Synergistic Mechanisms of DNA Demethylation during Transition to Ground-State Pluripotency

Jamie A. Hackett, Sabine Dietmann, Kazuhiro Murakami, Thomas A. Down, Harry G. Leitch, and M. Azim Surani

Supplementary Inventory

Supplementary Experimental Procedures

(hydroxy)methylated DNA Immunoprecipitaion-seq ((h)meDIP-seq)

To facilitate comparison with PGC (h)meDIP-seq, analysis of PSC 5mC and 5hmC was performed as described (Hackett et al, 2012; Harris et al, 2010) using a high resolution and low-input protocol. Purified genomic DNA was sonicated to an average size of 150bp (range 100-500bp) with a bioruptor (Diagenode) and an 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). 250ng 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 hmeDIP 1µl of 5hmC antibody was diluted 1/10. 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 cycles using adapter-specific primers before cluster generation and next generation sequencing. Multiplexed paired-end 36 sequencing was performed on an Illumina HiSeq with between 33 million and 56 million reads generated per 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 and duplicate paired-end mappings were excluded ("samtools rmdup"). Fragments with very low mapping quality (BAM q<10) were also excluded from subsequent analyses. Sequencing data has been deposited in the sequence read archive under accession SRA111995.

Quantitative RT-PCR

For gene expression analysis, RNA was purified from cell lines using the RNAeasy Mini kit (Qiagen), with on column DNase treatment included. cDNA was reverse transcribed with SuperScriptIII RT (Invitrogen) using random priming of 1µg of purified RNA. cDNA was diluted 1:10 and used in 1x Jumpstart SYBRgreen qPCR mastermix (Sigma) in triplicate. Amplification was achieved using gene-specific primers and cycling conditions of 2min at 95˚C; 40 cycles of 15sec at 95˚C and 1min at 60˚C; followed by a melting curve of 60 to 95˚C (0.3˚C increments) on a StepOne Plus qPCR cycler (Applied Biosystems). Primer sequences have been reported previously (Gillich et al, 2012). Standard curves were constructed and gene expression was normalised to *Gapdh* using the Pfaffl method.

Methylation-Sensitive Glucosyltransferase-qPCR (GluMS-qPCR)

Quantitative DNA modification levels of 5hmC were determined using the Quest 5hmC Detection Kit (Zymo Research), utilising an additional HpaII digestion to establish 5mC levels. Briefly, the 0% baseline for the assay was set by digesting unglucosylated DNA with MspI, while the 100% threshold was set by mock-modified and undigested DNA. Four reactions for each sample were set up (glucosylated+MspI; glucosylated+HpaII; unglucosylated+MspI; unglucosylated undigested) according to the manufacturer's instructions, each containing 250ng of DNA. After heat inactivation (15 min at 80℃) digestion resistant DNA was determined by quantitative-PCR (qPCR) using a StepOnePlus system (Applied Biosystems) and specific primers that spanned a single CCGG site of interest. Primer sequences were: *Esrrb*, F: CCCCTCTCCACAGACCTGTACT, R: ATGGCTTCTTGAGTGGCGTAG; *Cul1 (intron)*, F: CAGACCACTAACAGGTGTCTAAATGA, R: AAACAAACATGGCATTCAAAAG; *Capn9*, F: CCACTGTGCATCAACAGAGGTA, R: GTGAGACCATTGCCCACTCA; IAP, F: TGGTCTGTGGTGTTCTTCCT, R: AATTCGGCACCAATTGTTATT; *Dazl*, F: GGTCGCCGAGTCACTGAG, R: CGCCTATTGGCTGTAGCAC; *Asz1*, F GGCTCACAGCCACTTCCTC, R: AGCCTTCGTGAGATGTGACC; *Rhox9*, F: TTTTCTTCTGGCGCCTCAC, R: AGTCAGGACAGACCGCCTCT; *Foxo1 enh*, F: TGCAGAATAAAACACTCTTTGCA, R: AAGGGTCAATGGTTCACTGGA; *Klf4 enh*, F: AGTGGATAGGCTGGTTCTGGG, R: AATCAGCAGAGACGACCAGGG; *Gck enh*, F: CTGAAAACCACGGGCCTCTAG, R: TAGGGGTGTTGTGCTGTTCCA; *Nanog enh*, F: ACATTTGCTCCCTTGGGTTCA, R: CAAGTGCTGGCTGGGATTAAA; *Chr18*, F: AACCTCACACACAACAAGCTG, R: TGTGATAGGGAGAATGCTTGC. DNA loading was controlled for by using a primer set that spans a region lacking HpaII/MspI restriction site (Chr18F, Chr18R). Error bars show standard deviation from duplicate biological quantifications.

Western blotting

Embryonic stem cell extracts were isolated using RIPA buffer and western blots were performed as described (Dunican et al, 2008). Antibodies were α-PRDM14, 1:500 (a kind gift of Prof. D Reinberg); α-DNMT3A, 1:250 (Imigenex); α-DNMT3B, 1:250 (Imigenex); α-TUBULIN, 1:1000 (Sigma).

5mC and 5hmC Enzyme-Linked ImmunoSorbant Assay (ELISA)

Global 5mC and 5hmC levels were determined using the MethylFlash methylated DNA and Methylflash hydroxymethylated DNA quantification kits (Epigentek) according to the manufacturer's instructions. Briefly, sample and control DNA (100ng) were immobilised in assay wells in triplicate and captured using 5mC- or 5hmC- specific antibodies and subsequent colorimetric detection antibody. After incubation OD values were measured using a Victor3V fluorometer (Perkin Elmer).

Absolute quantitative levels were determined by constructing a standard curve of known 5mC and 5hmC values. Assays were performed in biological triplicate.

Bisulfite Sequencing

Genomic DNA was purified from cells using the DNeasy blood and tissue kit (Qiagen). Bisulphite conversion and desulphonation was Nested primer sets were used to amplify a specific region of interest with Hotstar *Taq* (Invitrogen). A single band was excised, gel extracted, and cloned into pGEM-T-easy (Promega) for sequencing. Bisulfite converted reads were analysed using the quantification tool for methylation analysis (QUMA) with default quality control settings and exclusion of identical bisulfite sequences enabled, to prevent skewing due to clonal amplicons.

Bioinformatics and Data Analysis

The Bioconductor MEDIPS package (PMID 20802089) was used to compute normalized (h)MEDIP tag counts (RPM) in selected genomic regions (imprinted DMRs, CGIs), and to detect differentially methylated regions of 300 nt in size genome-wide (step = 150); fragments were extended to 160 nt. Profiles around the transcripion start site and across selected genomic features were generated based on MEDIPS normalized values within 200 bp sliding windows (step=100).

To quantify enrichment of 5mC and 5hmC in binding sites of transcription factors and chromatin-associated proteins, a previously compiled set of ChIP-seq peaks was obtained (Martello et al, 2012). Enrichment of 5mC and 5hmC was given as the fraction of the average RPM normalized tag counts for the binding sites of a given factor versus the average RPM normalized tag counts for the binding sites of all selected factors (log2).

CpG islands mapped by CXXC Affinity Purification (CAP) were obtained (Illingworth et al, 2010). Tet1, Tet2 binding and histone modification ChIP-seq data sets were downloaded from GEO (Tet1: GSE24841; Tet2: GSE41722; H3K4me3 and H3K27me3: GSE23943), and aligned using the bowtie software tool. RPM normalized tag counts per CGI were computed for (h)MEDIP-seq and for ChIP-seq data sets. To allow for a comparison, the values were divided by their maximum number after removal of outliers, and the profiles were clustered across CGIs using average linkage clustering.

For analysis of gene networks, classification of genes were: Pluripotency-associated: *Pou5f1, Nanog, Esrrb, Tcl1, Tet1, Dppa3, Klf4, Klf2, Nr5a1, Sox2, Tbx1, Prdm14, Zfp42, Tbx3, Tcf3, Stat3, Mbd3, Lin28b.* Primed-associated: *Fgf5, Lefty2, Lefty1, Dnmt3a, Dnmt3b, Dnmt3l, Snai1, Nodal, Bmp4, Wnt8a, Gata3, Gdf3, Pitx2, Thy1, Ncam1, Cer1, Foxa1, Eomes.* Germline-associated: *Dazl, Tex19.1,*

Piwil2, Piwil1, Sycp3, Sycp1, Syce1, Asz1, Mael, Dppa4, Tex11, Rhox6, Rhox9, Tex19.2, Rhox4d, Tex12.

AUTHOR CONTRIBUTIONS

J.A.H designed and performed experiments and wrote the manuscript. S.D, J.A.H and T.A.D performed bioinformatics analysis. H.G.L & K.M generated and cultured cell lines. M.A.S supervised the study and wrote the manuscript.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

Dunican DS, Ruzov A, Hackett JA, Meehan RR (2008) xDnmt1 regulates transcriptional silencing in pre-MBT Xenopus embryos independently of its catalytic function. *Development* **135:** 1295-302

Gillich A, Bao S, Grabole N, Hayashi K, Trotter MW, Pasque V, Magnusdottir E et al (2012) Epiblast stem cell-based system reveals reprogramming synergy of germline factors. *Cell Stem Cell* **10:** 425-39

Hackett JA, Sengupta R, Zylicz JJ, Murakami K, Lee C, Down TA, Surani MA (2012) Germline DNA Demethylation Dynamics and Imprint Erasure Through 5-Hydroxymethylcytosine. *Science*

Harris RA, Wang T, Coarfa C, Nagarajan RP, Hong C, Downey SL, Johnson BE et al (2010) Comparison of sequencing-based methods to profile DNA methylation and identification of monoallelic epigenetic modifications. *Nat Biotech* **28:** 1097-1105

Illingworth RS, Gruenewald-Schneider U, Webb S, Kerr ARW, James KD, Turner DJ, Smith C et al (2010) Orphan CpG Islands Identify Numerous Conserved Promoters in the Mammalian Genome. *PLoS Genetics* **6:** e1001134-e1001134

Martello G, Sugimoto T, Diamanti E, Joshi A, Hannah R, Ohtsuka S, Gottgens B et al (2012) Esrrb is a pivotal target of the Gsk3/Tcf3 axis regulating embryonic stem cell self-renewal. *Cell Stem Cell* **11:** 491-504

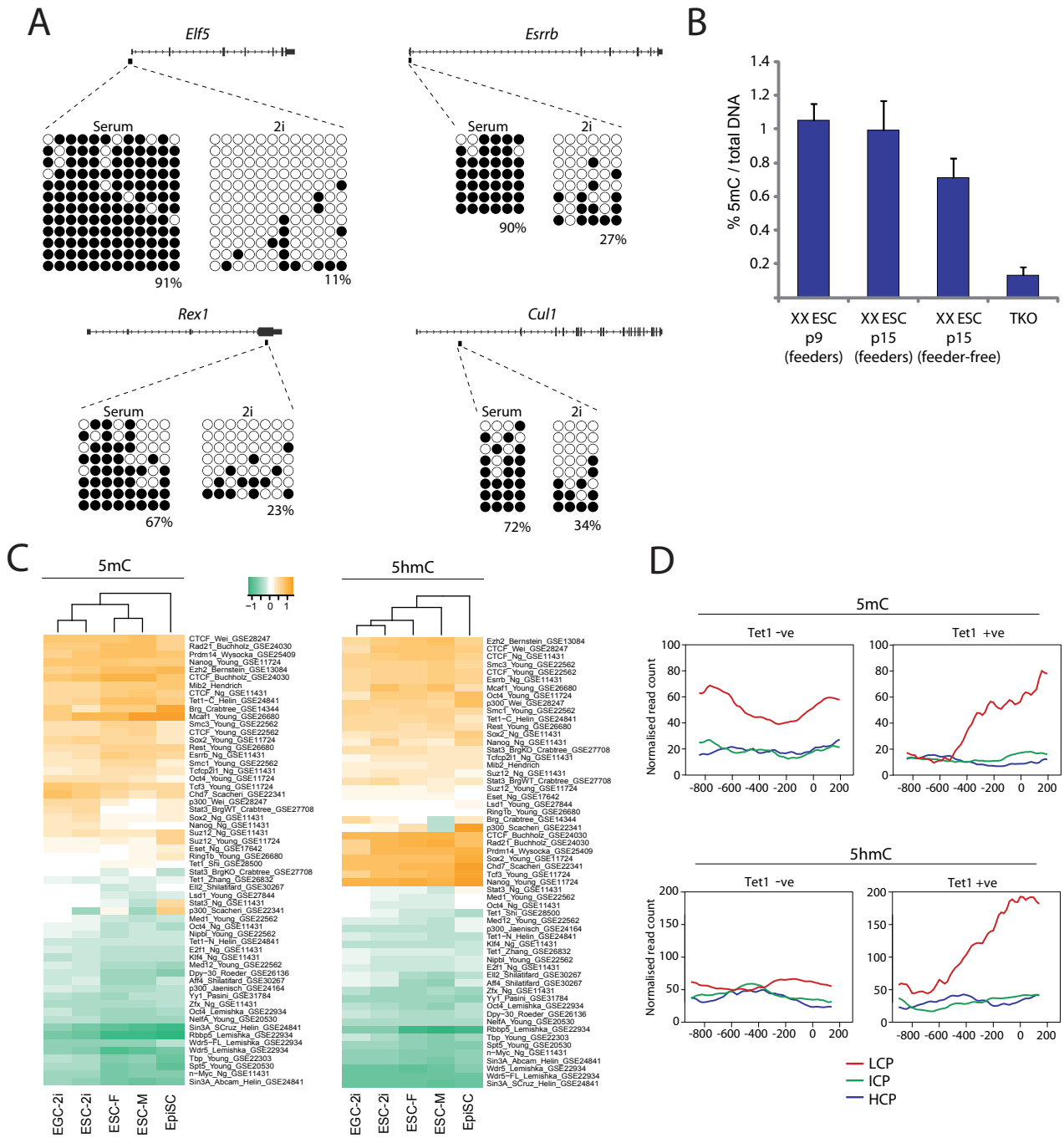

Figure S1. Enrichment of 5mC and 5hmC at genomic landmarks in 2i or serum. A.) Bisulfite sequencing of the promoter regions at *Elf5* and *Esrrb* (inactive in ESC-serum), an exonic region of *Rex1* and an intronic region in *Cul1* from ESC (XY) cultured in serum or 2i conditions. Filled circles represent methylated CpG dinucleotides and empty circles represent unmethylated CpG dicnucleotides. Each horizontal line corresponds to allelic methylation. **B.)** ELISA showing global levels of 5mC in early-passage (p9) XX ESC on feeders and +/- transition to feeder-free culture (passage 15), with negative control DNA methylation-deficient ESC (TKO). Error bars represent S.D of technical triplicate quantifications. **C.)** Clustering of 5mC and 5hmC according to enrichment at the binding sites of multiple chromatin-associated proteins in pluripotent stem cells. In general enrichment clusters into each of the three distinct pluripotent states: ground-state, metastable, and primed. **D.)** Profiles of 5mC and 5hmC around the TSS according to CpG density - low, intermediate and high CpG density promoters (LCP, ICP, HCP). Profiles are also separated according to whether they are TET1-negative or TET1 positive binding sites. Note 5mC is enriched at the TSS, but only at LCPs where TET1 binds, likely owing to the fact they are methylated (most LCPs are) and therefore competent for 5hmC conversion upon TET1 binding.

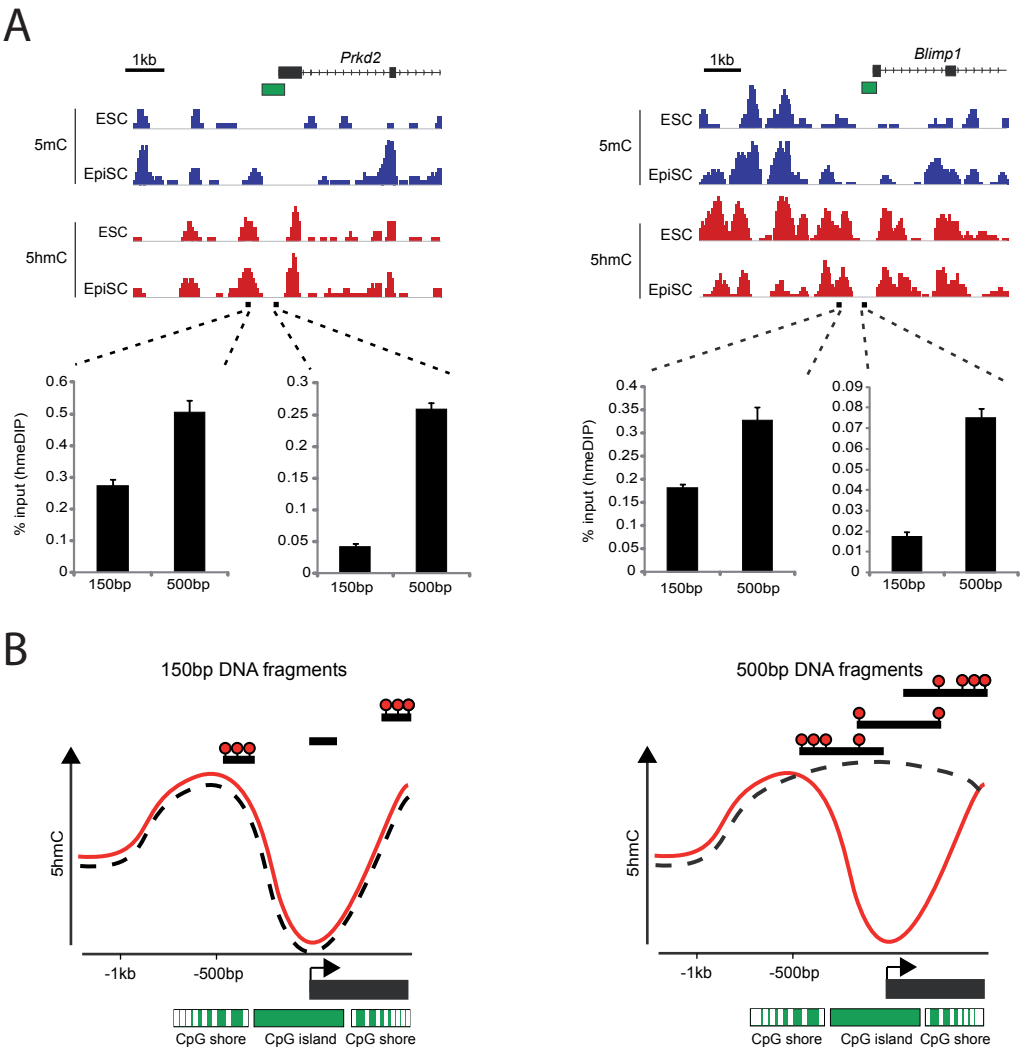

Figure S2. Effect of fragment size on hmeDIP enrichment at distinct genomic landmarks. A.) Quantitative hmeDIP-qPCR using mean DNA fragment sizes (resolution) of 150bp or 500bp. Analysis was performed at the transcriptional start site (TSS) or proximal upstream region of unmethylated CpG-dense promoters (*Prkd2 & Blimp1*) that are representative of not having 5hmC enrichment directly over the TSS by 150bp hmeDIP-seq. At the TSS, 500bp resolution resulted in a 5-7 fold greater enrichment (over input) relative to 150bp. In contrast at proximal promoter regions 5hmC was only enriched <2-fold in 500bp resolution relative to 150bp. This implies that the lower (500bp) resolution may give higher relative enrichment at the TSS as compared to other genomic regions, which may reflect bridging of larger fragments over the TSS and adjacent 5hmC enriched regions (see B). Error bars represent S.D of qPCR triplicate quantifications. **B.)** Schematic depicting potential 'bridging' of the TSS by larger hmeDIP DNA fragment sizes. Shown is the possible actual 5hmC distribution (red line) at CpG-dense promoters, based on bisulfite data and base-resolition 5hmC-sequencing (Yu et al., 2012), and the implied distribution (dashed line) by hmeDIP with 150bp resolution (left) or 500bp (right). While 150bp fragments will reveal a localised depletion of 5hmC around the TSS, consistent with bisulfite data, 500bp resolution may generate fragments that overlap both 5hmC enriched regions (promoter/CpG shores/1st-exon) and 5hmC-depleted regions (TSS). These bridging fragments will be immunoprecipitated and subsequently mapped back to the TSS, despite a possible localised depletion of 5hmC at the TSS.

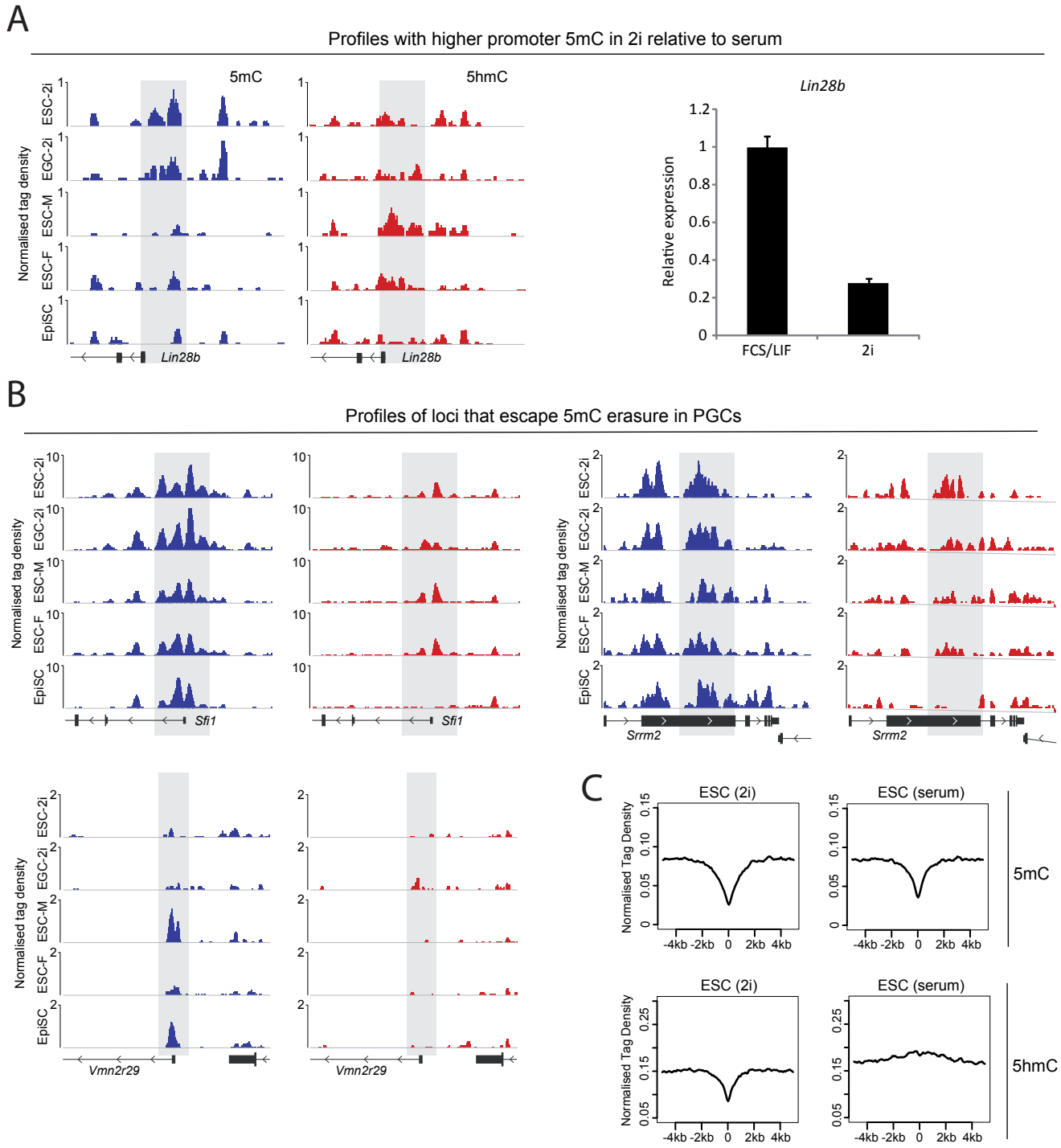

Figure S3. Genomic sites resistant to hypomethylation in 2i conditions. A.) 19 promoters showed higher relative 5mC in 2i relatie to serum, apparently including *Lin28b*, which was also transcriptionally repressed in 2i (right). Error bars represent S.D of biological duplicate quantifications. **B.)** Genomic loci such as the *Sfi1* CGI promoter and *Srrm2* exonic CGI, which escape DNA demethylation during reprogramming in PGCs (Hackett et al., 2012), also escape demethylation in 2i conditions. However, another resistant site in PGCs, *Vmn2r29*, does become demethylated in 2i and in female XX ESC in serum. Grey shading signifies region of interest. **C.)** Profiles of 5mC (upper) and 5hmC (lower) over active enhancers (DNase HSS, H3K4me1, H3K27ac) in ESC (XY) cultured in either 2i or serum conditions. Unlike 5hmC at poised enhancers (Fig 3F), 5hmC at active enhancers is neutral in serum and shows no clear peak of enrichment.

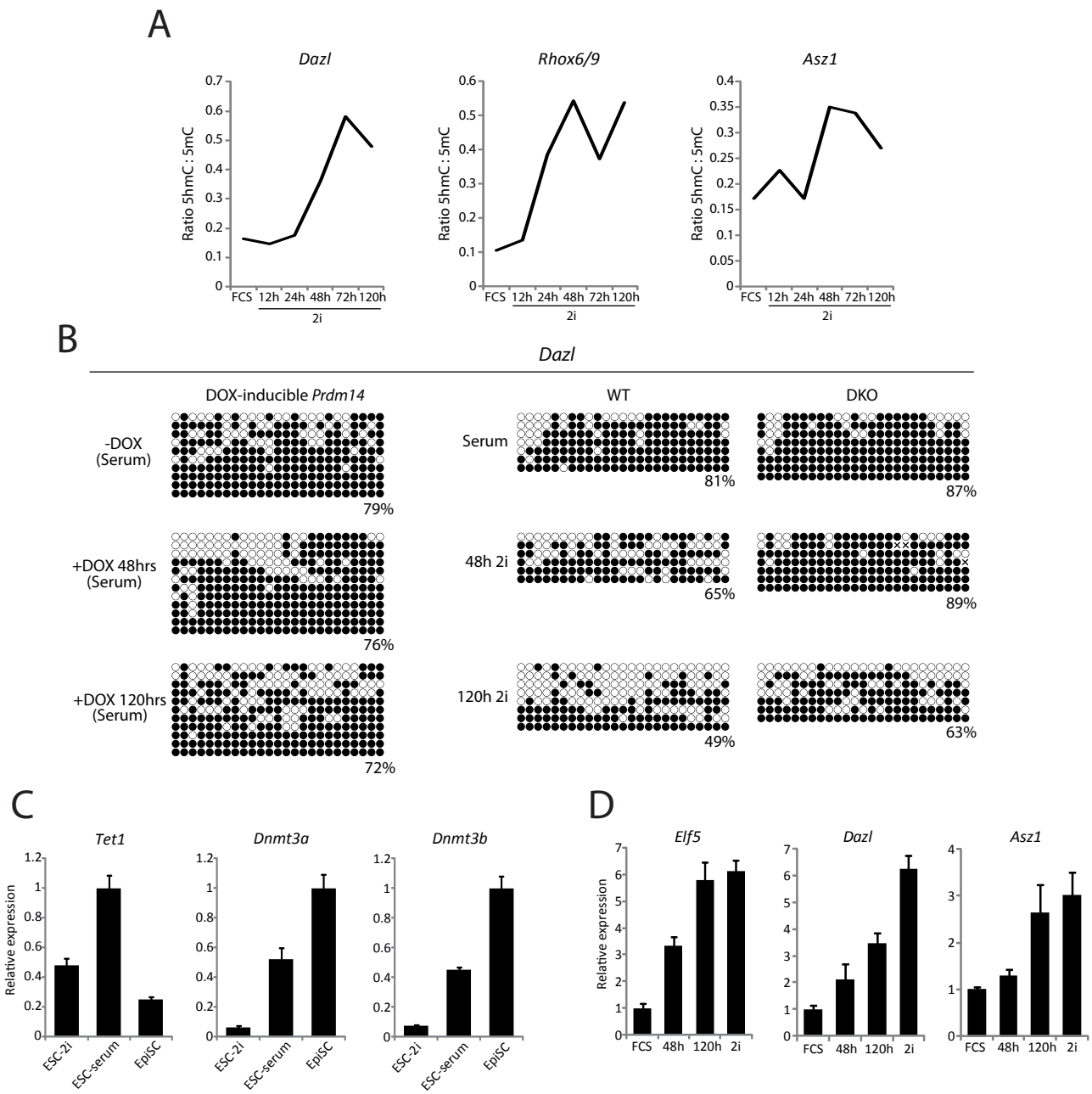

Figure S4. DNA demethylation & expression dynamics in WT, DKO and *Prdm14* overexpression ESC. A.) Ratio of 5hmC and 5mC during transition of WT ESC from serum (FCS) to ground-state conditions (2i) (relating to Fig 4C). There is a peak of 5hmC as a percentage of 5mC at 48-72hrs after switch to 2i, albiet the absolute increase in 5hmC is small. **B.)** Left: Bisulfite sequencing of *Dazl* in uninduced (-DOX) and *Prdm14*-induced (+DOX) ESC in serum for 48hrs or 120hrs shows a modest degree of demethylation by PRDM14 during this period. Right: Bisulfite sequencing of WT and *Tet1/Tet2* knockout (DKO) ESC during transfer to 2i conditions reveal a substantial delay in demethylation specifically in DKO ESC. **C.)** Relative gene expression of *Tet1, Dnmt3a* and *Dnmt3b* in ground state ESC in 2i, in ESC in serum, and in primed EpiSC. **D.)** Gene expression analysis of *Elf5* and DNA methylation-dependent genes *Dazl* and *Asz1* during transition from serum (FCS) to 2i conditions. Error bars represent S.D of biological duplicate quantifications.