

## Supplementary figure legends

**Figure S1. Antibody specificity test using ELISA.** The specificity of three H3K9me3 antibodies against various methylated or un-methylated histone peptides was tested using an ELISA assay. The plates were coated with 1 $\mu$ g/ml peptides overnight. Then the plates were blocked with 3% BSA in PBS at room temperature for 30 minutes and incubated with the primary antibodies (1:3000) for 3 hours at room temperature. The primary H3K9me3 antibodies tested are 39161 (Active Motif), 07-442 (Millipore), and 8898 (Abcam). The primary antibodies were diluted in ChIP reaction buffer (Millipore ChIP Kit, 90% Sample dilution buffer with 10% Lysis buffer). After rinses with PBS, plates were incubated with HRP-conjugated secondary antibodies (Promega) diluted in PBS for 1 hour at room temperature. After then rinses with PBS, the plates were incubated with the substrate and stop solution (R&D) and lastly read the absorbance at A450 (ELISA signal) and A540 (correction for plate background). The Y-axis represents the relative activity value determined using A450-A540.

**Figure S2. Comparison of H3K9me3 signal around SETDB1 peaks based on different H3K9me3 ChIP-seq experiments.** (A-H) Density plot showing the enrichment of H3K9me3 within the  $\pm 1$  kb regions of the SETDB1 binding sites (from Yuan *et al.*) based on eight different ChIP-seq experiments. Respective H3K9me3 ChIP-seq data for mouse ES cells generated in current research, and previous researches from Yuan *et al.* (GSE17642), Bilodeau *et al.* (GSE18371), Karimi *et al.* (two ES cell lines: J1 and TT2, GSE29413) were utilized. Kernel density estimation, a non-parametric way to estimate the probability density function of a random variable, was used here. The scale of y-axis indicates the possibility to observe H3K9me3 ChIP-seq reads intensity ( $\log_{10}$ -transformed) as a certain value (for example, x) in SETDB1 binding sites. (I-P) Average profiles of H3K9me3 ChIP-seq signal around SETDB1 solo and

ensemble peaks. SETDB1 solo and ensemble peaks were defined using SETDB1 ChIP-seq data from Yuan *et al.* and H3K9me3 ChIP-seq data from current research. All profiles were centered at the middle 1 bp of the SETDB1 binding regions. The average signal represents the average value of wiggle files (generated by MACS2) across a category of binding regions per 50-bp interval.

**Figure S3. Confirmation of SETDB1 binding at SETDB1 targets by ChIP-qPCR.** ChIP-qPCR was applied to confirm the SETDB1 binding at the SETDB1 solo and ensemble peaks. *Setdb1* iKO mES cells were treated with or without Tamoxifen and subjected to ChIP with SETDB1 antibody. Anti-rabbit normal IgG was used as the negative control. The selected SETDB1 peak loci were marked by their nearby gene symbols (*Atf4*, *Lhx2*, *Ascl1* for SETDB1 solo peaks; *Fam21*, *Gm6792*, *Sfi1* for SETDB1 ensemble peaks). Deletion of *Setdb1* diminished SETDB1 binding at all of these sites.

**Figure S4. Examination of H3K9me1 and H3K9me2 levels at SETDB1 solo and ensemble peaks.** ChIP-qPCR was applied to assess H3K9me1 and H3K9me2 levels at selected SETDB1 solo and ensemble peaks, respectively. It appears that SETDB1 solo peaks are generally devoid of H3K9me1 and H3K9me2 as well. The selected SETDB1 peak loci were marked by their nearby gene symbols (*Atf4*, *Lhx2*, *Ascl1*, *Foxa1*, *Tbx3* for SETDB1 solo peaks; *Fam21*, *Gm6792*, *Sfi1* for SETDB1 ensemble peaks).

**Figure S5. Confirmation of *Setdb1* deletion in iKO mES cells before neural differentiation.** The iKO cells were culture in adherent and treated with Tamoxifen for three days to induce *Setdb1* deletion. The deletion efficiency was confirmed by western blot analysis.

**Figure S6. Neural differentiation upon Tamoxifen treatment in control D3**

**mouse ES cells.** Control D3 cells were treated and assessed for neural differentiation the same as in iKO cells (Fig. 2C). No increased neural differentiation was seen from the Tamoxifen treatment. Scale bar = 50  $\mu$ m (100  $\mu$ m in the inserts).

**Figure S7. Relative binding level for 43 TFs at all SETDB1 ensemble and neural unrelated solo peak loci. (A-B)** The calculation of relative binding level was in the same way described for Fig 3A and the horizontal line represented the average binding intensity across all SETDB1 peaks.

**Figure S8. Summary of GO biological process enrichment analysis for target genes of SETDB1 solo peaks. (A)** SETDB1 solo peaks were divided into 3 groups based on the peak calling cutoffs for H3K9me3 data from current research. The looser the cutoff for H3K9me3 peak calling was, the more stringent standard for SETDB1 solo peaks would be. Each group stands for the SETDB1 solo peaks defined with certain stringent standard. **(B)** GO biological process enrichment analysis of target genes for different groups of SETDB1 solo peaks via DAVID interface (<http://david.abcc.ncifcrf.gov/>). Fisher's exact test and Benjamini correction were performed.

**Figure S9. Average profile of H3K27me3, EZH2, SUZ12 and RNF2 round the SETDB1 solo peaks generated by different MACS cutoffs. (A-D)** Solo peak numbers in different categories were listed in Figure S7 (A). All binding sites were extended to 10,000 bp and aligned by the center of the peak region detected by MACS2.

**Figure S10. H3K9me3 and H3K27me3 enrichment profiles among SETDB1 targets.** X-axis stands for normalized signal of H3K9me3 and Y-axis represents the normalized signal of H3K27me3 in control ES cells. SETDB1 binding sites were divided into two groups, solo and ensemble peaks and then

extended to the same width (5 kb upstream and downstream of peak center). The red dot points represent SETDB1 ensemble peaks whilst the green ones stand for SETDB1 solo peaks. For most SETDB1 solo peak regions, signal of H3K9me3 is comparative lower than that in SETDB1 ensemble peak region but the H3K27me3 enrichment level is much higher at a subset of SETDB1 solo peaks.

**Figure S11. Confirmation of target knocking down efficiency of shEZH2.**

*Setdb1* iKO mES cells were cultured in adherent and infected with lentiviral shRNA against EZH2. 48 hours after the infection, the cells were selected in 1  $\mu$ g/ml puromycin for another 48 hours to enrich infected cells. The target knocking down efficiency was assessed by Western blot analysis. The shRNA efficiently knocked down EZH2 expression accompanied by a significant reduction of global H3K27me3 as compared with the control shRNA (scramble sequence) or the no virus infected cells.

**Figure S12. H3K27me3 level variation at all  $\pm$ 10-kb regions of EZH2 peak loci before and after *Setdb1* iKO.** Only EZH2 peaks with enriched H3K27me3 signal before Tam treatment were included in this analysis (biological replicate pair 2 of current research). Those EZH2 peaks were divided into two categories based on the overlap status with SETDB1 binding: with SETDB1 binding (475 peaks) and without SETDB1 binding (14,143 peaks). Two-sided Welch's *t* test was performed for the change of H3K27me3 between the two groups (\*\*  $P < 0.01$ ).

**Figure S13. *Setdb1* deletion does not affect overall protein complex stability.** *Setdb1* iKO mES cells were treated with Tamoxifen for three days and assayed for EZH2, SUZ12 expression and histone H3K27 methylation by Western blot analysis. No reduction of EZH2, SUZ12 protein or global H3K27 methylation was seen.

**Figure S14. Quality control of H3K9me3 and H3K27me3 ChIP-seq data in this study. (A-H)** Correlation between H3K9me3 or H3K27me3 ChIP-seq signals and gene expression level was evaluated. One public RNA-seq dataset in mES (GSM1499144) was used and all genes were divided into three categories according to their FPKM values: low (0~10), moderate (10~100), and high (>100). For each H3K9me3 (A-D) or H3K27me3 (E-H) ChIP-seq biological replicate generated in this study, the average signal profiles across gene bodies for each gene category were drawn. (I-J) Fold enrichment score was used to demonstrate the signal enrichment for the MACS2-detected peaks in each H3K9me3 (I) or H3K27me3 (J) ChIP-seq experiment.

**Figure S15. Correlation between ChIP-seq datasets generated in this study and public H3K9me3 or H3K27me3 ChIP-seq datasets in mES. (A)** H3K9me3 ChIP-seq data for mES cells generated in current research, and previous researches from Yuan *et al.* (GSE17642), Bilodeau *et al.* (GSE18371), Karimi *et al.* (two ES cell lines: J1 and TT2, GSE29413) were used. Pearson correlation coefficient between signals at promoter regions (from 2kb upstream to 2kb downstream of TSS) of any two H3K9me3 ChIP-seq datasets was calculated. (B) H3K27me3 ChIP-seq data for mES cells generated in current research (without Tam treatment), and previous researches from Lienert *et al.* (GSE27827) and Stadler *et al.* (GSE30203) were utilized. Pearson correlation coefficient between signals at promoter regions (from 2kb upstream to 2kb downstream of TSS) of any two H3K27me3 ChIP-seq datasets was calculated.

## Supplementary tables

**Table S1.** List of transcription regulators used in the measurement for the level

of co-binding between SETDB1 and other factors

Transcription regulator	Peak number	GEO Accession ID
ATF7IP	5064	GSE26680
CHD7	5352	GSE22341
MYC	6184	GSE11431
CTCF	30300	GSE25777
CTR9	2182	GSE20530
DPY30	36708	GSE26136
E2F1	29936	GSE11431
ESRRB	83541	GSE11431
EZH2	18301	GSE18776
JARID2	7292	GSE18776
KDM2A	28556	GSE18588
KDM5A	9291	GSE18776
KLF4	11987	GSE11431
LAMINB	52567	GSE28247
MED1	26148	GSE22557
MED12	28541	GSE22557
MTF2	6255	GSE16526
NANOG	19318	GSE11724
NIPBL	14357	GSE22557
MYCN	7742	GSE11431
NR5A2	3072	GSE19019
POU5F1	26804	GSE11724
EP300	3290	GSE22341
REST	9224	GSE27841
RBBP5	12656	GSE22934
RNAP2	18491	GSE12680
RNF2	8657	GSE26680
SETDB1	5882	GSE17642
SMAD3	2040	GSE21614
SMARCA4	3780	GSE14344
SMC1A	60633	GSE22557
SMC3	31443	GSE22562
SOX2	4516	GSE11431
STAT3	6355	GSE11431
SUPT5H	8215	GSE20530
SUZ12	38174	GSE18776
TAF1	7992	GSE30959
TAF3	51014	GSE30959
TCF3	6113	GSE11724
TCFCP2L1	43947	GSE11431
TET1	34697	GSE26832
WDR5	4877	GSE22934

WHSC2	21684	GSE20530
ZFX	13784	GSE11431

Forty-three transcription factors known to be involved in ES cell regulation were selected for co-binding analysis with SETDB1. Listed are the factor name, the number of detected peak and the GEO accession ID from which the data were generated.

**Table S2. List of primers used in this study**

ChIP qPCR primers	Sequences (5'-- 3')
<i>Ascl1</i>	GAACCCGCCATAGAGTTCAA GTTGGTCAACCTGGGTTTG
<i>Lhx2</i>	AGTGGAAAGGGGAGTGGAAAGT GGTCACGATCAACGGAAAGT
<i>Atf4</i>	CGTCAGAGCTTGGCTAGGT GGGGTAACTGTGGCGTTAGA
<i>Sfi1</i>	AGACTCCCAATACCCACACG GCCTCCCACAAATGACTGTT
<i>Gm6792</i>	CACTTCCAGGCAGCTAGGAC CAGAATTCCCAGGCTGACAT
<i>Fam21</i>	TGGGATACTTGCCCTTCCTG TTCTGGCAATGTGTGTGGT
<i>Faah</i>	AACGGTCAAGGTCAAGCCTA TTCTGAGGACCCCTGGCTTAG
<i>Pou3f1</i>	GTTCTCGCAGACCACCATCT GGCTTGGGACACTTGAGAAA
<i>Tcfap2e</i>	GGACCCGAGTCTACTCACCA ACCCCGTACCTCTGGACTC
<i>Foxo6</i>	GGTGCCAGGACTGGTTAAGA TAAGCCACCCAACCTCCAATC
<i>Rnf220</i>	AGGACAGCCAGTTAGCCTGA CTCCAGCCAGCCTGCTATAC
<i>Tbx3</i>	GCTGCTGCGAACTCTCTTCT ACCAATTGTGTGGCTGCATA
<i>Foxa1</i>	GCAACCGGGTCAAGTTGGT CAAGTCGTTGGAGTAGTTGGG
<i>Nr2f1</i>	CACCAAGGCGAGAGAGAAAG ACAGGAACTGTCCCATCGAC

RT qPCR primers	Sequences (5'--3')
<i>Foxa2</i>	CCCTACGCCAACATGAACTCG

		GTTCTGCCGGTAGAAAGGGA
<i>Pax7</i>		TCTCCAAGATTCTGTGCCGAT
		CGGGGTTCTCTCTTATACTCC
<i>Nr2f2</i>		TCAACTGCCACTCGTACCTG
		CCATGATGTTGTTAGGCTGCAT
<i>Nr2f1</i>		GCACTACGGCCAATTCACCT
		TTGGAGGCATTCTTCCTCGC
<i>Hmx2</i>		CCAAGCCCTAAGCACCATACC
		TCCTTAAAGTCCGGGTGAGAAG
<i>Zic1</i>		CAGTATCCCGCGATTGGTGT
		GCGAACTGGGGTTGAGCTT
<i>Lhx5</i>		AATGTGTTCAATGCTGCGAGT
		AAAGCGCCTGAAAAAGTCGTT
<i>Hoxc8</i>		CTTCGTCAACCCCCCTGTTTC
		GTCTTGGACGTGGTGCAG
<i>Evx1</i>		GAGAGCCGAAAGGACATGGTT
		CTGCCTGCTAGTCCATCGAC
<i>Ascl1</i>		GCAACCGGGTCAAGTTGGT
		CAAGTCGTTGGAGTAGTTGGG
<i>Uncx</i>		ACCCGCACCAACTTACCG
		TGAACTCGGGACTCGACCA
<i>Hoxd3</i>		ATACTACGAGAACCCAGGACTC
		CAGGTCTCATACAGCTACCATTG
<i>Fezf1</i>		ACGCGACCACCAAAATGCTA
		TCGACGAGTTGAGATGCAGAG
<i>Hoxb8</i>		TGCGCCCCAATTATTATGACTG
		AACTCCTGGATTGCGAAGGG
<i>Gfra1</i>		CACTCCTGGATTGCTGATGT
		AGTGTGCGGTACTTGGTGC
<i>Hoxd9</i>		GCACCCTCAGCAACTACTACG
		AAAATACACGAGGGCGAACTC
<i>Fev</i>		ACGCCTACCGCTTGACTTC
		AAGCTGCCATCAAGTTGAGTT
<i>Hoxb2</i>		ATTCGCCTTTCTACCGGACC
		GGGCTATCGAGAGAACCTG
<i>Irx5</i>		TACAGCACCGCGTCATTG
		GAGCCCACGTAAGAGAAGGC
<i>Barhl2</i>		GCGGGTCGAGTTTGGAAATAG
		GCTCCTAAAATCCGTTGTCCTC
<i>Lrp2</i>		GAAAGCTGGACAAGACTGAGTT
		GGCAGTGGTCTGTGAGAATT
<i>Foxa1</i>		ACATTCAAGCGCAGCTACCC
		TGCTGGTTCTGGCGGTAATAG
<i>Sema6a</i>		CGAGACGGGTATGAGTCTAGG

		CCTGGTGGTTATGAGAGGACAC
<i>Lmx1b</i>		TCCTGGCACGAGGAGTGT
		AACTCGGTAGGGCGATCTT
<i>Pax3</i>		TTTCACCTCAGGTAAATGGGACT
		GAACGTCCAAGGCTTACTTTGT
<i>Satb2</i>		GCCGTGGAGGTTGATGATT
		ACCAAGACGAACTCAGCGTG
<i>Foxg1</i>		CAAGGCTGACGCACTTGGA
		CTTGGCGTTCTTCTTGTGCG
<i>Hoxb3</i>		GCACCTGGAGGGTGAETAC
		CCCCCGTTATTGCTGTTGC
<i>Six1</i>		ATGCTGCCGTCGTTGGTT
		CCTTGAGCACGCTCTCGTT
<i>Otp</i>		CAGGCTAGGTATGAAAGATGCC
		GAAGCAGGGTAGAGCCCCA
<i>Lhx2</i>		CTGTTCCACAGTCTGTCGGG
		CAGCAGGTAGTAGCGGTCAG
<i>Gfra2</i>		TCTTTTAGACGAAACCCCTCCG
		CCAGCATGGTATTGCGATCC
<i>Egr3</i>		TTGCCTGACAATCTGTACCCC
		TAATGGGCTACCGAGTCGCT
<i>Nkx2-1</i>		AGGACACCATGCGGAACAG
		CCATGCCGCTCATATTCATGC
<i>Neurog2</i>		AACTCCACGTCCCCATACAG
		GAGGCGCATAACGATGCTTCT
<i>Pitx1</i>		GCCTTCAAGGGAGGCATGAG
		GATTGCTGGCGGAGTTCTC
<i>Sall3</i>		TGGAGGTATCTACGGACAAGG
		GGGTACATGGTACTAGGCA
<i>Tbx3</i>		AGATCCGGTTATCCCTGGGAC
		CAGCAGCCCCCACTAACTG
<i>Prdm6</i>		GGAGCCTAGTAAGTCGAGCTG
		GGACGTTCAAGTTTCATCCTGG
<i>Pou4f2</i>		TGGACATCGTCTCCCAGAGTA
		GTGTTCATGGTGTGGTAAGTGG
<i>Zic2</i>		CAAGGTCCGGGTGCTTACC
		ATTAAAGGGAGGCCCGAATA
<i>Hnf1b</i>		AGGGAGGTGGTCGATGTCA
		TCTGGACTGTCTGGTTGAECT
<i>Sema5b</i>		GAAGCCGTGGGTCTTTAECTT
		CAAGAGCAAGCTGGGAGAAAT
<i>Otx2</i>		TATCTAAAGCAACCGCCCTACG
		GCCCTAGTAAATGTCGTCCTCTC
<i>Rax</i>		TGGGCTTACCAAGGAAGACG

	GGTAGCAGGGCCTAGTAGCTT
<i>Eomes</i>	GCGCATGTTCTTCTTGAG
	GGTCGGCCAGAACCACTTC
<i>Six3</i>	GCTCCGGCTCTCTTACC
	CGGCGAAGTTGGCAACAAAG
<i>Ascl2</i>	AAGCACACCTGACTGGTACG
	AAGTGGACGTTGCACCTCA
<i>Sim1</i>	ACAGGAGTACGAGATCGAACG
	CTCTGTCACAGCACTCGGAG
<i>Faah</i>	CCCTGCTCCAAGTGGTACAG
	TCACAGTCAGTCAGATAGGAGG
<i>Bmp1</i>	ATGGGGTCTCTGGTTCTGC
	CAATACCATCTGCTCCGTGAA
<i>Pou3f1</i>	TACCGCGAAGTGCAGAACG
	CGTGGGTAGCCATTGAGGG
<i>Alx1</i>	GAGCCCGCCAAGTAAAAACAG
	GTCGCTTGCCGTAAAAGGAC
<i>Foxo6</i>	CCCCGGACAAGAGACTCAC
	CGACAGGTTGTGCCGAATG
<i>Fgf2</i>	CCAGGACAGTATAGTGGAGATCC
	AGTAGACCCCGCGACCCATAG
<i>Hes5</i>	AGTCCAAGGAGAAAAACCGA
	GCTGTGTTCAGGTAGCTGAC
<i>Hoxa1</i>	CCACCAGGGTTATGCTGGG
	CGTGGAGAGGGATAAGGAGTTA
<i>Cbx2</i>	GGCTGGCCTCCAAACACAA
	CCCTGGGTCTCTGCCCTCT
<i>Gbx2</i>	CAACTTCGACAAAGCCGAGG
	ACTCGTCTTCCCTGCCCT
<i>Setdb1</i>	GGAGGAACCTCGTCAGTACATTG
	TCTTCTGTAGTACCCACGTCTC
<i>Gapdh</i>	AATGGATTGGACGCATTGGT
	TTTGCACTGGTACGTGTTGAT

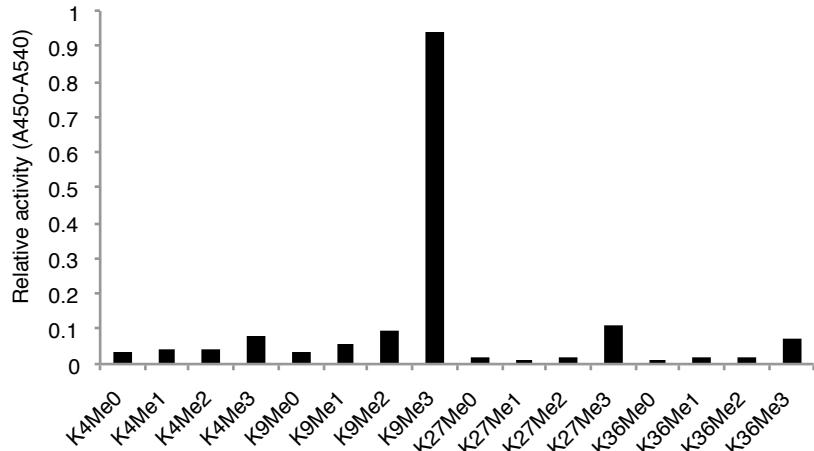
**Table S3.** Summary of ChIP-seq datasets generated in this study.

ChIP-seq experiment	Uniquely mapped reads	Library complexity
H3K9me3 Replicate 1 (single-end)	22.7 M	0.591
H3K9me3 Replicate 2 (paired-end)	14.3 M	0.810
H3K9me3 Replicate 3 (paired-end)	16.6 M	0.831

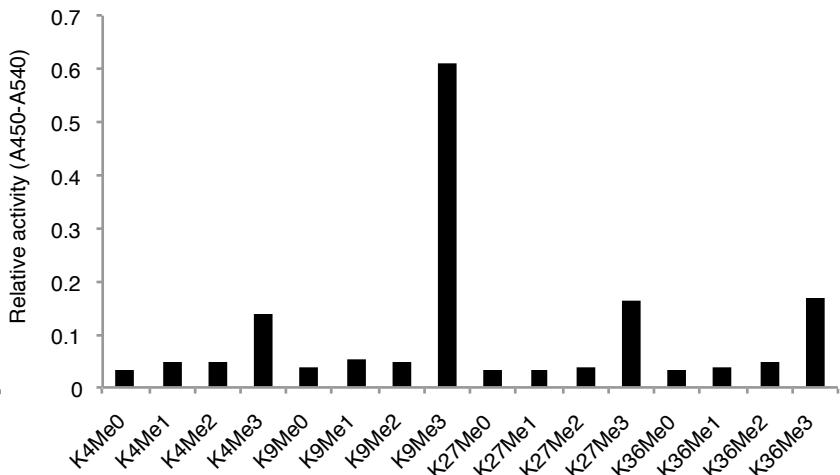
H3K9me3 Replicate 4 (paired-end)	15.4 M	0.838
H3K27me3 DMSO-treated Replicate 1 (single-end)	18.6 M	0.935
H3K27me3 Tam-treated Replicate 1 (single-end)	22.8 M	0.655
H3K27me3 DMSO-treated Replicate 2 (paired-end)	24.5 M	0.763
H3K27me3 Tam-treated Replicate 2 (paired-end)	24.7 M	0.587

Figure S1

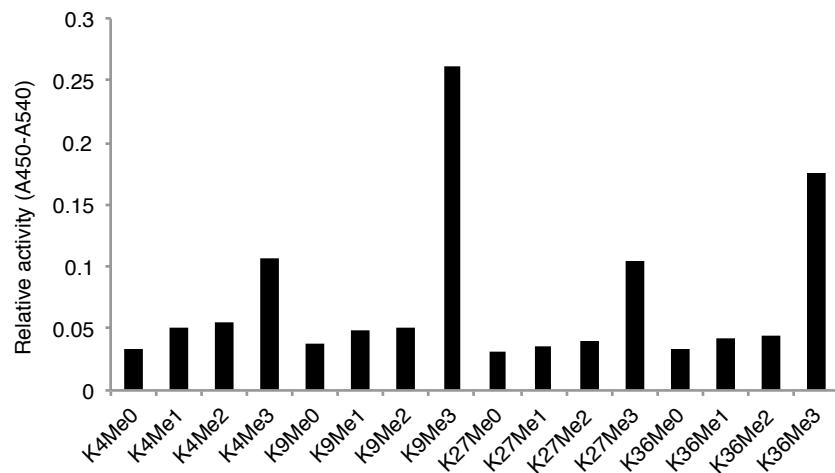
Active Motif-39161

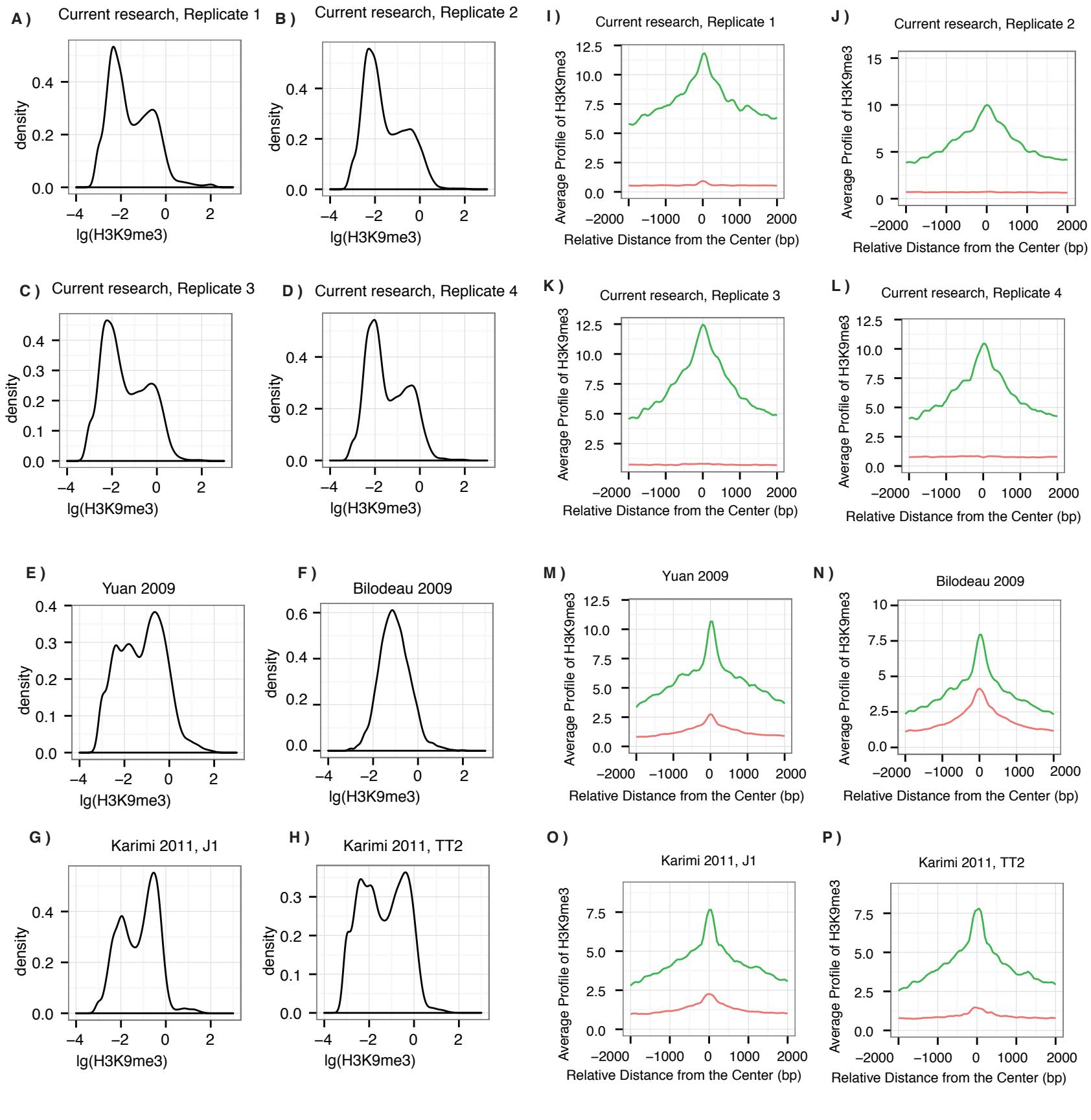


Abcam-8898



Millipore-07-442



**Figure S2**

**Figure S3**

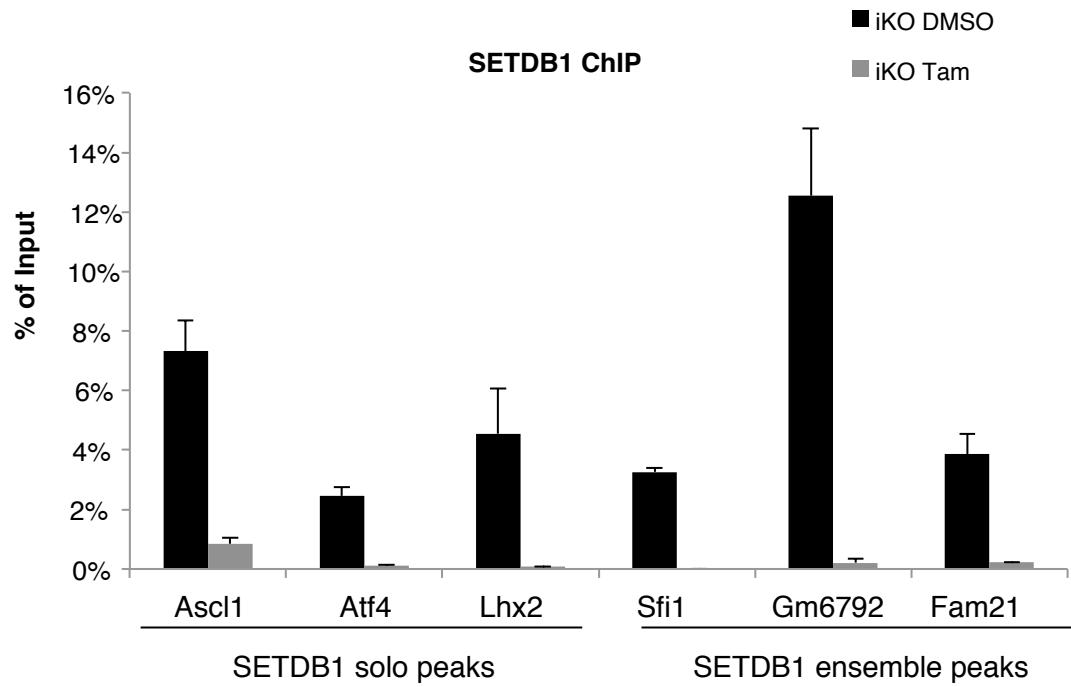
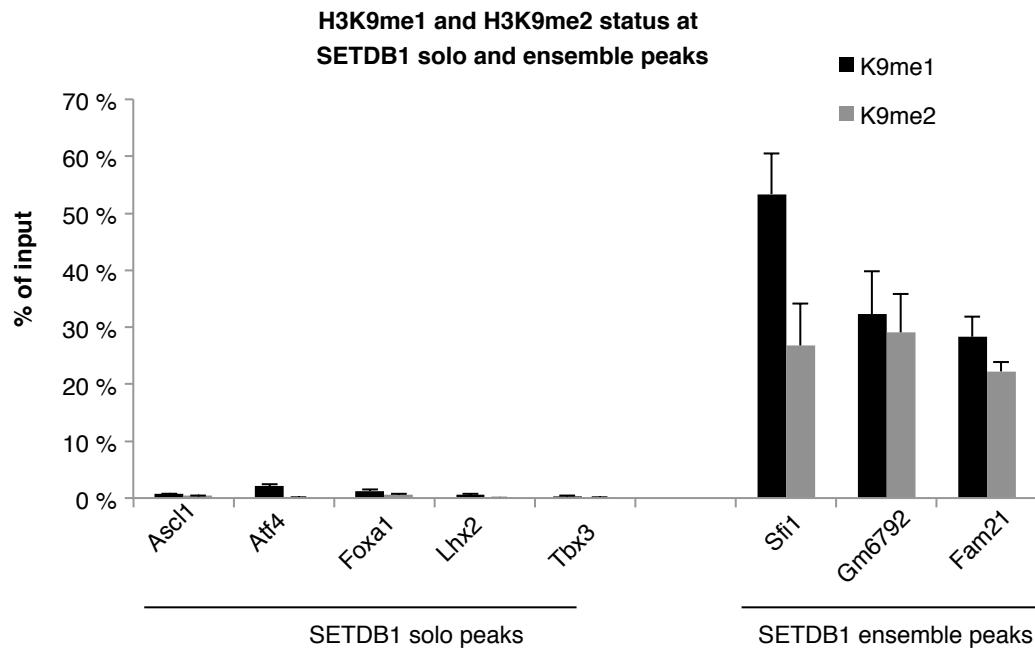


Figure S4



**Figure S5**

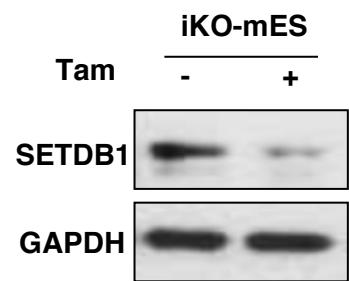
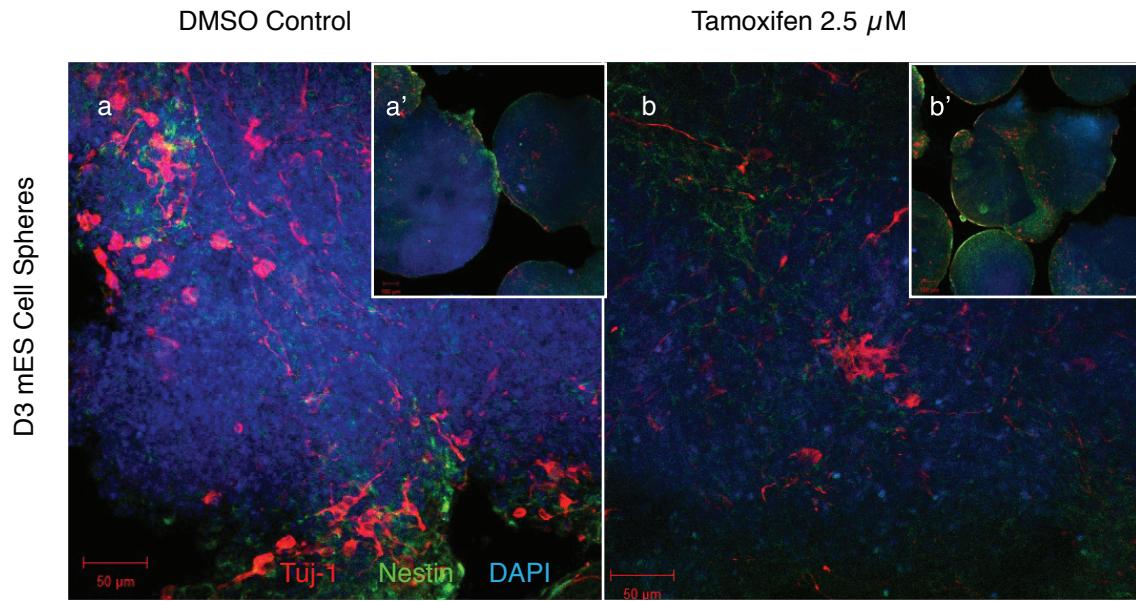
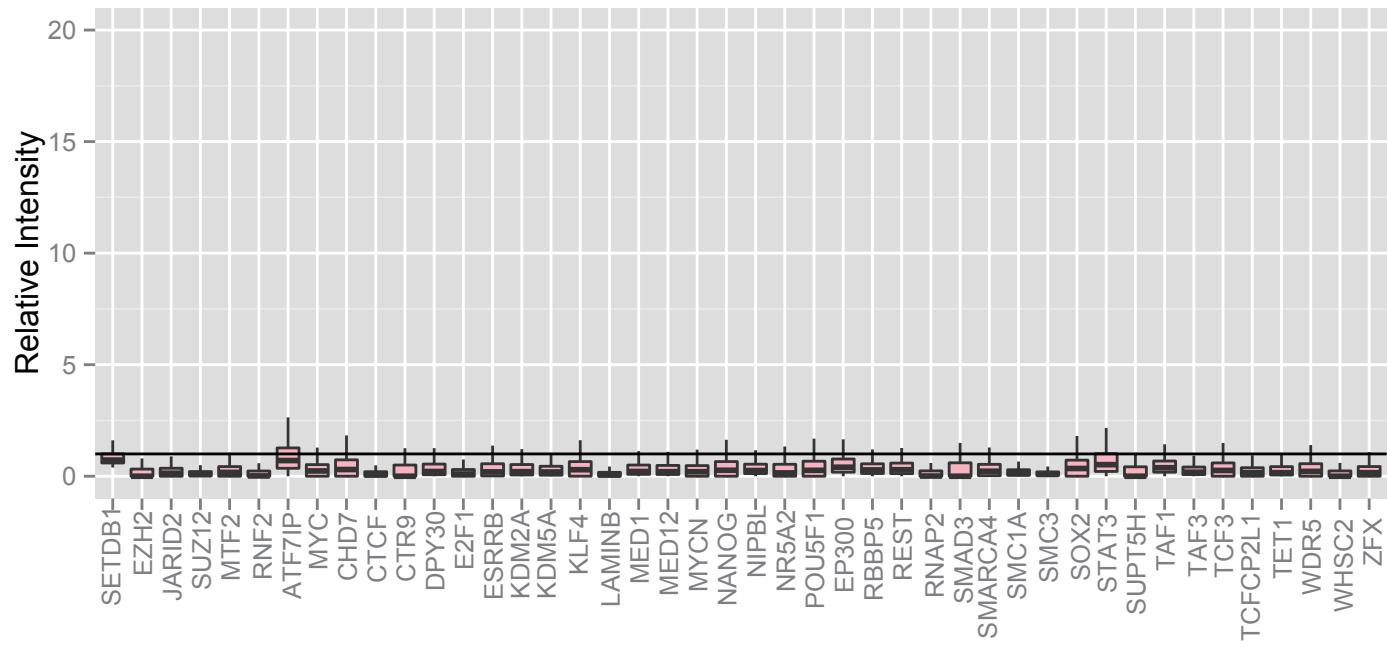


Figure S6

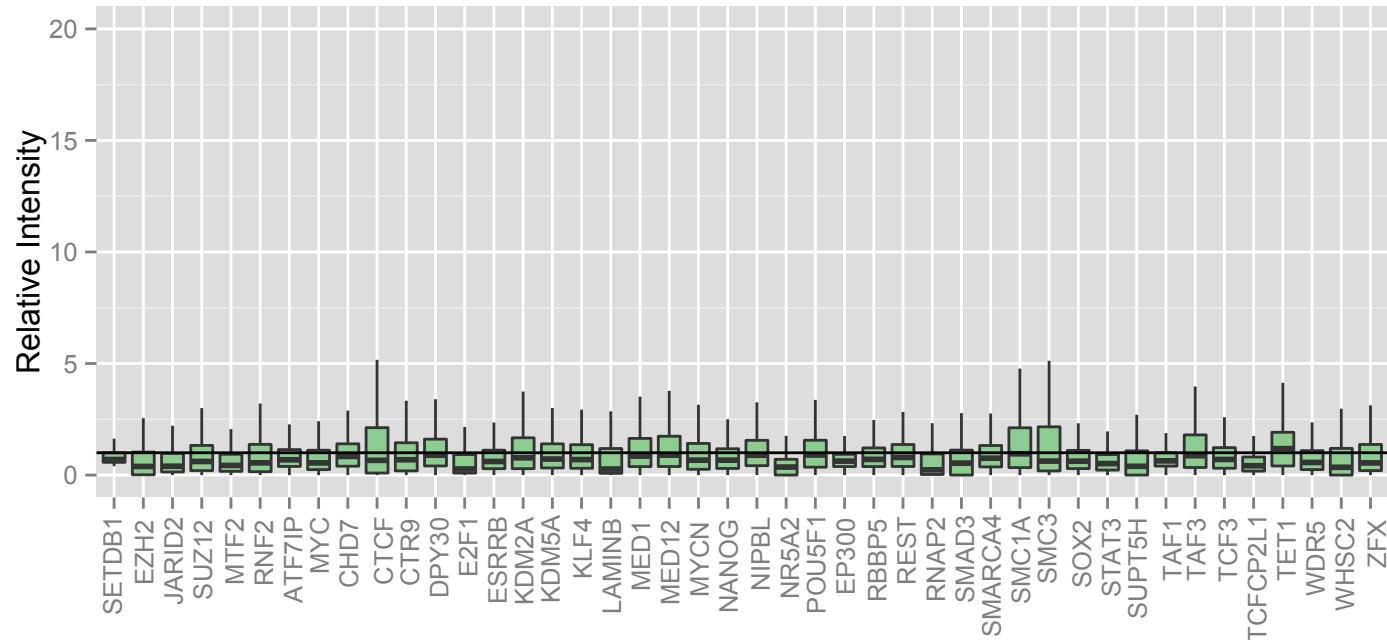


**Figure S7**

**A)**



**B)**



**Figure S8****A)**

SETDB1 solo peak Group	H3K9me3 MACS2 callpeak Cutoff	Solo Peak number	Gene number (Refseq ID)
Loose	q 0.01 ~ q 0.001	534	189
Moderate	p 0.001 ~ q 0.01	579	185
Stringent	p 0.001	2110	1260

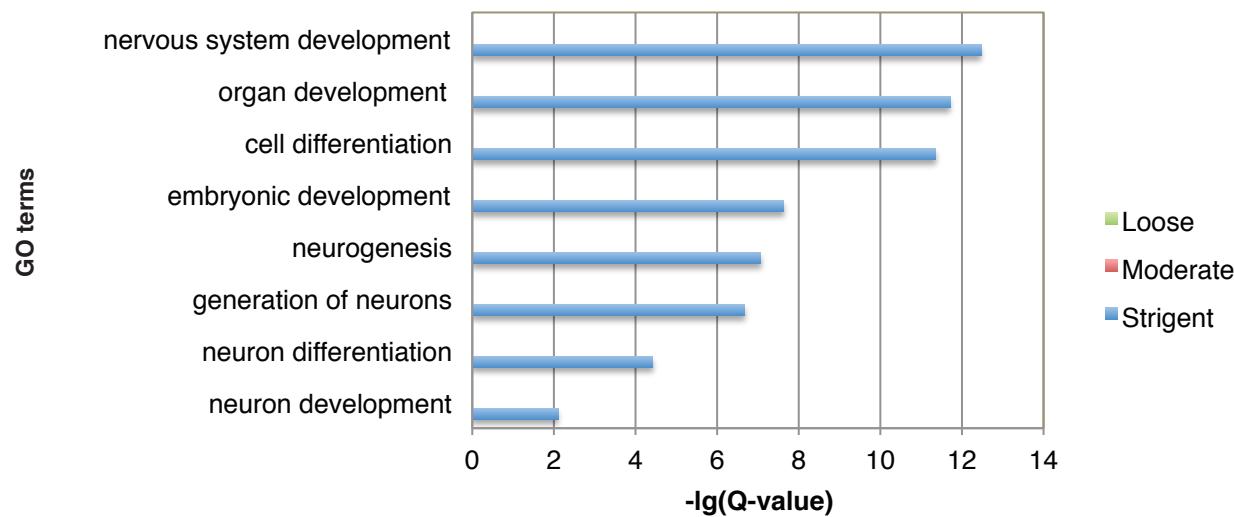
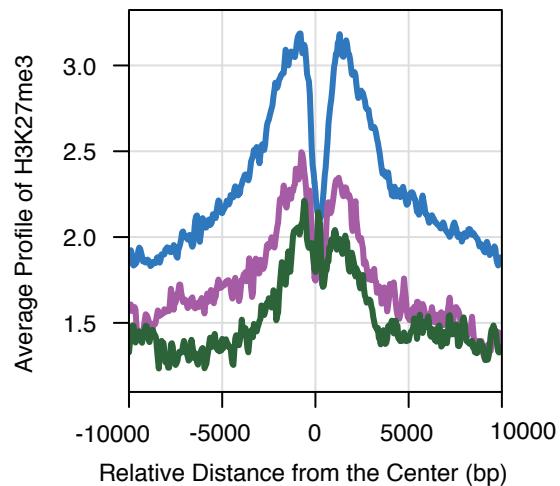
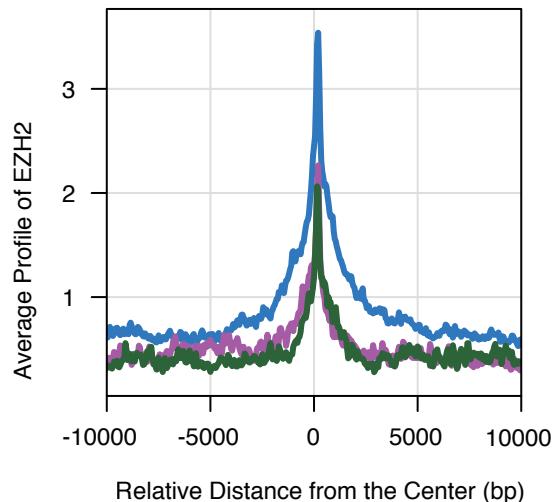
**B)**

Figure S9

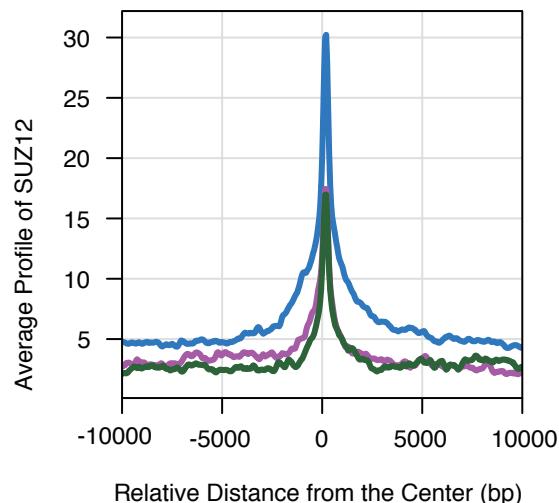
A )



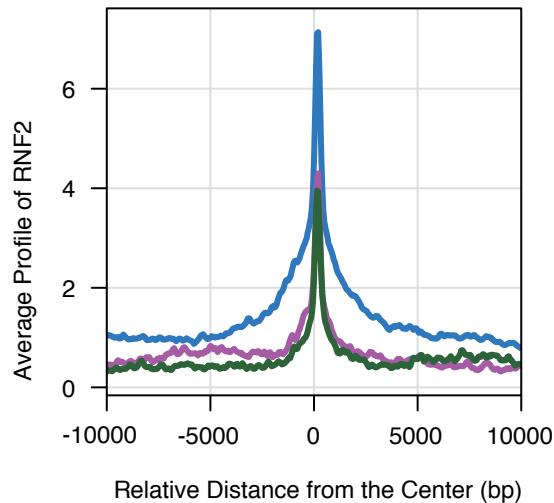
B )



C )

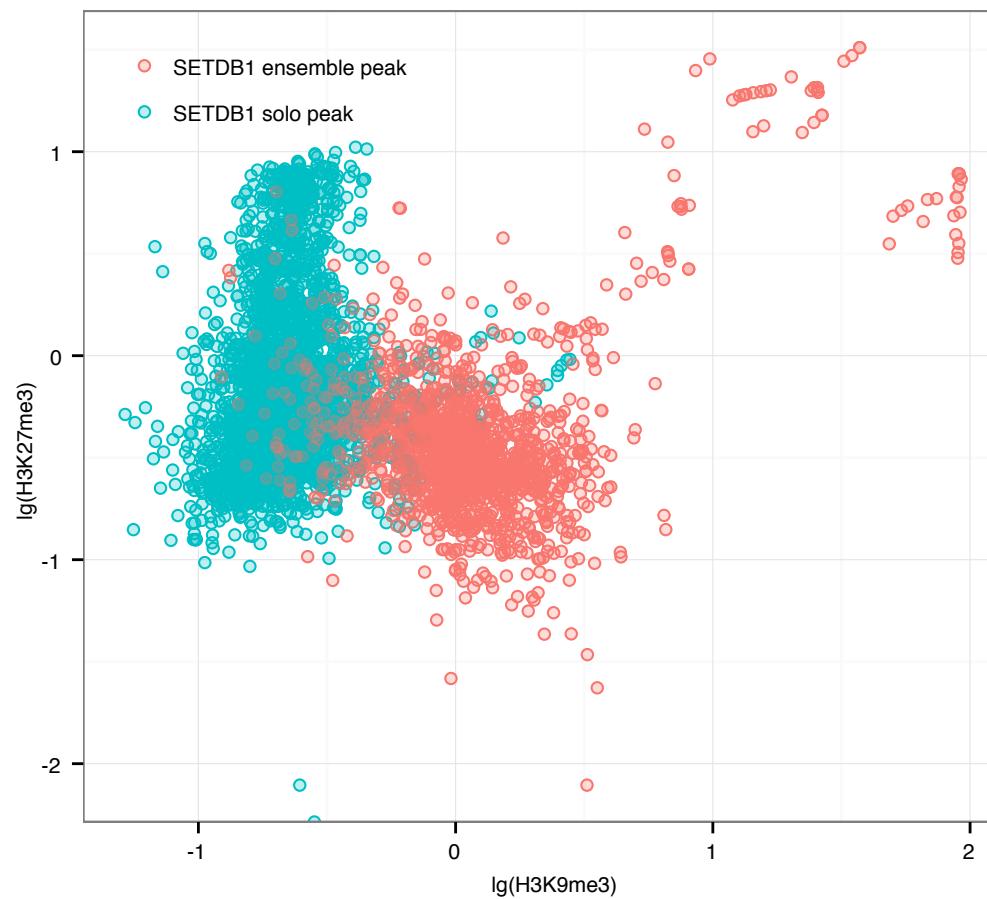


D )



Stringent ————— Moderate ————— Loose —————

Figure S10



**Figure S11**

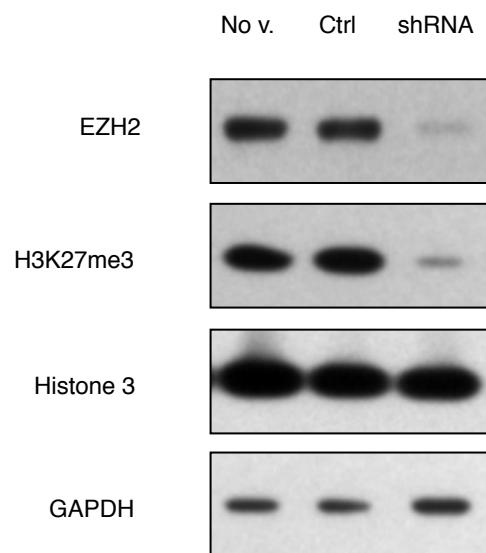
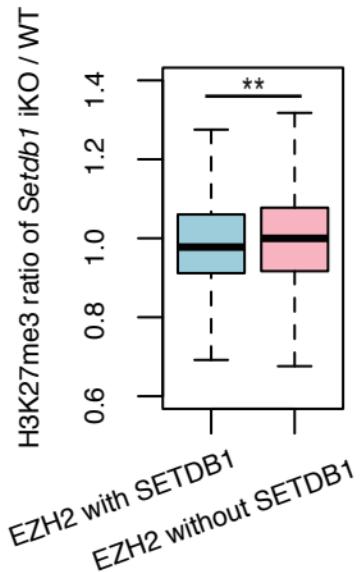


Figure S12



**Figure S13**

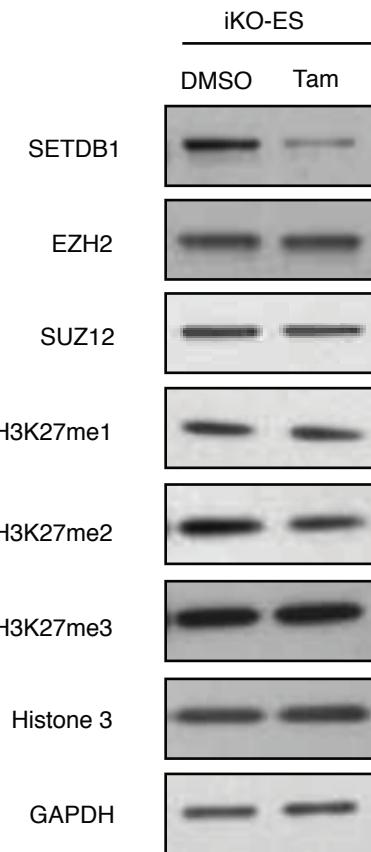


Figure S14

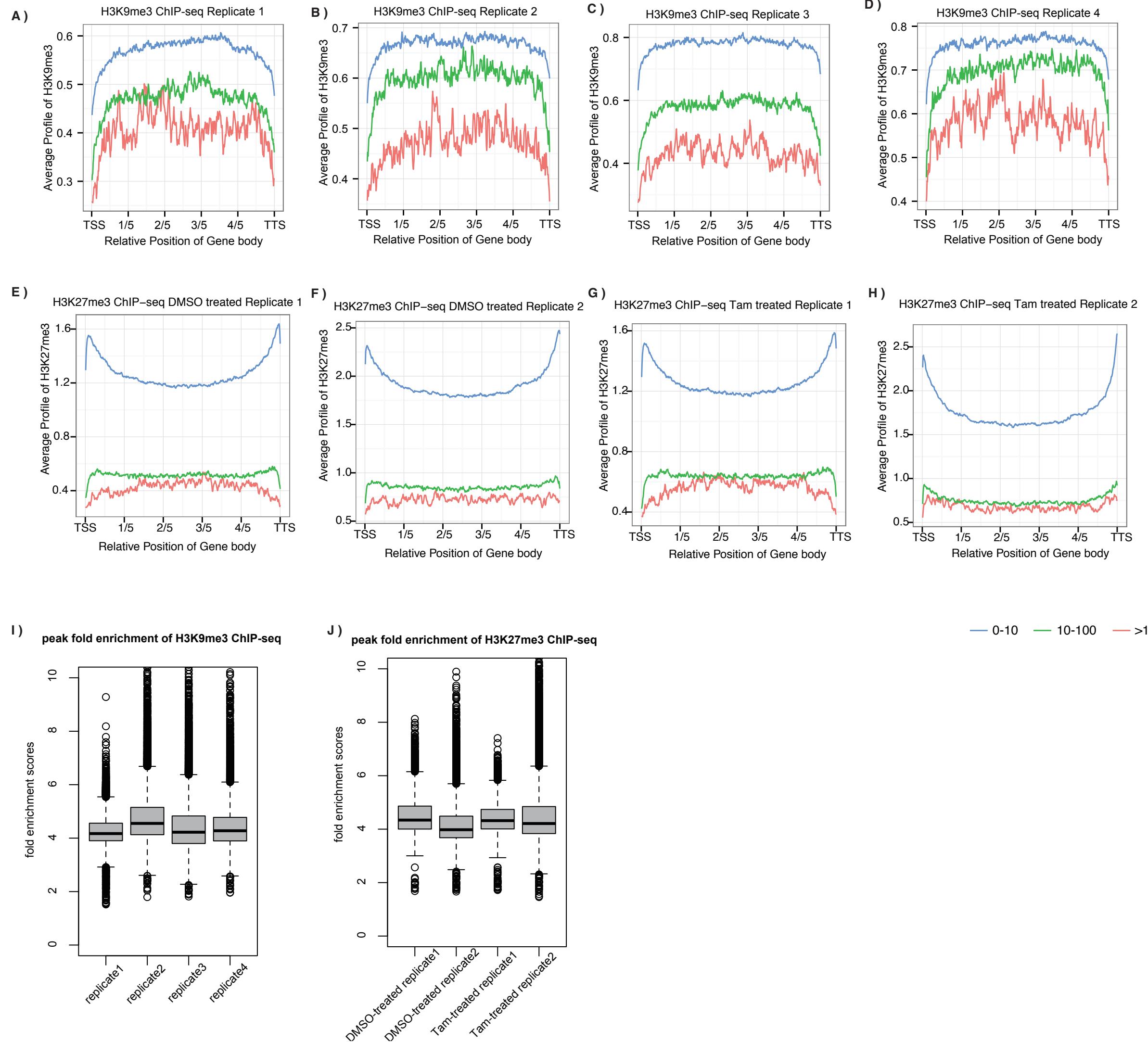


Figure S15

