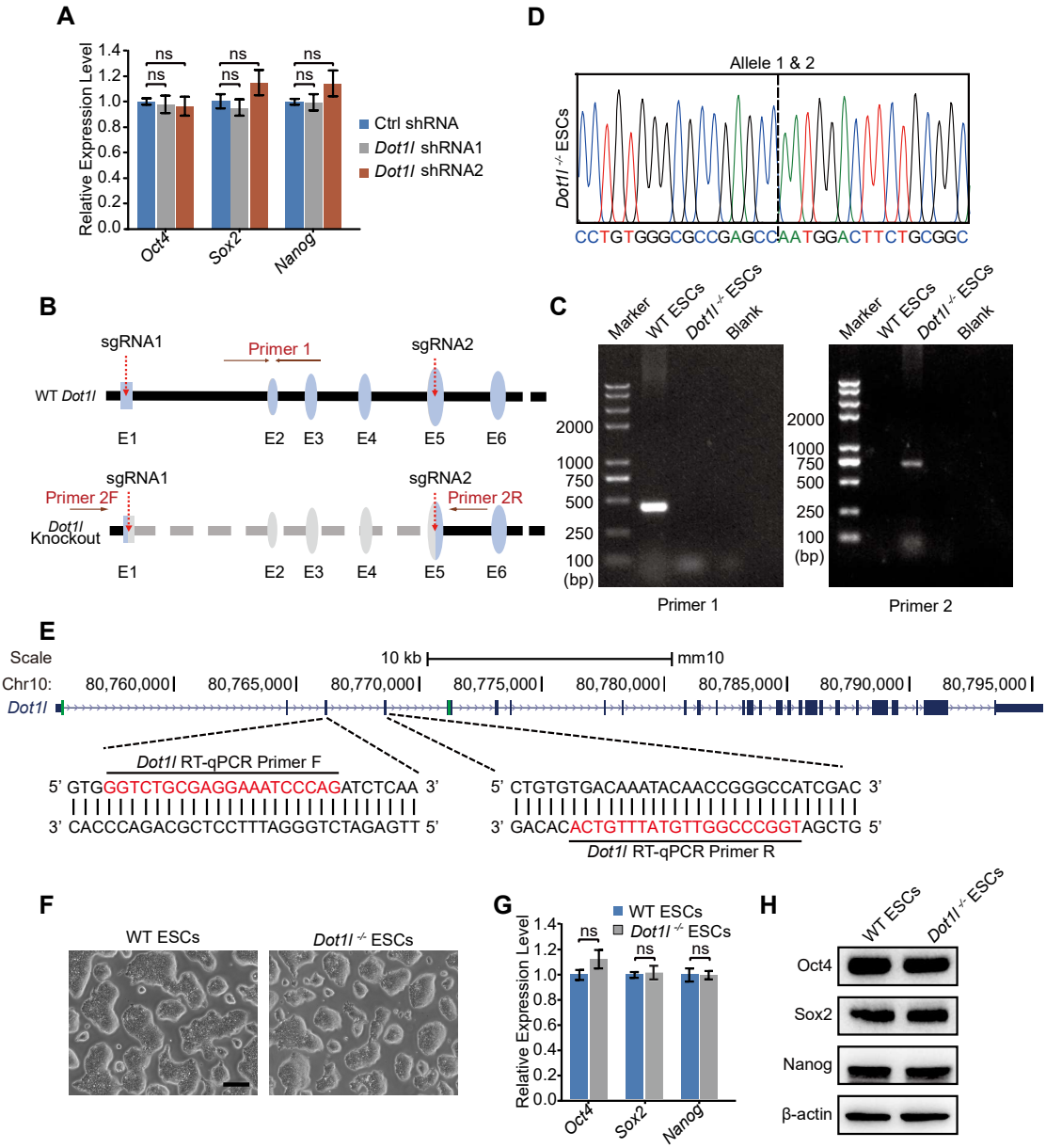
Supplementary Figures and Tables



Supplementary Figure S1. Knocking out *Dot1l* in ESCs

(A) qPCR analysis of the expression of pluripotency genes (*Oct4, Sox2* and *Nanog*) after the depletion of *Dot1l* in ESCs. qPCR data are presented as mean ± s.e.m. (n = 3 independent experiments).

(B) Schematic showing CRISPR/Cas9-mediated deletion of *Dot1l* using two sgRNAs. Locations of primers were used to distinguish *Dot1l*$^{-/-}$ clones from WT ESCs are shown as horizontal red arrows. WT: wild-type; E: exon.

(C) An agarose gel image illustrating the PCR-based genotyping assay of the *Dot1l*$^{-/-}$ ESCs with primers in (**B**).

(D) DNA sequencing results of mutation sites in *Dot1l*$^{-/-}$ ESCs.

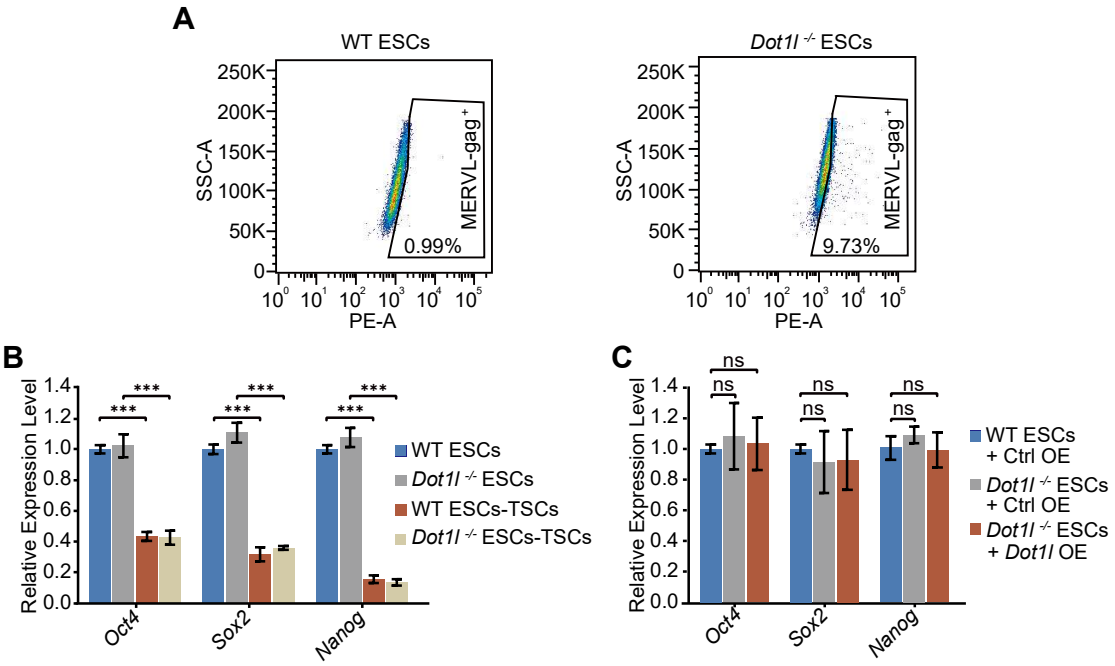
(E) Schematic of the location of *Dot1l* RT-qPCR primers. The sgRNA target sequences on exon 1 and exon 5 of the *Dot1l* gene are highlighted in green.

(F) Cell morphology of WT ESCs and *Dot1l*$^{-/-}$ ESCs. Scale bar, 100 μm.

(G) qPCR analysis of the expression of pluripotency genes in WT ESCs and *Dot1l*$^{-/-}$ ESCs.

Data are presented as mean ± s.e.m. (n = 3 independent experiments). ns: non-significant in Student's *t*-test.

(H) Western blot analysis of pluripotency genes in WT ESCs and *Dot1l*$^{-/-}$ ESCs. β-actin was used as a loading control.

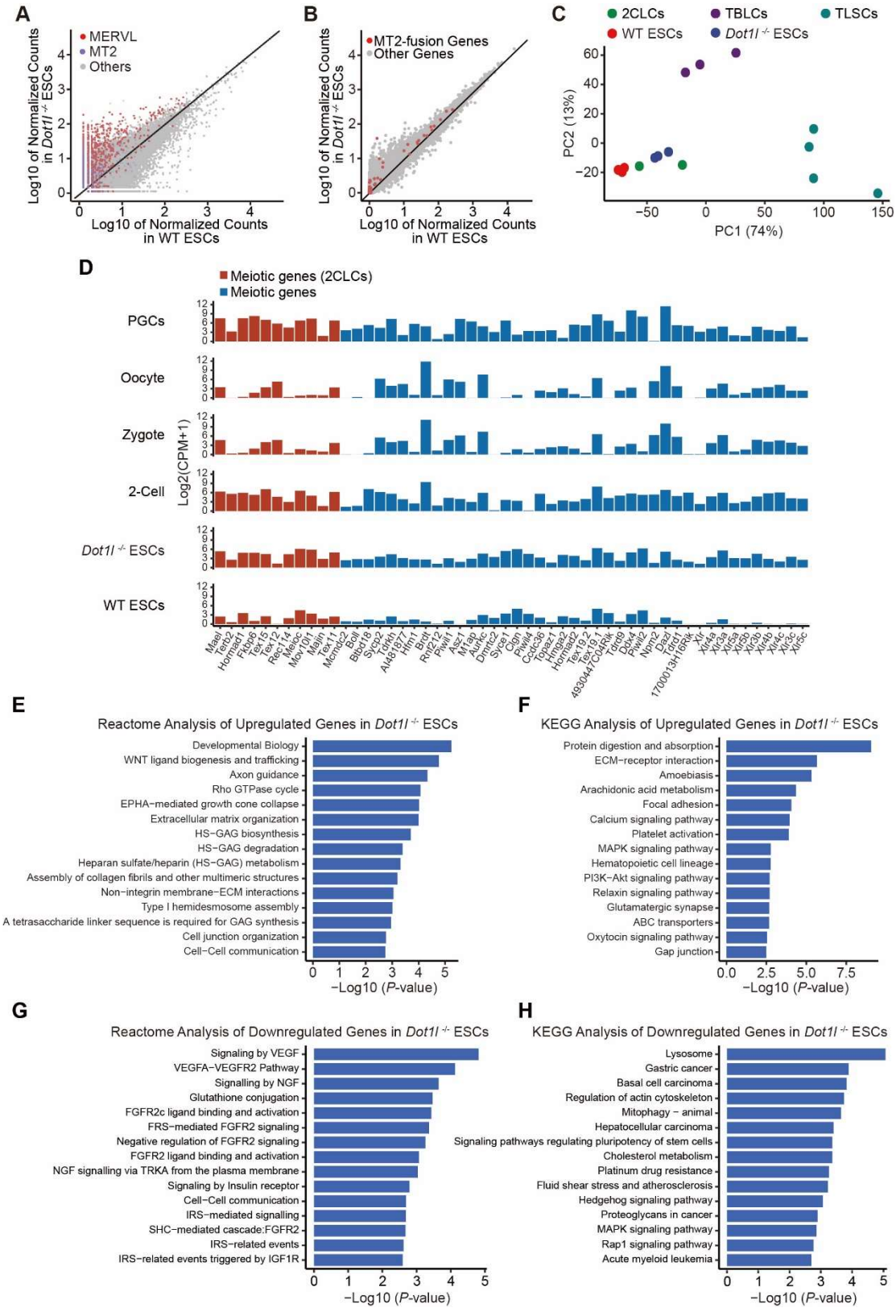


Supplementary Figure S2. Expression of pluripotency genes after differentiation and rescue

(A) Representative flow cytometry scatter diagram analysis of the MERVL-gag$^+$ population in WT ESCs and *Dot1l* $^{-/-}$ ESCs.

(B) qPCR analysis of the expression of pluripotency genes in WT ESCs and *Dot1l* $^{-/-}$ ESCs differentiated into TSCs respectively. qPCR data are presented as mean ± s.e.m. (n = 3 independent experiments).

(C) qPCR analysis of the expression of pluripotency genes after Dot1l overexpression (OE) in *Dot1l* $^{-/-}$ ESCs. Data are presented as mean ± s.e.m. (n = 3 independent experiments) for the qPCR results. ns: non-significant, ***$p < 0.001$ in Student's *t*-test.



Supplementary Figure S3. Expression of MERVL/MT2 loci and fusion genes in *Dot1l*$^{-/-}$ ESCs

(A) Dot plot of all TEs in WT ESCs and *Dot1l*$^{-/-}$ ESCs. Red dots represent MERVL, purple dots represent MT2.

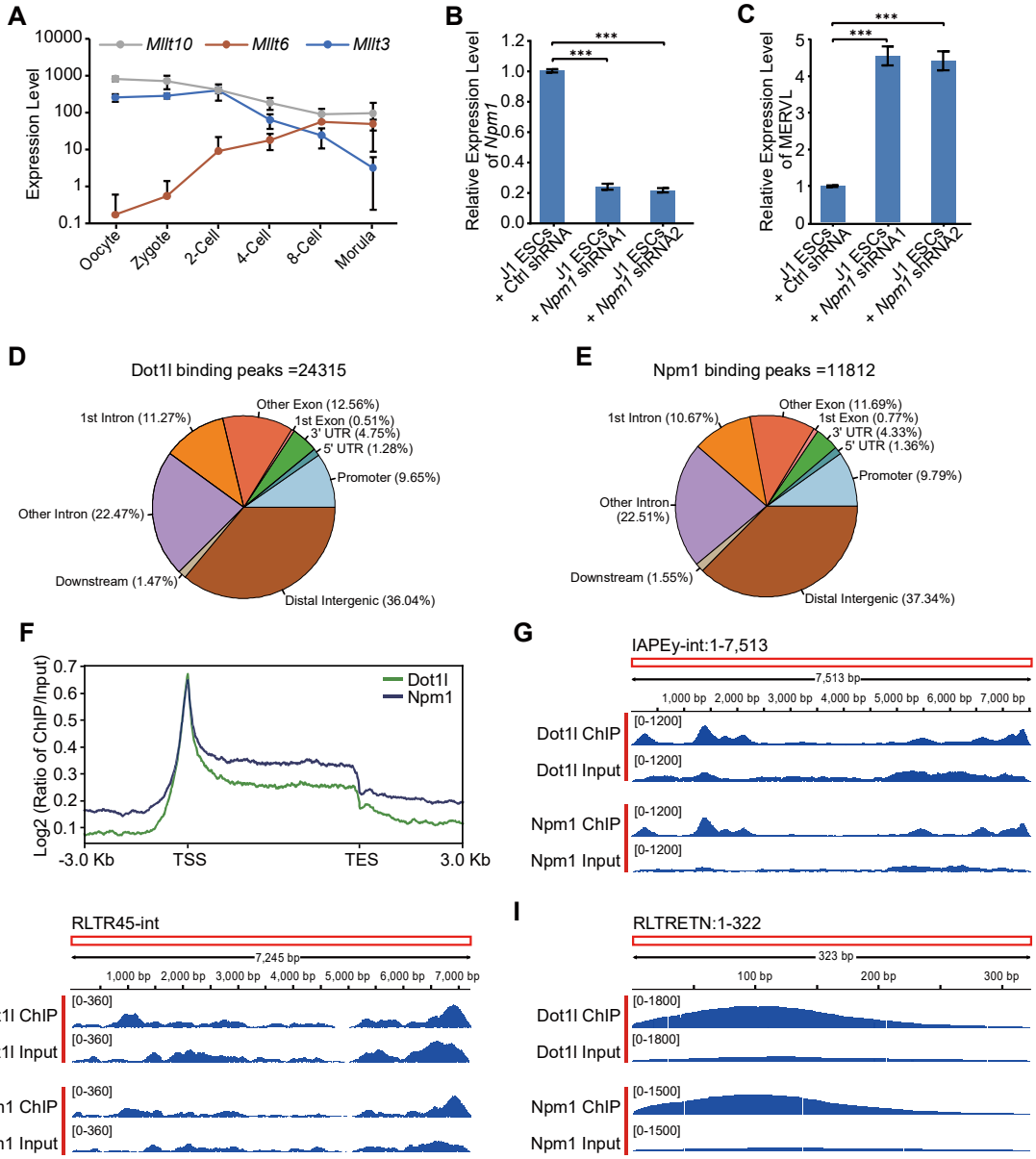
(B) Dot plot of all expressed genes in WT ESCs and *Dot1l*$^{-/-}$ ESCs. Genes with alternative transcript(s) overlapped with MT2 are labeled in red.

(C) Principal component analysis of *Dot1l*$^{-/-}$ ESCs, TBLCs, TLSCs and published 2CLCs based on gene expression. WT, wild type; TBLCs, totipotent blastomere-like cells; TLSCs, totipotent-like stem cells; 2CLCs, 2-cell like cells.

(D) Upregulated meiosis genes enriched in the GO term in *Dot1l*$^{-/-}$ ESCs.

(E-F) Reactome (E) and KEGG (F) pathway analysis of upregulated genes in *Dot1l*$^{-/-}$ ESCs.

(G-H) Reactome (G) and KEGG (H) pathway analysis of downregulated genes in *Dot1l*$^{-/-}$ ESCs.



Supplementary Figure S4. Binding profile of Npm1 and Dot1l in ESCs

(A) The expression level of *Mllt10*, *Mllt6* and *Mllt3* during early embryogenesis according to published RNA-seq data.

(B) qPCR analysis of the expression of *Npm1* in J1 ESCs treated with control (Ctrl) shRNA or *Npm1* shRNAs. Data are presented as mean ± s.e.m. (n = 3 independent experiments).

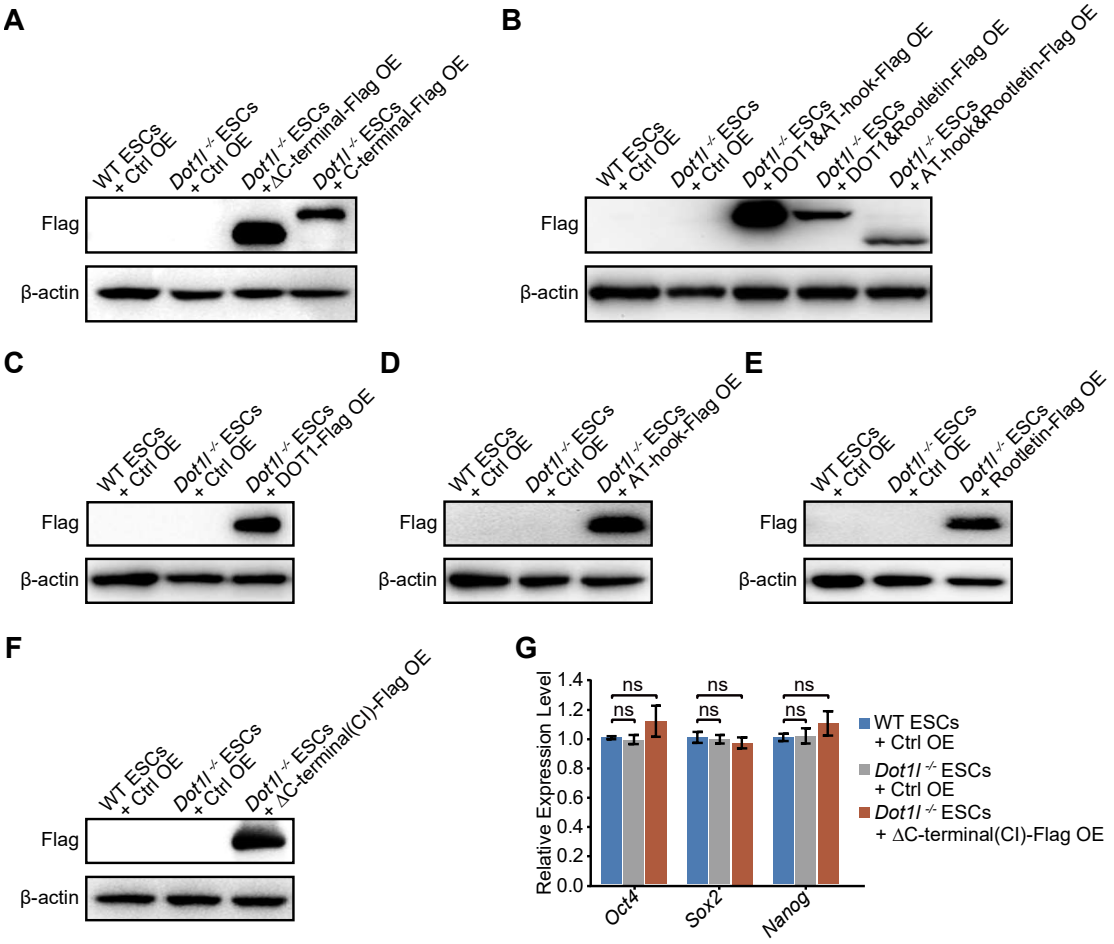
(C) qPCR analysis of the expression of MERVL after the depletion of *Npm1* in J1 ESCs. Data are presented as mean ± s.e.m. (n = 3 independent experiments). ***$p < 0.001$ in Student's *t*-test.

(D-E) Locations of Dot1l (D) and Npm1 (E) binding peaks relative to the nearest transcription units (Promoter, 2 kb around transcriptional start sites).

(F) PlotProfile of the distribution of Dot1l and Npm1 binding peaks on gene characteristics. The ChIP-seq signal was calculated as the log2 ratio of normalized reads relative to the input. TSS, transcription start site; TES, transcription termination site; -3 Kb, TSS

upstream 3 Kb; 3 Kb, TES downstream 3 Kb.

(G-I) Integrative Genomics Viewer (IGV) visualized the Dot1l and Npm1 ChIP-seq signals mapped to the consensus sequence of IAPEy-int (G), RLTR45-int (H) and RLTRETN (I) respectively.



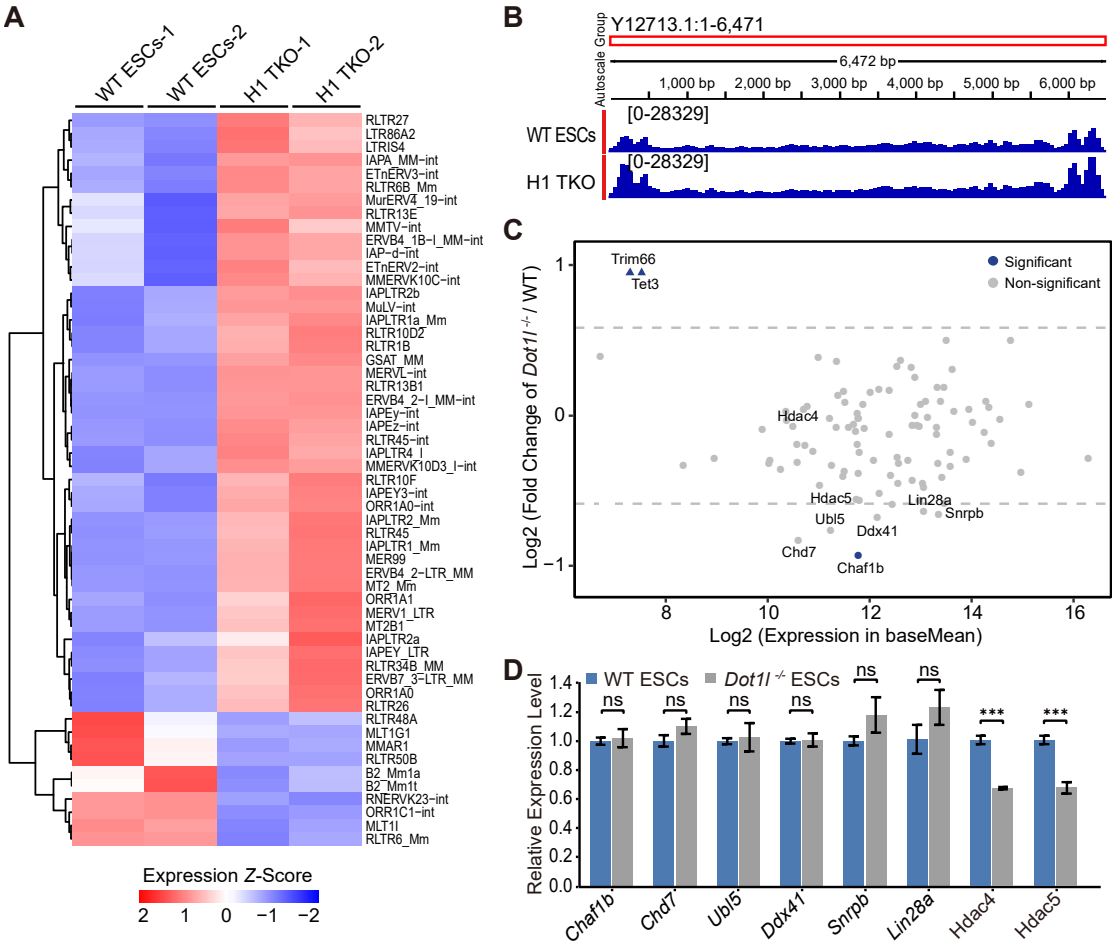
Supplementary Figure S5. Generation of Dot1l-mutant cell lines

(A-B) Western blot analysis of the expression of Dot1l mutants using Flag antibody in *Dot1l* $^{-/-}$ ESCs. β-actin was used as a loading control. Ctrl OE: control vector overexpression; Δ: deletion.

(C-E) Western blot analysis of the expression of Dot1l DOT1 (C), Dot1l AT-hook (D), Dot1l Rootletin (E) mutants using Flag antibody in *Dot1l* $^{-/-}$ ESCs. β-actin was used as a loading control. Ctrl OE: control vector overexpression.

(F) Western blot analysis of Dot1l ΔC-terminal (CI) mutant using Flag antibody in *Dot1l* $^{-/-}$ ESCs. β-actin was used as a loading control. Ctrl OE: control vector overexpression; CI: catalytic inactive.

(G) qPCR analysis of the expression of pluripotency genes after rescue of Dot1l ΔC-terminal (CI) mutant in *Dot1l* $^{-/-}$ ESCs. Ctrl OE: control vector overexpression; CI: catalytic inactive. qPCR data are presented as mean ± s.e.m. (n = 3 independent experiments). ns: non-significant in Student's *t*-test.



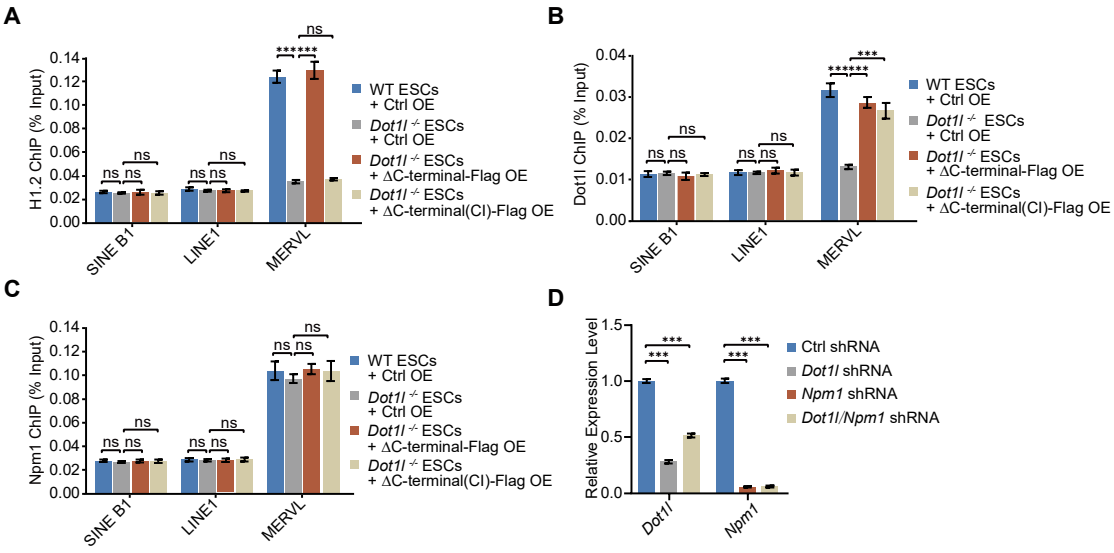
Supplementary Figure S6. Retrotransposon expression after the knockout of H1 variants

(A) Expression heatmap of retrotransposons after H1 TKO (triple-knockout). Differentially expressed TEs were defined as TEs with fold change > 1.5 and FDR adjusted $P < 0.05$ (Wald test).

(B) The ATAC-seq reads mapped to MERVL in WT ESCs and H1 TKO ESCs was visualized using Integrative Genomics Viewer (IGV).

(C) A scatter diagram shows transcriptome analysis of the expression of 100 reported MERVL repressors after *Dot1l* knockout. Blue dots or triangles indicate repressors with significant expression change (adjusted $p < 0.05$, Wald test). Triangles represent repressors with log2 (fold change) >1.

(D) qPCR validation of the expression of indicated MERVL repressors in *Dot1l* $^{-/-}$ ESCs. qPCR data are presented as mean ± s.e.m. (n = 3 independent experiments). ns: non-significant, ***$p < 0.001$ in Student's *t*-test.



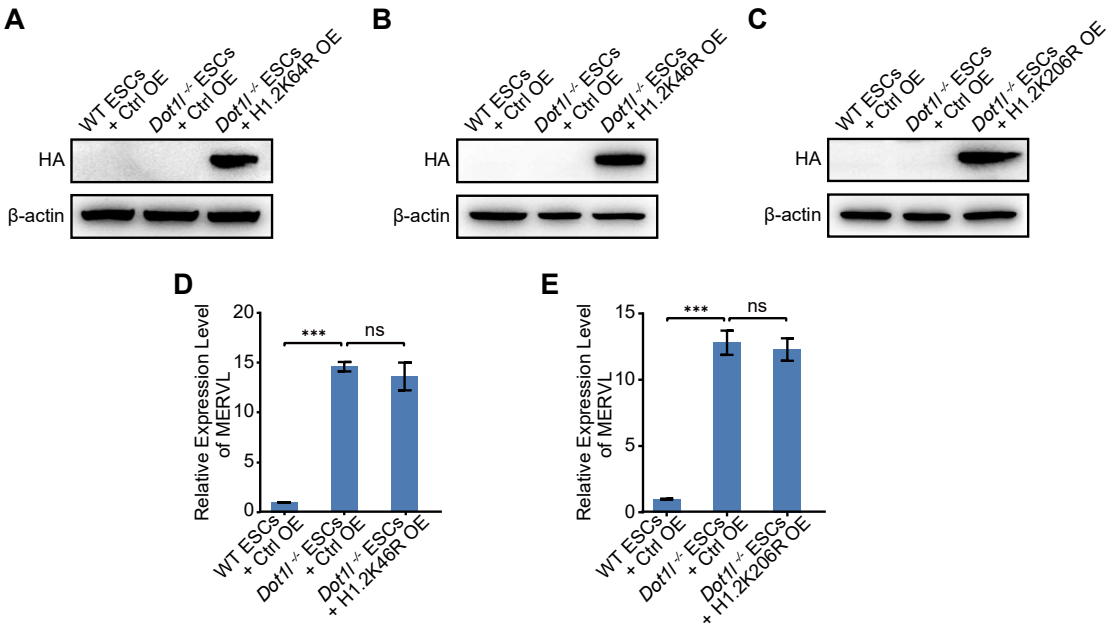
Supplementary Figure S7. Dot1l catalytic activity is required for H1.2 enrichment on chromatin

(A) ChIP-qPCR analysis of H1.2 binding on MERVL. ChIP-qPCR data were normalized to input and that of the control region. Data are presented as mean ± s.e.m. (n = 3 independent experiments).

(B) ChIP-qPCR analysis of Dot1l and catalytic-inactivated Dot1l mutant binding on MERVL. ChIP-qPCR data were normalized to input and that of the control region. Data are presented as mean ± s.e.m. (n = 3 independent experiments).

(C) ChIP-qPCR analysis of Npm1 binding on MERVL. ChIP-qPCR data were normalized to input and that of the control region. Data are presented as mean ± s.e.m. (n = 3 independent experiments).

(D) qPCR analysis of the expression levels of *Dot1l* and *Npm1* following their individual depletion as well as simultaneous depletion. Data are presented as mean ± s.e.m. (n = 3 independent experiments). ns: non-significant, ***$p < 0.001$ in Student's *t*-test.



Supplementary Figure S8. Screening histone H1.2 ubiquitination sites

(A-C) Western blot analysis of the overexpression of H1.2K64R (A), H1.2K46R (B), H1.2K206R (C) in *Dot1l* $^{-/-}$ ESCs. Ctrl OE, control vector overexpression; K, Lysine; R, Arginine. β-actin was used as a loading control.

(D-E) qPCR analysis of the expression of MERVL after overexpression of H1.2K46R (D) and H1.2K206R (E) in *Dot1l* $^{-/-}$ ESCs. Ctrl OE, control vector overexpression; K, Lysine; R, Arginine. Data are presented as mean ± s.e.m. (n = 3 independent experiments). ns: non-significant, ***$p < 0.001$ in Student's *t*-test.

Supplementary Table S1. Sequences of primers, shRNAs and guide RNAs		
Gene	**Sequence F (5'-3')**	**Sequence R (5'-3')**
Dot1l	GGTCTGCGAGGAAATCCCAG	TGGCCCGGTTGTATTTGTCA
Oct4	GTGGAAAGCAACTCAGAGG	GGTTCCACCTTCTCCAACT
Sox2	GCGGAGTGGAAACTTTTGTCC	CGGGAAGCGTGTACTTATCCTT
Nanog	TTGCTTACAAGGGTCTGCTACT	ACTGGTAGAAGAATCAGGGCT
Cdx2	AGGCTGAGCCATGAGGAGTA	TGAGGTCCATAATTCCACTCA
Eomes	CAATGTTTTCGTGGAAGTGG	GTTAGGAGATTCTGGGTGAA
Plet1	AACGATTCAGTCAGTGCCGT	TGACTTTGAGGCTGTGCGAT
Ascl2	AAGCACACCTTGACTGGTACG	AAGTGGACGTTTGCACCTTCA
Tcstv1	GGTGCTCCAAATCTGAGACACTT	ATAGATCCCAATCGGCAATCC
Zfp352	AAGTCCCACATCTGAAGAAACAC	GGGTATGAGGATTCACCCACA
Usp17lb	TCTCCTTCCCAGAAGATCCAG	ACTCTCCCAACTCAGACTGT
Gm8300	TACTCACCAGGTCAATGCAGG	GTCCTGGCTCCTGATAGTTAC
Gm4027	AGATGGTACTCACCAGGTCAA	ATGTCCCAGAGTACTGGCTT
Gm4340	TTGTTGGGAATTTGGCTGCC	CATGGGTGAAAGCTGGCCTA
Sp110	ATGAAGGTGAACATCGCCTATG	GGACAGAGGGACCAGATTTTG
Zscan4	GAGATTCATGGAGAGTCTGACTGATGAGTG	GCTGTTGTTTCAAAAGCTTGATGACTTC
Tcstv3	GATCCTGCATCTTATAGTGCCA	TGACTTTCTCACTTCTGGCG
Npm1	CATGTCTGGAAAGCGATC	CCTTTGATCTCGGTGTTG
Mllt3	CGTCTTCCACTTGCACGAAAG	CCCGGACTCTTCTACCTTGTAA
Mllt6	TAGCCTGACAGCGAAGAAGG	CAGCTTGAGGAAGCAGTCCA
Mllt10	AGGACGAGGTCTCCCATAGTA	GTGCCCGTCGCAATAAACC
H1.0	CATCAAGCGCCTAGTGACCA	TCTTGACAGGGGTGGCTTTG
H1.2	CGCGTCTAAAGCCGTAAAGC	CTTGGCTGCAACCTTCTTGG
Gapdh	AGAAACCTGCCAAGTATGATGAC	GTCATTGAGAGCAATGCCAG
IAPEz	AAATCAATCTGTTGTGTTTCCAC	ACCACATAACAGGAATCTGACAC
ERVK10C	TTCGCCTCTGCAATCAAGCTCTC	TCGCTCGTGCCTGAAGATGTTTC
IAPEY-int	TTCCCTCAAGCAGTAGATAATGA	GGCAGAGGTCCTTTATAGTCAGT
ERVB4-2	ACTTGATACCCAATGAATGG	AGATTTTGTCAGCAGTCCAG
MERVL	AAGAGCCAAGACCTGCTGAG	TCCTCGTTTCTGCAACTGGT
MT2	GGCTACACCTTCTGCTGGAG	TCGCAGCTGTGAATGGAAGT
MTA	TGGGTTCTATAAGAGAGCAGGC	TGTTCTTTACTGGCTTGCCTCC
SINEB1	GTGGCGCACGCCTTTAATC	GACAGGGTTTCTCTGTGTAG
Gene	**shRNA Sequence**	
Control shRNA	GATGAAATGGGTAAGTACA	***Dot1l* gRNA1**
Dot1l shRNA1	CCTCGGTTTACACAGCTTCAA	AGCCCGCCGTCTACCCGTGG
Dot1l shRNA2	CGGCAGAATCGTATCCTCAAA	***Dot1l* gRNA2**
*Npm1*shRNA1	GGAAGATGCAGAGTCTGAA	CTGAATACTCGGCCGTCCAA
*Npm1*shRNA2	GCAGAAGCAATGAACTATGAA	
Mllt3 shRNA	CCCGCCACCATTATTGAAA	
Mllt6 shRNA	GCTTGCTATGGCATCGTCCA	
Mllt10 shRNA	GCAGTATCGACATGATGGA	

Supplementary Table S2. primary antibodies	
primary antibodies	**Catalog**
Anti-Dot1l	D4O2T, Cell Signaling Technology
Anti-H3K79me1	ab2886, Abcam
Anti-H3K79me2	ab3594, Abcam
Anti-H3K79me3	C15410068, Diagenode
Anti-Oct4	sc-5279, Santa Cruz
Anti-Sox2	sc-365964, Santa Cruz
Anti-Nanog	sc-293121, Santa Cruz
Anti-MERVL-gag	A-2801, EpiGentek
Anti-HA	30701ES60, Yeasen
Anti-Npm1	sc-271737, Santa Cruz
Anti-Flag	F1804, Sigma
Anti-H1.0	A3298, ABclonal
Anti-H1.2	19649-1-AP，Proteintech
Anti-H3	17168-1-AP, Proteintech
Anti-Ubiquitin	10201-2-AP, Proteintech
Anti-β-actin	AC026, ABclonal