SUPPLEMENTAL DATA

Supplemental Figure Legends

Supplemental Fig. 1: Resolution of the ESC-bivalent chromatin to monovalent H3K4Me3 or H3K27Me3 marks in MSC and osteoblasts. Venn diagram shows the overlaps of the genes that are H3K4Me3, H3K27Me3 or bivalent marked in ESC with the corresponding marks in MSC and osteoblasts. Percentages of modified ESC genes that retain the corresponding mark in MSC and osteoblasts are shown below the venn plots. Related to Figure 3.

Supplemental Fig. 2: Distribution of chromatin marks and expression for genes methylated in U2OS. (A) Pie chart indicates the proportion of genes methylated in U2OS that have the different chromatin marks or none of the marks in U2OS cells. (B) H3K4Me3-enrichment distribution around the transcription start site (TSS) and (C) distribution of gene expression intensities for following different sets of genes: 384 genes methylated in U2OS (All), the set of methylated genes that are H3K4Me3-marked (H3K4Me3), set of genes that are not H3K4Me3 marked (non-H3K4Me3; majority of methylated genes belong to this category), genes in the 0-20, 21-40, 41-60, 61-80 and 81-100 percentile groups of gene expression intensities in the Agilent gene expression array. In B, y-axis represents the average of ChIP-seq sequence counts at defined positions along 200 bp intervals across -5000/+5000 bp around the TSS. In C, y-axis represents the Log2 intesities of genes in the Agilent gene expression array. (D) H3K4Me3 and H3K27Me3 ChIP enrichement patterns (+/- 5000 bp around TSS) in

MSC, osteoblasts and U2OS of genes hypermethylated in U2OS cells. The H3K4Me3 ChIP enrichment values (seqence counts) were normalised to the average sequence counts of high expressing genes in the respective cell lines (81-100th percentile gene expression group) while the H3K27Me3 ChIP enrichment values (seqence counts) were normalised to the average sequence counts of low expressing genes in the respective cell lines (0-20th percentile gene expression group). Related to Figure 2 and 3.

Supplemental Fig. 3: Hypermethylation in U2OS occurs specifically at ESC bivalent genes and not silenced ESC genes. Genes in ESC were stratified into 5 groups based on their expression percentiles in (0-20, 21-40, 41-60, 61-80, 81-100 percentiles) (A). Genes exclusively in the 0-20th percentile or the remaining (21st to 100 percentile) (B) and 81-100th percentile or the remaining (0 to 79th percentile) (C) are analyzed for the distribution of overlap with U2OS-hypermethylated genes. In D, the ESC bivalent genes are analyzed for the distribution of overlap with U2OS-hypermethylated genes. Fisher's p-values and odds ratio were calculated to determine the significnace of the overlaps. (E) The number of U2OS-hypermethylated in each of the five ESC gene expression categories is shown. The number of genes that are ESC-bivalent or non-bivalent are shown in orange and grey respectively. (F) Expression in ESCs of the whole set of bivalent genes, a random sample of 384 bivalent genes or the 384 U2OS-hypemethylated genes. Expression of U2OS-hypermethylated genes in ESCs is similar to the group of bivalent genes since majority of the U2OS-hypermethylated genes are bivalent in ESCs.

Supplemental Fig. 4: Genes identified as methylated in lung, colon and breast cancer are methylated in primary tumors. Probes identified as methylated in lung, colon and breast tumor cell lines were analyzed for their methylation levels in corresponding primary tumor samples in the Cancer Genome Atlas project as analyzed by the infinium platform. Heatmap plots show the beta-values for these probes in lung (A), colon (B) and breast (C) primary tumors. Percentage probes that have beta-value > 0.5 in atleast 5% of the primary tumors are are shown in (D). Related to Figure 4.

Supplemental Fig. 5: DAC treatment causes re-expression of methylated genes in U2OS cells while TSA treatment causes slight re-expression of these genes in MSCs (and osteoblasts; osteoblasts data not shown). The x and y-axes in the scatterplot respectively shows the gene expression changes, plotted as log2-fold ratios, in TSA-treated to mock-treated and DAC-treated to mock-treated cells respectively. Expression changes in U2OS and MSC are shown in the left and right panels respectively. The points in grey are all genes in the Agilent gene expression array while points in red are U2OS-hypermethylated genes. Related to Figure 6.

Supplemental Fig. 6: Hierarchical cluster analysis of primary colon and breast cancer samples. Samples were clustered based on Infinium b-values of genes identified as methylated in colon (A and B) or breast (C and D) cancer cell lines and grouped according to their chromatin lineage in ESC, viz. H3K4Me3-marked (A and C), bivalent-marked (B and D). TCGA primary colon tumor and normal samples (A and B) were

classified as CIMP+ (red) or CIMP- (gray). The key at the top represents the CIMP status and tumor or normal status (T/N; tumor is in red and normal in gray). Similarly breast tumor samples were classified as estrogen (ER)/progestrone (PR) positive or negative; and also classified according to their gene expression (PAM50) classes. Related to Figure 7.

Supplemental Table Legends

Supplemental Table 1: Infinium methylation array data. Infinium probes with β-value greater than 0.75 in any one of the cancer cell lines and less that 0.25 in all the normal cells were selected as hypermethylated genes. Normal cells are labeled in green. Related to Figure 1B.

Supplemental Table 2: Analysis of the GO categories. Genelists used in analyzing GO categories and significant biological processes identified in methylated genes that are H3K4Me3- or bivalent-marked in ESC are listed. Related to Figure 5.

Supplemental Table 3: Molecular functions associated with the top 10 significant biological process categories. Related to Figure 5.

Supplemental Table 4: Cancer cell lines used in this study. Related to Figure 4.

Supplemental Table 5: File names of methylation data for lung, colon and breast primary tumors from the Cancer Genome Atlas (TCGA) used for analyzing the methylation status in primary tumors of genes identified as methylated in the corresponding cell lines. Related to Figure 4.

Supplemental Table 6: ChIP-seq peak calls in the different cell types. Related to Figure 3 and 4.

Supplemental Table 7: Genes showing enrichment of H3K4Me3 or H3K27Me3 at their promoters. Related to Figure 3 and 4.

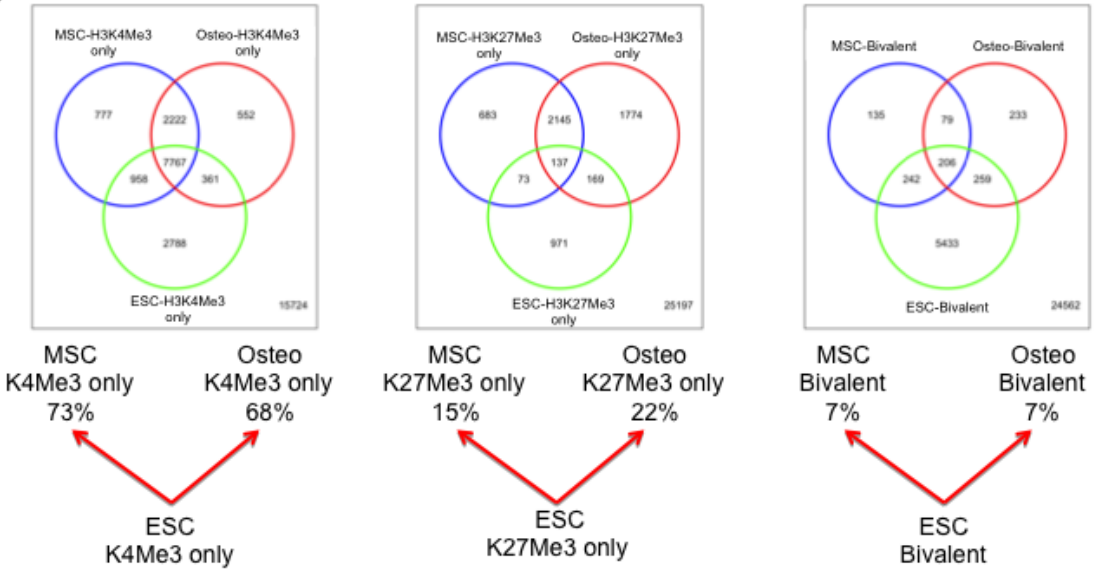
Supplemental Figures

Supplemental Fig. 1

A

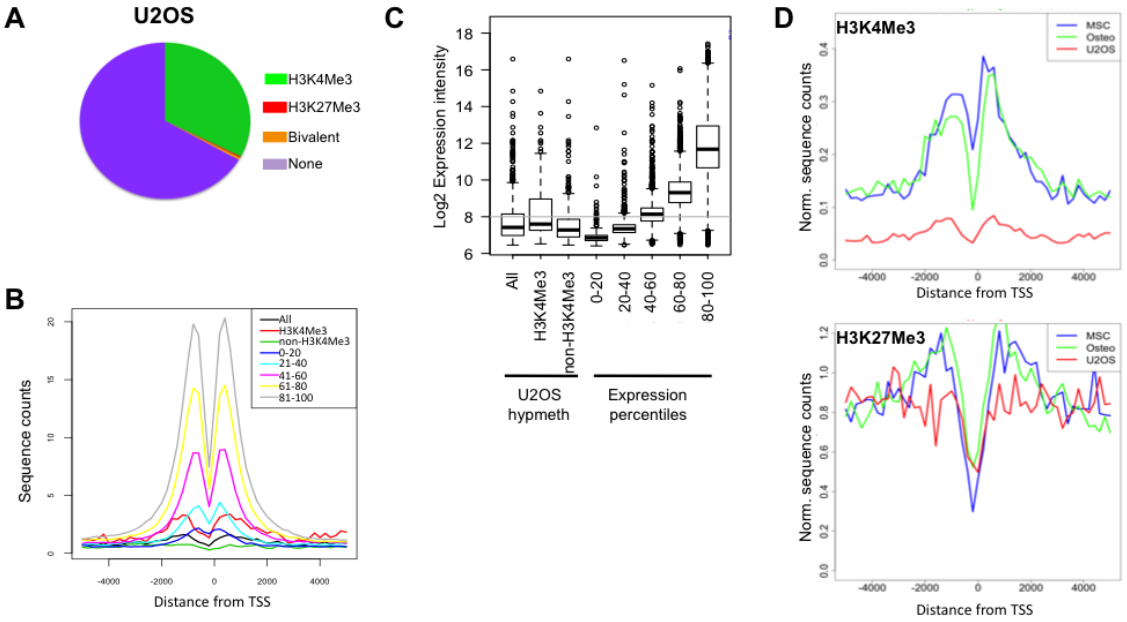
	U2OS hypermethylated	U2OS non-methylated	p value
ESC H3K4Me3 marked	20.8 %	66.2 %	9.7E-71
ESC Bivalent marked	78.9 %	32.9 %	8.1E-72

B

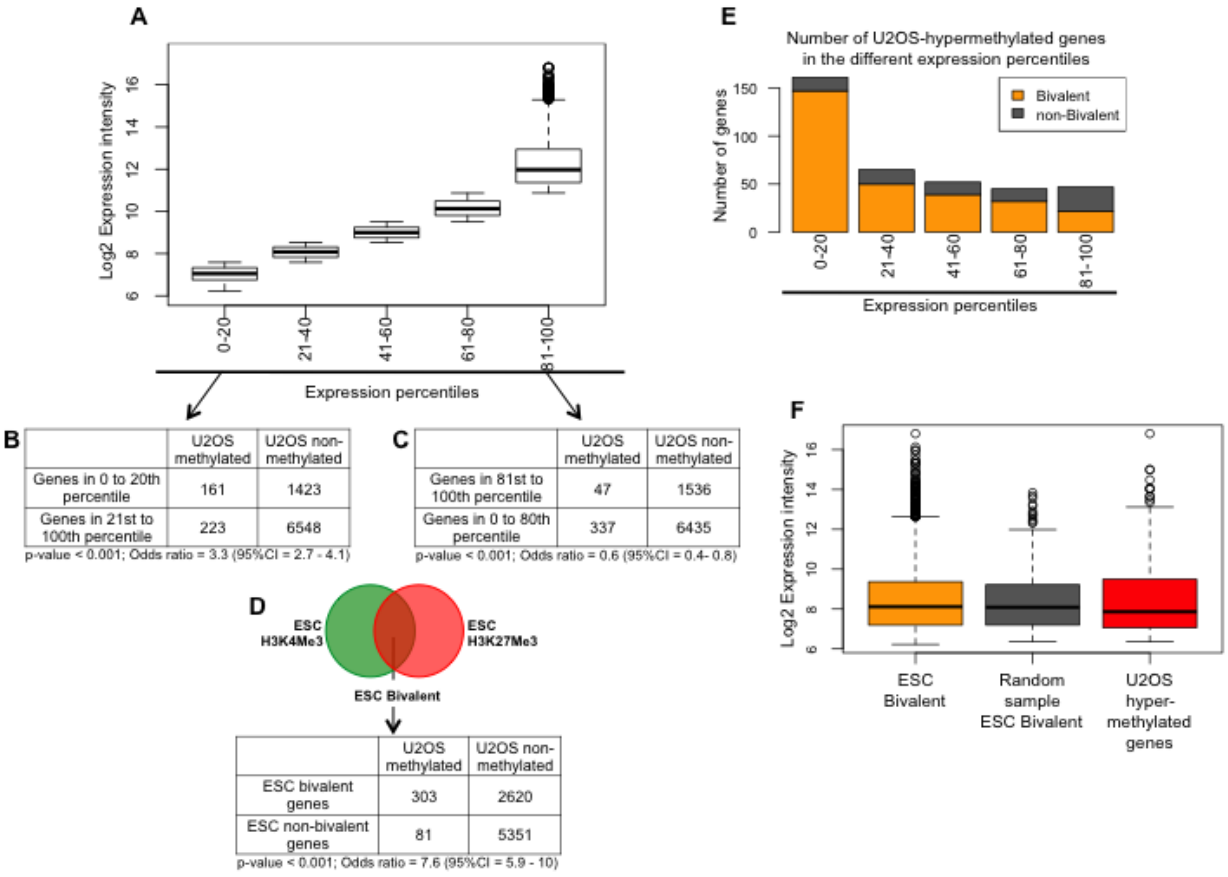


Percentages of modified ESC genes that retain the corresponding mark in MSC and Osteoblasts

Supplemental Fig. 2



Supplemental Fig. 3



B

	U2OS methylated	U2OS non-methylated
Genes in 0 to 20th percentile	161	1423
Genes in 21st to 100th percentile	223	6548

p-value < 0.001; Odds ratio = 3.3 (95%CI = 2.7 - 4.1)

C

	U2OS methylated	U2OS non-methylated
Genes in 81st to 100th percentile	47	1536
Genes in 0 to 80th percentile	337	6435

p-value < 0.001; Odds ratio = 0.6 (95%CI = 0.4- 0.8)

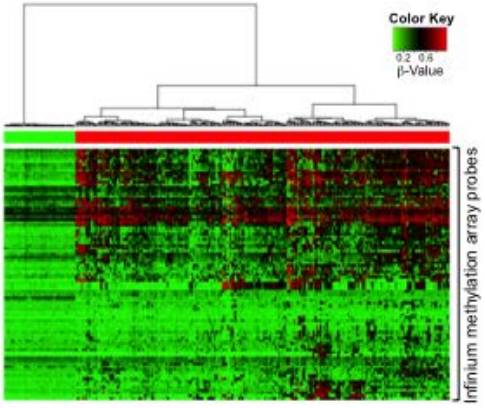
D

	U2OS methylated	U2OS non-methylated
ESC bivalent genes	303	2620
ESC non-bivalent genes	81	5351

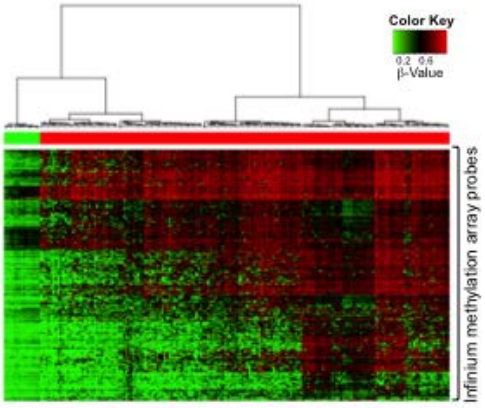
p-value < 0.001; Odds ratio = 7.6 (95%CI = 5.9 - 10)

Supplemental Fig. 4

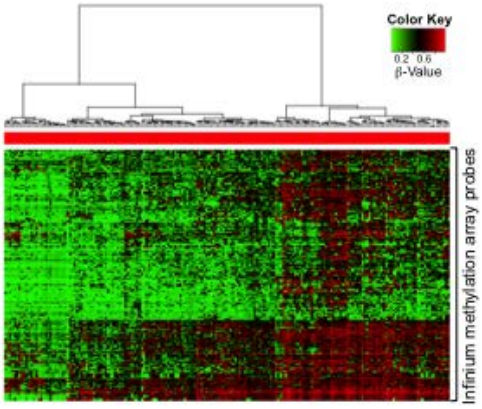
A



B



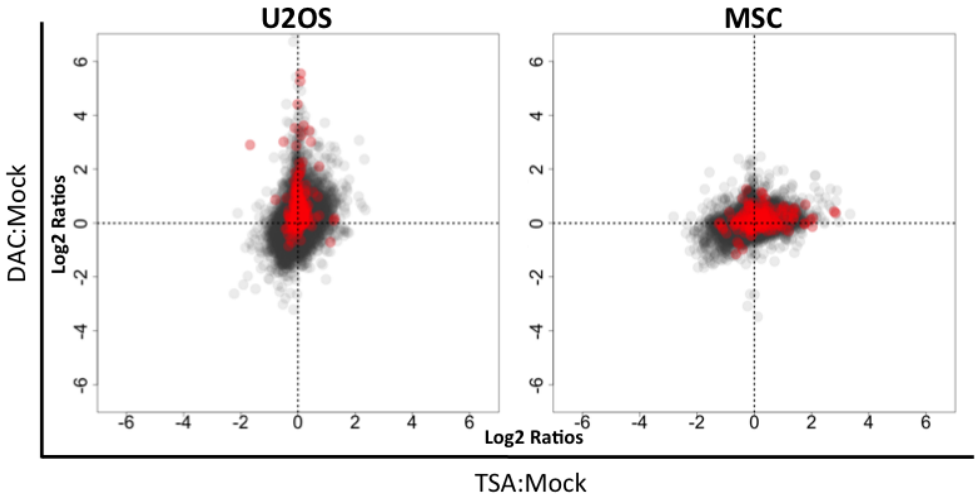
C



D

	Percent cell line methylated probes that are methylated ≥5% of TCGA samples	Percent cell line non-methylated probes that are methylated ≥5% of TCGA samples	p-value
Colon	99	16	$< 1.52 \times 10^{-164}$
Breast	99	18	1.52×10^{-164}
Lung	80	13	1.56×10^{-84}

Supplemental Fig. 5



Supplemental Fig. 6