Supplementary data for:

Human histone H3K79 methyltransferase DOT1L binds actively transcribing RNA polymerase II to regulate gene expression

Seung-Kyoon Kim[‡,1], Inkyung Jung[§,1], Hosuk Lee[‡], Keunsoo Kang[‡], Mirang Kim[¶], Kwiwan Jeong[1], Chang Seob Kwon[†], Yong-Mahn Han[‡], Yong Sung Kim[¶], Dongsup Kim[§], and Daeyoup Lee[‡,2]

From the [‡] Department of Biological Sciences and the [§]Department of Bio and Brain Engineering, Korea Advanced Institute of Science and Technology, Daejeon 305-701, Korea, [†]Department of Chemistry and Biology, Korea Science Academy of KAIST, Busan 614-822, Korea, and the [¶]Medical Genomics Research Center, Korea Research Institute of Bioscience and Biotechnology, Daejeon 305-806, Korea

*Running title: *Interaction of hDOT1L with actively transcribing RNAPII*

To whom correspondence should be addressed: Department of Biological Sciences, Korea Advanced Institute of Science and Technology, 291 Daehak-ro, Yuseong-gu, Daejeon 305-701, Korea. Tel: +82-42-350-2623; Fax: +82-42-350-2610; E-mail: daeyoup@kaist.ac.kr

Keywords: Transcription; hDOT1L; RNA polymerase II; H2B monoubiquitylation; H3K79 methylation.

SUPPLEMENTARY PROCEDURES

Cell lines and construction of expression vectors–NCCIT and HEK293T cells were cultured in RPMI 1640 (Gibco) and DMEM (Hyclone), respectively, supplemented with 10% fetal bovine serum (FBS, Hyclone) and 1% antibiotic-antimycotic solution (Hyclone) at 37ºC in a humidified atmosphere containing 5% CO_2.

The cDNA for *hDot1l* was PCR amplified from the MegaMan library (Stratagene, La Jolla, CA, USA) and inserted into the mammalian expression vector, pME18SZ-FLAG. All sequences were verified by DNA sequencing. The primers for *hDot1l* were as follows:

hDot1l _F-EcoRI: 5'-GGGGAATTCATGGGGGAGAAGCTGGAG

hDot1l _R-XhoI: 5'-GGGCTCGAGCTAGTTACCTCCAACTGT

Illumina library construction and sequencing–The purified ChIP DNA fragments were ligated to paired adaptors designed for sequencing on the Illumina Genome Analyzer. The ligation products were size-fractioned on an agarose gel to obtain 200- to 300-bp fragments, and these fragments were subjected to 18 cycles of PCR amplification. Cluster generation and 72 cycles of single-read sequencing were performed.

Modification peak detection–Modification peaks were determined using MACS-1.3.7.1 (59) with default parameters, except that we changed the -mfold option to 5 and the –pvalue option to 1E-4. For the control, the un-treated input sequencing result (IgG) was used. The number of overlapping peaks between hDOT1L and RNAPII was calculated by counting the number of hDOT1L peaks located within 2 kb upstream and downstream of RNAPII peaks.

High-quality hDOT1L binding sites–To get a more stringent set of hDOT1L binding sites, we used the following thresholds: -mfold 5,25 and –pvalue 1E-5. A total of 8,953 hDOT1L binding sites were re-obtained and sorted by 'FDR' and 'enrichment folds' categories, which are attributes of each binding site. High-quality binding sites (the top 500) were defined as the regions containing lower FDRs and higher enrichment folds. To draw the average tag density graphs of hDOT1L, the RNAPII, as well as IgG (supplementary Fig. S2), flanking regions (+/- 10,000 bp) of each high-quality hDOT1L binding site were divided into 100-bp bins, and tags in each bin were summed. The average number of tags in each bin position around the high-quality hDOT1L binding sites was calculated and plotted. In case of the fold-enrichment graph (Fig. 6), the flanking regions (+/- 10,000 bp) of 223 high-quality hDOT1L binding sites containing the predicted hDOT1L motif were examined.

Ontology analysis using GREAT (version 1.5.1)–Ontology analysis was performed by GREAT with the following settings: Top Rows/Table, 40; Statistical Significance, *Raw P-value*; and View, *Significant By Region-based Binomial* (http://great.stanford.edu) (60). For input genomic sites, we used the defined

high-quality hDOT1L binding sites. Pathway commons category contains numerous pathways, including biochemical reactions, complex assembly, transport, and catalysis events, as well as physical interactions. The result describes significantly-associated pathways based on attributes of the gene products around binding sites. The test set of 500 hDOT1L binding sites picked 722 genes (4%) out of 17,506 tested genes. The pathway commons category contained 1,400 terms covering 4,665 (27%) of the 17,506 total genes.

Generation of hDOT1L expression constructs–Full-length hDOT1L cDNA was amplified by PCR from the MegaMan Human Transcriptome Libraries (Stratagene) and cloned into the pGEM-T Easy vector (Promega). The resulting hDOT1L cDNA clone was confirmed by sequence analysis and then subcloned into the pME18S-FLAG vector (addgene). The expression of FLAG-hDOT1L from the resulting pME18S-FLAG-hDOT1L plasmid was confirmed by western blotting using an anti- FLAG antibody (Sigma F1804). Next, we generated hDOT1L (WT), hDOT1L-si^R (si^R), hDOT1L-4K/618-627/4A (4K/4A), hDOT1L-N241A-si^R (N241A-si^R) in both pME18S-FLAG and pFastBac-FLAG vector (Invitrogen) for co-immunoprecipitation analyses and *in vitro* assay, respectively. Also, we generated hDOT1L deletion constructs; hDOT1L (1-467), (468-1002), and (1003-1537) in pFastBac-FLAG vector and overlapping truncations; hDOT1L (468-665), (666-820), (821-1002), (555-820), (555-820)-4K/618-627/4A, and (555-820)-Δ4K/618-627 in pET28a vector (Novagen) for peptide pull down assay.

Antibodies–The utilized antibodies recognizing hSNF5/INI1, yDot1, H3, NANOG, OCT4, and SOX2 were polyclonal rabbit antibodies produced in-house as previously described (31,61). The utilized commercial antibodies included those against H3K79me1 (abcam, ab2886), H3K79me2 (abcam, ab3594), H3K79me3 (abcam, ab2621), α-Rpb1 (Millipore, ABE30), α-RNAPII (abcam, ab817), α-RNAPII Ser5/2-P (Cell Signaling, 4735), α-RNAPII Ser5-P (abcam, ab5131), α-RNAPII Ser2-P (abcam, ab5095), H2Bmono-ub (Millipore, 17-650), hDOT1L (abcam, ab64077), ENL (abcam, 49052), FLAG (Sigma, F1804), HIS (Santa Cruz Biotechnology, sc-8036), and β-actin (Santa Cruz Biotechnology, sc-47778).

RNA isolation and reverse transcription (RT)–Total RNA was isolated from triplicate cultures of NCCIT cells using the TRIZOL® Reagent (Invitrogen) according to the manufacturer's protocols. Briefly, NCCIT cells grown in 100-mm diameter dishes were harvested. One ml of TRIZOL was added, and the cell lysate was passed through a pipette several times. The homogenized samples were incubated for 5 min at 25°C to permit the complete dissociation of nucleoprotein complexes, and 200 µl of chloroform was added. The sample tubes were shaken vigorously by hand for 15 sec and incubated at 25°C for 2 ~ 3 min. The samples were cleared by centrifugation at 12,000 × g for 15 min at 4°C, and the aqueous phase was transferred to a fresh tube. To precipitate the RNA from the aqueous phase, the

sample was mixed with 500 µl isopropyl alcohol and incubated at 25°C for 10 min. The RNA samples were cleared by centrifugation at 12,000 × g for 15 min at 4°C and washed once with 75% ethanol, and the RNA pellet was briefly dried and dissolved in RNase-free water. Finally, cDNA was synthesized from 1 µg of DNase-treated total RNA, using the Improme Kit (Promega) according to the manufacturer's guidelines.

Real-time PCR–qPCR analysis was performed with CFX96 (Bio-Rad). The DNA was amplified in a 20-µl reaction with 2X h-Taq real time mix (Solgent), 20X Evagreen (Biotium), 20X tetraethylammonium chloride, 10 pmol of each primer, and the template DNA. PCR reactions were carried out using the following conditions: an initial melting step of 95°C for 12 min, followed by 40 cycles of 95°C for 20 sec (denaturation), 57°C for 30 sec (annealing), and 72°C for 30 sec (extension), and then a final extension at 72°C for 5 min. The primer specificity for PCR amplification was tested by agarose gel electrophoresis. After PCR amplification, melting curve analysis was used to verify the amplification of a single PCR product. The threshold cycle (Ct) was determined for each reaction. For calculation purposes, the average Ct values of two copies of the IP DNA and input DNA from three independent reactions were termed A and B, respectively. The following formula was used to calculate fold enrichment (F) for each IP sample: $F = 1/2^{\wedge(A-B)}$. The relative mRNA levels were quantified with respect to that of β-actin.

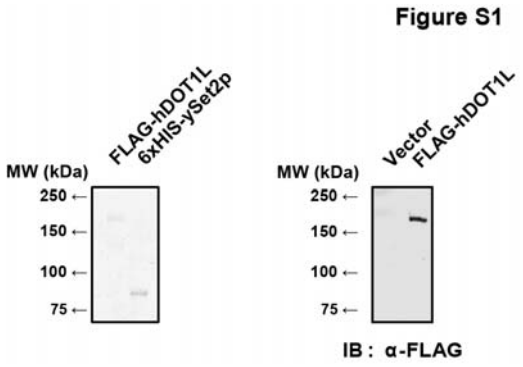
Histone methyltransferase (HMT) assay–In vitro histone methyltransferase assays were performed as previously described (62). Briefly, recombinant hDOT1L proteins were generated using a baculovirus expression system, and 500 ng purified recombinant protein was incubated at 30°C for 6 h in 20 µl containing 50 mM Tris-HCl (pH 8.0), 5% glycerol, 20 mM KCl, 5 mM $MgCl_2$, 1 mM dithiothreitol, 1 mM PMSF, 1 µl of [^{3}H] methyl S-adenosyl-methionine (1µCi, Amersham Biosciences), and 500 ng of HeLa long oligonucleosomes (LON) as a substrate. The reaction was stopped by the addition of SDS sample buffer, the samples were separated by 14% SDS-PAGE, and the results were visualized by radiography.

Alkaline phosphatase assays–Alkaline phosphatase activity was measured using an alkaline phosphatase detection kit (Millipore, SCR004) according to the manufacturer's guidelines, with minor modifications. Briefly, NCCIT cells were plated in 12-well plates, incubated overnight, and transfected with 60 nM *Egfp*- and 100 nM *hDot1l*-siRNA. For rescue approaches, si^R, 4K/4A, and N241A-si^R were transfected into NCCIT cells that had been depleted of endogenous hDOT1L. After three days of siRNA transfection, the cells were fixed, stained, and washed. Images were obtained under a microscope.

SUPPLEMENTARY REFERENCES

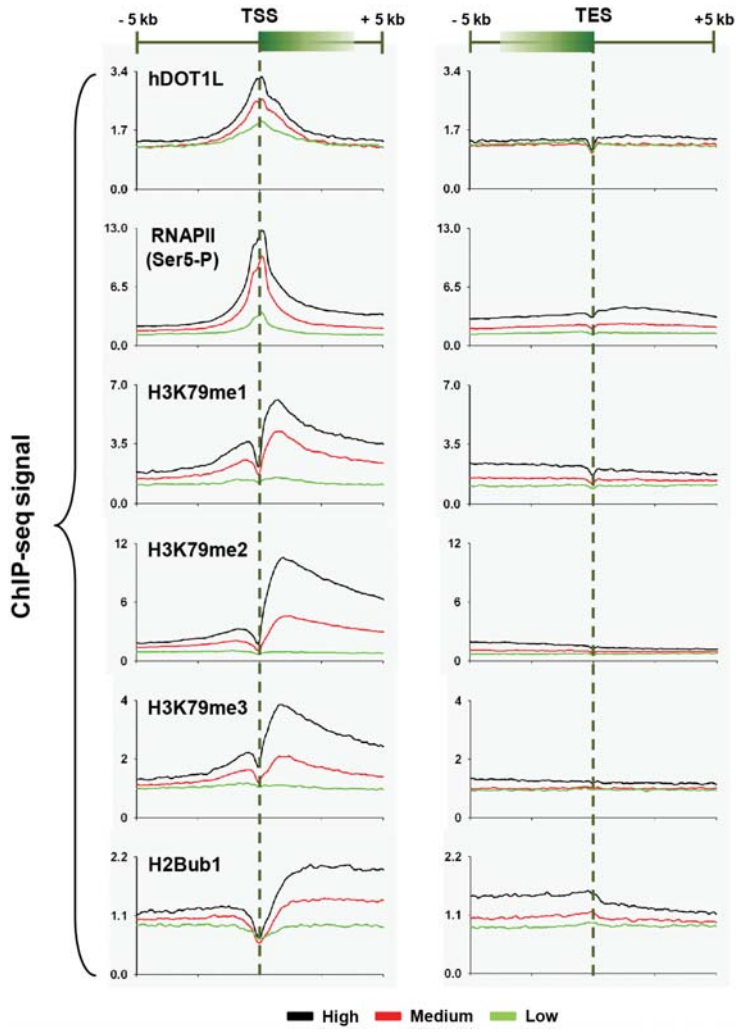
60. McLean, C. Y., Bristor, D., Hiller, M., Clarke, S. L., Schaar, B. T., Lowe, C. B., Wenger, A. M., and Bejerano, G. (2010) GREAT improves functional interpretation of cis-regulatory regions. *Nat Biotechnol* **28**, 495-501
61. Hwang, S., Lee, D., Gwack, Y., Min, H., and Choe, J. (2003) Kaposi's sarcoma-associated herpesvirus K8 protein interacts with hSNF5. *J Gen Virol* **84**, 665
62. Shi, X., Hong, T., Walter, K. L., Ewalt, M., Michishita, E., Hung, T., Carney, D., Lan, F., Kaadige, M. R., and Lacoste, N. (2006) ING2 PHD domain links histone H3 lysine 4 methylation to active gene repression. *Nature* **442**, 96-99

SUPPLEMENTARY FIGURES AND TABLES

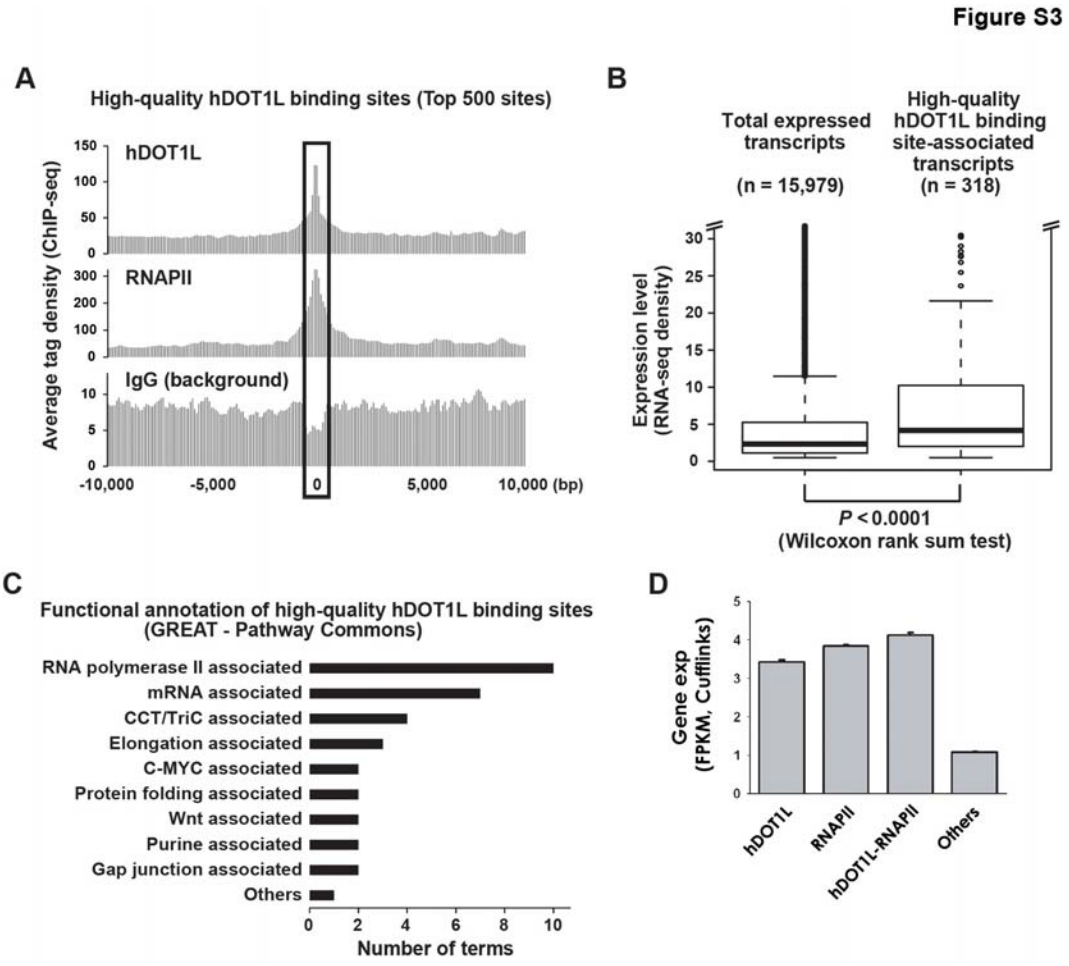


Supplementary FIGURE 1. The purified recombinant proteins used in this study. Left panel, Full-length FLAG-hDOT1L was purified from a Sf-21 insect cell-based baculovirus system, and 6xHIS-ySet2p was purified from a bacterial system. The two proteins were resolved by SDS-PAGE and stained with Coomassie blue. Right panel, The full-length FLAG-hDOT1L was resolved by SDS-PAGE and immunoblotted using the indicated antibody.

Figure S2



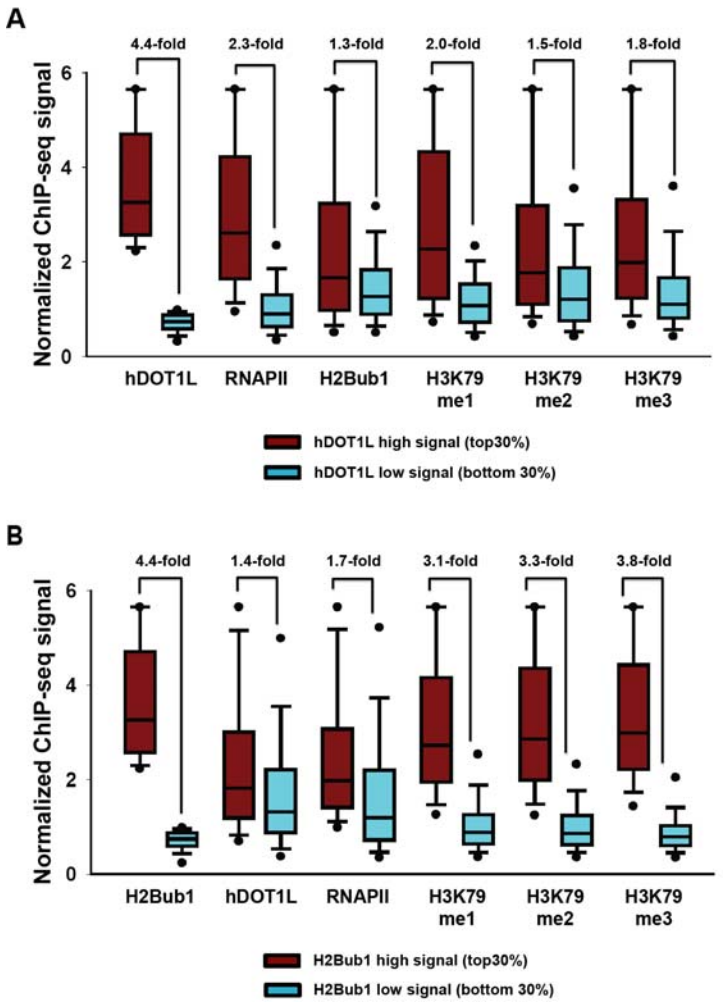
Supplementary FIGURE 2. Complete data from the genome-wide analysis of hDOT1L, RNAPII, and histone modifications near the TSS and TES. ChIP-seq results for hDOT1L, RNAPII, H3K79me1, H3K79me2, H3K79me3, and H2BUb1 are shown. The average intensity of the mapped sequence read counts is presented for the high (top 30%), middle, and low (bottom 30%) levels of gene expression. The profiles show 5 kb upstream and downstream of the TSS and TES.



Supplementary FIGURE 3. Comprehensive analysis of high-quality hDOT1L binding sites. *A-C*, A total of 8,953 hDOT1L-enriched regions were obtained from a re-analysis of our ChIP-seq results using MACS (see Methods) (59). Among the 8,953 hDOT1L-enriched regions, the top 500 sites that showed low FDRs (false discovery rates) and high-fold enrichment were regarded as high-quality hDOT1L binding sites. *A*, The high-quality hDOT1L binding sites are highly occupied by actively transcribing RNAPII (ChIP using anti-RNAPII phospho CTD Ser-5). *B*, The expression level of whole transcripts (n = 15,979) in the NCCIT cell line were measured using RNA-seq, and compared with the high-quality hDOT1L binding site-associated transcripts (n = 318) (see Methods) (60). *C*, Functional annotation analysis of high-quality hDOT1L binding sites in the promoter regions was performed with GREAT (see Supplementary Methods), which infers the functions of binding sites by exploiting available gene annotation databases. The top 40 terms in the 'Pathway Commons' category, which were significantly associated with the binding sites (binomial P-value < 4.3×10^{-3}), were clustered by keywords, and the number of terms in each cluster was counted. *D*, Genes targeted by both hDOT1L and RNAPII tend to be highly expressed. The median gene expression levels and standard deviations of hDOT1L-targeted genes, RNAPII-targeted genes, and genes targeted by both hDOT1L and RNAPII

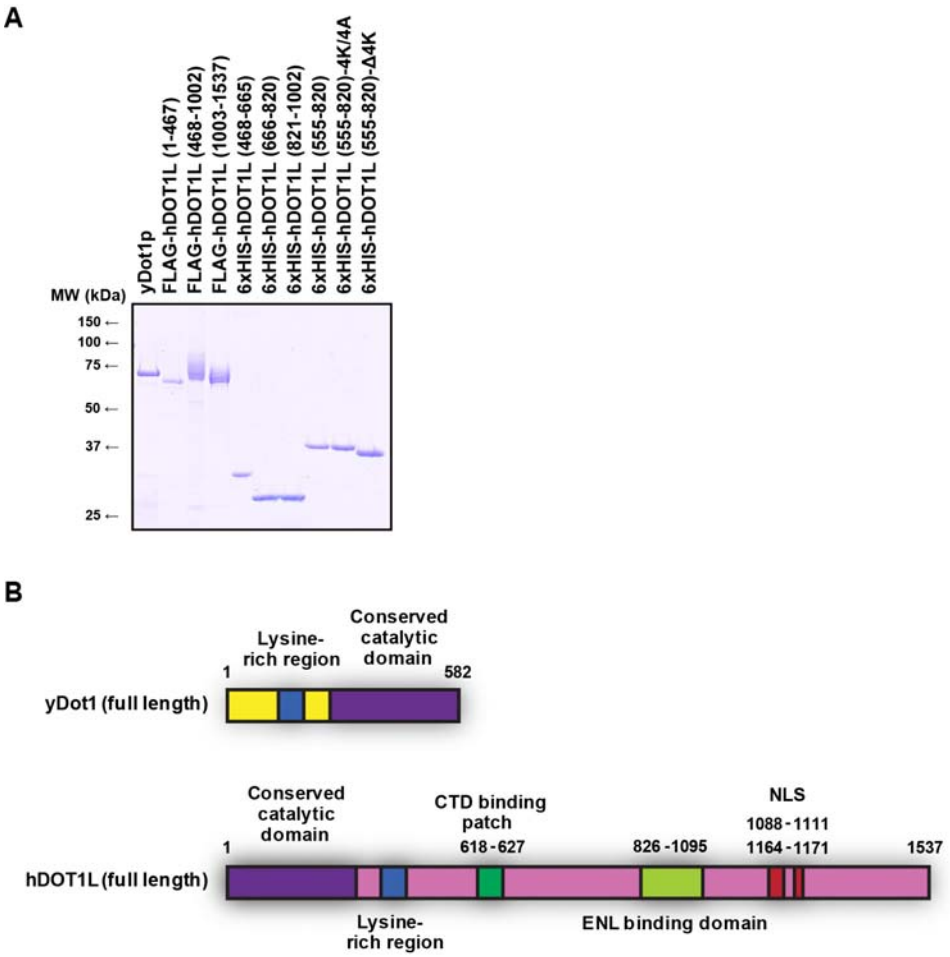
were plotted using 100 iterations randomly selecting 90% of the genes in each group (y-axis). The P-values between the co-occupied gene set and hDOT1L or RNAPII occupied genes were 0.0015 and 0.001, respectively (KS-test). ‘Others’ indicates genes that were targeted by neither hDOT1L nor RNAPII (n = 12,010).

Figure S4



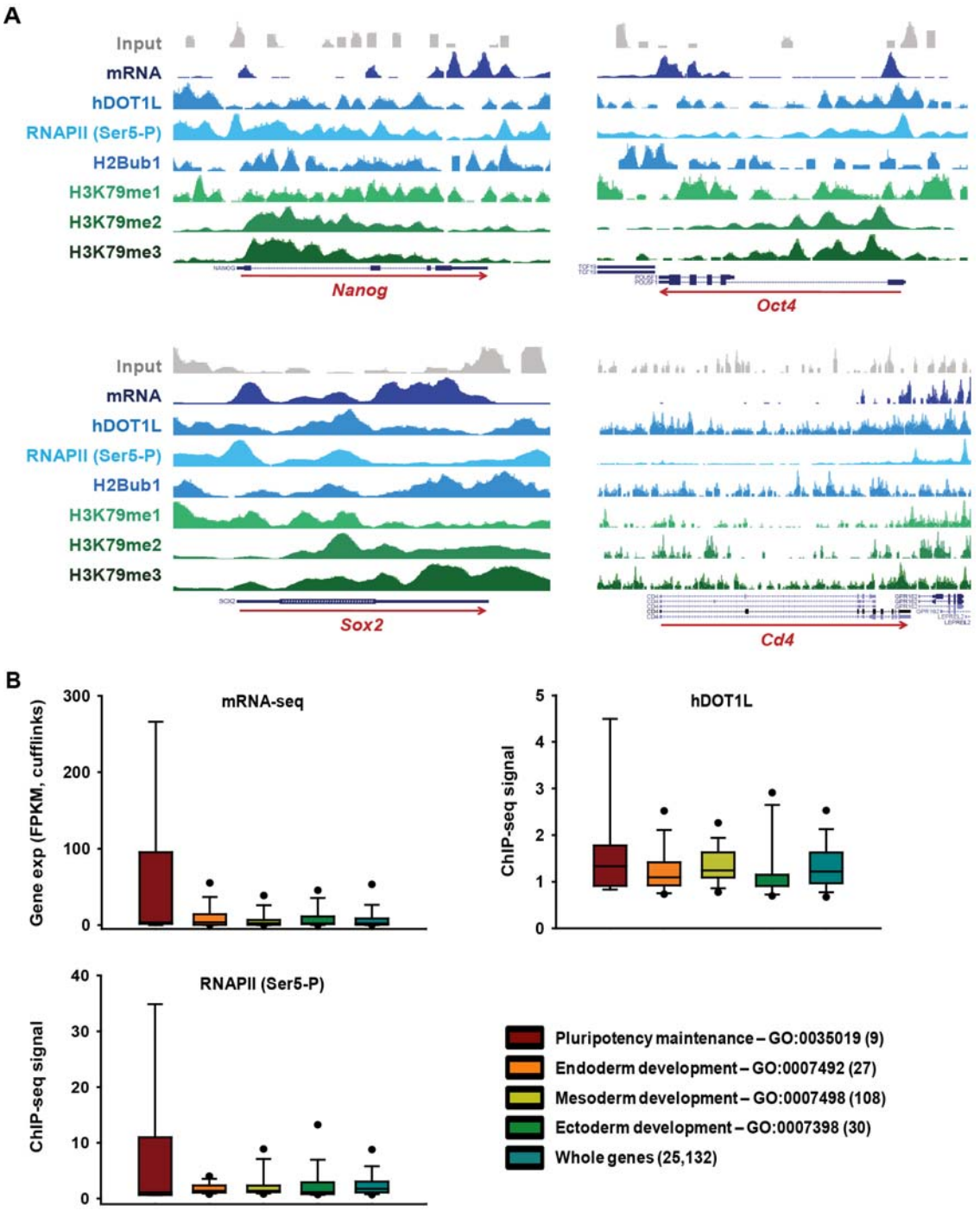
Supplementary FIGURE 4. Patterns of genome-wide co-occurrence between hDOT1L and other modification patterns in gene body regions. *A*, Box plots illustrate modification signals in hDOT1L-enriched (red, top 30%) and -depleted genes (cyan, bottom 30%), with the fold enrichment between gene groups shown. Due to scale differences between the modifications, the ChIP-seq signals were quintile-normalized. *B*, Boxplots of modification signals in H2BUb1-enriched (red, top 30%) and -depleted genes (cyan, bottom 30%) with the fold enrichment between gene groups.

Figure S5



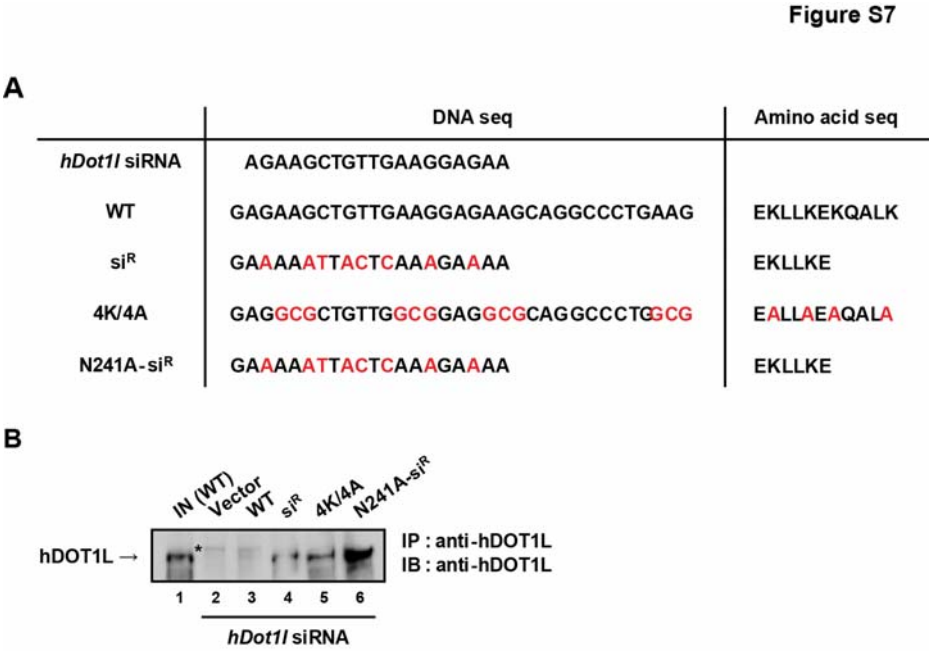
Supplementary FIGURE 5. Schematic representation of the full-length yDot1 and human hDOT1L constructs used in the experiment. *A*, Several purified recombinant proteins were used in this study. FLAG-hDOT1L (1-467), (468-1002) and (1003-1537) were purified from an Sf-21 insect cell-based baculovirus expression system, while the others were purified from a bacterial system. All proteins were resolved by SDS-PAGE and stained with Coomassie blue. FLAG-hDOT1L (468-1002) and (1003-1537) take on multiple forms, possibly due to a post-translational modification such as glycosylation. *B*, Full-length yDot1p and hDOT1L are compared. The CTD binding patch of hDOT1L identified in this study is shown within the long undefined middle and C-terminal regions of hDOT1L.

Figure S6



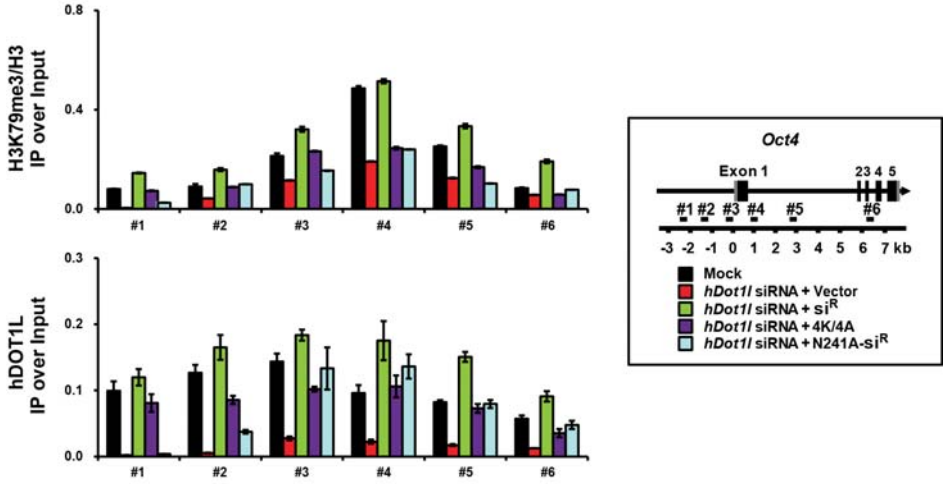
Supplementary FIGURE 6. Actively-transcribed pluripotency maintenance-related genes are co-occupied by both hDOT1L and RNAPII. *A*, All ChIP-seq data were mapped to the human genome (UCSC hg18 assembly based on NCBI build 36.1). The y-axis indicates the read counts of sequence tags in 10-bp increments. The figure shows the results from the RNA-seq and six different ChIP-seq results for hDOT1L, RNAPII, H3K79me1, H3K79me2, H3K79me3, and H2BUb1. The random distribution of IgG along the genome and its strong enrichment in specific regions may reflect

sequencing bias and the normalization process. *B*, top left, Box plot of gene expression levels in pluripotency-related genes. The utilized genes were extracted from the Gene Ontology database. We focused on nine pluripotency maintenance-related genes along with 25,132 total genes and representative development-related genes (27 endoderm genes, 108 mesoderm genes, and 30 ectoderm genes). *B*, top right and bottom left, hDOT1L and RNAPII were enriched at pluripotency maintenance-related genes compared to development-related genes.



Supplementary FIGURE 7. SiRNA-mediated knock-down of hDOT1L. *A*, DNA sequence comparison of the *hDot1l* siRNA used in this study versus its target sequences in the WT, si^R, 4K/4A, and N241A-si^R constructs. Amino acid sequences for the corresponding DNA sequences of the *hDot1l* constructs are also shown. *B*, The *hDot1l* siRNA was transiently transfected into NCCIT cells, followed by transfection with control vector, WT, si^R, 4K/4A, and N241A-si^R. Whole-cell extracts were immunoprecipitated overnight with anti-DOT1L at 4°C. IP samples were resolved by SDS-PAGE and immunoblotted with the indicated antibodies. A nonspecific band in the hDOT1L immunoblot is indicated by an asterisk.

Figure S8



Supplementary FIGURE 8. The decrease in H3K79me3 was rescued by exogenous hDOT1L but not by an enzymatic activity-defective mutant. Schematic representation of the *Oct4* locus, along with the location of the primers and ChIP analyses. Control vector, si^R, 4K/4A, and N241A-si^R were transfected into hDOT1L-knockdown NCCIT cells, and ChIP analyses were carried out with indicated antibodies. All signals were normalized with respect to the input and total histone H3. Error bars denote standard deviations from three biological replicates, each consisting of three qPCR reactions.

Supplementary Table 1. Real-time qPCR primers

Gene	Seq (5' → 3')	Product size
hDot1l-RT-F	CAGCTTCAGTGATGGTGCTTCT	132
hDot1l-RT-R	AGCTGCCTCTTGCTCAGGAAA	
Nanog-RT-F	ACCTTCCAATGTGGAGCAACCA	184
Nanog-RT-R	TGCATGCAGGACTGCAGAGATT	
Oct4-RT-F	AAACCCACACTGCAGCAGATCA	126
Oct4-RT-R	TCGTTGTGCATAGTCGCTGCTT	
Sox2-RT-F	TGTGGTTACCTCTTCCTCCCACT	138
Sox2-RT-R	TGGTAGTGCTGGGACATGTGAA	

Supplementary Table 1. Primer sets used in real-time qPCR. Primer sets were designed using the online IDT program (http://eu.idtdna.com/Scitools/Applications/Primerquest/).

Supplementary Table 2. Conventional ChIP primers

Gene	Seq (5' → 3')	Product size
*Nanog*_-2,698_F	TGGTCAGCTCCTTTACTGCAACCT	231
*Nanog*_-2,698_R	TGGGAGAGAACACGGCAACTAA	
*Nanog*_-1,506_F	AAGAGAGACAGGAGGGCAAGTT	254
*Nanog*_-1,506_R	TGCTCCCACACAAGCTGACTTT	
*Nanog*_-576_F	TTTGAGACGTAGTCCCGCTCTGTT	211
*Nanog*_-576_R	GAGTTTGAAACCAGCCTGGCCAACAT	
*Nanog*_+636_F	ATGGGCTTAGGCATGGTGGAAA	287
*Nanog*_+636_R	AGCTGCCCAGTAACATCCACAA	
*Nanog*_+2,931_F	TAACACCCAGTGTGGGAGCTTT	268
*Nanog*_+2,931_R	ATCTCCTGGGTTCAAGCGATTCTCCT	
*Nanog*_+6,184_F	AACAGCTGGGATTTACAGGCGT	205
*Nanog*_+6,184_R	AGCCTCCCAATCCCAAACAATACG	
*Oct4*_-2,549_F	GAGGATGGCAAGCTGAGAAA	148
*Oct4*_-2,549_R	CTCAATCCCCAGGACAGAAC	
*Oct4*_-1,672_F	AGCCCCACTAAACAAAGCAC	162
*Oct4*_-1,672_R	GCAATCCCCTCAAAGACTGA	
*Oct4*_-264_F	AGTCTGGGCAACAAAGTGAGA	169
*Oct4*_-264_R	AGAAACTGAGGCGAAGGATG	
*Oct4*_+990_F	ATCTCCTGTTTGGGCTGTGTGT	250
*Oct4*_+990_R	TTCAATCCCGCAGCAGCTCTAT	
*Oct4*_+2,825_F	ACTTCAGGGCCTTGCCACATTA	207
*Oct4*_+2,825_R	TTTGCAGTGAGCCAAGATCGCA	
*Oct4*_+6,131_F	AACCTGGAGTTTGTGCCAGGGTTT	202
*Oct4*_+6,131_R	TGTCTATCTACTGTGTCCCAGGCT	
*Sox2*_-2,133_F	TCTCCAGGTCCGTGTTTACCTT	192
*Sox2*_-2,133_R	TTTCCAATCAACCTTCCTGCCC	
*Sox2*_-227_F	AAAGGTTTCTCAGTGGCTGGCA	190
*Sox2*_-227_R	TGGGTTTCTAGCGACCAATCAG	
*Sox2*_+561_F	TGAATGCCTTCATGGTGTGGTC	84
*Sox2*_+561_R	TGCTGATCTCCGAGTTGTGCAT	
*Sox2*_+1,581_F	AAACCGCGATGCCGACAAGAAA	198

*Sox2*_+1,581_R	TGCAAACTTCCTGCAAAGCTCC	
*Cd4*_-1,897_F	AATACCTAGGCTTTCTCGGGCA	119
*Cd4*_-1,897_R	AAGTTCTTGCAGGCAGTGTGCT	
*Cd4*_-416_F	TTGCAGTGGTGCAATCTTGGCT	163
*Cd4*_-416_R	AGCCTGGCCAACATGGTGAAA	
*Cd4*_+272_F	TGATCTCGGCTCACTGCAATCT	152
*Cd4*_+272_R	AGCCTGGCCAACATGGTGAAA	
*Cd4*_+19,337_F	TCTTTGTTCCTGCAGCCGGTTT	176
*Cd4*_+19,337_R	TGGCAAATTGTAGAGGAGGCGA	

Supplementary Table 2. Primer sets used in conventional ChIP. Primer sets were designed using the online IDT program (http://eu.idtdna.com/Scitools/Applications/Primerquest/).

Supplementary Table 3. Pluripotency maintenance-related genes

Accession number	Gene	GO number	Function
NM_003106	*SOX2*	GO:0035019	Transcription factor SOX-2
NM_000222	*KIT*	GO:0035019	Mast/stem cell growth factor receptor
NM_000038	*APC*	GO:0035019	Adenomatous polyposis coli protein
NM_002701	*POU5F1*	GO:0035019	POU domain, class 5, transcription factor 1
NM_006766	*MYST3*	GO:0035019	Histone acetyltransferase MYST3
NM_005170	*ASCL2*	GO:0035019	Achaete-scute homolog 2
NM_024865	*NANOG*	GO:0035019	Homeobox protein NANOG
NM_001265	*CDX2*	GO:0035019	Homeobox protein CDX-2
NM_206937	*LIG4*	GO:0035019	DNA ligase 4

Supplementary Table 4. Endoderm development-related genes

Accession number	Gene	GO number	Function
NM_030761	*WNT4*	GO:0007492	Protein Wnt-4
NM_002293	*LAMC1*	GO:0007492	Laminin subunit gamma-1
NM_020909	*EPB41L5*	GO:0007492	Band 4.1-like protein 5
NM_001904	*CTNNB1*	GO:0001711	Catenin beta-1
NM_022454	*SOX17*	GO:0001706	Transcription factor SOX-17
NM_000127	*EXT1*	GO:0007492	Exostosin-1
NM_015193	*ARC*	GO:0007492	Activity-regulated cytoskeleton-associated protein
NM_017617	*NOTCH1*	GO:0007492	Neurogenic locus notch homolog protein 1
NM_012242	*DKK1*	GO:0007492	Dickkopf-related protein 1
NM_004329	*BMPR1A*	GO:0048382	Bone morphogenetic protein receptor type-1A
NM_002729	*HHEX*	GO:0007492	Hematopoietically-expressed homeobox protein HHEX
NM_003393	*WNT8B*	GO:0007492	Protein Wnt-8b
NM_004626	*WNT11*	GO:0007492	Protein Wnt-11
NM_014212	*HOXC11*	GO:0007492	Homeobox protein Hox-C11
NM_003317	*NKX2-1*	GO:0007492	Homeobox protein Nkx-2.1
NM_006194	*PAX9*	GO:0007492	Paired box protein Pax-9
NM_004498	*ONECUT1*	GO:0007492	Hepatocyte nuclear factor 6
NM_005902	*SMAD3*	GO:0007492	Mothers against decapentaplegic homolog 3
NM_005568	*LHX1*	GO:0001706	LIM/homeobox protein Lhx1
NM_000458	*HNF1B*	GO:0042663	Hepatocyte nuclear factor 1-beta
NM_004459	*BPTF*	GO:0007492	Nucleosome-remodeling factor subunit BPTF
NM_182641	*BPTF*	GO:0007492	Nucleosome-remodeling factor subunit BPTF
NM_005257	*GATA6*	GO:0007492	Transcription factor GATA-6
NM_005901	*SMAD2*	GO:0007492	Mothers against decapentaplegic homolog 2
NM_001003652	*SMAD2*	GO:0007492	Mothers against decapentaplegic homolog 2
NM_005359	*SMAD4*	GO:0007492	Mothers against decapentaplegic homolog 4
NM_000660	*TGFB1*	GO:0007492	Transforming growth factor beta-1
NM_021784	*FOXA2*	GO:0007492	Hepatocyte nuclear factor 3-beta
NM_153675	*FOXA2*	GO:0007492	Hepatocyte nuclear factor 3-beta

Supplementary Table 5. Mesoderm development-related genes

Accession number	Gene	GO number	Function
NM_004431	*EPHA2*	GO:0048320	Ephrin type-A receptor 2
NM_012090	*MACF1*	GO:0001707	Microtubule-actin cross-linking factor 1, isoform 4
NM_033044	*MACF1*	GO:0001707	Microtubule-actin cross-linking factor 1, isoform 4
NM_005424	*TIE1*	GO:0007498	Tyrosine-protein kinase receptor Tie-1
NM_018230	*NUP133*	GO:0048339	Nuclear pore complex protein Nup133
NM_016170	*TLX2*	GO:0001707	T-cell leukemia homeobox protein 2
NM_031283	*TCF7L1*	GO:0048319	Transcription factor 7-like 1
NM_020909	*EPB41L5*	GO:0048339	Band 4.1-like protein 5
NM_001616	*ACVR2A*	GO:0007498	Activin receptor type-2A
NM_001105	*ACVR1*	GO:0007498	Activin receptor type-1
NM_001204	*BMPR2*	GO:0001707	Bone morphogenetic protein receptor type-2
NM_001106	*ACVR2B*	GO:0007498	Activin receptor type-2B
NM_003212	*TDGF3*	GO:0007500	Putative teratocarcinoma-derived growth factor 2
NM_003741	*CHRD*	GO:0001707	Chordin
NM_003722	*TP63*	GO:0007499	Tumor protein 63
NM_002111	*HTT*	GO:0048341	Huntingtin
NM_016269	*LEF1*	GO:0048341	Lymphoid enhancer-binding factor 1
NM_005900	*SMAD1*	GO:0001710	Mothers against decapentaplegic homolog 1
NM_001003688	*SMAD1*	GO:0001710	Mothers against decapentaplegic homolog 1
NM_002715	*PPP2CA*	GO:0007498	Serine/threonine-protein phosphatase 2A catalytic subunit alpha isoform
NM_001453	*FOXC1*	GO:0048343	Forkhead box protein C1
NM_033229	*TRIM15*	GO:0007500	Tripartite motif-containing protein 15
NM_003131	*SRF*	GO:0001707	Serum response factor
NM_003376	*VEGFA*	GO:0007498	Vascular endothelial growth factor A
NM_198392	*TCF21*	GO:0007498	Transcription factor 21
NM_003206	*TCF21*	GO:0007498	Transcription factor 21
NM_003181	*T*	GO:0007498	Brachyury protein

NM_002192	*INHBA*	GO:0048333	Inhibin beta A chain
NM_006060	*IKZF1*	GO:0007498	DNA-binding protein Ikaros
NM_177524	*MEST*	GO:0007498	Mesoderm-specific transcript homolog protein
NM_177525	*MEST*	GO:0007498	Mesoderm-specific transcript homolog protein
NM_002402	*MEST*	GO:0007498	Mesoderm-specific transcript homolog protein
NM_000193	*SHH*	GO:0007502	Sonic hedgehog protein
NM_023106	*FGFR1*	GO:0048378	Basic fibroblast growth factor receptor 1
NM_023105	*FGFR1*	GO:0048378	Basic fibroblast growth factor receptor 1
NM_023111	*FGFR1*	GO:0048378	Basic fibroblast growth factor receptor 1
NM_015850	*FGFR1*	GO:0048378	Basic fibroblast growth factor receptor 1
NM_023110	*FGFR1*	GO:0048378	Basic fibroblast growth factor receptor 1
NM_023107	*FGFR1*	GO:0048378	Basic fibroblast growth factor receptor 1
NM_023108	*FGFR1*	GO:0048378	Basic fibroblast growth factor receptor 1
NM_003068	*SNAI2*	GO:0007499	Zinc finger protein SNAI2
NM_000127	*EXT1*	GO:0007498	Exostosin-1
NM_003923	*FOXH1*	GO:0048318	Forkhead box protein H1
NM_004972	*JAK2*	GO:0007498	Tyrosine-protein kinase JAK2
NM_173198	*NR4A3*	GO:0001707	Nuclear receptor subfamily 4 group A member 3
NM_173199	*NR4A3*	GO:0001707	Nuclear receptor subfamily 4 group A member 3
NM_006981	*NR4A3*	GO:0001707	Nuclear receptor subfamily 4 group A member 3
NM_173200	*NR4A3*	GO:0001707	Nuclear receptor subfamily 4 group A member 3
NM_004235	*KLF4*	GO:0007500	Krueppel-like factor 4
NM_004329	*BMPR1A*	GO:0048378	Bone morphogenetic protein receptor type-1A
NM_000401	*EXT2*	GO:0001707	Exostosin-2
NM_207122	*EXT2*	GO:0001707	Exostosin-2
NM_004561	*OVOL1*	GO:0007498	Putative transcription factor Ovo-like 1
NM_005430	*WNT1*	GO:0048332	Proto-oncogene Wnt-1
NM_005811	*GDF11*	GO:0007498	Growth/differentiation factor 11
NM_182742	*TXNRD1*	GO:0001707	Thioredoxin reductase 1, cytoplasmic
NM_182743	*TXNRD1*	GO:0001707	Thioredoxin reductase 1, cytoplasmic

NM_003330	*TXNRD1*	GO:0001707	Thioredoxin reductase 1, cytoplasmic
NM_182729	*TXNRD1*	GO:0001707	Thioredoxin reductase 1, cytoplasmic
NM_016569	*TBX3*	GO:0048332	T-box transcription factor TBX3
NM_005996	*TBX3*	GO:0048332	T-box transcription factor TBX3
NM_000545	*HNF1A*	GO:0048341	Hepatocyte nuclear factor 1-alpha
NM_006237	*POU4F1*	GO:0007498	POU domain, class 4, transcription factor 1
NM_130850	*BMP4*	GO:0048392	Bone morphogenetic protein 4
NM_130851	*BMP4*	GO:0048392	Bone morphogenetic protein 4
NM_001202	*BMP4*	GO:0048392	Bone morphogenetic protein 4
NM_005902	*SMAD3*	GO:0001707	Mothers against decapentaplegic homolog 3
NM_003502	*AXIN1*	GO:0048320	Axin-1
NM_181050	*AXIN1*	GO:0048320	Axin-1
NM_004608	*TBX6*	GO:0007498	T-box transcription factor TBX6
NM_005251	*FOXC2*	GO:0048343	Forkhead box protein C2
NM_183231	*IKZF3*	GO:0007498	Zinc finger protein Aiolos
NM_183230	*IKZF3*	GO:0007498	Zinc finger protein Aiolos
NM_183232	*IKZF3*	GO:0007498	Zinc finger protein Aiolos
NM_012481	*IKZF3*	GO:0007498	Zinc finger protein Aiolos
NM_183229	*IKZF3*	GO:0007498	Zinc finger protein Aiolos
NM_183228	*IKZF3*	GO:0007498	Zinc finger protein Aiolos
NM_030753	*WNT3*	GO:0001707	Proto-oncogene Wnt-3
NM_212471	*PRKAR1A*	GO:0001707	cAMP-dependent protein kinase type I-alpha regulatory subunit
NM_002734	*PRKAR1A*	GO:0001707	cAMP-dependent protein kinase type I-alpha regulatory subunit
NM_212472	*PRKAR1A*	GO:0001707	cAMP-dependent protein kinase type I-alpha regulatory subunit
NM_003004	*SECTM1*	GO:0007498	Secreted and transmembrane protein 1
NM_020648	*TWSG1*	GO:0001707	Twisted gastrulation protein homolog 1
NM_005901	*SMAD2*	GO:0001707	Mothers against decapentaplegic homolog 2
NM_001003652	*SMAD2*	GO:0001707	Mothers against decapentaplegic homolog 2
NM_005359	*SMAD4*	GO:0007498	Mothers against decapentaplegic homolog 4
NM_000479	*AMH*	GO:0007506	Muellerian-inhibiting factor
NM_002378	*MATK*	GO:0007498	Megakaryocyte-associated tyrosine-protein

			kinase
NM_139354	*MATK*	GO:0007498	Megakaryocyte-associated tyrosine-protein kinase
NM_139355	*MATK*	GO:0007498	Megakaryocyte-associated tyrosine-protein kinase
NM_016581	*ECSIT*	GO:0001707	Evolutionarily conserved signaling intermediate in Toll pathway, mitochondrial
NM_002730	*PRKACA*	GO:0001707	cAMP-dependent protein kinase catalytic subunit alpha
NM_207518	*PRKACA*	GO:0001707	cAMP-dependent protein kinase catalytic subunit alpha
NM_203486	*DLL3*	GO:0048339	Delta-like protein 3
NM_016941	*DLL3*	GO:0048339	Delta-like protein 3
NM_004609	*TCF15*	GO:0007498	Transcription factor 15
NM_002110	*HCK*	GO:0007498	Tyrosine-protein kinase HCK
NM_172110	*EYA2*	GO:0007501	Eyes absent homolog 2
NM_005244	*EYA2*	GO:0007501	Eyes absent homolog 2
NM_005985	*SNAI1*	GO:0001707	Zinc finger protein SNAI1
NM_001719	*BMP7*	GO:0001707	Bone morphogenetic protein 7
NM_181832	*NF2*	GO:0001707	Merlin
NM_181831	*NF2*	GO:0001707	Merlin
NM_181833	*NF2*	GO:0001707	Merlin
NM_181828	*NF2*	GO:0001707	Merlin
NM_181825	*NF2*	GO:0001707	Merlin
NM_000268	*NF2*	GO:0001707	Merlin
NM_181829	*NF2*	GO:0001707	Merlin
NM_016418	*NF2*	GO:0001707	Merlin
NM_181830	*NF2*	GO:0001707	Merlin
NM_203281	*BMX*	GO:0007498	Cytoplasmic tyrosine-protein kinase BMX
NM_001721	*BMX*	GO:0007498	Cytoplasmic tyrosine-protein kinase BMX
NM_000061	*BTK*	GO:0007498	Tyrosine-protein kinase BTK
NM_003308	*TSPY1*	GO:0007506	Testis-specific Y-encoded protein 1

Supplementary Table 6. Ectoderm development-related genes

Accession number	Gene	GO number	Function
NM_003443	*ZBTB17*	GO:0007398	Zinc finger and BTB domain-containing protein 17
NM_015872	*ZBTB7B*	GO:0007398	Zinc finger and BTB domain-containing protein 7B
NM_012476	*VAX2*	GO:0007398	Ventral anterior homeobox 2
NM_020909	*EPB41L5*	GO:0007398	Band 4.1-like protein 5
NM_003507	*FZD7*	GO:0042666	Frizzled-7
NM_001904	*CTNNB1*	GO:0007398	Catenin beta-1
NM_003220	*TFAP2A*	GO:0007398	Transcription factor AP-2-alpha
NM_181349	*SMURF1*	GO:0007398	E3 ubiquitin-protein ligase SMURF1
NM_020429	*SMURF1*	GO:0007398	E3 ubiquitin-protein ligase SMURF1
NM_000193	*SHH*	GO:0007398	Sonic hedgehog protein
NM_004329	*BMPR1A*	GO:0007398	Bone morphogenetic protein receptor type-1A
NM_001422	*ELF5*	GO:0007398	ETS-related transcription factor Elf-5
NM_198381	*ELF5*	GO:0007398	ETS-related transcription factor Elf-5
NM_005555	*KRT6B*	GO:0007398	Keratin, type II cytoskeletal 6B
NM_005554	*KRT6A*	GO:0007398	Keratin, type II cytoskeletal 6A
NM_001980	*STX2*	GO:0007398	Syntaxin-2
NM_194356	*STX2*	GO:0007398	Syntaxin-2
NM_005568	*LHX1*	GO:0001705	LIM/homeobox protein Lhx1
NM_021784	*FOXA2*	GO:0001705	Hepatocyte nuclear factor 3-beta
NM_153675	*FOXA2*	GO:0001705	Hepatocyte nuclear factor 3-beta
NM_181832	*NF2*	GO:0007398	Merlin
NM_181831	*NF2*	GO:0007398	Merlin
NM_181833	*NF2*	GO:0007398	Merlin
NM_181828	*NF2*	GO:0007398	Merlin
NM_181825	*NF2*	GO:0007398	Merlin
NM_000268	*NF2*	GO:0007398	Merlin
NM_181829	*NF2*	GO:0007398	Merlin
NM_016418	*NF2*	GO:0007398	Merlin

NM_181830	*NF2*	GO:0007398	Merlin
NM_001399	*EDA*	GO:0007398	Ectodysplasin-A
NM_001005615	*EDA*	GO:0007398	Ectodysplasin-A
NM_001005612	*EDA*	GO:0007398	Ectodysplasin-A
NM_001005611	*EDA*	GO:0007398	Ectodysplasin-A
NM_001005609	*EDA*	GO:0007398	Ectodysplasin-A
NM_001005610	*EDA*	GO:0007398	Ectodysplasin-A