Supplemental Methods

Alkaline phosphatase staining and imaging. Alkaline-phosphatase staining was performed on ESCs using the Red AP Substrate Kit (Vector Laboratories) following the manufacturer's directions. ESCs and EBs were captured using the Nikon Eclipse Ts2R microscope and DS-Qi2 camera. EB sizes were measured using Nikon's NIS-Elements Basic Research software.

Quantitative RT-PCR (qRT-PCR). Total RNA was extracted using RNeasy mini kit (Qiagen) following the manufacturer's instructions, and cDNA was synthesized using the High Capacity RNA-to-cDNA Kit (Applied Biosystems). Resulting cDNA levels were measured on a CFX connect Real-Time PCR detection system (Bio-Rad) using the Maxima SYBR Green/ROX qPCR Master Mix (ThermoFisher), and relative expression to *Actin* was calculated. The following primers used in the qRT-PCR assays: *Hoxa* genes (Lin et al. 2011; Zhang et al. 2012; Cao et al. 2017); *Sox2* forward: GAACGCCTTCATGGTATGGT; *Sox2* reverse: TCTCGGTCTCGGACAAAAGT; *Oct4* forward: AATGCCGTGAAGTTGGAGAA; *Oct4* reverse: CCTTCTGCAGGGCTTTCAT; *Nanog* forward: TGCTTACAAGGGTCTGCTACTG; *Nanog* reverse: GAGGCAGGTCTTCAGAGGAA; *Bmp5* forward: CCACAGAACAATTTGGGCTTA; *Bmp5* reverse: AGTACCTCGCTTGCCTTGAA; *Epor* forward: GTCCTCATCTCGCTGTTGCT; *Epor* reverse: ATGCCAGGCCAGATCTTCT.

ChIP-seq, RNA-seq, and next generation sequencing data processing. ChIP was performed as described previously (Lee et al. 2006; Cao et al. 2017; Piunti et al. 2017). In brief, ESCs were fixed in 1% formaldehyde followed by quenching, cell lysis, and chromatin shearing with an

E220 focused ultrasonicator (Covaris). Sonicated chromatin was subsequently subjected to immunoprecipitation. Immunoprecipitated DNA was then washed, eluted, reverse-crosslinked, and purified prior to library preparation using the KAPA HTP library preparation kit (KAPA Biosystems). ChIP-seq experiments in EBs were performed similarly except EBs were dounce-homogenized prior to sonication. RNA was isolated using RNeasy mini kit (Qiagen) following the manufacturer's instructions, and RNA-seq libraries were made using the TruSeq Stranded Total RNA Preparation Kit (Illumina). ChIP-seq and RNA-seq libraries were single-read-sequenced on the NextSeq 500 Sequencing System (Illumina). The raw BCL output files were processed using bcl2fastq (Illumina, version 2.17.1.14), followed by quality trimming using Trimmomatic (Bolger et al. 2014). Trimmed reads were then aligned to the mouse genome (UCSC mm9) using TopHat version 2.1.0 (Trapnell et al. 2009) for RNA-seq reads and Bowtie version 1.1.2 (Langmead et al. 2009) for ChIP-seq reads. Only uniquely mapped reads with a maximum of two mismatches over the whole length of the gene were considered for ensuing analyses. The gene annotations used came from Ensembl release 72. Mapped ChIP-seq reads were extended to 150 base pairs (bp) to represent sequenced fragments. Raw read counts from both ChIP-seq and RNA-seq were normalized to total reads per million (RPM), and were formatted as bigwig coverage plots to generate UCSC genome browser tracks. For RNA-seq, exonic reads were assigned to specific genes from Ensembl release 72 using htseq-count script from the Python package HTSeq 0.6.1 (Anders et al. 2015).

ChIP-seq analysis. Peaks were called using MACS version 1.4.2 with default parameters. Heat maps and metaplots were generated using ngsplot (Shen et al. 2014). ChIP-seq occupancy levels in RPM were aligned to peaks as indicated, and were either sorted by peak width or partitioned

into groups by K-means clustering. Differential heat maps show $\log_2$ fold changes in peak occupancy of mutant cells relative to wild-type cells. Metaplots illustrate average ChIP-seq occupancy in RPM. Genome-wide Set1A distribution was determined by ChIP-seq and calculated by HOMER (Heinz et al. 2010). For the evaluation of differential H3K4me3 occupancy levels shown in the MA plot (Supplemental Fig. S4B), BEDTools (Quinlan and Hall 2010) was used to quantify the read counts within peaks, edgeR version 3.0.8 (Robinson et al. 2010) was used to evaluate statistical differences, and custom R scripts were used for data plotting.

RNA-seq analysis. Gene count tables were quantified according to Ensembl gene annotations and used as input for edgeR version 3.0.8 (Robinson et al. 2010). Genes with Benjamini-Hochberg-adjusted *P*-values of <0.01 were considered differentially expressed. Custom Perl and R scripts were used to generate MA plots (log ratio and mean average). The heat map in Supplemental Figure S3A features gene expression levels from differentially expressed genes identified by edgeR that were normalized and converted into Z-scores, and the results were visualized using the pheatmap R package, where the genes and samples were subjected to unsupervised hierarchical clustering. GO functional analysis in Supplemental Figure S3A was conducted using Metascape (http://metascape.org/) (Tripathi et al. 2015) with default parameters.

Supplemental Figure Legends

Supplemental Figure 1. CRISPR/Cas9-generated Set1A$^{\Delta SET}$ ESCs (related to Figure 1).

(A) Diagram of the known domain organization of mouse Set1A protein.

(B) PCR genotyping results of WT vs. Set1A$^{\Delta SET}$ ESCs. Arrowheads indicate base pair (bp) size of PCR products.

(C) Top: Schematic of the Set1A genomic locus and the two CRISPR/Cas9 cut sites targeting the SET domain (green). gRNA sequences are in orange, and PAM sequences are in red. Bottom: Sanger sequencing of genomic DNA (gDNA) and complementary DNA (cDNA) revealed the precise sequences deleted (indicated in bp) in Set1A$^{\Delta SET}$ ESCs. Early STOP codon introduced (highlighted in red) as a result of SET domain deletion. Start and stop codons of SET domain are indicated in green.

(D) qRT-PCR analysis of expression levels of pluripotency factors *Sox2*, *Nanog*, and Oct4 in ESCs.

Supplemental Figure 2. Set1A$^{\Delta SET}$ ESCs exhibit decreased H3K4me3 at certain sites corresponding with highest nearest gene expression despite no change in Pol II occupancy (related to Figure 2).

(A) Genome-wide distribution of Set1A binding in ESCs relative to gene structure as determined by ChIP-seq and HOMER annotation.

(B) Heatmaps of Set1A binding in WT and Set1A$^{\Delta SET}$ cells, with occupancy levels centered at WT peaks sorted by decreasing peak width.

(C) Log2 fold changes in H3K4me3 occupancy were determined in Set1A$^{\Delta SET}$ relative to WT cells for cluster 1 peaks identified and ordered in Fig. 2C.

(D) Box plot showing distribution of expression level for genes nearest to Set1A peaks corresponding to clusters presented in Fig. 2C.

(E) GREAT analysis used to display distribution of H3K4me3 peak distances relative to TSSs for cluster 1 peaks identified in Fig. 2C.

Supplemental Figure 3. Genes downregulated in Set1A$^{\Delta SET}$ day 6 EBs compared to WT EBs linked to differentiation and development (related to Figure 3).

(A) Expression heatmap of genes downregulated in Set1A$^{\Delta SET}$ EBs relative to WT EBs. Two replicates of day 6 EBs were generated and harvested for RNA-seq for WT and mutant EBs.

(B) Venn diagram illustrating that of the 3,314 genes upregulated during wildtype differentiation (yellow and green sections), 447 genes are significantly downregulated (green section) in Set1A$^{\Delta SET}$ EBs compared to WT EBs. Differentially expressed genes were identified using the criteria as shown. Colors also correspond to categories described in Fig. 3C.

(C) GO analysis for the 447 differentially expressed genes identified in panel (B) and in Fig. 3C as determined by Metascape (Tripathi et al. 2015).

(D) qRT-PCR analysis of expression levels of the two example genes (see Fig. 3D) downregulated in Set1A$^{\Delta SET}$ EBs compared to WT EBs. Both ESC and day 6 EB expressions were examined. P-value was determined by the Student's t-test, with significance level of 0.001 (denoted by ***).

(E) Expressions of pluripotency factors *Nanog*, *Oct4*, and *Sox2* are noticeably increased in the mutant EBs vs. WT EBs.

(F) qRT-PCR analysis of expression levels of pluripotency factors in day 6 EBs. P-value was determined by the Student's t-test, with significance level of 0.001 (denoted by ***).

(G) WT and Set1A$^{\Delta SET}$ ESCs were cultured in N2B27 media without 2i/LIF and induced towards neuronal lineage. Images were taken at day 1, day 2, and day 3 of N2B27 culturing. Black scale bar = 100μm.

(H) qRT-PCR analysis of expression levels of *Hoxa4*, *Hoxa5*, and *Hoxa7* genes in cells treated for 24 hours with RA. Cyp26a1 served as a control for RA induction. P-value was determined by the Student's t-test, with significance level of 0.05 (denoted by *).

Supplemental Figure 4. Set1A$^{\Delta SET}$ EBs have decreased H3K4me3 relative to WT EBs.

(A) MA plot showing differential H3K4me3 occupancy levels during differentiation. Log fold changes of H3K4me3 signals between ESCs and EBs are plotted against average expression levels in all samples. 2,319 H3K4me3 peaks (purple) increased during differentiation were defined by adj.p<0.01.

(B) Metaplot of H3K4me3 levels in Set1A$^{\Delta SET}$ EBs vs. WT EBs aligned to peaks where H3K4me3 occupancy significantly increased during differentiation identified as purple dots in panel (A).

(C) *Set1A* and *Mll2* mRNA levels in ESCs vs. day 6 EBs. CPM (counts per million) was determined using edgeR, and averaged among replicates for corresponding cell state. P-values were determined using the Student's t-test.

(D) Box plot gene expression (left) and H3K4me3 (right) analyses of the 991 activated bivalent genes for Set1A$^{\Delta SET}$ EBs (top) and Mll2-KO EBs (bottom). P-values were determined by the t-test.

Supplemental References

Anders S, Pyl PT, Huber W. 2015. HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**: 166-169.

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**: 2114-2120.

Cao K, Collings CK, Marshall SA, Morgan MA, Rendleman EJ, Wang L, Sze CC, Sun T, Bartom ET, Shilatifard A. 2017. SET1A/COMPASS and shadow enhancers in the regulation of homeotic gene expression. *Genes Dev* **31**: 787-801.

Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass CK. 2010. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell* **38**: 576-589.

Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10**: R25.

Lee TI, Johnstone SE, Young RA. 2006. Chromatin immunoprecipitation and microarray-based analysis of protein location. *Nat Protoc* **1**: 729-748.

Lin C, Garrett AS, De Kumar B, Smith ER, Gogol M, Seidel C, Krumlauf R, Shilatifard A. 2011. Dynamic transcriptional events in embryonic stem cells mediated by the super elongation complex (SEC). *Genes Dev* **25**: 1486-1498.

Piunti A, Hashizume R, Morgan MA, Bartom ET, Horbinski CM, Marshall SA, Rendleman EJ, Ma Q, Takahashi YH, Woodfin AR et al. 2017. Therapeutic targeting of polycomb and BET bromodomain proteins in diffuse intrinsic pontine gliomas. *Nat Med* **23**: 493-500.

Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**: 841-842.

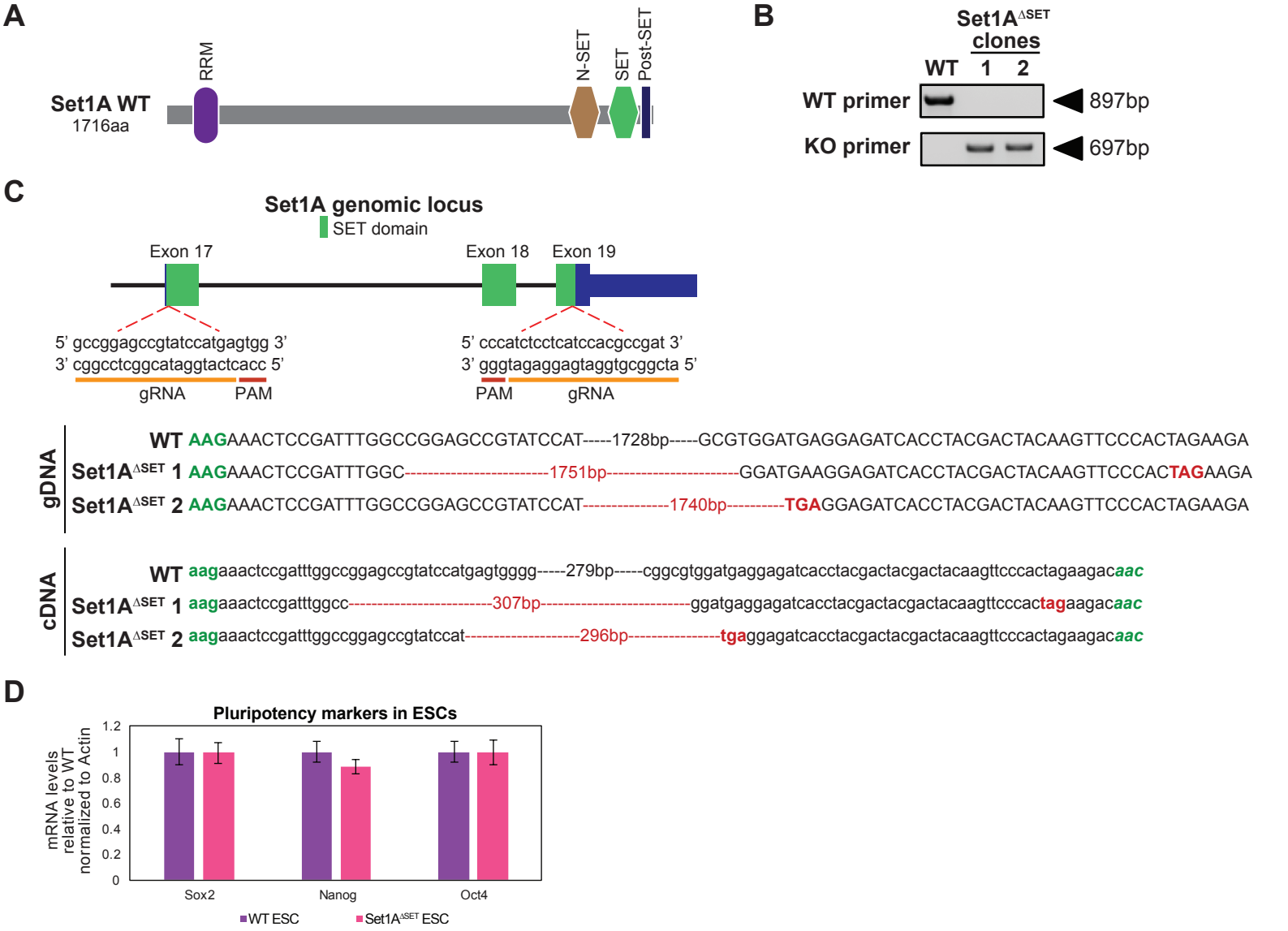
Robinson MD, McCarthy DJ, Smyth GK. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**: 139-140.

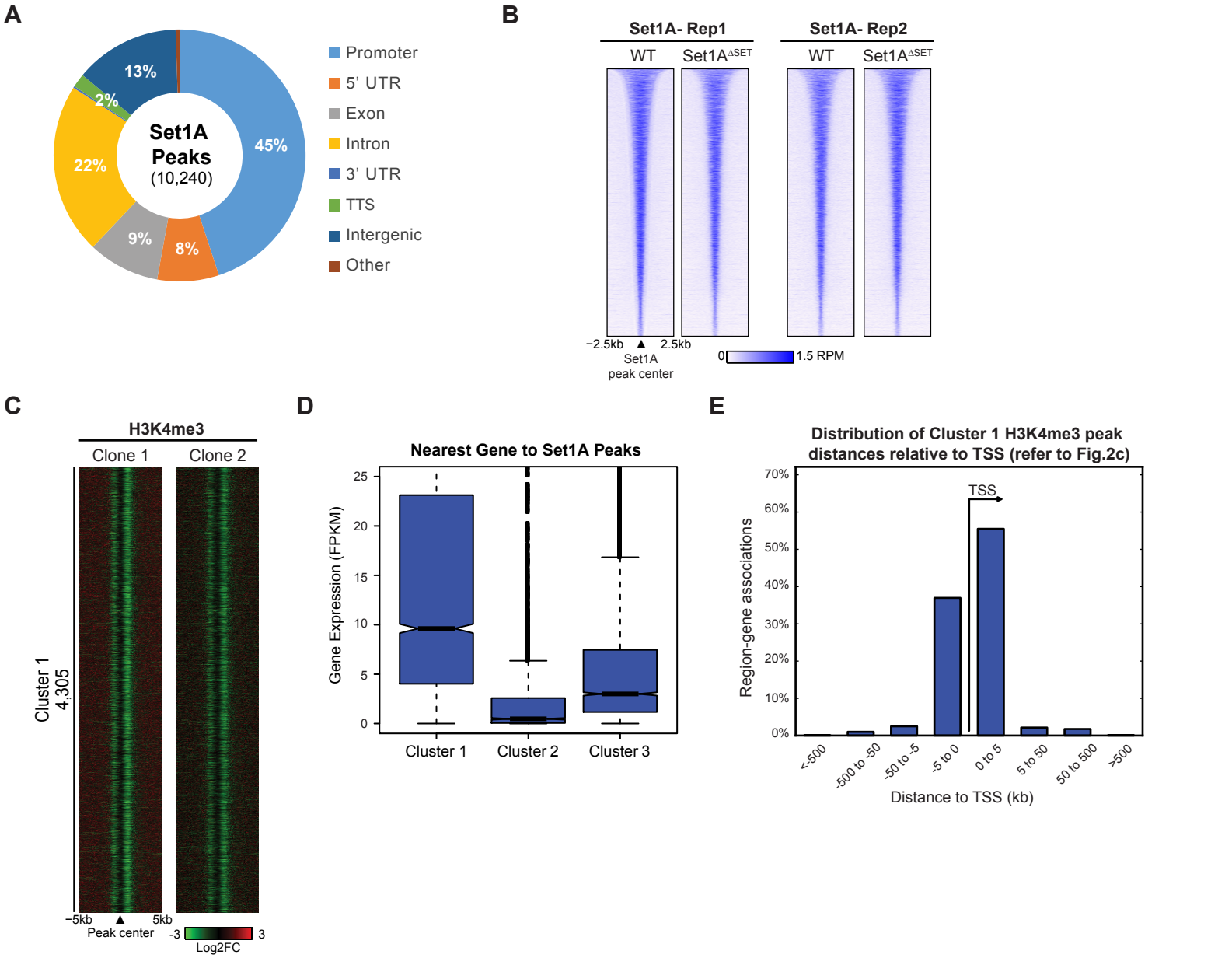
Shen L, Shao N, Liu X, Nestler E. 2014. ngs.plot: Quick mining and visualization of next-generation sequencing data by integrating genomic databases. *BMC genomics* **15**: 284.

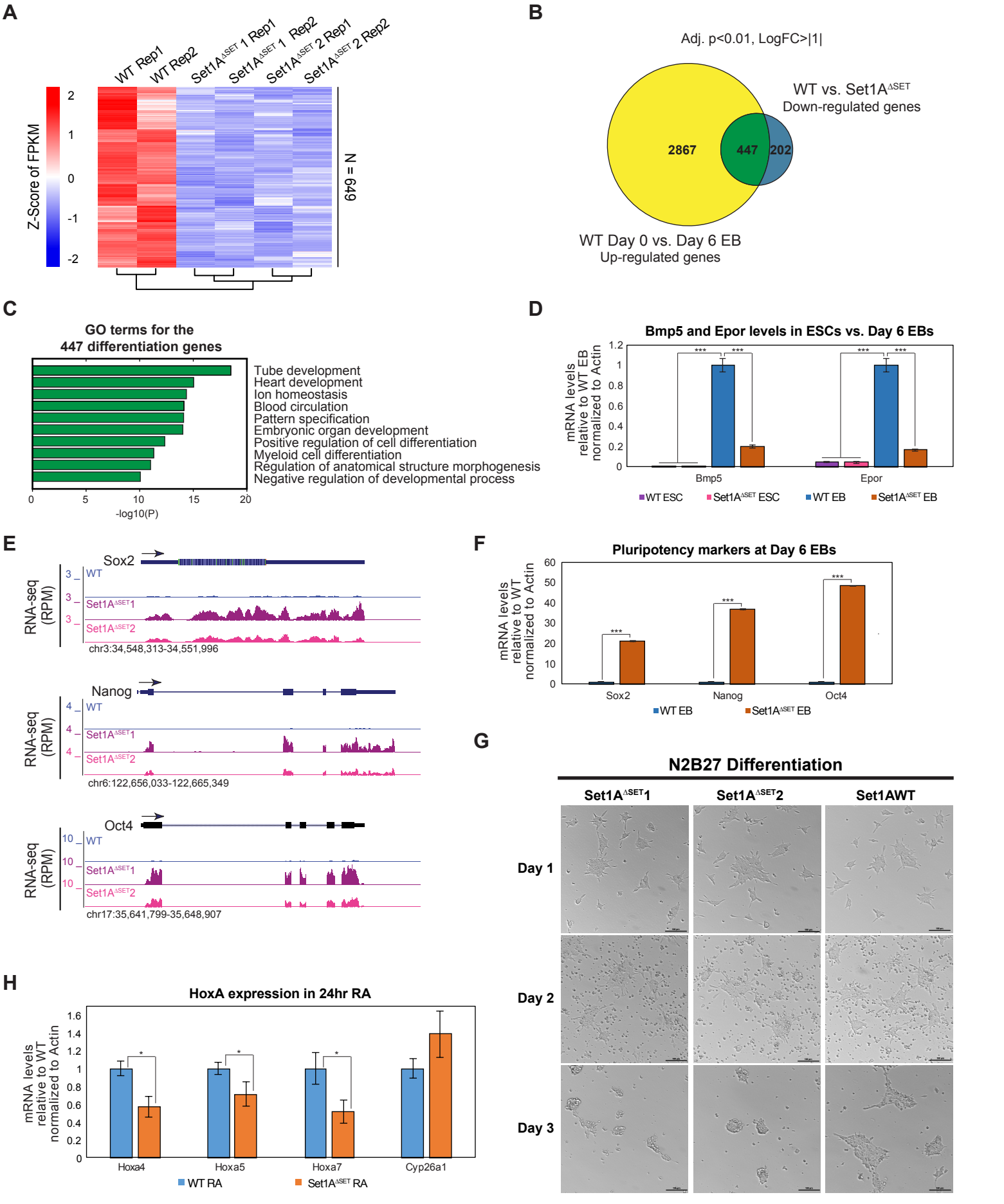
Trapnell C, Pachter L, Salzberg SL. 2009. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25**: 1105-1111.

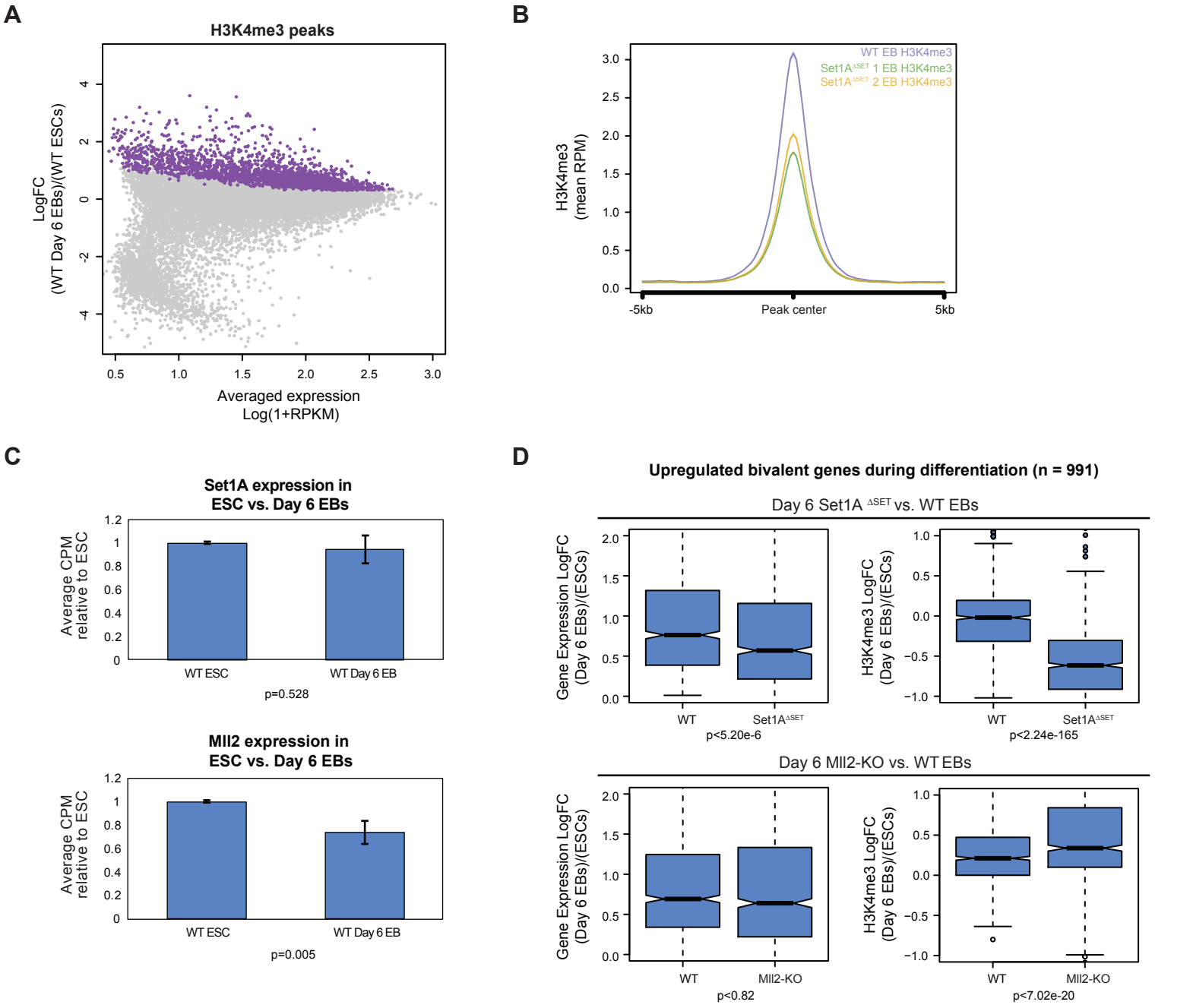
Tripathi S, Pohl MO, Zhou Y, Rodriguez-Frandsen A, Wang G, Stein DA, Moulton HM, DeJesus P, Che J, Mulder LC et al. 2015. Meta- and Orthogonal Integration of Influenza "OMICs" Data Defines a Role for UBR4 in Virus Budding. *Cell Host Microbe* **18**: 723-735.

Zhang Y, Liu Z, Medrzycki M, Cao K, Fan Y. 2012. Reduction of Hox gene expression by histone H1 depletion. *PLoS One* **7**: e38829.


A
Set1A WT
1716aa
RRM
N-SET
SET
Post-SET
B
Set1A^ΔSET clones
WT
1
2
WT primer
KO primer
897bp
697bp
C
Set1A genomic locus
SET domain
Exon 17
Exon 18
Exon 19
5' gccggagccgtatccatgagtgg 3'
3' cggcctcggcataggtactcacc 5'
gRNA
PAM
5' cccatctcctcatccacgccgat 3'
3' gggtagaggagtaggtgcggcta 5'
PAM
gRNA
gDNA
WT AAGAAACTCCGATTTGGCCGGAGCCGTATCCAT-----1728bp-----GCGTGGATGAGGAGATCACCTACGACTACAAGTTCCCACTAGAAGA
Set1A^ΔSET 1 AAGAAACTCCGATTTGGC----------1751bp----------GGATGAAGGAGATCACCTACGACTACAAGTTCCCACTAGAAGA
Set1A^ΔSET 2 AAGAAACTCCGATTTGGCCGGAGCCGTATCCAT----------1740bp----------TGAGGAGATCACCTACGACTACAAGTTCCCACTAGAAGA
cDNA
WT aagaaactccgatttggccggagccgtatccatgagtgggg-----279bp-----cggcgtggatgaggagatcacctacgactacgactacaagttcccactagaagacaac
Set1A^ΔSET 1 aagaaactccgatttggcc----------307bp----------ggatgaggagatcacctacgactacgactacaagttcccactagaagacaac
Set1A^ΔSET 2 aagaaactccgatttggccggagccgtatccat----------296bp----------tgaggagatcacctacgactacgactacaagttcccactagaagacaac
D
Pluripotency markers in ESCs
mRNA levels relative to WT normalized to Actin
1.2
1
0.8
0.6
0.4
0.2
0
Sox2
Nanog
Oct4
WT ESC
Set1A^ΔSET ESC


A
Set1A Peaks (10,240)
45%
8%
9%
22%
2%
13%
Promoter
5' UTR
Exon
Intron
3' UTR
TTS
Intergenic
Other
B
Set1A- Rep1
Set1A- Rep2
WT
Set1A^ΔSET
−2.5kb
2.5kb
Set1A peak center
0
1.5 RPM
C
H3K4me3
Clone 1
Clone 2
Cluster 1 4,305
−5kb
5kb
Peak center
-3
3
Log2FC
D
Nearest Gene to Set1A Peaks
Gene Expression (FPKM)
Cluster 1
Cluster 2
Cluster 3
E
Distribution of Cluster 1 H3K4me3 peak distances relative to TSS (refer to Fig.2c)
Region-gene associations
TSS
<-500
-500 to -50
-50 to -5
-5 to 0
0 to 5
5 to 50
50 to 500
>500
Distance to TSS (kb)


A
WT Rep1
WT Rep2
Set1A^ΔSET 1 Rep1
Set1A^ΔSET 1 Rep2
Set1A^ΔSET 2 Rep1
Set1A^ΔSET 2 Rep2
Z-Score of FPKM
2
1
0
-1
-2
N = 649
B
Adj. p<0.01, LogFC>|1|
WT vs. Set1A^ΔSET
Down-regulated genes
2867
447
202
WT Day 0 vs. Day 6 EB
Up-regulated genes
C
GO terms for the
447 differentiation genes
Tube development
Heart development
Ion homeostasis
Blood circulation
Pattern specification
Embryonic organ development
Positive regulation of cell differentiation
Myeloid cell differentiation
Regulation of anatomical structure morphogenesis
Negative regulation of developmental process
0
5
10
15
20
-log10(P)
D
Bmp5 and Epor levels in ESCs vs. Day 6 EBs
mRNA levels relative to WT EB normalized to Actin
1.2
1
0.8
0.6
0.4
0.2
0
***
Bmp5
Epor
WT ESC
Set1A^ΔSET ESC
WT EB
Set1A^ΔSET EB
E
Sox2
RNA-seq (RPM)
WT
Set1A^ΔSET1
Set1A^ΔSET2
chr3:34,548,313-34,551,996
Nanog
chr6:122,656,033-122,665,349
Oct4
chr17:35,641,799-35,648,907
F
Pluripotency markers at Day 6 EBs
mRNA levels relative to WT normalized to Actin
60
50
40
30
20
10
0
***
Sox2
Nanog
Oct4
WT EB
Set1A^ΔSET EB
G
N2B27 Differentiation
Set1A^ΔSET1
Set1A^ΔSET2
Set1AWT
Day 1
Day 2
Day 3
H
HoxA expression in 24hr RA
mRNA levels relative to WT normalized to Actin
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
*
Hoxa4
Hoxa5
Hoxa7
Cyp26a1
WT RA
Set1A^ΔSET RA


A
H3K4me3 peaks
LogFC
(WT Day 6 EBs)/(WT ESCs)
Averaged expression
Log(1+RPKM)
B
WT EB H3K4me3
Set1A^ΔSET 1 EB H3K4me3
Set1A^ΔSET 2 EB H3K4me3
H3K4me3
(mean RPM)
-5kb
Peak center
5kb
C
Set1A expression in
ESC vs. Day 6 EBs
Average CPM
relative to ESC
WT ESC
WT Day 6 EB
p=0.528
Mll2 expression in
ESC vs. Day 6 EBs
Average CPM
relative to ESC
WT ESC
WT Day 6 EB
p=0.005
D
Upregulated bivalent genes during differentiation (n = 991)
Day 6 Set1A ΔSET vs. WT EBs
Gene Expression LogFC
(Day 6 EBs)/(ESCs)
WT
Set1A^ΔSET
p<5.20e-6
H3K4me3 LogFC
(Day 6 EBs)/(ESCs)
WT
Set1A^ΔSET
p<2.24e-165
Day 6 Mll2-KO vs. WT EBs
Gene Expression LogFC
(Day 6 EBs)/(ESCs)
WT
Mll2-KO
p<0.82
H3K4me3 LogFC
(Day 6 EBs)/(ESCs)
WT
Mll2-KO
p<7.02e-20