Cell Reports, Volume *28*

Supplemental Information

Quantitative Multiplexed ChIP Reveals Global Alterations that Shape Promoter Bivalency in Ground State Embryonic Stem Cells

Banushree Kumar and Simon J. Elsässer

Figure S1

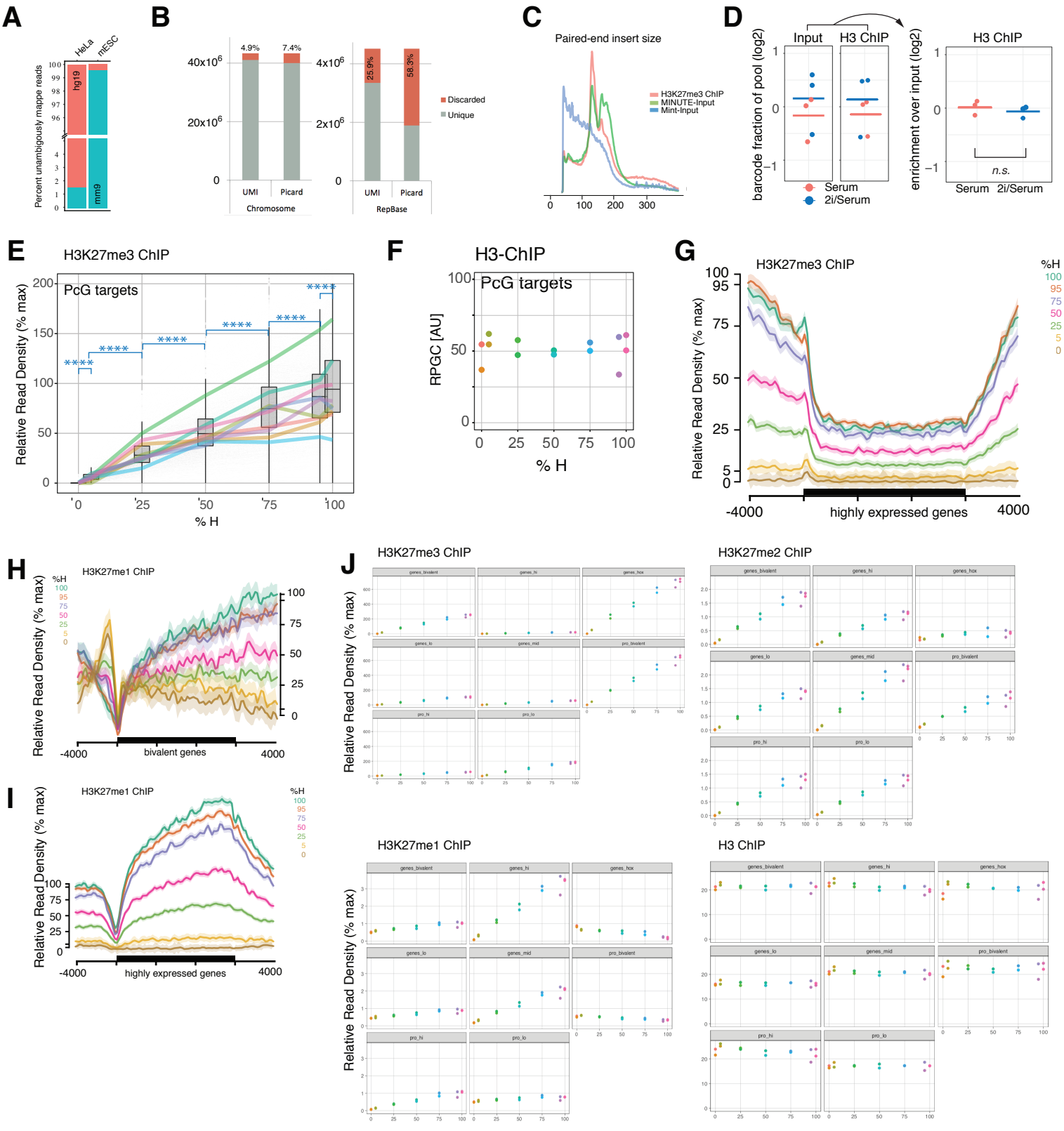
A
HeLa
mESC
hg19
mm9
Percent unambiguously mapped reads
B
4.9%
7.4%
25.9%
58.3%
Discarded
Unique
UMI
Picard
Chromosome
RepBase
C
Paired-end insert size
H3K27me3 ChIP
MINUTE-Input
Mint-Input
D
Input
H3 ChIP
barcode fraction of pool (log2)
enrichment over input (log2)
Serum
2i/Serum
n.s.
E
H3K27me3 ChIP
PcG targets
Relative Read Density (% max)
% H
F
H3-ChIP
PcG targets
RPGC [AU]
% H
G
H3K27me3 ChIP
Relative Read Density (% max)
%H
-4000
highly expressed genes
4000
H
H3K27me1 ChIP
Relative Read Density (% max)
%H
-4000
bivalent genes
4000
I
H3K27me1 ChIP
Relative Read Density (% max)
%H
-4000
highly expressed genes
4000
J
H3K27me3 ChIP
H3K27me2 ChIP
H3K27me1 ChIP
H3 ChIP
Relative Read Density (% max)
genes_bivalent
genes_hi
genes_hox
genes_lo
genes_mid
pro_bivalent
pro_hi
pro_lo

Figure S1. Related to Figure 1. A Analysis of barcode cross-contamination using a human HeLa cell sample in Pool1. One HeLa and one mESC sample were aligned to a hg19/mm9 metagenome and only uniquely mapped reads to either human or mouse chromosomes were counted. **B** Deduplication using unique molecular identifier. Total mapped read counts of the Input (Pool 1, see Supplementary Table 2 for details) before and after demultiplexing with Picard (http://broadinstitute.github.io/picard/, relies on location of read1 one and read2) or with JE (Girardot et al., 2016) using the UMI information. Reads were mapped to mm9 genome (left) and a metagenome with all murine repetitive sequences from RepBase (Bao et al., 2015) (right). **C** Fragment insert size distribution determined from aligned BAM file, comparing the input library prepared using RA3 cDNA primer (see Figure 1A, MINUTE-ChIP) or random hexanucleotide priming according to original Mint-ChIP protocol. **D** Example of quantification. H3 ChIP was performed on a pool (Pool 2, see Supplementary Table 2 for details) of two condition with three biological replicates each. Barcode representation before and after ChIP (left), and resulting quantitative comparison (right). **E** Boxplot of H3K27me3 at 2731 PcG target regions for each calibration point. Pairwise differences between all calibration points are significant (paired two-tailed t-test, **** = p<2e-16). Individual calibration curves for 2731 PcG targets (1-2kb peak region). Ten trajectories were chosen randomly for highlighting in color. **F** H3 Quantification at PcG targets relating to Figure 1F and 1H. **G** H3K27me3 average plot of different mixing ratios across highly expressed genes (average of two replicates). Standard error is rendered as shaded area around lines. **H** H3K27me1 average plot of different mixing ratios across bivalent genes (average of two replicates). Standard error is rendered as shaded area around lines. **I** H3K27me1 average plot of different mixing ratios across highly expressed genes (average of two replicates). Standard error is rendered as shaded area around lines. **J** Replicate Quantification of H3K27me1, me2, me3, H3 ChIP signal at bivalent, low, medium and high expressed genes and promoters.

Figure S2

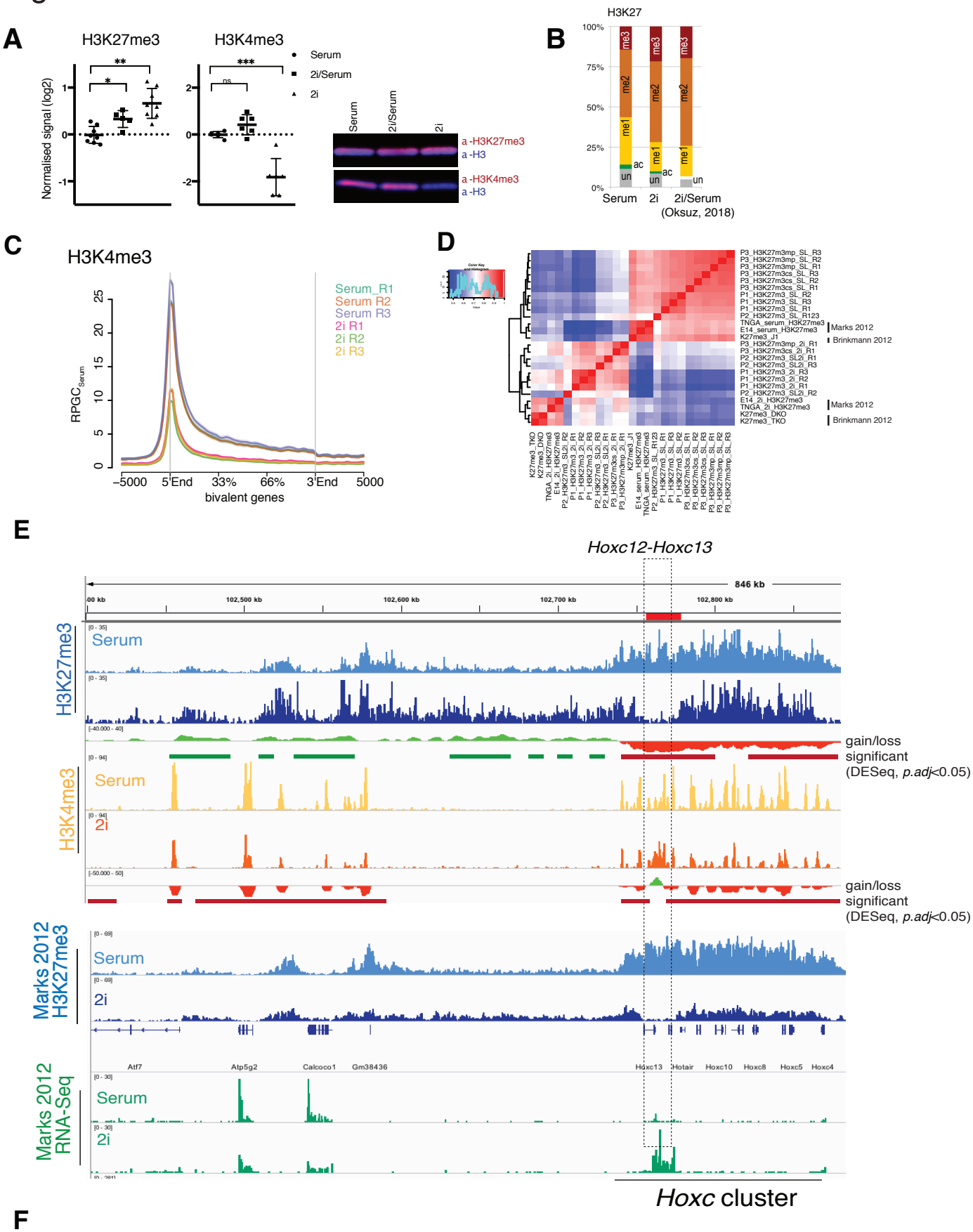
A
H3K27me3
H3K4me3
Normalised signal (log2)
Serum
2i/Serum
2i
a -H3K27me3
a -H3
a -H3K4me3
a -H3
B
H3K27
Serum
2i
2i/Serum
(Oksuz, 2018)
C
H3K4me3
Serum_R1
Serum R2
Serum R3
2i R1
2i R2
2i R3
bivalent genes
D
Marks 2012
Brinkmann 2012
E
Hoxc12-Hoxc13
846 kb
H3K27me3
H3K4me3
Serum
2i
gain/loss
significant
(DESeq, p.adj<0.05)
Marks 2012 H3K27me3
Marks 2012 RNA-Seq
Atf7
Atp5g2
Calcoco1
Gm38436
Hoxc13
Hotair
Hoxc10
Hoxc8
Hoxc5
Hoxc4
Hoxc cluster
F
number of 10kb windows with p.adj<0.05
Pool2 H3(2i/Serum vs Serum)
Pool2 H3K4me3 (2i/Serum vs Serum)
Pool2 H3K27me3 (2i/Serum vs Serum)
Pool1 H3K4me3 (2i vs Serum)
Pool1 H3K27me3 (2i vs Serum)
up
non-significant
down

Figure S2. Related to Figure 2. Validation of quantitative differences between Serum and 2i condition using western blot and mass spectrometry. **A** Quantitation of H3K27me3 and H3K4me3 using two-color IR western blot, in Serum, 2i and 2i/Serum conditions. Data is represented as mean and standard deviation. T-test was performed on triplicate samples (* = $p<0.05$; ** = $p<0.005$). Representative image of the replicates is shown. **B** relative quantitative assessment of H3K27me3 levels using mass spectrometry (ModSpec, ActiveMotif) in Serum and 2i (one sample each). K27me3 on histone H3.1 is estimated to be 14% in serum and 22% in 2i conditions. A recent study performed a similar quantitation in 2i/Serum conditions (right) and reported 20% (Oksuz et al., 2018). H3K27ac was not reported in the latter study. Note that H3K4me3 does not produce adequate peptides for quantification with this protocol. **C** Example of three Serum and 2i replicates in Pool 1, summarized as mean graph in Figure 3C. Standard error is rendered as shaded area around lines. **D** Hierarchical clustering shows that present datasets cluster with prior H3K27me3 data from Serum vs 2i comparison (Marks et al., 2012). Pearsson correlation was calculated on 10kb bins of the mm9 genome after removing outliers and blacklisted regions. **E** Comparison of data from ours study with published H3K27me3 dataset (Marks et al., 2012). Tracks are scaled as Reads Per Genomic Content relative to Serum condition ($RPGC_{Serum}$) or RPGC for non-quantitative data. Additional tracks show gain (green) and losses (red) for each quantitative comparison, as well as 10kb windows deemed to be significantly ($p.adj < 0.05$) increased (green) or decreased (red) amongst replicates.
Qualitative similarities and differences in global scaling are apparent. Nevertheless. previously reported region in Hoxc cluster shows dramatic loss of H3K27me3 and transcriptional induction, as reported (Marks et al., 2012), also a concomitant gain in H3K4me3. **F** Results of negative binomial test (DESeq) comparing the biological triplicate measurements across the genome in 10kb windows. The number of significantly different 10kb windows

Figure S3

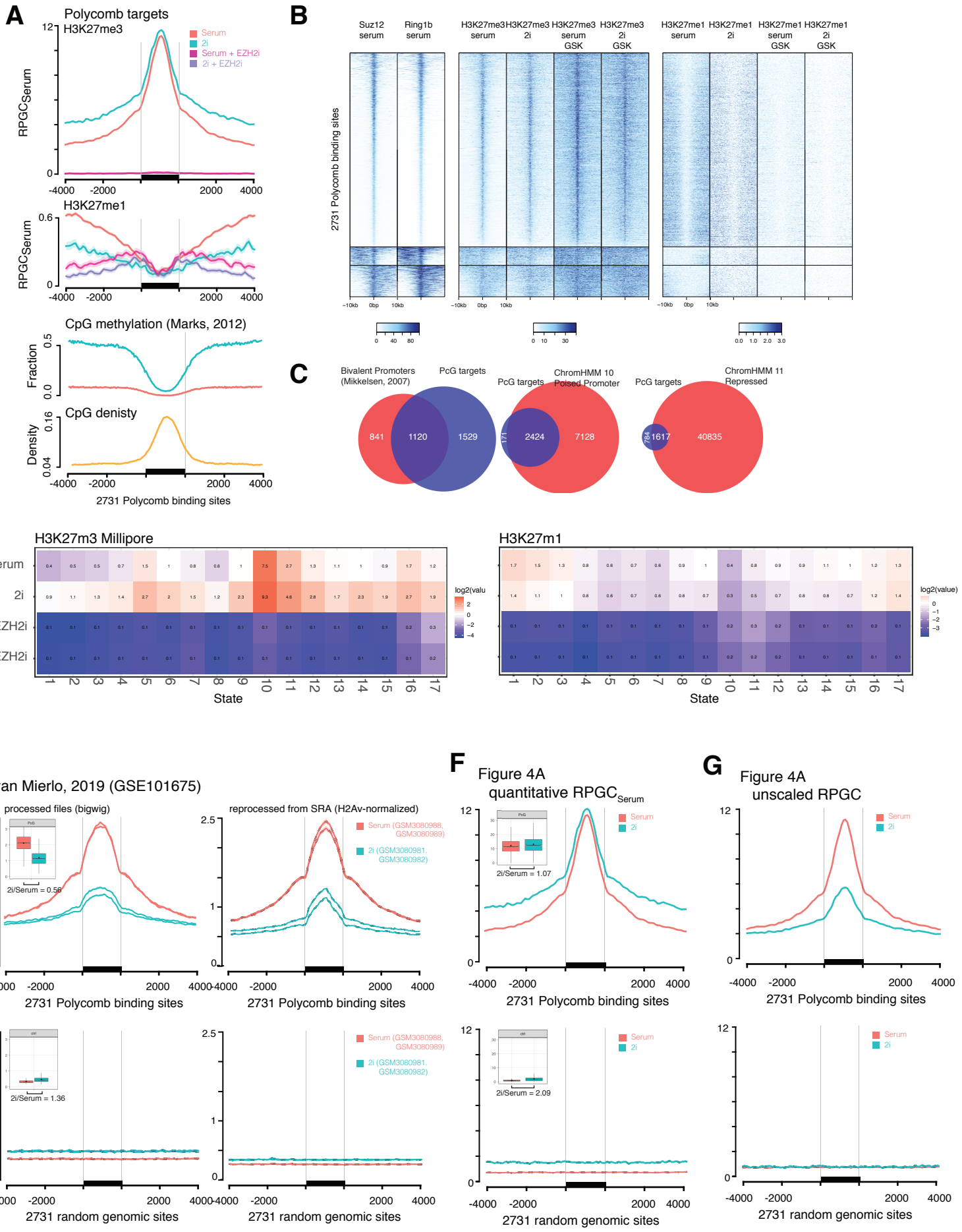
A
Polycomb targets
H3K27me3
Serum
2i
Serum + EZH2i
2i + EZH2i
RPGCSerum
H3K27me1
CpG methylation (Marks, 2012)
Fraction
CpG denisty
Density
2731 Polycomb binding sites
B
Suz12 serum
Ring1b serum
H3K27me3 serum
H3K27me3 2i
H3K27me3 serum GSK
H3K27me3 2i GSK
H3K27me1 serum
H3K27me1 2i
H3K27me1 serum GSK
H3K27me1 2i GSK
2731 Polycomb binding sites
C
Bivalent Promoters (Mikkelsen, 2007)
PcG targets
ChromHMM 10 Poised Promoter
ChromHMM 11 Repressed
841
1120
1529
171
2424
7128
784
1617
40835
D
H3K27m3 Millipore
H3K27m1
Serum
2i
Serum EZH2i
2i EZH2i
log2(value)
State
E
van Mierlo, 2019 (GSE101675)
processed files (bigwig)
reprocessed from SRA (H2Av-normalized)
Serum (GSM3080988, GSM3080989)
2i (GSM3080981, GSM3080982)
2i/Serum = 0.56
2i/Serum = 1.36
2731 Polycomb binding sites
2731 random genomic sites
F
Figure 4A
quantitative RPGC Serum
2i/Serum = 1.07
2i/Serum = 2.09
G
Figure 4A
unscaled RPGC

Figure S3. Related to Figure 4. A Quantitative H3K27me3 and H3K27me1 profiles at 2731 PcG target sites defined by EZH2 (ENCODE) or Ring1b (Joshi et al., 2015) binding, in Serum and 2i condition, as well as after treatment with EZH2 inhibitor EPZ-6438 (EZH2i) for 7 days. Y axis shows Reads Per Genomic Content relative to Serum condition (RPGCSerum), where Serum condition is scaled to 1x coverage. Additionally CpG methylation (% methylation per CpG, Marks 2012) and underlying genomic CpG density is shown **B** Heatmap of 2731 PcG target sites, showing Ring1b and Suz12 binding (Joshi et al., 2015), as well as H3K27me3 and H3K27me1 as in A. **C** Overlap of PcG target sites (defined by PRC1 or PRC2 binding), bivalent promoters (defined by H3K27me3, H3K4me3 co-occurence, Mikkelsen et. al. 2007) and ChromHMM States 10 and 11 (H3K27me3-enriched). **D** Heatmap of H3K27me3, H3K27me1 across 17 chromatin states. Quantitative comparison between 2i and Serum, with and without EZH2i, are shown. Color scale represents log2 of each in comparison to Serum. **E** Comparison with published data (van Mierlo et. al., 2019): H3K27me3 ChIP average profiles over 2731 PcG target sites or random genomic regions (background control) in Serum vs. 2i conditions. Quantitative ChIP was performed using Drosophila H2Av Spike-in controls in the latter study. Left: Quantitatively scaled bigwig tracks were downloaded from GEO and plotted using SeqPlot (GSM3080981 2i-H3K27me3-rep1-spikeIn-13612, GSM3080982 2i-H3K27me3-rep2-spikeIn-13613, GSM3080988 FCS-H3K27me3-rep1-spikeIn-13615, GSM3080989 FCS-H3K27me3-rep2-spikeIn-13616). Inserted graph shows median and quartiles, as well as mean of the H3K27me3 levels at 2731 the PcG target sites or control regions. The mean fold-difference between for Serum and 2i is indicated. Right: as above but regenerated from primary source files. FASTQ files were downloaded from GEO, including corresponding input files, and mapped to mm9 and dm3. After deduplication with Picard, total mapped read counts were determined for mm9 and H2Av-bound regions of dm3 (result using total dm3 reads were similar, not shown). mm9 tracks were generated using deepTools and scaled according to the ratio of mm9 to H2Av reads in the ChIP, over the corresponding ratio in the respective input sample. **F** H3K27me3 MINUTE-ChIP in Serum vs. 2i conditions, scaled according to INRC at 2731 the PcG target sites (same as Figure 4A) or control regions. The mean fold-difference between for Serum and 2i is indicated. **G** same data as in F processed with standard ChIP-Seq normalization (1x genome coverage). Since standard normalization assumes stable background levels, the peak size appears dramatically reduced.

Figure S4

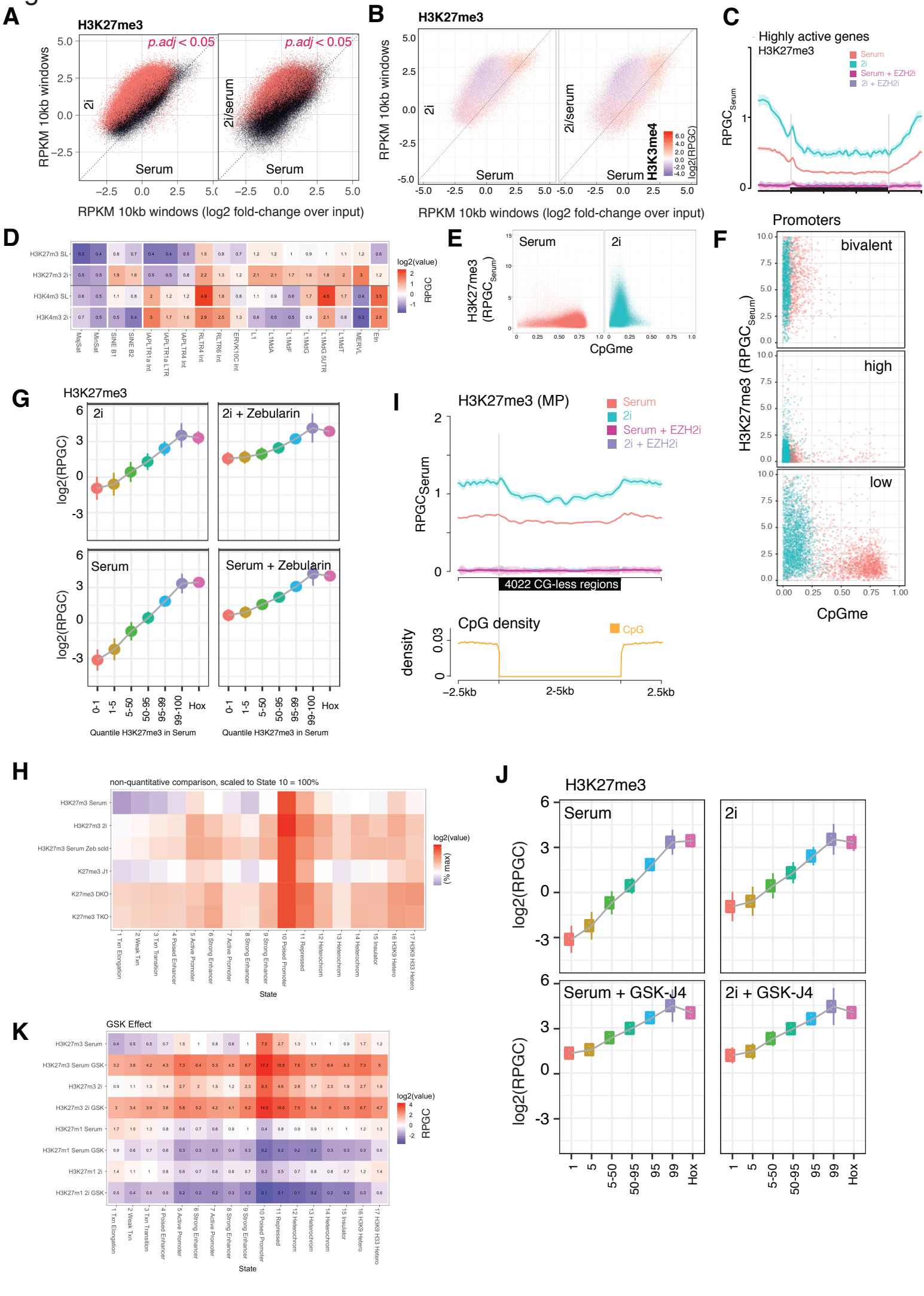
A
H3K27me3
p.adj < 0.05
p.adj < 0.05
2i
2i/serum
Serum
Serum
RPKM 10kb windows
RPKM 10kb windows (log2 fold-change over input)
B
H3K27me3
2i
2i/serum
Serum
Serum
H3K3me4
log2(RPGC)
RPKM 10kb windows
RPKM 10kb windows (log2 fold-change over input)
C
Highly active genes
H3K27me3
Serum
2i
Serum + EZH2i
2i + EZH2i
RPGC_Serum
D
H3K27m3 SL
H3K27m3 2i
H3K4m3 SL
H3K4m3 2i
log2(value)
RPGC
MajSat
MinSat
SINE B1
SINE B2
IAPLTR1a Int
IAPLTR1a LTR
IAPLTR4 Int
RLTR4 Int
RLTR6 Int
ERVK10C Int
L1
L1MdA
L1MdF
L1MdG
L1MdG 5UTR
L1MdT
MERVL
Etn
E
Serum
2i
H3K27me3 (RPGC_Serum)
CpGme
F
Promoters
bivalent
high
low
H3K27me3 (RPGC_Serum)
CpGme
G
H3K27me3
2i
2i + Zebularin
Serum
Serum + Zebularin
log2(RPGC)
0-1
1-5
5-50
50-95
95-99
99-100
Hox
Quantile H3K27me3 in Serum
Quantile H3K27me3 in Serum
I
H3K27me3 (MP)
Serum
2i
Serum + EZH2i
2i + EZH2i
RPGCSerum
4022 CG-less regions
CpG density
CpG
density
0.03
−2.5kb
2-5kb
2.5kb
H
non-quantitative comparison, scaled to State 10 = 100%
H3K27m3 Serum
H3K27m3 2i
H3K27m3 Serum Zeb scld
K27me3 J1
K27me3 DKO
K27me3 TKO
log2(value)
(% max)
1 Txn Elongation
2 Weak Txn
3 Txn Transition
4 Poised Enhancer
5 Active Promoter
6 Strong Enhancer
7 Active Promoter
8 Strong Enhancer
9 Strong Enhancer
10 Poised Promoter
11 Repressed
12 Heterochrom
13 Heterochrom
14 Heterochrom
15 Insulator
16 H3K9 Hetero
17 H3K9 H33 Hetero
State
J
H3K27me3
Serum
2i
Serum + GSK-J4
2i + GSK-J4
log2(RPGC)
1
5
5-50
50-95
95
99
Hox
K
GSK Effect
H3K27m3 Serum
H3K27m3 Serum GSK
H3K27m3 2i
H3K27m3 2i GSK
H3K27m1 Serum
H3K27m1 Serum GSK
H3K27m1 2i
H3K27m1 2i GSK
log2(value)
RPGC
State

Figure S4. Related to Figure 4. A Levels of H3K27me3 determined in 10kb windows (RPGC, mean of three biological replicates, log2-fold enrichment over input) in 2i or 2i/Serum condition were plotted against Serum condition. Significant different regions (p.adj < 0.05) across the three biological replicates were determined using a negative binomial test (DESeq). **B** H3K27me3 bins as in A were overlaid with H3K4me3 levels of each bin. H3K27me3 increased almost universally across the genome in 2i/serum or 2i. As an exception, the most extreme high regions of H3K27me3 (which coincided with high H3K4me3, thus representing bivalent domains) did not further increase in H3K27me3. **C** H3K27me3 average plot over highly expressed genes (associated with high H3K36me3 in the gene body) shows that H3K27me3 is present in Serum above background levels and further increased in 2i.
D Relative enrichment of H3K27me3 and H3K4me3, in Serum and 2i conditions (Pool 1), over repetitive regions of the genome using reads mapped to a metagenome of repetitive sequences from RepBase (Bao et al., 2015). **E** H3K27me3 and CpG methylation (Habibi et al., 2013) levels in Serum and 2i conditions were determined in 10kb windows and plotted against each other. The gain of H3K27me3 observed in the bulk of 10kb windows in 2i is concomitant to a loss of CpG methylation, which may be a direct or indirect effect. **F** H3K27me3 and CpG methylation (Habibi et al., 2013) levels at promoters of bivalent, high and lowly expressed genes, in Serum and 2i conditions. **G** 10kb bins were partitioned in 1, 5, 50, 95, 99th quantiles according to H3K27me3 levels in Serum (log2 of RPGC). Regions overlapping with hox genes were analyzed separately. As compared to Serum, 2i shows an increase inversely related to the original level, with the lowest quantiles showing the largest fold-change. Zebularin treatment promotes hypomethylation across all quantiles, and a particularly strong increase in H3K27me3 at the lowest quantiles. **H** Comparison of Zebularin treatment and DNMT3a/b knockout (DKO), DNMT1/3a/3b knockout (TKO), in modulating H3K27me3 levels accross functional chromatin states. Since available data for DNMT KO was not performed quantitative, we normalized data according to State 10 (% max). Zebularin treatment and DNMT DKO/TKO exhibit similar shifts in H3K27me3, which is qualitatively similar but not identical to 2i. Note that DNMT DKO and TKO profiles also clustered with 2i samples in a correlation analysis (Figure S2C). **I** 4022 regions larger than 2kb devoid of any CpG dinucleotide were identified in the mouse mm9 genome. Quantitative H3K27me3 average was plotted across CG-devoid and flanking regions. Average levels accross CpG-devoid regions were higher than in highly expressed/H3K36me3 high gene bodies (see panel C), and regions gained H3K27me3 at similar proportion to CpG-containing flanking regions. **J** 10kb bins were partitioned in 1, 5, 50, 95, 99th quantiles according to H3K27me3 levels in Serum. Regions overlapping with hox genes were analyzed separately. Side-by-side comparison of Serum and 2i condition with and without GSK-J4 treatment (4 days) shows that H3K27me3 levels are increased across all quantiles. **K** GSK-J4 effect on H3K27me3 and H3K27me1 levels analyzed by 17 ChromHMM chromatin states. Increase in H3K27me3 upon GSK-J4 treatment is mirrored by a reduction of H3K27me1, the product of H3K27me3 demethylases.

Figure S5

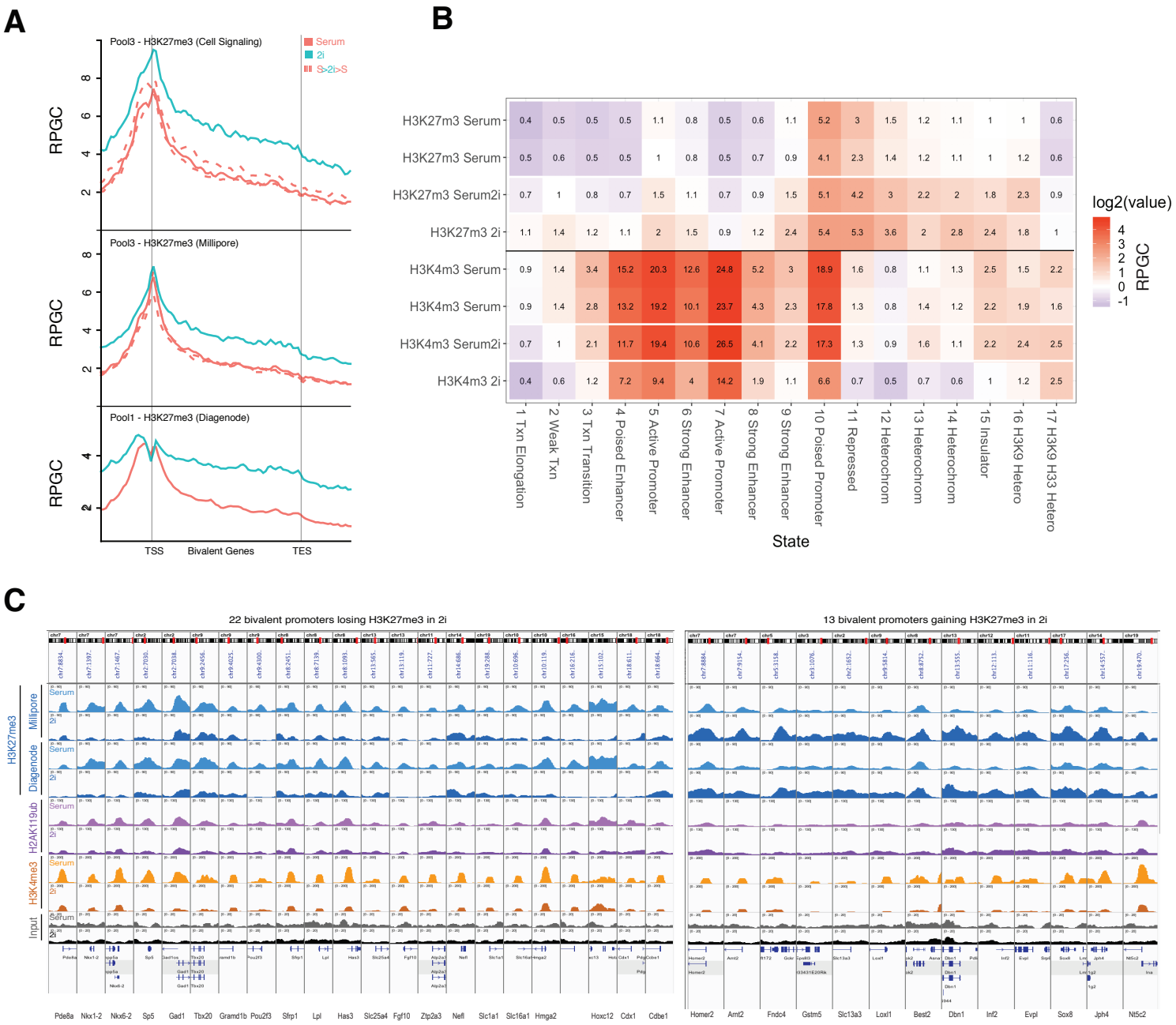

Figure S5. Related to Figure 5. A Average plots of H3K27me3 levels at bivalent genes produced by three different antibodies (and two fully independent biological experiments, Pool 1 and Pool 3). **B** H3K27me3 and H3K4me3 levels in Serum, 2i, 2i/Serum by ChromHMM chromatin state. It is apparent that changes to the H3K27me3 landscape in 2i are largely mirrored by 2i addition to Serum condition (top). In contrast, 2i/Serum is very similar to both Serum datasets, and does not exhibit the strong reduction of H3K4me3 in 2i. H3K4me3 is lost most dramatically (>2.5-fold reduction) from State 10 "Poised_Promoters" that are bivalent. Active promoters are reduced ~2-fold. **C** Bivalent genes that lose or gain H3K27me3 in 2i condition more than two-fold. Notably, all promoters also lose H3K4me3, independent of the gain or loss of H3K27me3.