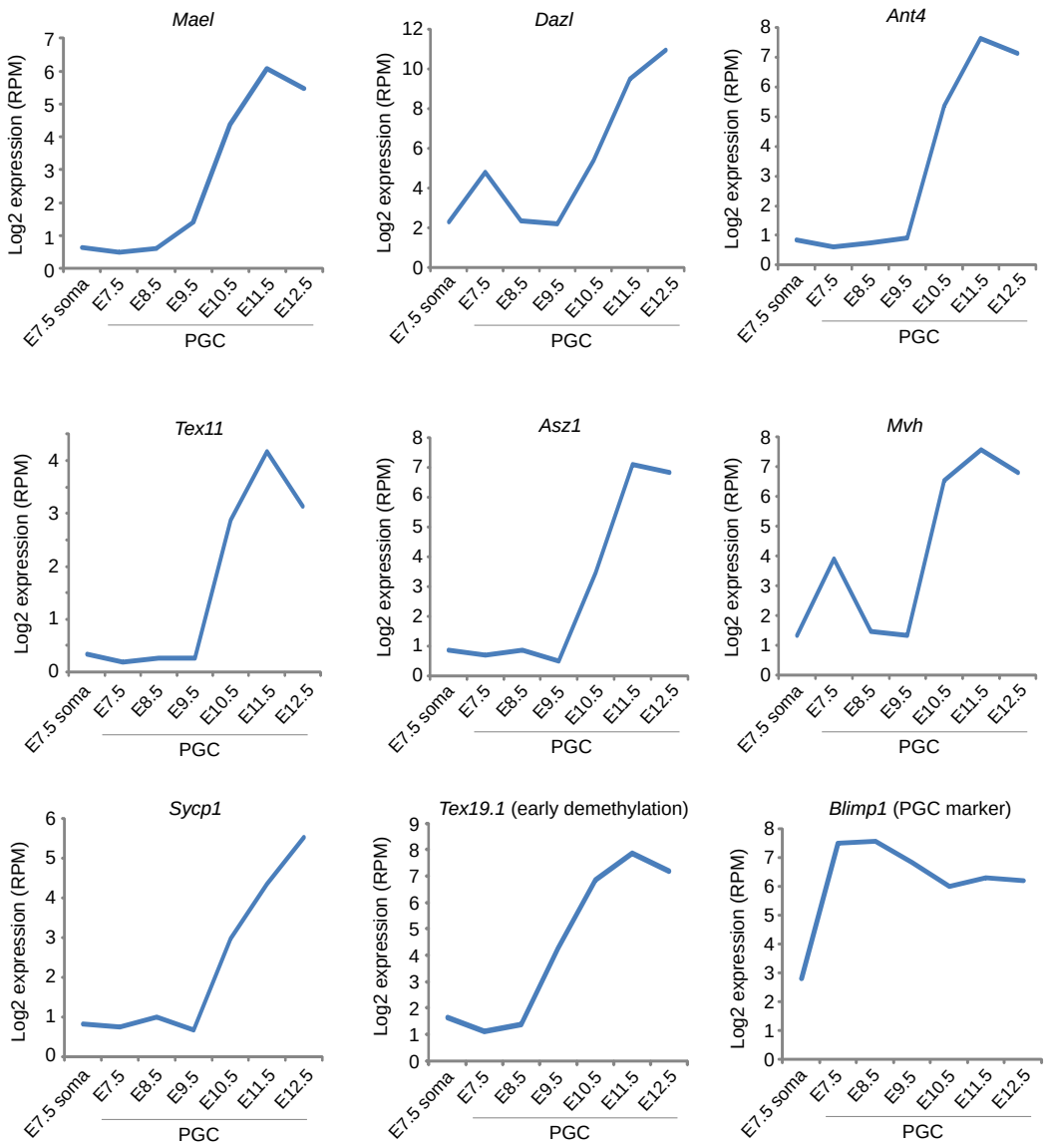


Figure S15. Single cell RNA-seq showing that activation of methylation-dependent germline genes correlates with DNA demethylation in PGCs. Germline-specific genes that are silenced by DNA methylation of their CpG-dense promoters *(8, 9)* are initially de-repressed between E9.5-E10.5 and fully expressed by E11.5. This is consistent with erasure of 5mC initiating at E9.5 in PGCs and being complete by E11.5, and thereby provides a functional validation of the DNA demethylation dynamics we report in PGCs (Fig. 4B). Note that consistent with previous reports *(9)*, *Tex19.1* is activated by demethylation slightly earlier than other methylation-depedent genes during PGC development (prior to E9.5), and is thus one of the first loci to be demethylated. *Blimp1* (also known as *Prdm1*) is a marker of PGC specification that is not known to be regulated by promoter DNA methylation. Shown is Log2 reads per million (RPM) of two replicate experiments.

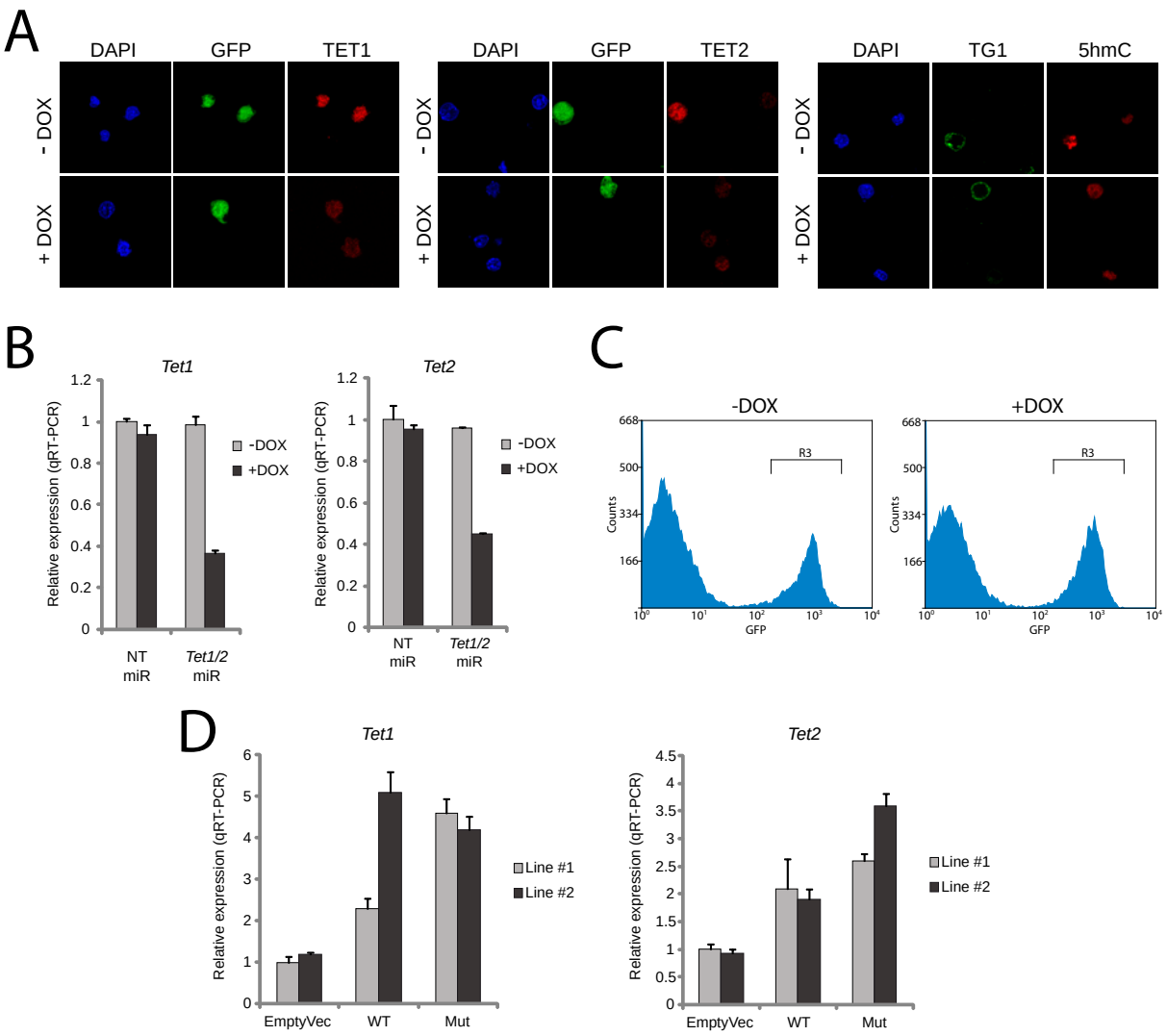


Figure S16. Validation of inducible *Tet1/Tet2* microRNA knockdown and *Tet1/Tet2* constitutive overexpression stable cell lines. (A) Immunofluorescence for 5hmC, TET1 and TET2 in PGCLCs (green) before (-DOX) and after (+DOX) *Tet1/Tet2* miR induction. This confirms that *Tet1/Tet2* miR leads to reduced levels of TET1 and TET2 protein and inhibits enrichment of 5hmC in PGCLCs (compare to adjacent GFP-negative cells). (B) qRT-PCR validating knockdown of *Tet1* and *Tet2* before (-DOX) and after induction (+DOX) of *Tet1/2* miR or non-targeting (NT) miR for 48hrs in stable ES cells (in 2i medium). Note mammalian microRNAs primarily repress targets at the translational level (see A) (C) Representative FACS-profiles demonstrating that knockdown of *Tet1/Tet2* (+DOX) does not reduce the efficiency of PGCLC specification (GFP-positive; R3) from EpiLCs compared to uninduced (-DOX) PGCLCs (-DOX= 24.3%, +DOX= 31.4%). GFP-positive *Tet*-miR PGCLCs +/-DOX exhibited expression of multiple germline-specific markers including TG1 (SSEA1), STELLA and BLIMP1 (data not shown) (D) qRT-PCR showing total *Tet1* and *Tet2* levels (endogenous and exogenous) in stable ES cell lines (in 2i medium) expressing catalytically active (WT) or catalytically inactive (Mut) *Tet1* and *Tet2* or with empty vector (endogenous only). PGCLC induction was performed with 2 independent stable cell lines (#1 & #2) and DNA methylation results averaged (Fig 2G). Error bars represent s.e.m.

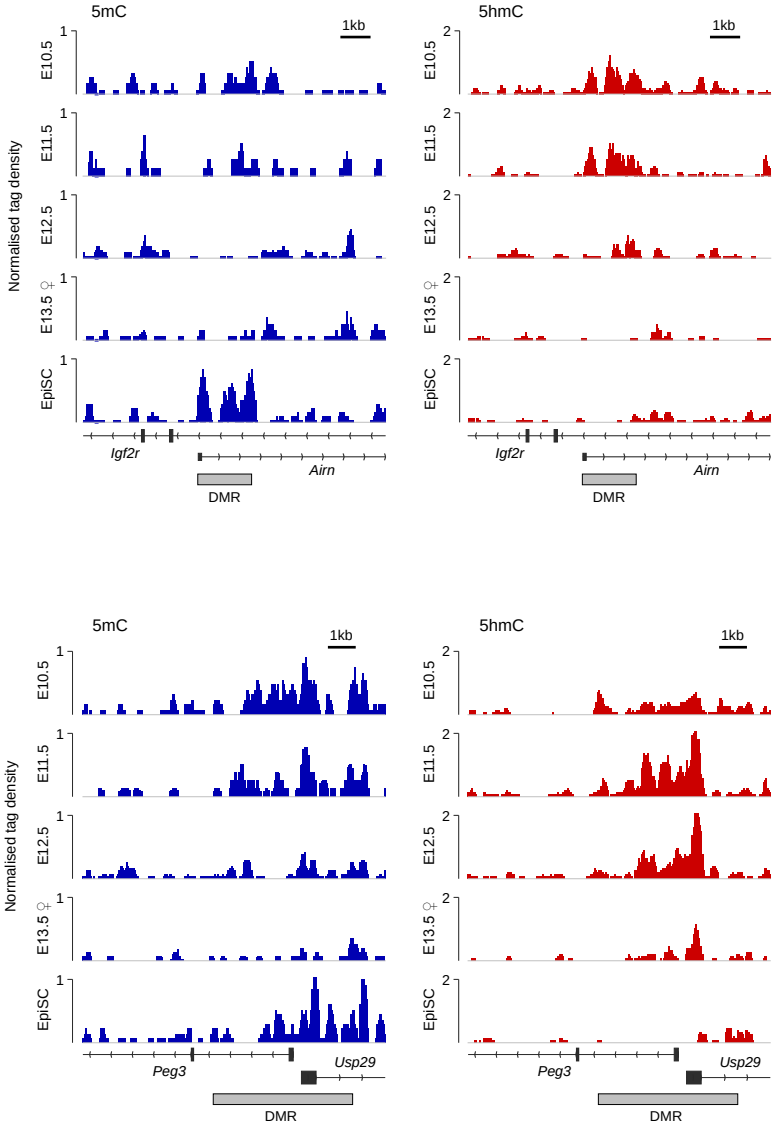


Figure S17. Early and late reprogramming of imprinted genes. Profiles of 5mC and 5hmC at the *Igf2r* imprinted DMR (early) and the *Peg3* DMR (late). Erasure of 5mC at *Igf2r* (upper panel) has commenced prior to E10.5 (compare to EpiSC; which represents ~50% allelic methylation) and is complete by E11.5. In contrast, erasure of 5mC at *Peg3* (lower panel) is delayed until at least E11.5 and there is a concomitant delay in enrichment of 5hmC at this locus.

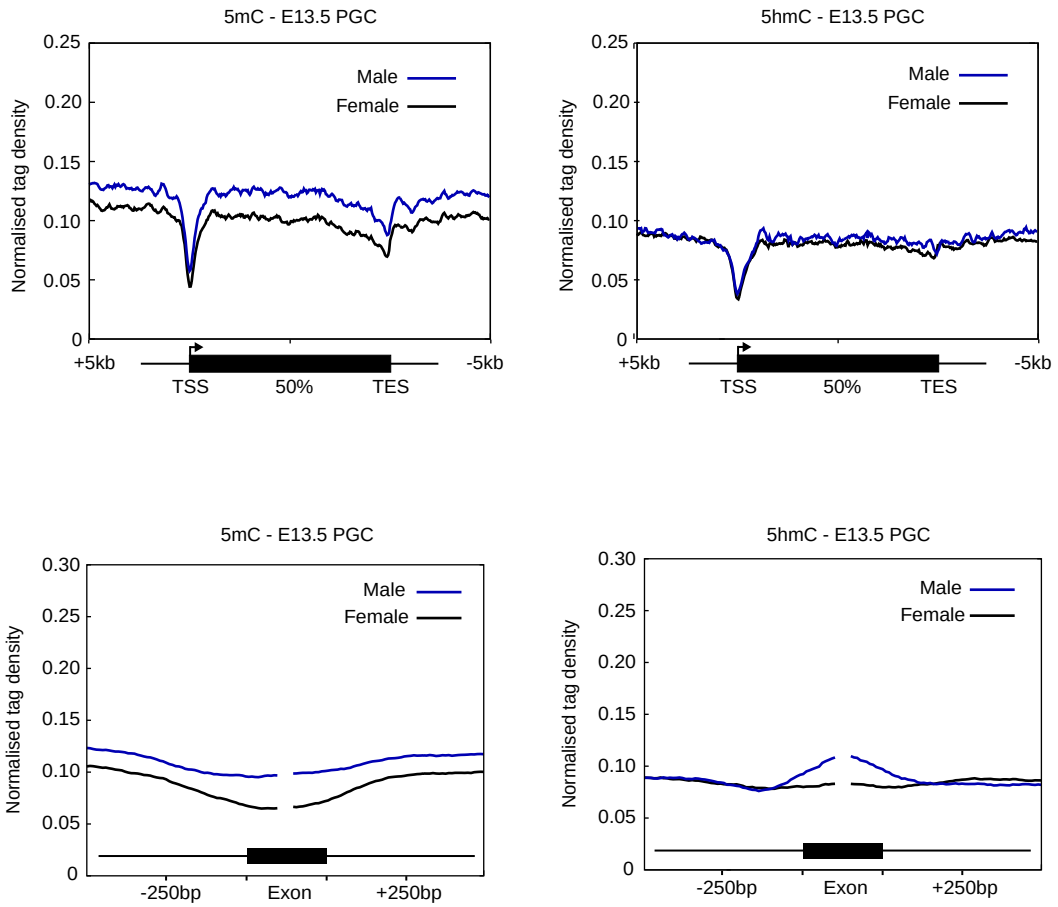


Figure S18. Comparison of male and female PGC 5mC and 5hmC profiles after reprogramming. Profiles of 5mC reveal that male PGCs have higher genic levels of 5mC at E13.5, consistent with previous reports *(14)*. However, it is unclear whether this represents (i) a slower rate of DNA demethylation or (ii) a lesser extent of DNA demethylation, or (iii) initiation of *de novo* remethylation in male PGCs from E13.5. The presence of some 5hmC enrichment in male E13.5 PGCs (lower right) indicates that the difference in 5mC may be due to a slower rate of DNA demethylation via 5hmC. Indeed, we observed consistently fewer male PGCs than female PGCs (up to 2.5-fold) at E13.5 by FACS, suggesting a slower proliferation rate, and thus a potentially slower rate of replication-coupled 5hmC dilution.

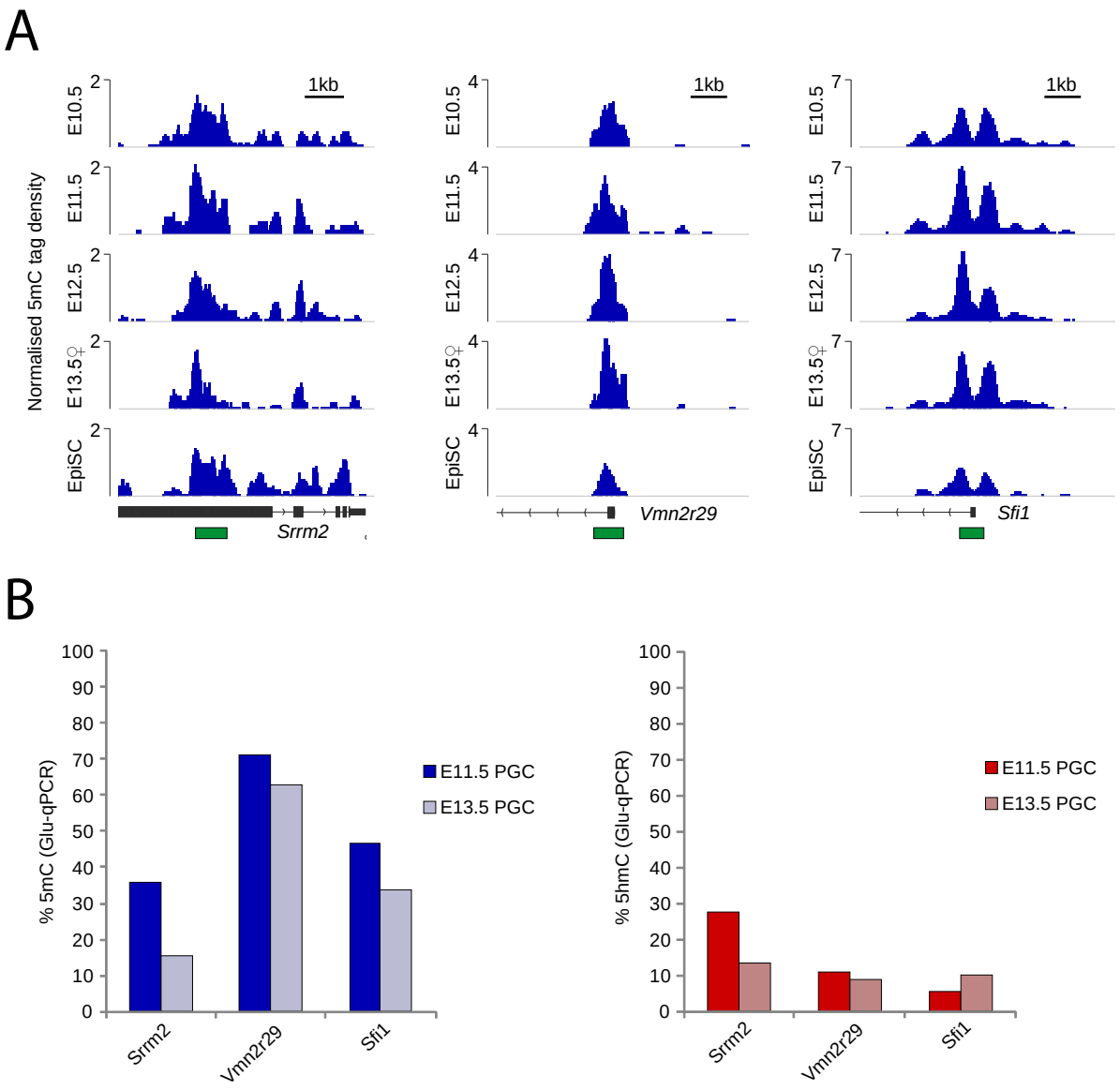


Figure S19. meDIP profiles and Glu-qPCR validation of DNA methylation at loci that escape reprogramming in PGCs. (A) Examples of 5mC profiles at single-copy 'escapees' of PGC reprogramming. Green boxes represent CpG islands (CGI) (B) Quantitative analysis using glucosyltransferase-qPCR confirms that *Srrm2*, *Vmn2r29* or *Sfi1* all retain significant 5mC signal at CCGG motifs in PGCs at E11.5 and E13.5 (female), when the rest of the genome is highly demethylated (~2%). Notably these loci acquire some 5hmC, which is inversely correlated with the levels of 5mC, suggesting that maintenance of 5mC in PGCs is not exclusively due to exclusion of TET proteins and may involve other mechanisms such as reiterative *de novo* methylation.

Associated gene	CGI landmark	CGI size	Chromosome	CGI coverage fraction
Mid1	Intronic	949	X	1
Vmn2r29	Promoter	789	7	1
Sfi1	Promoter	703	11	1
Gm14461	Intergenic	529	2	0.88
mir-715	Intergenic	604	17	0.88
mir-715	Intergenic	3179	17	0.78
Srrm2	Exonic	1029	17	0.45
Pisd-ps2	Promoter	628	17	0.42
Usp11	Intronic	695	X	0.40
Ppp4r1l-ps	Intronic	1344	2	0.40
Gm7120	Intronic	1150	13	0.40

Figure S20. CGIs that maintain significant 5mC meDIP signal after PGC reprogramming. Table of CpG islands (CGI) that exhibit 5mC enrichment in E13.5 female PGC meDIP-seq, shown in reference to the nearest annotated gene. To identify sites of 5mC enrichment at E13.5 we searched for CGIs (USSC CGI set) with >10 tags and a CGI coverage fraction >0.4, to prevent 'bleed in' enrichment from adjacent repeat elements that have partially evaded reprogramming (such as IAP). CGIs are usually associated with promoter or regulatory elements and are thus potentially of functional importance for epigenetic inheritance. We attempted to screen for methylated sites with lower CpG-density but the dual problems of high numbers of partially reprogrammed repeat elements and increasingly reduced meDIP-sensitivity at low CpG-density meant there was a high propensity for false-positives. Nevertheless, the identification of these sites provides proof of principle that the epigenomic state of genomic loci can escape reprogramming and potentially be inherited.

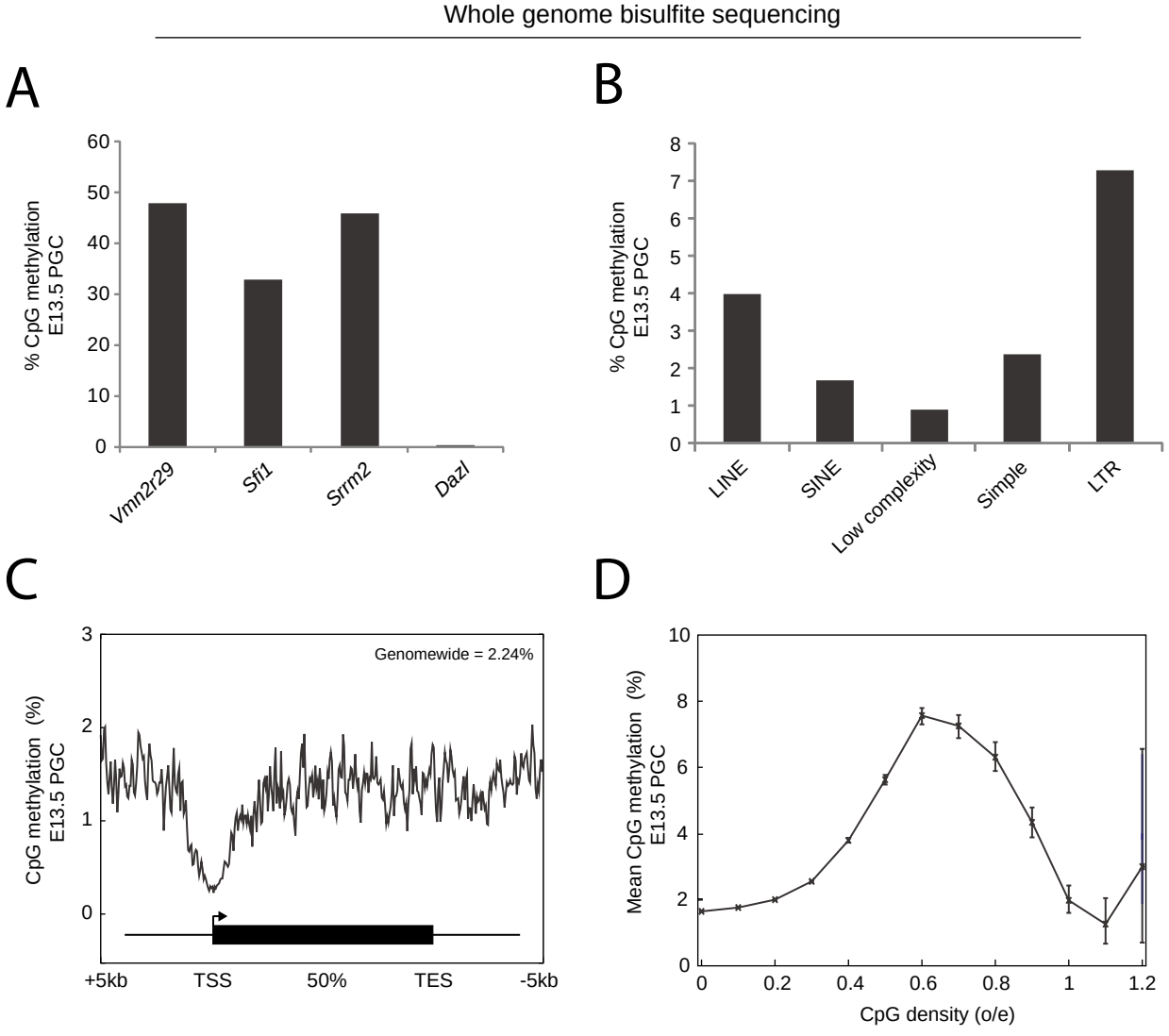


Figure S21. Whole genome bisulphite sequencing of E13.5 PGCs. (A) Validation that the 'escapees' (*Vmn2r29*, *Sfi1*, *Srrm2*) identified by meDIP and locus-specific bisulfite sequencing (Fig 4A & S19) retain high CpG methylation levels in PGCs after reprogramming by WGBS. (B) Global level of CpG methylation at distinct higher classes of repetitive elements in E13.5 PGCs. (C) Metagene showing the distribution of 5mC across coding regions (exon, intron & adjacent) in E13.5 PGCs. The genome-wide methylation is 2.24% compared to ~80% in somatic tissue suggesting a comprehensive erasure of 5mC in PGCs. Notably our analysis suggests PGCs reach a lower overall level of genomic methylation (2.24%) than the previously reported 8-14% *(14)*. We speculate that our higher sequencing depth and/or purer preparations of PGCs may account for this difference. (D) CpG methylation level in E13.5 PGCs according to the underlying CpG-density (500bp windows). Notably an 'intermediate' CpG density is associated with relative resistance to demethylation. This at least in part represents the fact that many repetitive elements (including IAP) have a CpG density ~0.6 but may also suggest an underlying propensity for intermediate CpG-density regions to retain methylation during reprogramming.

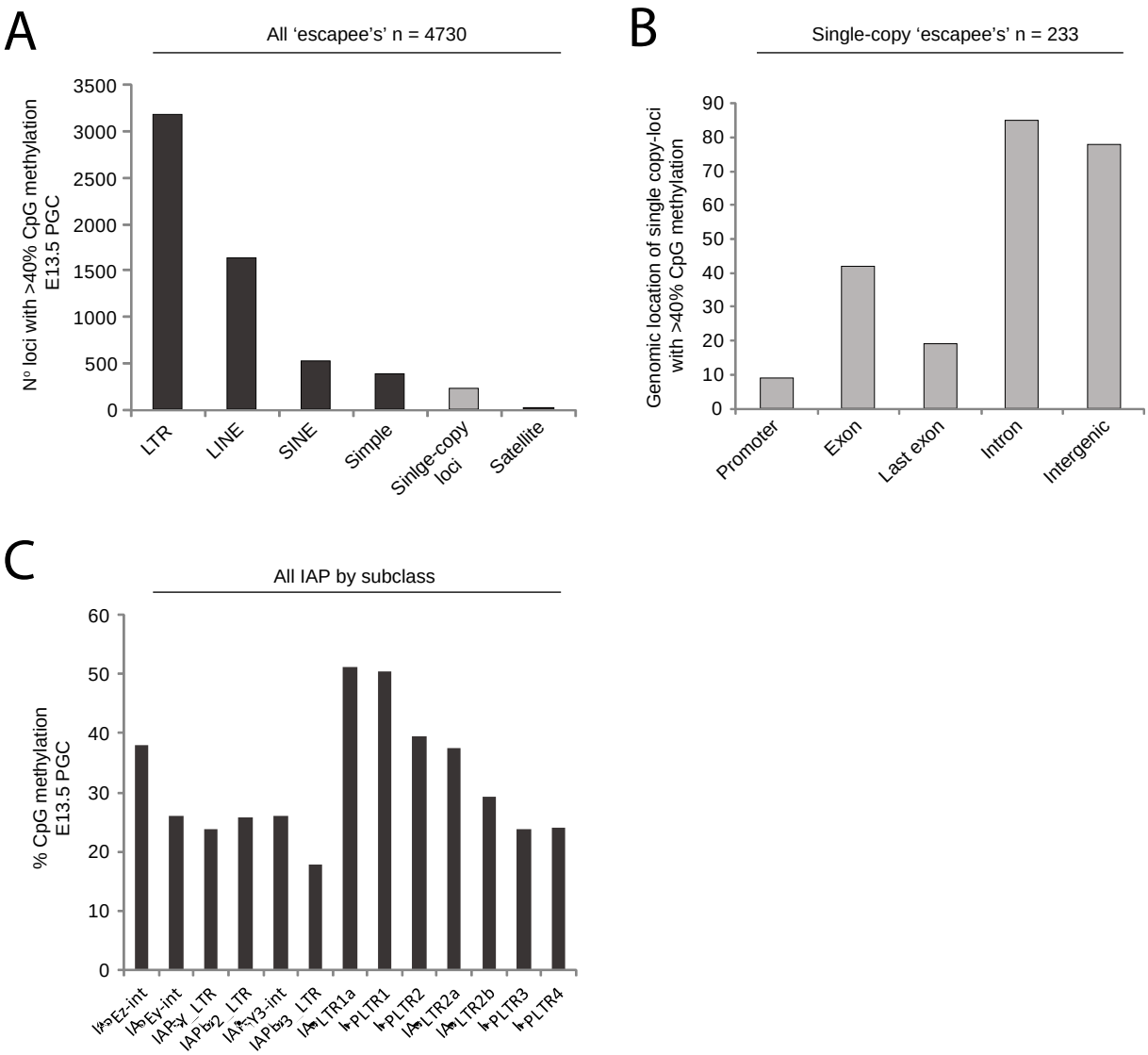


Figure S22. Identification of 'escapees' and the genomic properties associated with evasion from 5mC erasure in PGCs. (A) 4730 loci (sliding 500bp window) retain >40% CpG methylation in E13.5 PGCs after reprogramming. Of these the majority (3176) overlap with LTR-type genomic elements which mainly correspond to intracisternal-A-particle (IAP) retroelements. The 4730 'escapees' include 233 single-copy loci . (B) Genomic landmarks associated with the 233 single-copy loci that escape reprogramming. (C) % CpG methylation at the different subclasses of IAPs in E13.5 PGCs. Notably IAPLTR1 and IAPLTR1a maintain the highest level of methylation. As these are the most active, and hence hazardous IAP subclass to genomic integrity *(15)*, the maintenance of 5mC at these elements may represent an adaptive system to ensure their transcriptional silencing during reprogramming.

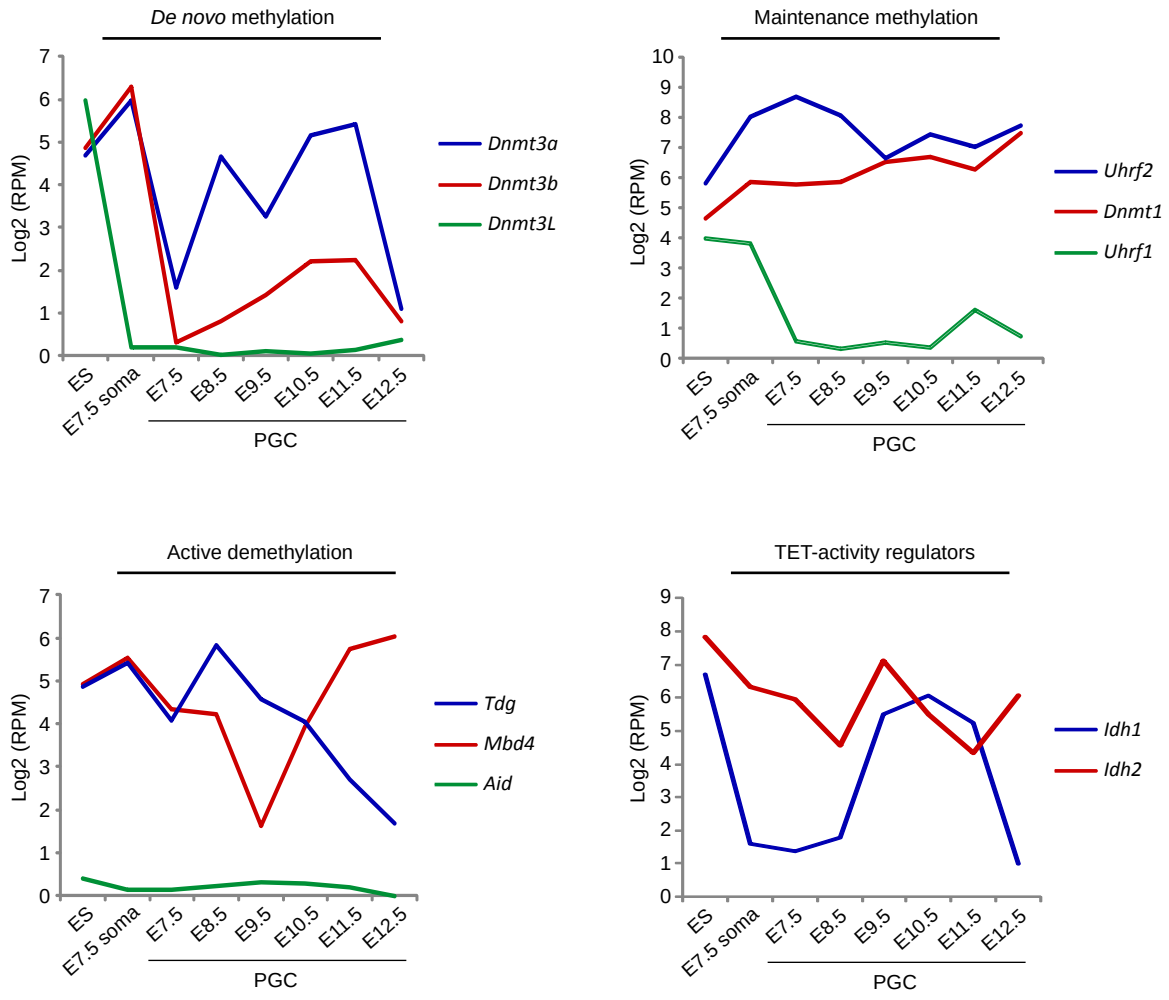
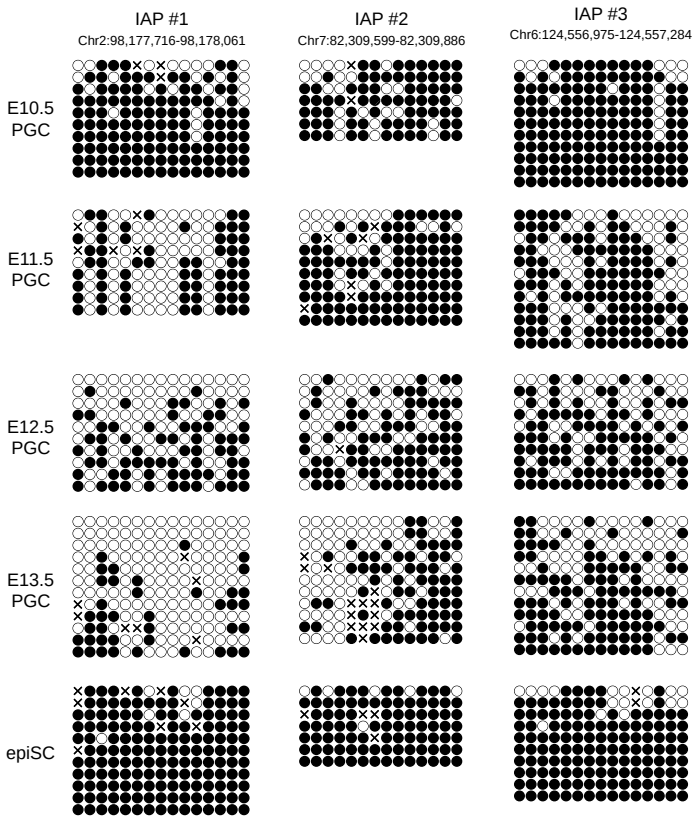


Figure S23. Expression of *de novo* methylation, maintenance methylation and active demethylation components in PGCs. Upper left: key components of the *de novo* methylation machinery (particularly *Dnmt3b and Dnmt3L*) are repressed in PGCs from E7.5 (relative to ES or soma). All *de novo* components are strongly downregulated at E11.5. The absence of *de novo* methylation in PGCs may be a key parallel mechanism to 5hmC conversion that prevents re-methylation following 5hmC depletion. Upper right: The maintenance methylase *Dnmt1* is expressed throughout PGC development but crucially *Uhrf1* is specifically repressed in PGCs. In the absence of *Uhrf1*, cells are unable to effectively maintain DNA methylation. The lack of maintenance methylation may be partly redundant with 5hmC conversion in PGCs, by promoting direct passive loss of 5mC. Notably, the upregulation of *Uhrf2* may partially compensate for loss of *Uhrf1 and* may explain how 5mC is maintained at specific heterochromatinised loci and IAP elements, as *Uhrf2* preferentially binds H3K9me3 marked regions. Lower left: Potential components of a parallel active demethylation system in PGCs. Lower right; Expression of *Idh* genes in PGCs. IDH proteins metabolise isocitrate to α-ketoglutarate, an essential cofactor required for TET enzymatic activity. Note *Idh1* is strongly upregulated coincident with initiation of 5hmC conversion by TET1/TET2 in PGCs (E9.5). *Idh2* expression also peaks at E9.5.



	Bisulfite (% methylation)			meDIP (Normalised tag density)			Average ratio (meDIP/Bisulfite)	Normalised scaling ratio
	IAP#1	IAP#2	IAP#3	IAP#1	IAP#2	IAP#3		
E10.5 PGC	83%	76%	89%	0.37	0.46	0.97	0.71	**1.37**
E11.5 PGC	49%	79%	67%	0.46	0.65	1.51	1.34	**2.57**
E12.5 PGC	44%	61%	61%	0.49	0.71	2.19	1.96	**3.75**
E13.5 PGC	23%	51%	57%	0.63	0.56	1.69	2.28	**4.36**
EpiSC	96%	94%	89%	0.33	0.53	0.59	0.52	**1.00**

Figure S24. Locus-specific IAP CpG methylation in PGCs & scaling ratios for meDIP at repeat elements. meDIP-seq generates a distribution of 5mC and therefore regions that are unmethylated in PGCs at both E10.5 and at E13.5 will appear to gain enrichment of 5mC by E13.5, as the region is no longer depleted of 5mC relative to the rest of the (demethylated) genome. Similarly a region that becomes demethylated between E10.5 and E13.5, will not exhibit demethylation in an meDIP-seq profile unless the extent of demethylation is greater than the genomic average, for example at CpG-rich exons. To account for this at genome-wide repeat elements, we normalised 5mC levels using single-locus IAP elements, which escape 5mC erasure. Thus, while quantitatively measured IAP CpG methylation (bisulfite sequencing) goes down slightly between E10.5 and E13.5, the meDIP-seq signal goes up, as IAPs are relatively more methylated than the rest of the genome at E13.5. The ratio between these two measurements generates a scaling factor, which can normalise meDIP data, analogously to a reference gene in qRT-PCR. This scaling factor was only applied to Figure S12, to determine the relative rate and extent of genome-wide 5mC loss in meDIP-seq distributions. Genome-wide bisulfite sequencing (Fig S21-S22) quantitatively validates the extent of DNA demethylation in PGCs.