Chd5 Regulates MuERV-L/MERVL Expression in Mouse Embryonic Stem Cells Via H3K27me3 Modification and Histone H3.1/H3.2

林, 正康

https://doi.org/10.15017/1654730

出版情報：九州大学, 2015, 博士（医学）, 課程博士
バージョン：
権利関係：やむを得ない事由により本文ファイル非公開（2）

Chd5 regulates MuERV-L/MERVL expression in mouse embryonic stem cells via H3K27me3 modification and histone H3.1/H3.2.

Masayasu Hayashi[1,2], Kazumitsu Maehara[1], Akihito Harada[1], Yuichiro Semba[1,2], Kensuke Kudo[1,3], Hidehisa Takahashi[4], Shinya Oki[5], Chikara Meno[5], Kenji Ichiyanagi[6], Koichi Akashi[2], and Yasuyuki Ohkawa[1]*

[1]Department of Advanced Medical Initiatives, Faculty of Medicine, Kyushu University, Fukuoka 812-8582, Japan

[2]Department of Medicine and Biosystemic Science, Faculty of Medicine, Kyushu University, Fukuoka 812-8582, Japan

[3]Department of Surgery and Science, Graduate School of Medical Sciences, Kyushu University, Fukuoka 812-8582, Japan

[4]Department of Biochemistry, Hokkaido University Graduate School of Medicine, Hokkaido 060-8638, Japan

[5]Department of Developmental Biology, Graduate School of Medical Sciences, Kyushu University, Fukuoka 812-8582, Japan

[6]Division of Epigenomics and Development, Medical Institute of Bioregulation, Kyushu University,

Fukuoka 812-8582, Japan

***Corresponding author**

Yasuyuki Ohkawa

Kyushu University, Faculty of Medicine, 3-1-1 Maidashi, Higashi, Collabo-S. I, 3F, Fukuoka

812-8582, Japan

E-mail: yohkawa@epigenetics.med.kyushu-u.ac.jp

Tel: +81-92-642-6216

Fax: +81-92-642-6099

Running head: Chd5 regulates MERVL expression

Keywords:

- **chromatin remodeling enzyme,**
- **Chd5**
- **Retrotransposon**
- **stem cell biology**

This article contains 8 figures and 2 tables.

Contract grant sponsor: The Japan Society for the Promotion of Science (JSPS) ; Contract Grant number: 25116010.

Contract grant sponsor: The Japan Society for the Promotion of Science (JSPS) ; Contract Grant number: 26290064.

Abstract

Chd5 is an essential factor for neuronal differentiation and spermatogenesis, and known as a tumor suppressor. H3K27me3 and H3K4un are modifications recognized by Chd5; however, it remains unclear how Chd5 remodels chromatin structure. We completely disrupted the Chd5 locus using the CRISPR-Cas9 system to generate a 52 kbp long deletion, and analyzed Chd5 function in mouse embryonic stem cells. Our findings show that Chd5 represses murine endogenous retrovirus-L (MuERV-L/MERVL), an endogenous retrovirus-derived retrotransposon, by regulating H3K27me3 and H3.1/H3.2 function.

Introduction

Gene expression is regulated by changes in chromatin structure due to nucleosome positioning on the genome. Nucleosomes are the base units of chromatin structure, and consist of two core histones: H2A, H2B, H3, or H4. Each histone has a tail region and various modifications that functionally correlate to genomic nucleosome position, in particular, the N-terminal tail of histone H3 [Strahl and Allis, 2000]. Histone modifications, including H3K4me3 and H3K27me3, are recognized by chromatin remodeling enzymes such as the SWI/SNF ATP dependent chromatin remodeling factors, and are critical for gene-regulated changes in chromatin structure [de la Serna et al., 2006]. Defects in SWI/SNF function interrupt gene expression followed by chromatin remodeling and results in incomplete differentiation or tumorigenesis [Reisman et al., 2009].

SWI/SNF family members (e.g., Brahma-related gene 1 (Brg1)/Brahma (Brm)) has been demonstrated to be expressed ubiquitously in various tissues, except chromodomain helicase DNA-binding proteins (Chds) family, which is one of the SWI/SNF family members, is preferentially expressed in specific tissues [Marfella and Imbalzano, 2007]. Essential functions for Chd1 in maintenance of embryonic stem cell pluripotency [Gaspar-Maia et al., 2009], and Chd2 in skeletal muscle differentiation [Harada et al., 2012] have been reported. In addition, Chd5 is essential for neuronal differentiation and spermatogenesis [Egan et al., 2013; Li et al., 2014; Zhuang et al., 2014].

Chd5 is also a tumor suppressor [Bagchi et al., 2007; Paul et al., 2013], and was identified as the

corresponding gene of Del 1p36 [Fujita et al., 2008], a poor prognosis factor in human neuroblastomas. High Chd5 expression has been detected in the nervous system and testis [Thompson et al., 2003; Zhuang et al., 2014]. Interestingly, loss of Chd5 function in various cancers with low Chd5 expression has been reported [Kolla et al., 2014], suggesting that Chd5 plays a critical role in ubiquitous gene expression.

Gene regulation by Chd5 has been examined in areas of high Chd5 expression [Bagchi et al., 2007; Egan et al., 2013; Paul et al., 2013]. Chd5 function is necessary to recognize H3K4un and H3K27me3, and disruption of its PHD domain or chromodomain results in tumor progression or ineffective neural differentiation [Bagchi et al., 2007; Egan et al., 2013; Oliver et al., 2012; Paul et al., 2013]. Additionally, Chd5 functions in nucleosome removal in spermatids, consistent with loss of nucleosome conversion to protamine during spermatogenesis in Chd5 knock-out (KO) mice [Li et al., 2014; Zhuang et al., 2014]. However, which genes Chd5 targets on chromatin or how Chd5 functions to regulate target gene in various tissues, remains unclear.

Here, we investigated the direct function of Chd5 on chromatin structure, and show increased expression of two-cell (2C) specific genes, and a retrotransposon, murine endogenous retrovirus-L (MuERV-L/MERVL), in Chd5 KO mouse embryonic stem cells (mESCs). Specifically, Chd5 converts facultative heterochromatin to euchromatin, which correlates with H3K27me3 and H3.1/H3.2 levels.

Materials and Methods

Mouse embryonic stem cell culture

EB5 [Ogawa et al., 2004], a mESC line, was cultured in Glasgow minimum essential medium (Sigma-Aldrich, Saint Louis, MO, USA) supplemented with 10% fetal bovine serum, 1000 U/ml recombinant leukemia inhibitory factor (Nacalai, Japan), 1 mM sodium pyruvate, 0.1 mM non-essential amino acids (Nacalai, Japan), 0.1 mM β-mercaptoethanol (Nacalai, Japan), and penicillin/streptomycin (50 U/50 µg/ml; Nacalai, Japan), at 37°C in a 5% CO2 incubator. The medium was changed daily, and cells routinely passaged every 2-3 days.

Generation of Chd5−/− mESC lines

To completely disrupt the Chd5 locus, the CRISPR-Cas9 system (Mali et al., 2013) was used. Target sequences were designed upstream of the transcription start site (TSS), 5′-AAG TGC CTT CGC GGG CAG GGT GG-3′, and downstream of the transcription end site (TES), 5′-GTT CTT AGC CTC CGT GAT TTA GG-3′. To address off-target sites, manual alignment using Bowtie [Langmead et al., 2009] with a 2 nucleotide (nt) mismatch allowance was performed. No identical sequences were identified with the above conditions. Thus, these sequences were inserted into pSpCas9 (BB)-2A-GFP (PX458, Addgene), which was modified to also express a guide RNA (gRNA) scaffold. The obtained plasmids were transfected into mESCs using Xfect (ClonTech, USA), in accordance with the manufacturer's

instructions. Twenty-four hours after transfection, GFP-positive cells were sorted into 96-well plates for single-cell cloning using a cell sorter, SH800 (SONY, Japan). Genotyping PCR for Chd5−/− mESC selection was performed as described below.

Genomic PCR genotyping

DNA from each clone was recovered using DNeasy (Qiagen, Netherlands), in accordance with the manufacturer's instructions. Primers were designed as follows: PCR region 1F, 5′-TCT GTG TTT CCT CCA TGT TTC TCA GGA ATG-3′; PCR region 1R, 5′-ATT CTG CCT CAC AGA AAA CTA GCA TAG GTG-3′; PCR region 2F, 5′-AGG CGA TAA CAA AGA TGT AGA CAG CAG-3′; PCR region 2R, 5′-ACT GTT CTT GGG GTT CAT GGT TTT AAA GGA-3′; PCR region 3F, 5′-TGT GGG ACA CAT TTT CTT GGA CTT CTC TTT-3′; and PCR region 3R; 5′-TCT ACA GCT AAG GAC AAG ATG AGA GGA AGG-3′; Intra-Gapdh F, 5'- CTT CAG AGT GGA ATA CTG TTG CAC CAT AGG-3'; Intra-Gapdh R, 5'-GAG ACT TAG AAT GAC TTG GAG GAG GTT TGC-3'. Recovered DNA was amplified for 30 cycles using these primers and KOD FX Neo (TOYOBO, Japan).

Total RNA isolation and reverse transcription

Total RNA from each clone was isolated using Sepazol (Nacalai, Japan), precipitated using isopropanol, and then dissolved in RNase-free H2O. One microgram total RNA and random 6-mer

oligonucleotides were reverse-transcribed using the Prime Script II 1st cDNA synthesis Kit (Takara, Japan) in 20 μl reactions.

Reverse-transcription quantitative PCR

Appropriate primers for reverse-transcription quantitative polymerase chain reaction (RT-qPCR) were designed: chd5 F; 5'- GAC CAG GAG TGG CAG GAT GA-3′, chd5 R, 5′-CCT CCT GGA TTG TCG TC-3′; pou5f1 F, 5′-GAA GCA GAA GAG GAT CAC CTT G-3′; pou5f1 R, 5′-TTC TTA AGG CTG AGC TGC AAG-3′; nanog F, 5′-CCT CAG CCT CCA GCA GAT GC-3′; nanog R, 5′-CCG CTT GCA CTT CAC CCT TTG-3′; sox2 F, 5′-GCG GAG TGG AAA CTT TTG TCC-3′; sox2 R, 5′-GGG AAG CGT GTA CTT ATC CTT CT-3′; klf4 F, 5′-ATC CTT TCC AAC TCG CTA ACC C-3′; klf4 R, 5′-CGG ATC GGA TAG CTG AAG CTG-3′; gapdh F, 5′-ATG AAT ACG GCT ACA GCA ACA GGG-3′; and gapdh R, 5′-GTC TGG GAT GGA AAT TGT GAG GGA-3′. Primers for MERVL and Zscan4 were obtained from [Maksakova et al., 2013] and [Zalzman et al., 2010], respectively. cDNAs for each clone were amplified for 40 cycles with these primers and the THUNDERBIRD SYBR qPCR Mix (TOYOBO, Japan), and amplicons detected by PikoReal (Thermo Scientific, USA). For quantification, *Gapdh* was used as the internal control.

Western blotting

Each clone was directly lysed in sodium dodecyl sulfate (SDS) polyacrylamide gel electrophoresis sample buffer (125 mM Tris (pH 6.8), 4% SDS, 20% glycerol, and 0.01% bromophenol blue). Eluted DNA was sheared by pipetting with a 25G needle syringe. Next, 10% β-mercaptoethanol was added and incubated for 5 min at 95°C. The primary antibodies used were rabbit anti-Chd5 (1:500) (sc68389; Santa Cruz, USA) and rabbit anti-Hsp90 (1:1000) (H114; Santa Cruz, USA). Membranes with primary antibodies were incubated for 2 hours at room temperature. The secondary antibody used was horseradish peroxidase-conjugated anti-rabbit IgG (1:5000) (GE Healthcare, USA), and membranes with secondary antibody were incubated for 30 minutes at room temperature. Signals were detected using Chemilumi Ultra- One (Nacalai, Japan).

Chromatin immunoprecipitation

Samples fixed by formaldehyde (0.5%, 5 min at room temperature) were lysed into 1 ml chromatin immunoprecipitation (ChIP) buffer (10 mM Tris-HCl (pH 8.0), 200 mM KCl, 5 mM MgCl2, 1 mM CaCl2, and 0.5% NP40). Sonication was performed using a VCX130 (SONICS, USA). Next, 200 U/ml micrococcal nuclease (MNase) was added to sonicated lysates and incubated for 40 min at 37°C. To stop MNase digestion, 10 mM EDTA was added. Next, 1 ml modified RIPA buffer (50 mM Tris-HCl (pH 8.0), 200 mM KCl, 2 mM EDTA, 1% NP40, 0.5% sodium deoxycholate, and 0.1% SDS) was added to lysates before sonicating again, aiming to finally achieve approximately 500 bp fragments.

A portion of each sample was stored as input control. For ChIP, samples were then incubated with antibody-conjugated beads (Dynabeads; VERITAS, Japan) for 6 hours at 4°C, the beads washed, and DNA eluted using elution buffer (50 mM Tris-HCl (pH 8.0), 10 mM EDTA, and 1% SDS). Input and immunoprecipitates were incubated overnight at 65°C for reverse-crosslinking, and eluted samples purified using the MinElute PCR Purification Kit (Qiagen, Netherlands). For immunoprecipitation, used antibodies were described below; anti-H3K4me3, anti-H3K27me3, anti-H3.1/H3.2, anti-H3.3 (in house raised each) and anti-Chd5 (sc68389; Santa Cruz, USA).

mRNASeq library preparation

One microgram of total RNA was purified using the NEBNext Poly(A) mRNA Magnetic Isolation Module (NEB, UK), in accordance with the manufacturer's instructions. A messenger RNA sequencing (mRNA-Seq) library was prepared from polyA-purified RNA using the NEBNext Ultra Directional RNA Library Prep Kit for Illumina (NEB), in accordance with the manufacturer's instructions.

ChIPSeq library preparation

A chromatin immunoprecipitation sequencing (ChIP-Seq) library was prepared from the ChIP samples described above using the NEBNext Ultra DNA Library Prep Kit for Illumina (NEB), in

accordance with the manufacturer's instructions.

mRNASeq data analysis

We performed two types of transcript abundance quantification using the mRNA-seq data. Tophat (version 2.0.12) [Kim et al., 2013] and Cufflinks (version 2.1.1) [Trapnell et al., 2010], with default parameters, were utilized to quantify expression levels of Refseq genes (mm9) and to detect the differentially expressed genes (type I). Bowtie2 (version 2.2.5) [Langmead and Salzberg, 2012] and eXpress (version 1.5.1) [Roberts and Pachter, 2013] was utilized for accurate quantification of mutually resemble murine retroviral sequences (type II). Bowtie2 was run with the option: -a --rdg 6,5 --rfg 6,5 --score-min L,-.6,-.4, recommended by eXpress. The eXpress was run with the option: -B 10, which performs additional ten rounds of parameter optimization steps. The expression levels (TPMs) of mouse retroviral genes were quantified using consensus sequences FASTA downloaded from Repbase (type II-a) [Bao et al., 2015], and individual genomic sources of retroviral genes definition downloaded from UCSC (RepeatMasker track (A.F.A. Smit, R. Hubley & P. Green RepeatMasker at http://repeatmasker.org)) (type II-b). In addition, the simple (naive) count of mapped reads on the retroviral genes using bowtie2 (default option; select each best scored candidate in multi-mapped reads) was utilized to show the improvements of retroviral gene expression estimation by using eXpress.

ChIPSeq data analysis

The sequenced reads were mapped onto mouse genome (mm9) using bowtie2 with default options. To compare the intensities of ChIP-seq signals between different samples prepared from different input (e.g. KO vs. WT cells), we performed between-sample normalization after input normalization. First, log-transformed mapped-read counts on genomic windows (2kb upstream of TSSs or gene-body region) were calculated for all samples. Next the corresponding input sample for a (log-transformed) ChIP read counts were subtracted from (log-transformed) input read counts with some constant c. The constant c was determined so that the Pearson's correlation coefficient between ChIP and c · Input becomes 0 (taking the residuals of linear regression of Input vs. ChIP counts). Finally, between-sample normalization utilizing median of ratios of geometric means (geometric normalization) [Anders and Huber, 2010]was applied to correct the different library sizes of the input normalized ChIP read counts. The aggregation plot of ChIP-seq signals were generated by agplus [Maehara and Ohkawa, 2015].

Data access

The deep-sequencing data were deposited with the accession number [DDBJ: DRA003957].

Results

Complete Chd5 disruption

To address the function of Chd5 on gene regulation in mESCs, we first knocked out *Chd5* gene loci in the mouse embryonic stem cell, EB5. We used the CRIPSR-Cas9 system [Mali et al., 2013] for complete disruption of *Chd5* since insertional inactivation [Carlson and Largaespada, 2005] and deletion of partial exons [Carlson and Largaespada, 2005], which have been effectively used to knock out certain genes, sometimes leave residual functional transcripts. Thus, to disrupt the *Chd5* locus, two target sequences were designed upstream of the TSS and downstream of the TES of *Chd5* (Fig. 1A). With proper design of gRNAs, the Cas9 protein causes double-strand breaks (DSBs) at arbitrary DNA sequences. In the case of two DSBs, the upstream and downstream margins get re-connected, and the sequence between both DSBs is eliminated [Fujii et al., 2013]. Genotyping was performed by genomic PCR to determine the truncated genomic structure (Fig. 1A). The interval between the TSS and TES of the *Chd5* gene was 52 kbp. PCR primer sets 1 and 2 were designed inside the *Chd5* locus, and are only amplified in wild-type (WT) and heterozygous knocked-out KO lines, with amplicon sizes of 455 bp and 461 bp, respectively. With PCR region 3, the primers were designed outside of both target sequences and only amplify when *Chd5* is homologously or heterozygously disrupted, producing an amplicon size of approximately 600 bp, depending on where the non-homologous end joining (NHEJ) has occurred. Homozygous *Chd5*-disrupted lines (*Chd5*−/− lines) generate an appropriate band in PCR

region 3 only. Internal Sequence of *Gapdh* was also amplified as control of PCR (Supplementary Fig. 1A). To confirm where the NHEJ has occurred, PCR products in PCR region 3 were sequenced by the Sanger method (Fig. 1C). In *Chd5*−/− line #1, NHEJ occurred between overlapping adenines in each target sequence, while in *Chd5*−/− line #2, NHEJ occurred after insertion of two cytosines between each target sequence. Complete disruption of *Chd5* was shown, with RT-qPCR and western blotting signals detected in WT only (Fig. 2A, 2B).

To comprehensively determine the functional effect of Chd5 disruption in mESCs, mRNA sequencing (mRNA-Seq) was performed to examine gene expression in both *Chd5*−/− lines (#1 and #2). In *Chd5*−/− #1 and #2, mRNA-Seq data signals, visualized using the Integrative Genomics Viewer (IGV), were not observed within the entire *Chd5* locus (Fig. 2C), while WT showed signal along the gene body. *Kcnb2* and *Nphp4*, genes proximal to *Chd5* gene loci, had consistently low expression in WT, and *Chd5*−/− #1 and #2. Therefore, *Chd5* genes are substantially expressed in WT compared to these control signals. These findings show that two independent *Chd5*−/− clones lost the entire *Chd5* gene loci using the CRISPR-Cas9 system.

Chd5 disruption in mESCs minimally affects gene expression

To identify the gene specifically affected by loss of the *Chd5* gene in mESCs, we identified differentially expressed genes (DEGs) between WT and *Chd5*-/- mESCs using mRNA-Seq data.

Calculation and analysis of FPKM was performed using Cufflinks [Trapnell et al., 2010]. FPKM of protein-coding genes in WT vs *Chd5*−/− #1 and WT vs *Chd5*−/− #2 is shown by scatterplots (Fig. 3A). Both the X- and Y-axes are represented logarithmically, with 1 added to each FPKM to avoid log0. Red dots indicate DEGs. There were 64 DEGs between WT and *Chd5*−/− #1, and 56 DEGs between WT and *Chd5*−/− #2. Venn diagrams of overlapping downregulated and upregulated DEGs in *Chd5*−/− #1 and #2 are shown (Fig. 3B). Almost all DEGs overlap between *Chd5*−/− #1 and #2 (overlapping genes were listed in Table 1), therefore these expression changes are consistent and likely due to *Chd5* disruption. Moreover, it suggests that off-target effects should have been minimized.

To further confirm that both *Chd5*−/− lines maintain mESC characteristics, and because the number of DEGs is limited, expression levels of pluripotent markers (*Pou5f1*, *Nanog*, *Sox2*, and *Klf4*) and housekeeping genes (*Eef1a1*, *Actb*, and *Gapdh*) were compared by mRNA-Seq and RT-qPCR between WT and *Chd5*−/− lines (Fig. 3C). No significant expression changes were detected by either mRNA-Seq or RT-qPCR. Thus, *Chd5*-/- #1 and #2 maintain mESC characteristics.

Chd5 disruption in mESCs minimally affects the chromatin structure of gene loci

Next, to determine if Chd5 disruption changes the chromatin structure of gene loci, chromatin immunoprecipitation sequencing (ChIP-Seq) was performed to detect representative histone

modifications. H3K4me3 was used to represent the active chromatin state and H3K27me3 the inactive state. The histone H3 variants H3.1/H3.2 and H3.3 were used to examine histone exchange. Each ChIP-Seq signal peak around gene was calculated using MACS [Zhang et al., 2008], a ChIP-Seq peak-caller, and the distance of these peaks from the TSS determined, followed by clustering (Fig. 4A). We examined distribution changes in histone modifications and histone H3 variants between WT and *Chd5*−/−, and found that the distribution pattern was rarely changed around the TSS. The distribution tendencies of H3K4me3, H3K27me3, H3.1/H3.2, and H3.3 in WT were consistent with previous reports [Ray-Gallet et al., 2011; Won et al., 2015; Young et al., 2011]. To further examine the Chd5 distribution pattern dependency around the TSS of each histone modification and histone H3 variants, we plotted line charts of ChIP-Seq signals around TSS for each gene set, categorized into 10 groups by WT expression levels (expression levels were obtained from mRNA-Seq data; Fig. 4B). By comparing WT and *Chd5*-/- cells, distribution changes of each histone modification and histone H3 variant between WT and *Chd5*-/- were rarely observed around the TSS in each gene set. Higher expressed genes in WT were associated with increased H3K4me3 signal around the TSS in both WT and *Chd5*−/− cells (Fig. 4B, upper panels), and lower expressed genes in WT also showed correlation with reduced H3K27me3 signal around the TSS in both WT and *Chd5*-/- cells (Fig. 4B, 2nd panels). Histone H3.1/H3.2 and H3.3 signals showed similar pattern in as well (Fig. 4B, 3rd and lower panels). These findings suggests that the depletion of Chd5 does not affect in global histone modification

change for transcription.

2C-specific genes are upregulated in Chd5-/- mESCs

A previous report has demonstrated that two-cell (2C)-specific genes are upregulated in HP1α/β, KAP1, and G9a KO mESCs [Maksakova et al., 2013]. In *Chd5*-/- mESCs, 2C-specific genes were also detected as upregulated genes. Overlapping DEGs between *Chd5*-/- mESCs and these KO mESCs are shown (Fig. 5A, Table 2). *Zscan4a/d/e*, *Tdpoz3/4*, and *Usp17lc* (known as *Dub2a*) were detected as commonly upregulated genes in the mRNA-Seq data sets deposited in the Gene Expression Omnibus (GEO) (http://www.ncbi.nlm.nih.gov/geo/). These 2C genes were upregulated over 3-fold by KO of each factor compared with the DEGs identified in this study. 2C genes are specifically expressed in two-cell embryos, with approximately 1% of the mESC population expressing these genes [Tang et al., 2011]. In particular, Zscan4 plays a key role in maintenance of mESC proliferation [Amano et al., 2013; Falco et al., 2007; Zalzman et al., 2010]. Thus, we focused on *Zscan4* as a representative gene, and confirmed our FPKM results by RT-qPCR (Fig. 5B). FPKM bar graphs showed upregulation of all *Zscan4* transcripts in *Chd5*−/− mESCs. Even *Zscan4b/c/f*, which were not detected as DEGs, were upregulated at least over 4-fold. Upregulation of *Zscan4* by approximately 4-fold was confirmed by RT-qPCR. However, primers that individually distinguish all *Zscan4* paralogs were hard to design because of their high sequence similarity, therefore we referenced published

primers [Zalzman et al., 2010]. Expression levels of 2C genes, such as *Zscan4*, may be affected by Chd5.

Apart from 2C genes, other genes were also identified as DEGs, therefore we further screened the molecular pathway involved *Chd5* loss in mESCs. Ontology analysis of DEGs was performed using the Database for Annotation, Visualization and Integrated Discovery DAVID (https://david.ncicrf.gov/) and the Panther pathway (http://www.pantherdb.org/pathway/) (data not shown). No significant ontologies in either downregulated or upregulated genes were detected, except for Nodal signaling (including *Lefty1* and *Lefty2)*. However, *Lefty1* KO mice exhibit severe heart defects [Meno et al., 1998] and *Lefty2* KO mice are embryonic lethal [Meno et al., 1999], whereas *Chd5* KO mice have no defects except in spermatogenesis [Li et al., 2014; Zhuang et al., 2014], therefore Nodal signaling could not be the primary defect in *Chd5*-/- mESCs, and it was omitted from further analysis.

Upregulated MuERV-L/MERVL expression in Chd5-/- mESCs

Zscan4 expression is coupled to expression of MERVL, a long terminal repeat (LTR) retrotransposon derived from an endogenous retrovirus (ERV), with its transcript specifically expressed in two-cell embryos [Svoboda et al., 2004]. MERVL is transcribed as a chimeric RNA with neighboring genes (e.g., Zscan4), and can substitute its own LTR with a promoter to regulate

expression of those genes [Schoorlemmer et al., 2014]. A previous study also showed that HP1 α/β, KAP1, and G9a regulate 2C-genes indirectly via direct regulation of MuERV-L/MERVL [Maksakova et al., 2013]. Therefore, we next hypothesized that Chd5 functions to regulate retrotransposons such as MERVL, which cannot be detected as gene transcripts by regular mRNA-Seq. The default settings of Tophat2 [Kim et al., 2013] and Cufflinks are not appropriate for mapping reads onto retrotransposons because of the repetitive sequence, therefore for MERVL expression level analysis, the obtained mRNA-Seq fastq files were mapped to the whole genome using Bowtie2 and the retrotransposon reference sequence available in UCSC (https://genome.ucsc.edu/). Additionally, because the reference published in UCSC shares many identical sequences, the best option in Bowtie2 was used to avoid read abrogation due to excess multi-hit reads.

First, to evaluate if mRNA-Seq data mapped to MERVL, the data around MERVL near *Zscan4* were observed in IGV (Fig. 5C). An obvious increase in MERVL expression was detected in *Chd5*−/− mESCs. To examine expression of other retrotransposons, the ratio of all retrotransposons registered in UCSC was compared between WT and *Chd5*-/- lines (represented as a scatterplot in Fig. 6A). The top 50-most upregulated retrotransposons in *Chd5*-/- mESCs are shown, and of all these retrotransposons, MERVL was exceptionally upregulated. MERVL was upregulated about 35-fold in *Chd5*−/− mESCs (FPKM results are shown in Fig. 6B). RT-qPCR also revealed MERVL upregulation, yet although there is at most 2-fold upregulation, significance was detected ($n = 3$, $P = 0.0021$).

We further performed ChIP-Seq analysis for both WT and *Chd5*-/- to identify chromatin structural changes around MERVL. To examine repetitive elements such as retrotransposons, ChIP-Seq data was again mapped using the best option, as commonly performed [Elsasser et al., 2015]. The heat map shows clustered mRNA-Seq data and H3K4me3 data around/on MERVL (Fig. 6C). Green boxes indicate H3K4me3 signal in *Chd5*−/− was increased (in accordance with MERVL upregulation). Red boxes indicate in *Chd5*−/− was decreased (regardless of MERVL upregulation). No correlation between WT and *Chd5*−/− is shown in yellow boxes. Despite MERVL upregulation, there was no correlation in distribution changes of H3K4me3 signal between WT and *Chd5*−/−. To address the cause of this discrepancy, each ChIP-Seq data was visualized in IGV (Fig. 6D), which showed that the mappability around MERVL is almost 0. Namely, individual MERVL loci cannot be determined even with the best option, indicating that correspondence between reasonably expressed MERVL and the sequences around them is random, and causing this discrepancy. These findings suggest that MERVEL is significantly expressed but determination of individual MERVL loci is required to address MERVL regulation by *Chd5*.

Increased MERVL around H3.1/H3.2 and reduced MERVL around H3K27me3 in Chd5−/− mESCs

To quantify expression of individual MERVL, we distinguished multi-hit reads on MERVL by scoring overlapping reads on individual MERVL. We used eXpress

(http://bio.math.berkeley.edu/eXpress/) to detect commonly upregulated MERVL loci individually in two *Chd5*-/- mESCs. Each estimation by eXpress and Bowtie2 with the best option is shown (Fig. 7A). The X- and Y-axes represent MERVL expression in *Chd5*-/- #1 and #2, respectively. Estimation by eXpress shows high correlation between *Chd5*-/- #1 and #2, but few correlations with Bowtie2, indicating that eXpress can individually detect common upregulated MERVL between *Chd5*-/- #1 and #2, i.e., individual MERVL are detected.

For analysis, the top 10% highest expressed MERVL (over 2 kbp) were extracted from Fig. 7A, because highly upregulated MERVL in *Chd5*-/- mESCs are considered to be particularly regulated by Chd5, and MERVL of longer lengths are functional with transposable elements and/or LTRs. These MERVL were defined as long MERVL. Analysis 2 kbp upstream of long MERVL and other retrotransposons was performed because mappability was assured in this condition. To determine if Chd5 affects long MERVL expression specifically, first, we analyzed Chd5 accumulation around long MERVL, MERVL, and MMERVK10C (Fig. 7B). MMERVK10C is the same as MERVL, i.e., an ERV-derived retrotransposon with stable expression between WT and *Chd5*−/− mESCs. The representative locus for each retrotransposon is shown in IGV (Fig. 7B, left panel). Chd5 signals were abundant around long MERVL. Chd5 accumulation around long MERVL was significantly higher than MERVL, and MMERVK10C (Fig. 7B, right panel).

Next, to investigate changes in chromatin structure in *Chd5*-/- compared with WT mESCs, we re-

analyzed ChIP-Seq data based on the MERVL position determined in Fig. 7A. A representative long MERVL locus (shown in IGV) indicates H3.1/H3.2 accumulation increases in *Chd5*-/-, in accordance with Chd5 accumulation (Fig. 8A). To determine if the same pattern is observed at other loci, ChIP-Seq data of both WT and *Chd5*-/- were compared around long MERVL, MERVL, and MMERVK10C (Fig. 8B). To directly compare WT and *Chd5*-/- signals, ChIP-Seq data were normalized to each other, and signal fluctuations in the objective area (2 kbp upstream of each TSS) examined by Z-score. As expected, increased H3.1/H3.2 was observed, and additionally, H3K27me3 reduction also observed with long MERVL. These events only occurred around long MERVL, which were thought to have functional transposable elements and/or LTRs, and likely reflect *Chd5* function, as suggested below.

The suggested mechanism is summarized in Fig. 8C, taking together our data in this study and previous reports. Chd5 can remove H3.1/H3.2 from the chromatin structure, making the chromatin relatively H3.3-rich. In addition, H3K27me3 modifications will be preferentially added to histone H3.3, and then MERVL repressed.

Discussion

In this study, we showed MERVL, one of retrotransposons, was selectively induced, and increased H3.1/H3.2 and decreased H3K27me3 around MERVL was observed, in Chd5-/- mESCs. Conversely, Chd5 decreases H3.1/H3.2 and increases H3K27me3 around MERVL, suggesting that Chd5 regulates MERVL in mESCs. Removal of H3.1/H3.2 around MERVL by Chd5 in mESCs is consistent with previous reports; Chd5 removes nucleosomes during spermatogenesis [Li et al., 2014; Zhuang et al., 2014], and remodels nucleosomes by unwrapping [Quan and Yusufzai, 2014]. Moreover, Maksakova et al. (2013) demonstrated addition of H3K9me2/3 to chromatin around MERVL, which was followed by MERVL repression by HP1 α/β, KAP1, and G9a in mESCs [Maksakova et al., 2013]. Curiously, CAF1 knock-down in mESCs exhibits the similar phenotype as Chd5-/- mESCs, with regards 2C-genes and MERVL upregulation [Ishiuchi et al., 2015]. Additionally, CAF1 and HP1 form a heterodimer [Brasher et al., 2000], and both H3K27me3 and H3K9me3 are preferentially added to H3.3 [Banaszynski et al., 2013; Elsasser et al., 2015] and are essential for stable binding of HP1 to chromatin [Boros et al., 2014]. Functional interaction between these factors and Chd5 for MERVL regulation might be explained as follows: defective H3.1/H3.2 removal by Chd5 KO causes loss of H3.3 incorporation, H3K27me3 reduction, and loss of H3K9me3 marking by G9a and KAP1, with subsequent loss of heterochromatin formation by CAF1 and HP1 around MERVL, and MERVL upregulation. Thus, Chd5 may function as the most upstream factor in the MERVL repression cascade.

In contrast, Chd5 recognizes H3K27me3 through its chromodomain [Egan et al., 2013], and H3K4un through its PHD domain [Oliver et al., 2012; Paul et al., 2013]. Although H3K27me3 recognition appears to contradict H3K27me3 reduction in Chd5-/- mESCs, Chd5 may only recognize H3K27me3 and remove histones recognized by the PHD domain.

The means of which Chd5 plays a key role in tumor suppression is also suggested by our results. Retrotransposon upregulation is shown in various cancers [Criscione et al., 2014; Kassiotis, 2014]. Retrotransposons encode transposases and reverse transcriptases themselves, and can replicate into other loci [Dombroski et al., 1994]. Some retrotransposons, such as MERVL, have LTRs at both ends. Replication to other genes causes insertional inactivation of those genes and LTRs activate oncogenes by substituting with the original promoter [Howard et al., 2008]. Thus, Chd5 may function as a tumor suppressor through MERVL suppression.

Chd5 accumulation and changes in chromatin structure were only observed with long MERVL extracted from mRNA-Seq data and calculated by eXpress. The longer MERVL preferentially has MERVL_LTR/MT2_Mm, LTR of MERVL with promoter activity [Schoorlemmer et al., 2014]. Thus, Chd5 may remodel chromatin structure upstream of MERVL which is long enough and includes MERVL_LTR/MT2_Mm. Additionally, all MERVL transcripts will be detected either in ChIPqPCR or in RTqPCR because of its highly repetitive sequence. The regulation of Individual MERVL_LTR/MT2_Mm will be addressed its function in the future.

We also identified upregulated 2C-genes in Chd5-/- mESCs. mESCs that highly express 2C-genes and MERVL are in a 2C-like state, and suggested to be totipotent [Macfarlan et al., 2012]. However, loss of Chd5 function interrupts neural differentiation in mESC and SH-SH5Y [Egan et al., 2013], and it is possible that 2C-like state in mESCs induced by Chd5 KO may be in a more undifferentiated state than ordinary mESCs, but its differentiation potency might be lost because the undifferentiated state requires less heterochromatin. In contrast, heterochromatin is formed upon differentiation in a tissue-specific manner, and subsequently loss of Chd5 may cause loss of heterochromatin formation according to the above hypothesis. Thus, the 2C-like state induced in mESCs by Chd5 KO represents the undifferentiated state only. This could be applied to another tumorigenesis mechanism by Chd5 deficiency, specifically that loss of Chd5 transforms tumor cells into the undifferentiated state, similar to the 2C-like state in mESCs, and then causes malignant progression of tumors.

In summary, we have shown that Chd5 regulates MERVL, and also chromatin structural changes that are caused by Chd5, at least, around MERVL, to our knowledge, we have shown those for the first time.

Acknowledgments

We would like to thank T. Tachibana, Osaka City University for supplying antibodies, H. Kimura, Tokyo Institute of Technology for supplying antibodies, H. Niwa and D. Konno, RIKEN for supplying mESC (EB5), J. Katahira, Osaka University for discussion, and the Research for Information Technology, Kyushu University (tatara) and the National Institute of Genetics (NIG) for providing high-performance computing resources.

This work was supported by JSPS KAKENHI Grant Numbers 25116010, 26290064.

The authors have no conflict of interests to disclose.

References

Amano T, Hirata T, Falco G, Monti M, Sharova LV, Amano M, Sheer S, Hoang HG, Piao Y, Stagg CA, Yamamizu K, Akiyama T, Ko MS. 2013. Zscan4 restores the developmental potency of embryonic stem cells. Nat Commun 4:1966.

Anders S, Huber W. 2010. Differential expression analysis for sequence count data. Genome Biol 11.

Bagchi A, Papazoglu C, Wu Y, Capurso D, Brodt M, Francis D, Bredel M, Vogel H, Mills AA. 2007. CHD5 is a tumor suppressor at human 1p36. Cell 128:459-75.

Banaszynski LA, Wen D, Dewell S, Whitcomb SJ, Lin M, Diaz N, Elsasser SJ, Chapgier A, Goldberg AD, Canaani E, Rafii S, Zheng D, Allis CD. 2013. Hira-dependent histone H3.3 deposition facilitates PRC2 recruitment at developmental loci in ES cells. Cell 155:107-20.

Bao WD, Kojima KK, Kohany O. 2015. Repbase Update, a database of repetitive elements in eukaryotic genomes. Mobile DNA 6.

Boros J, Arnoult N, Stroobant V, Collet JF, Decottignies A. 2014. Polycomb repressive complex 2 and H3K27me3 cooperate with H3K9 methylation to maintain heterochromatin protein 1alpha at chromatin. Mol Cell Biol 34:3662-74.

Brasher SV, Smith BO, Fogh RH, Nietlispach D, Thiru A, Nielsen PR, Broadhurst RW, Ball LJ, Murzina NV, Laue ED. 2000. The structure of mouse HP1 suggests a unique mode of single peptide recognition by the shadow chromo domain dimer. EMBO J 19:1587-1597.

Carlson CM, Largaespada DA. 2005. Insertional mutagenesis in mice: new perspectives and tools. Nat Rev Genet 6:568-80.

Criscione SW, Zhang Y, Thompson W, Sedivy JM, Neretti N. 2014. Transcriptional landscape of repetitive elements in normal and cancer human cells. BMC Genomics 15:583.

de la Serna IL, Ohkawa Y, Imbalzano AN. 2006. Chromatin remodelling in mammalian differentiation: lessons from ATP-dependent remodellers. Nat Rev Genet 7:461-73.

Dombroski BA, Feng Q, Mathias SL, Sassaman DM, Scott AF, Kazazian HH, Boeke JD. 1994. An in vivo assay for the reverse transcriptase of human retrotransposon L1 in Saccharomyces cerevisiae. Mol Cell Biol 14:4485-4492.

Egan CM, Nyman U, Skotte J, Streubel G, Turner S, O'Connell DJ, Rraklli V, Dolan MJ, Chadderton N, Hansen K, Farrar GJ, Helin K, Holmberg J, Bracken AP. 2013. CHD5 is required for neurogenesis and has a dual role in facilitating gene expression and polycomb gene repression. Dev Cell 26:223-36.

Elsasser SJ, Noh KM, Diaz N, Allis CD, Banaszynski LA. 2015. Histone H3.3 is required for endogenous retroviral element silencing in embryonic stem cells. Nature 522:240-4.

Falco G, Lee SL, Stanghellini I, Bassey UC, Hamatani T, Ko MS. 2007. Zscan4: a novel gene expressed exclusively in late 2-cell embryos and embryonic stem cells. Dev Biol 307:539-50.

Fujii W, Kawasaki K, Sugiura K, Naito K. 2013. Efficient generation of large-scale genome-modified mice using gRNA and CAS9 endonuclease. Nucleic Acids Res 41:e187.

Fujita T, Igarashi J, Okawa ER, Gotoh T, Manne J, Kolla V, Kim J, Zhao H, Pawel BR, London WB, Maris JM, White PS, Brodeur GM. 2008. CHD5, a tumor suppressor gene deleted from 1p36.31 in neuroblastomas. J Natl Cancer Inst 100:940-9.

Gaspar-Maia A, Alajem A, Polesso F, Sridharan R, Mason MJ, Heidersbach A, Ramalho-Santos J, McManus MT, Plath K, Meshorer E, Ramalho-Santos M. 2009. Chd1 regulates open chromatin and pluripotency of embryonic stem cells. Nature 460:863-8.

Harada A, Okada S, Konno D, Odawara J, Yoshimi T, Yoshimura S, Kumamaru H, Saiwai H, Tsubota T, Kurumizaka H, Akashi K, Tachibana T, Imbalzano AN, Ohkawa Y. 2012. Chd2 interacts with H3.3 to determine myogenic cell fate. EMBO J 31:2994-3007.

Howard G, Eiges R, Gaudet F, Jaenisch R, Eden A. 2008. Activation and transposition of endogenous retroviral elements in hypomethylation induced tumors in mice. Oncogene 27:404-8.

Ishiuchi T, Enriquez-Gasca R, Mizutani E, Boskovic A, Ziegler-Birling C, Rodriguez-Terrones D, Wakayama T, Vaquerizas JM, Torres-Padilla ME. 2015. Early embryonic-like cells are induced by downregulating replication-dependent chromatin assembly. Nat Struct Mol Biol.

Kassiotis G. 2014. Endogenous retroviruses and the development of cancer. J Immunol 192:1343-9.

Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. 2013. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol 14:R36.

Kolla V, Zhuang T, Higashi M, Naraparaju K, Brodeur GM. 2014. Role of CHD5 in human cancers: 10

years later. Cancer Res 74:652-8.

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods 9:357-U54.

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.

Li W, Wu J, Kim SY, Zhao M, Hearn SA, Zhang MQ, Meistrich ML, Mills AA. 2014. Chd5 orchestrates chromatin remodelling during sperm development. Nat Commun 5:3812.

Macfarlan TS, Gifford WD, Driscoll S, Lettieri K, Rowe HM, Bonanomi D, Firth A, Singer O, Trono D, Pfaff SL. 2012. Embryonic stem cell potency fluctuates with endogenous retrovirus activity. Nature 487:57-63.

Maehara K, Ohkawa Y. 2015. agplus: a rapid and flexible tool for aggregation plots. Bioinformatics (Oxford, England).

Maksakova IA, Thompson PJ, Goyal P, Jones SJ, Singh PB, Karimi MM, Lorincz MC. 2013. Distinct roles of KAP1, HP1 and G9a/GLP in silencing of the two-cell-specific retrotransposon MERVL in mouse ES cells. Epigenetics Chromatin 6:15.

Mali P, Yang L, Esvelt KM, Aach J, Guell M, DiCarlo JE, Norville JE, Church GM. 2013. RNA-guided human genome engineering via Cas9. Science (New York, N.Y.) 339:823-826.

Marfella CG, Imbalzano AN. 2007. The Chd family of chromatin remodelers. Mutat Res 618:30-40.

Meno C, Gritsman K, Ohishi S, Ohfuji Y, Heckscher E, Mochida K, Shimono A, Kondoh H, Talbot WS,

Robertson EJ, Schier AF, Hamada H. 1999. Mouse Lefty2 and zebrafish antivin are feedback inhibitors of nodal signaling during vertebrate gastrulation. Mol Cell 4:287-298.

Meno C, Shimono A, Saijoh Y, Yashiro K, Mochida K, Ohishi S, Noji S, Kondoh H, Hamada H. 1998. lefty-1 is required for left-right determination as a regulator of lefty-2 and nodal. Cell 94:287-297.

Ogawa K, Matsui H, Ohtsuka S, Niwa H. 2004. A novel mechanism for regulating clonal propagation of mouse ES cells. Genes Cells 9:471-477.

Oliver SS, Musselman CA, Srinivasan R, Svaren JP, Kutateladze TG, Denu JM. 2012. Multivalent recognition of histone tails by the PHD fingers of CHD5. Biochemistry 51:6534-44.

Paul S, Kuo A, Schalch T, Vogel H, Joshua-Tor L, McCombie WR, Gozani O, Hammell M, Mills AA. 2013. Chd5 requires PHD-mediated histone 3 binding for tumor suppression. Cell Rep 3:92-102.

Quan J, Yusufzai T. 2014. The tumor suppressor chromodomain helicase DNA-binding protein 5 (CHD5) remodels nucleosomes by unwrapping. J Biol Chem 289:20717-26.

Ray-Gallet D, Woolfe A, Vassias I, Pellentz C, Lacoste N, Puri A, Schultz DC, Pchelintsev NA, Adams PD, Jansen LE, Almouzni G. 2011. Dynamics of histone H3 deposition in vivo reveal a nucleosome gap-filling mechanism for H3.3 to maintain chromatin integrity. Mol Cell 44:928-41.

Reisman D, Glaros S, Thompson EA. 2009. The SWI/SNF complex and cancer. Oncogene 28:1653-68.

Roberts A, Pachter L. 2013. Streaming fragment assignment for real-time analysis of sequencing experiments. Nat Methods 10:71-3.

Schoorlemmer J, Perez-Palacios R, Climent M, Guallar D, Muniesa P. 2014. Regulation of Mouse Retroelement MuERV-L/MERVL Expression by REX1 and Epigenetic Control of Stem Cell Potency. Front Oncol 4:14.

Strahl BD, Allis CD. 2000. The language of covalent histone modifications. Nature 403:41-45.

Svoboda P, Stein P, Anger M, Bernstein E, Hannon GJ, Schultz RM. 2004. RNAi and expression of retrotransposons MuERV-L and IAP in preimplantation mouse embryos. Dev Biol 269:276-85.

Tang F, Barbacioru C, Nordman E, Bao S, Lee C, Wang X, Tuch BB, Heard E, Lao K, Surani MA. 2011. Deterministic and stochastic allele specific gene expression in single mouse blastomeres. PLoS One 6.

Thompson PM, Gotoh T, Kok M, White PS, Brodeur GM. 2003. CHD5, a new member of the chromodomain gene family, is preferentially expressed in the nervous system. Oncogene 22:1002-1011.

Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L. 2010. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol 28:511-5.

Won KJ, Choi I, LeRoy G, Zee BM, Sidoli S, Gonzales-Cope M, Garcia BA. 2015. Proteogenomics analysis reveals specific genomic orientations of distal regulatory regions composed by non-canonical histone variants. Epigenetics Chromatin 8:13.

Young MD, Willson TA, Wakefield MJ, Trounson E, Hilton DJ, Blewitt ME, Oshlack A, Majewski IJ. 2011. ChIP-seq analysis reveals distinct H3K27me3 profiles that correlate with transcriptional activity. Nucleic

Acids Res 39:7415-7427.

Zalzman M, Falco G, Sharova LV, Nishiyama A, Thomas M, Lee SL, Stagg CA, Hoang HG, Yang HT, Indig FE, Wersto RP, Ko MS. 2010. Zscan4 regulates telomere elongation and genomic stability in ES cells. Nature 464:858-63.

Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, Liu XS. 2008. Model-based analysis of ChIP-Seq (MACS). Genome Biol 9:R137.

Zhuang T, Hess RA, Kolla V, Higashi M, Raabe TD, Brodeur GM. 2014. CHD5 is required for spermiogenesis and chromatin condensation. Mech Dev 131:35-46.

Figure Legends

Fig. 1 Complete Chd5 gene disruption in mESCs

A) Schematic of the *Chd5* disruption strategy. Target sequence 1 (indicated by gray triangle) is located upstream of Chd5 TSS (Chr4:151712446-151712469). Target sequence 2 (indicated by white triangle) is located downstream of Chd5 TES (Chr4:151764305-151764327). PCR regions for genotyping (see B) are also represented.

B) Genomic PCR genotyping. PCR regions are indicated in A. In PCR regions 1 and 2, amplicons are detected in WT only (amplicon sizes of 455bp and 461bp, respectively), with no genomic *Chd5* sequence in the other two samples. In PCR region 3, an amplicon is detected in the Chd5-/- cells only (amplicon size of approximately 600bp, which differs depending on the clone), showing that the DSB loci are connected. Observed bands and sizes are shown by arrows. The *Chd5* gene is completely and homozygously disrupted in both lines.

C) An amplicon from PCR region 3 was sequenced by the Sanger method, then connection between target sequence 1 and 2 were detected. In Chd5−/− #1, the DSB locus appears to share one nucleotide and connect directly, while in Chd5−/− #2, there is a two nucleotide insertion (CC) between each DSB locus.

Fig. 2 Complete loss of mRNA and protein in Chd5-/- mESCs

A) RT-qPCR shows reduced *Chd5* mRNA in Chd5−/− clones. Error bars indicate SD (*n* = 3). *Gapdh* was used an internal control.

B) Chd5 protein is not detected in Chd5−/− clones. Hsp90 was used as the loading control.

C) Complete loss of *Chd5* expression in Chd5−/− cells. mRNASeq signal was calculated by binning whole read-fragments mapped onto coding sequences. *Kcnb2* and *Nphp4* represent control signals.

Fig. 3 Significant changes of gene expression between WT and Chd5-/- mESCs were rarely detected.

A) FPKM scatterplots comparing WT and Chd5-/-. Significantly changed genes are represented by red dots. FPKM of each gene is indicated logarithmically. To avoid log0, 1 was added to FPKM. Radix = 2.

B) Venn diagram showed overlapping genes between Chd5-/- #1 and #2, with 15 downregulated and 33 upregulated genes in common. Overlapped genes are listed in Table 1.

C) FPKM and RT-qPCR confirmation of pluripotent marker genes and housekeeping genes. There were no significant changes in FPKM. Also in RT-qPCR, no significant changes were detected between WT and Chd5-/- #1. RT-qPCR was performed independently three times, with Gapdh used as an internal control. The traits of each mESC strain were maintained.

Fig. 4 Distribution pattern of histone H3 modification and histone H3 variants were rarely changed around genes.

A) ChIP-Seq signals of H3K4me3, H3K27me3, H3.1/H3.2, and H3.3 were clustered in a heat map based on TSS ± 5 kbp. Distribution patterns of each signal barely changed between WT and Chd5−/−.

B) Line charts showing ChIP-Seq signal accumulation surrounding TSS for each gene's expression level (every 10-percentile of gene expression levels in WT mRNA-Seq). Accumulation patterns of each gene's expression level barely changed.

Fig. 5 Two-cell specific genes and theirs proximal retrotransposons, MERVL, were increased in Chd5-/- mESCs.

A) Venn diagrams representing overlapping upregulated genes between Chd5-/- and HP1α-/-, HP1β-/-, KAP1-/-, and G9a-/- mESCs. 2C-specific genes such as Tdpoz3/4, Zscan4a/d/e, and Usp17lc were extracted. Overlapped genes are listed in Table 2.

B) Left panel: bar plots represent FPKM of Zscan4 paralogs, representative 2C-genes. All paralogs were upregulated, even those not detected as DEGs. Right panel: RT-qPCR confirmed Zscan4 upregulation. Primers amplifying all Zscan4 paralogs were obtained from those referenced (Zalzman, 2010). PCR was performed independently three times.

C) IGV screenshots showing MERVL expression neighboring Zscan4a/d in Chd5−/− mESCs. Chd5

(and not 2C-genes such as Zscan4) may directly regulate MERVL expression.

Fig. 6 MERVL expression was genome-widely increased in Chd5-/- mESCs.

A) Scatterplots comparing fold increase in retrotransposons between WT and Chd5-/- mESCs. MERVL-int and MERVL_LTR are exclusively upregulated compared with other retrotransposons.

B) FPKM of MERVL in each clone, with RT-qPCR confirmation of MERVL upregulation. MERVL was only but significantly upregulated 2-fold by RT-qPCR. PCR was performed independently three times.

C) mRNA-Seq and H3K4me3 ChIP-Seq signals around MERVL in WT and Chd5-/- #1 were clustered in a heat map. Red boxes indicate H3K4me reduction regardless of MERVL upregulation, green boxes show increased H3K4me3 in accordance with MERVL upregulation, while yellow boxes show no correlation between MERVL expression and H3K4me3 signal.

D) IGV screenshot that visualized mRNA-Seq and H3K4me3 ChIP-Seq data in WT and Chd5-/- #1 provides a reasons for the discrepancy in C): MERVL mappability is almost 0. No read can distinguish an individual MERVL. That is, actual expression of specific MERVL is invalidated. The mappability score track on mm9 genome, visualized in IGV, was downloaded from UCSC (ENCODE CrgMappability 40mer).

Fig. 7 Chd5 accumulation around long MERVL extracted by eXpress

A) MERVL expression was compared between Chd5-/- #1 and #2. Left panel: compared using eXpress, with positive correlation. Right panel: compared using Bowtie2 with best option, with no correlations. Thus, eXpress can distinguish individual MERVL.

B) Left panel: representative IGV screenshots showing Chd5 accumulation around each retrotransposon. Right panel: box plot represents Chd5 accumulation around each retrotransposon. Long MERVL were extracted using eXpress. MMERVK10C has stable expression levels between WT and Chd5-/-, and is represented as a control. * represents significance calculated by Welch's t-test.

Fig. 8 Histone H3.1/H3.2 accumulation and H3K27me3 modification around MERVL were changed in Chd5-/- mESCs.

A) Representative IGV screenshot indicates H3.1/H3.2 accumulation around MERVL (in accordance with Chd5), which was increased in Chd5−/− mESCs.

B) Accumulation of each ChIP-Seq signal around long MERVL was compared by Z-score. Not only increased H3.1/H3.2, reduced H3K27me3 was also revealed.

C) Schematic representation of Chd5 function. Chd5 removes H3.1/H3.2, making a relatively H3.3-rich chromatin structure. H3K27me3 is preferentially added to this relatively H3.3-rich chromatin, and MERVL is repressed.

Supplementary Fig. 1

A) Whole agarose gel shown in Fig. 1B. Right lanes indicate amplicons of internal sequence of *Gapdh* as control of PCR.

B) Whole blotted PVDF membranes of Western blotting shown in Fig. 2B. Left, Chd5; Right, Hsp90.

Supplementary Fig.1