TITLE:

# Inherent genomic properties underlie the epigenomic heterogeneity of human induced pluripotent stem cells

AUTHOR(S):

Yokobayashi, Shihori; Yabuta, Yukihiro; Nakagawa, Masato; Okita, Keisuke; Hu, Bo; Murase, Yusuke; Nakamura, Tomonori; ... Majewski, Jacek; Yamamoto, Takuya; Saitou, Mitinori

京都大学
KYOTO UNIVERSITY

京都大学学術情報リポジトリ
KURENAI 紅
Kyoto University Research Information Repository

**Article**

# Cell Reports

## Inherent genomic properties underlie the epigenomic heterogeneity of human induced pluripotent stem cells

### Graphical abstract



### Highlights

- Two distinct epigenomic states stably maintain X inactivation in hiPSCs

- X-chromosome-wide propagation of H3K9me3 can occur in *XIST*-downregulated hiPSCs

- Epigenomic heterogeneity on autosomes is enriched in non-genic/repressive regions

- Heterogenous H3K9me3 domains are associated with nuclear lamina in somatic cells
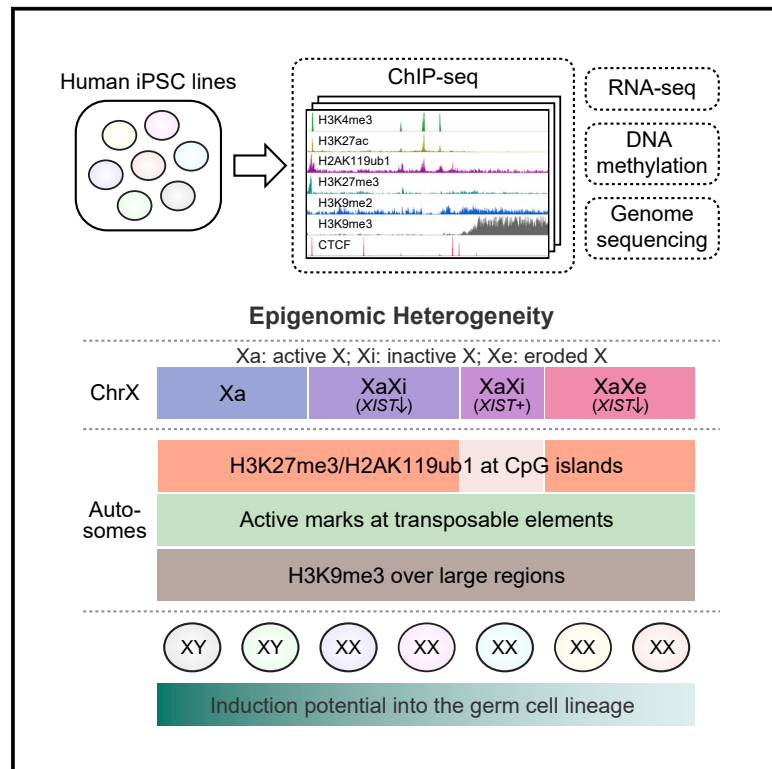
### Authors

Shihori Yokobayashi, Yukihiro Yabuta, Masato Nakagawa, ..., Jacek Majewski, Takuya Yamamoto, Mitinori Saitou

### Correspondence

yokobayashi@anat2.med.kyoto-u.ac.jp

### In brief

Yokobayashi et al. find that, in addition to imprinting and X-chromosome variations, CpG density, transposons, and H3K9me3-associated large domains contribute to the emergence or retention of epigenomic heterogeneity in hiPSCs. They identify differential epigenomic states that correlate with the *in vitro* germ cell differentiation potential of hiPSCs.

CellPress

# Cell Reports

CellPress
OPEN ACCESS

## Article

# Inherent genomic properties underlie the epigenomic heterogeneity of human induced pluripotent stem cells

Shihori Yokobayashi,[1,2,3,7,*] Yukihiro Yabuta,[2,3] Masato Nakagawa,[1] Keisuke Okita,[1] Bo Hu,[2,3,4] Yusuke Murase,[2,3] Tomonori Nakamura,[2,3] Guillaume Bourque,[3,4] Jacek Majewski,[4] Takuya Yamamoto,[1,3,5,6] and Mitinori Saitou[1,2,3]

[1]Center for iPS Cell Research and Application (CiRA), Kyoto University, 53 Kawahara-cho, Shogoin, Sakyo-ku, Kyoto 606-8507, Japan
[2]Department of Anatomy and Cell Biology, Graduate School of Medicine, Kyoto University, Yoshida-Knoe-cho, Sakyo-ku, Kyoto 606-8501, Japan
[3]Institute for the Advanced Study of Human Biology (ASHBi), Kyoto University, Yoshida-Knoe-cho, Sakyo-ku, Kyoto 606-8501, Japan
[4]Department of Human Genetics, McGill University, Montreal, QC H3A 1C7, Canada
[5]AMED-CREST, AMED, 1-7-1 Otemachi, Chiyoda-ku, Tokyo 100-0004, Japan
[6]Medical-risk Avoidance Based on iPS Cells Team, RIKEN Center for Advanced Intelligence Project (AIP), Kyoto 606-8507, Japan
[7]Lead contact
*Correspondence: yokobayashi@anat2.med.kyoto-u.ac.jp
https://doi.org/10.1016/j.celrep.2021.109909

## SUMMARY

Human induced pluripotent stem cells (hiPSCs) show variable differentiation potential due to their epigenomic heterogeneity, whose extent/attributes remain unclear, except for well-studied elements/chromosomes such as imprints and the X chromosomes. Here, we show that seven hiPSC lines with variable germline potential exhibit substantial epigenomic heterogeneity, despite their uniform transcriptomes. Nearly a quarter of autosomal regions bear potentially differential chromatin modifications, with promoters/CpG islands for H3K27me3/H2AK119ub1 and evolutionarily young retrotransposons for H3K4me3. We identify 145 large autosomal blocks ($\geq$ 100 kb) with differential H3K9me3 enrichment, many of which are lamina-associated domains (LADs) in somatic but not in embryonic stem cells. A majority of these epigenomic heterogeneities are independent of genetic variations. We identify an X chromosome state with chromosome-wide H3K9me3 that stably prevents X chromosome erosion. Importantly, the germline potential of female hiPSCs correlates with X chromosome inactivation. We propose that inherent genomic properties, including CpG density, transposons, and LADs, engender epigenomic heterogeneity in hiPSCs.

## INTRODUCTION

Pluripotent stem cells (PSCs), including embryonic stem cells (ESCs) and induced PSCs (iPSCs), have a robust capacity for self-renewal and differentiation into cells in three germ layers. Accordingly, human PSCs (hPSCs) (Takahashi et al., 2007; Thomson et al., 1998) are a valuable resource for investigating human development with *in vitro* reconstitution systems as well as for various applications aiming at medical interventions. Human iPSCs (hiPSCs) are especially advantageous, as they are generated through the reprogramming of differentiated somatic cells and thus bypass the ethical issues caused by the use of human embryos. On the other hand, hPSC lines show heterogeneity in their differentiation potentials (Yamanaka, 2020). Previous studies have reported that such heterogeneity is linked to the differences in their genetic background (Choi et al., 2015; Kajiwara et al., 2012; Ortmann et al., 2020) and differences in their epigenomic states including DNA methylation states (Bock et al., 2011; Koyanagi-Aoi et al., 2013; Nishizawa

et al., 2016); however, understanding of the molecular basis for such heterogeneity is still limited.

Epigenetic heterogeneity (we use the term "epigenetic" when referring to specific regions/chromosomes) associated with genomic imprinting and X chromosome inactivation (XCI) has been described in both human ESCs (hESCs) and hiPSCs. Genomic imprinting regulates the parent-of-origin-specific expression of particular genes ("imprinted" genes) by the differential DNA methylation of their regulatory elements between parental alleles; loss of parental-specific DNA methylation has been observed in a subset of imprinted regions in hPSCs (Bar et al., 2017; Adewumi et al., 2007; Ma et al., 2014; Nazor et al., 2012). XCI compensates for the dosage imbalances between the sexes by silencing one X chromosome in females. XCI is established in post-implantation embryos in eutherian mammals and maintained throughout life; the association of non-coding RNA *XIST* (X-inactive specific transcript) and repressive histone modifications on the inactive X chromosome (Xi) plays a crucial role in the establishment/maintenance of the XCI (Galupa and Heard, 2018). Notably, hPSCs often exhibit the downregulation
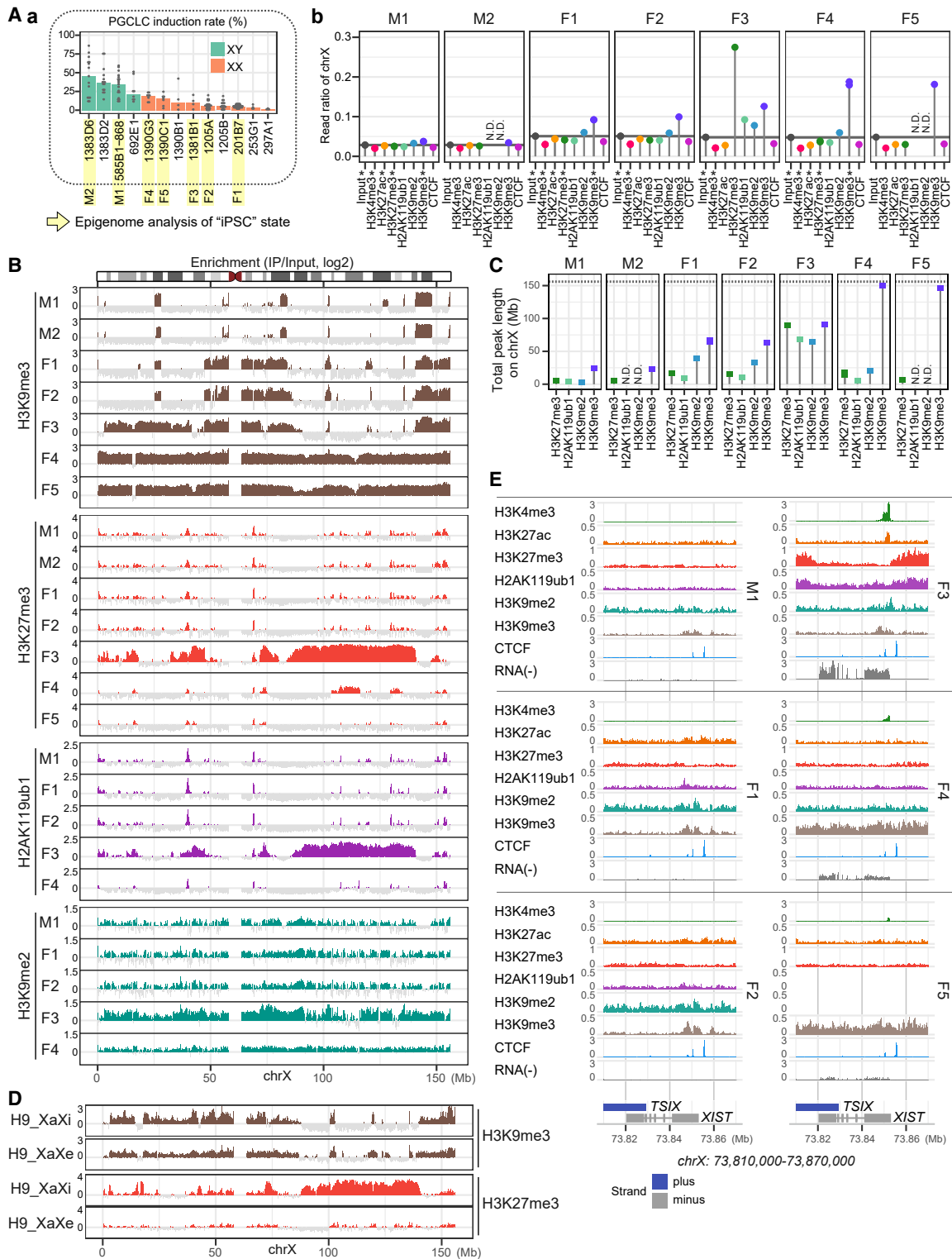
**Figure 1. Differential accumulation of repressive histone marks on the X chromosomes in female hiPSC lines**

(A) (a) Schematic of the concept of this study. Seven hiPSC lines were chosen for detailed epigenome analysis based on Yokobayashi et al. (2017). 585B1, the parental line of 585B1-868, was used as M1 in this study. (b) The ratio of the ChIP-seq reads of chromosome X, to the total reads. Horizontal lines indicate the ratio of input samples. N.D., not examined; *, samples with biological replicates (listed in Table S3), and these replicates are overlaid.

(B) Distribution of the indicated histone marks on chromosome X in the hiPSC lines. The y axis represents log₂-transformed enrichment values (IP/input).

*(legend continued on next page)*

# Cell Reports
## Article

of *XIST* RNA, the loss of the Xi-accumulated repressive histone marks, and the stochastic reactivation of X-linked genes with DNA de-methylation at their promoter regions. This phenomenon is called "the erosion of XCI" and progresses during cell passages (Bansal et al., 2021; Mekhoubad et al., 2012; Nazor et al., 2012; Patel et al., 2017; Vallot et al., 2015).

It has been reported that in human H9 ESCs, the Xi has a repressive chromatin organization partitioned into two types of domains, as follows: the one enriched by trimethylation of histone H3 lysine 27 (H3K27me3) and the other enriched by trimethylation of histone H3 lysine 9 (H3K9me3). When the erosion of XCI progresses, the eroded X chromosome (Xe) in H9 ESCs loses the H3K27me3-enriched domains, coupled with the loss of *XIST* expression, but maintains the H3K9me3 enrichment as observed on the Xi (Vallot et al., 2015). The extent of Xe appears to vary among hPSCs (Bansal et al., 2021; Patel et al., 2017), but a more comprehensive characterization of the Xe states, including histone modification states, remains to be performed. It has been suggested that two active X chromosomes (Xas) or progressed Xe interfere with cell differentiation in mouse PSCs and hPSCs (Patel et al., 2017; Salomonis et al., 2016; Schulz et al., 2014).

Comparative analyses between hiPSCs and hESCs have revealed that, although the genome-wide profiles of gene expression and DNA methylation are globally similar (Bock et al., 2011; Guenther et al., 2010; Lister et al., 2011), hiPSCs bear aberrant DNA methylation. Some of these differentially methylated regions (DMRs) originate from the donor cells ("somatic memory"), whereas others are independent of the donor cells, which have presumably arisen during the reprogramming process ("reprogramming-associated") (Lister et al., 2011; Ma et al., 2014; Ohi et al., 2011; Roost et al., 2017; Ruiz et al., 2012). In particular, whole-genome bisulfite sequencing analysis has revealed non-CG hypomethylation and significant enrichment of H3K9me3 in megabase-scale regions (Lister et al., 2011). This information indicates the existence of epigenomic variations in rather broad regions of the genome in hiPSCs. Therefore, it is important to carefully assess line-to-line variations in a genome-wide manner, including multiple epigenetic signatures in addition to DNA methylation.

We have previously shown clonal variations in the efficiency of induction of hiPSCs into human primordial germ-cell-like cells (hPGCLCs) (Yokobayashi et al., 2017). In the *in vitro* system used in that previous report, hiPSCs are first differentiated into incipient mesoderm-like cells (iMeLCs) by stimulation with activin and WNT signals and then are differentiated into hPGCLCs under a culture condition including BMP4 and other cytokines. The hPGCLCs show gene expression properties similar to those of *in vivo* hPGCs, which are the first precursors of the germ cell lineage specified during the second week after fertilization (Sasaki et al., 2015; Yokobayashi and Saitou, 2018). We have observed that the differences in the hPGCLC induction effi-

ciencies are reflected in the gene expression changes in the iMeLC state (Yokobayashi et al., 2017), suggesting that clonal variations exist in the responsivity to the differentiation signals. In the present study, we set out to understand the extent of the clonal epigenomic variations among the hiPSC lines by focusing on six histone modifications and a chromatin binding protein, the CCCTC-binding factor (CTCF). In addition, we performed an association analysis between the epigenomic variations and the efficiency of PGCLC induction to understand the possible impacts of epigenomic variations on gene regulations during the human germ cell differentiation process.

## RESULTS

### Quantitative comparison of epigenomic states among hiPSC lines

To understand the molecular basis of clone-to-clone differences in the differentiation potentials of hiPSC lines, we chose seven hiPSC lines as follows based on the previous study (Yokobayashi et al., 2017; Figure 1A; Table S1): two were males (585B1 and 1383D6, termed M1 and M2, respectively) and five were females (201B7, 1205A, 1381B1, 1390G3, and 1390C1, termed F1, F2, F3, F4, and F5, respectively). Six lines originated from the peripheral blood mononuclear cells (PBMCs) of four different donors (M1, M2, F2/F3, and F4/F5), and one originated from human dermal fibroblast cells (HDFs) (F1). These lines were generated through retroviral transduction under the SNL-feeder condition (F1) (Takahashi et al., 2007), through the transduction of non-integrating episomal plasmids under the mouse embryonic feeder (MEF) or SNL-feeder condition (M1 or F2, respectively) (Okita et al., 2013), and through the transduction of episomal plasmids under the feeder-free condition with a defined medium (M2, F3, F4, and F5) (Nakagawa et al., 2014). All the lines were maintained under the defined feeder-free condition (Nakagawa et al., 2014). They showed very similar gene expression profiles (Pearson's R, $\geq 0.97$) and proliferation rates (Figures S1A and S1B; Table S2).

We assessed the epigenomic states of these hiPSC lines by performing chromatin immunoprecipitation followed by sequencing (ChIP-seq) analysis. The ChIP-seq libraries for trimethylation of histone H3 lysine 4 (H3K4me3; promoter), acetylation of histone H3 lysine 27 (H3K27ac; enhancer), H3K27me3 (polycomb repressive complex 2 [PRC2]-mediated transcriptional repression), H3K9me3 (heterochromatin), and CTCF (insulator activity and/or the formation of higher-order chromatin structure) were prepared in all seven hiPSC lines, and those for dimethylation of histone H3 lysine 9 (H3K9me2; transcriptional repression) and monoubiquitylation of histone H2A lysine 119 (H2AK119ub1: PRC1-mediated transcriptional repression) were prepared in five hiPSC lines (M1, F1, F2, F3, and F4) (Table S3). The percentage of mapped reads, the distribution of GC content in the sequenced reads, and the score of the fraction of reads in peaks

(C) Total peak length of the indicated histone marks on chromosome X called by epic2 software. Horizontal dotted lines indicate the length of chromosome X.
(D) Distribution of H3K9me3 and H3K27me3 on chromosome X in H9 ESCs, reanalyzed from Vallot et al. (2015), as in (B) (Table S5).
(E) Distribution of the indicated epigenetic marks on the *XIST* locus. The y axis represents $\log_2$-transformed IP reads or stranded RNA sequencing (RNA-seq) reads normalized by total reads.
See also Figures S1 and S2.

京都大学
KYOTO UNIVERSITY

京都大学学術情報リポジトリ
KURENAI 紅
Kyoto University Research Information Repository

CellPress
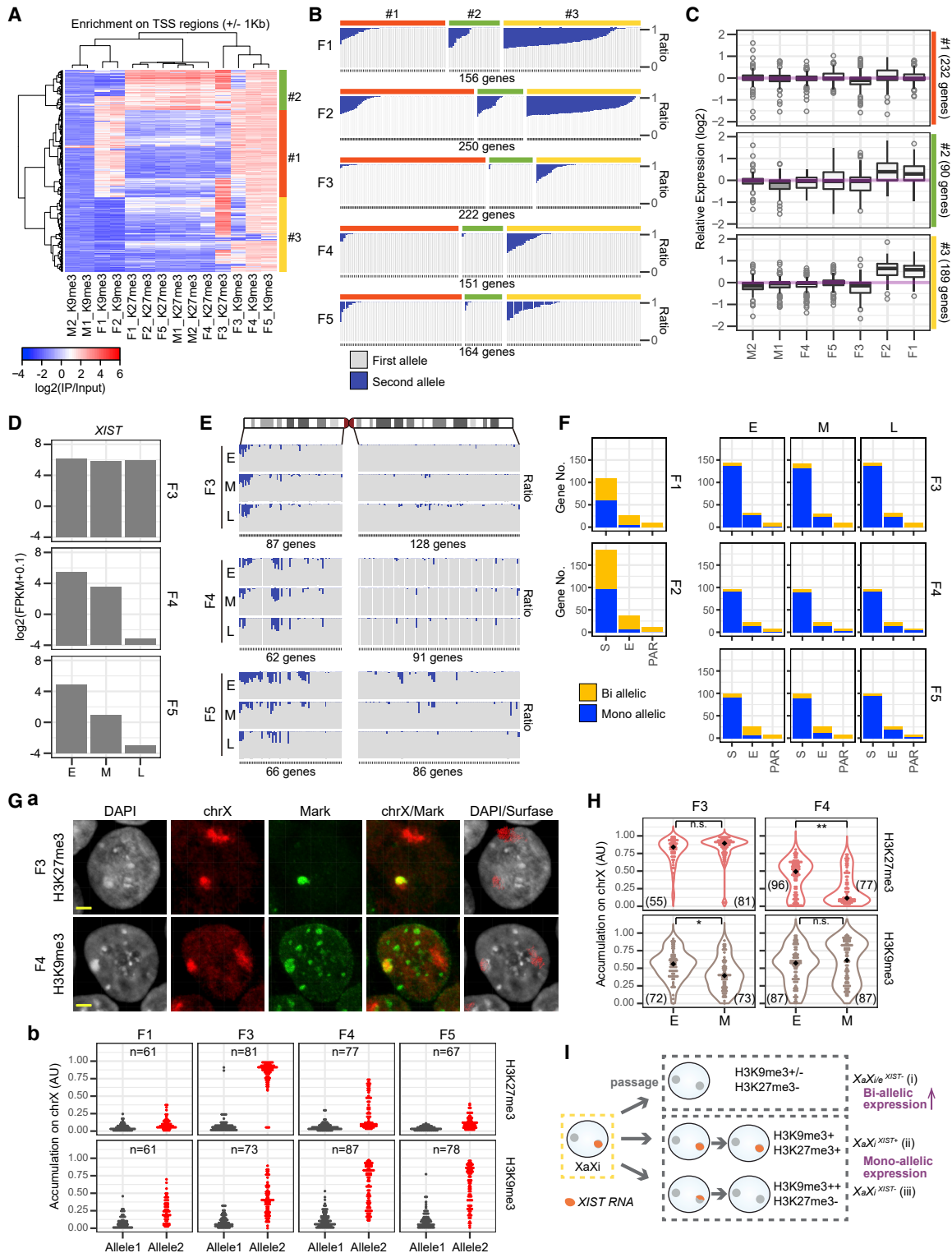OPEN ACCESS

Cell Reports
Article

**Figure 2. A dominant role of H3K9me3 in XCI in some female hiPSC lines**

(A) UHC and heatmap analyses of the enrichment of H3K9me3 or H3K27me3 at the TSSs of expressing X-linked genes.

(B) Allelic expression of X-linked genes. The y axis represents the ratio of two alleles, and each vertical bar in the x axis represents a gene that has informative SNPs. Genes were grouped by TSS classes as shown in (A) and sorted based on the frequency of "bi-allelic" expression.

(C) Boxplots of the gene expression values relative to the median of the seven hiPSC lines, grouped by TSS classes as shown in (A).

*(legend continued on next page)*

# Cell Reports
## Article

(FRiP) were highly consistent among samples (Figures S1C and S1D). Moreover, the biological replicates exhibited high reproducibility (Figure S1E). Together, these ChIP-seq data were of sufficient quality for use in a detailed comparative analysis.

### A dominant role of H3K9me3 in XCI in some female hiPSC lines

When examining the read distribution per chromosome, we first noticed that the read occupancy of X chromosomes was strikingly different among the hiPSC clones. Specifically, in males, the ratio of X chromosome reads was comparable to that of input samples in all ChIP samples, whereas this ratio varied greatly in females, especially for repressive histone marks, such as H3K27me3, H3K9me3, H3K9me2, and H2AK119ub1 (Figure 1A; Table S3). We then examined the distribution of these histone marks on the X chromosomes (Figures 1B and 1C). In males, only a small part of the X chromosome was enriched by H3K9me3, whereas H3K27me3 or H2AK119ub1 was present as small peaks, representing the state of the Xa. In contrast, in females, the enrichment of H3K9me3 was extended to proximal to the centromeres and the distal part of the long arm in F1 and F2 cells (~65 Mb), and it further spread to most of the short arm in F3 cells (~90 Mb) and, remarkably, to the entire chromosomes in F4 and F5 cells (~150 Mb). On the other hand, H3K27me3 and H2AK119ub1 were propagated to broad regions in F3 cells, which was not the case in the other female lines. F3 cells exhibited higher enrichment of H3K9me2 than the other female clones. The distribution patterns of H3K9me3 and H3K27me3 in F3 cells resembled those in the XaXi populations of H9 ESCs described previously (Vallot et al., 2015; Figure 1D; Figure S2A). In contrast, those patterns in the other female cell lines were distinct from the XaXi and also the XaXe populations of H9 ESCs (Vallot et al., 2015; Figures 1B and 1D; Figure S2A). Importantly, XIST was expressed in F3 cells, and active histone marks were associated with the XIST locus (Figure 1E). In contrast, XIST expression and H3K4me3 accumulation were lower in F4 and F5 cells and were barely detected in F1 and F2 cells or in the male hiPSCs. These findings demonstrate that the X chromosomes in the female hiPSC lines have epigenetic landscapes with multiple distinct modes.

We next investigated the effect of the distinct epigenetic states on the expression of the X-linked genes. First, we examined the enrichment of H3K9me3 and H3K27me3 around the transcription start sites (TSSs) of the X-linked genes and classified them into 3 classes by unsupervised hierarchical clustering (UHC). Class 1 contained TSSs associated with repressive histone marks (H3K9me3 or H3K27me3) in all female but not in male hiPSC lines ("repressive in females"); class 2 contained TSSs associated with repressive histone marks in both female and male lines ("repressive in all"); and class 3 contained TSSs associated with repressive histone marks in F3, F4, and F5 cells but not in F1, F2, or the male lines ("variably repressive in females") (Figure 2A). We determined the allelic expression based on single-nucleotide polymorphisms (SNPs) defined by whole-genome sequencing analysis (see STAR Methods) and observed an increase in the number of bi-allelically expressed genes in F1 and F2 cells in all three classes, but particularly in class 3 (Figure 2B). In agreement with this result, in F1 and F2 cells, the X-linked gene expression levels were elevated most remarkably in class 3, indicating that the absence of repressive marks, i.e., erosion of XCI, leads to the increase in gene dosages in F1 and F2 cells (Figure 2C). On the other hand, F3, F4, and F5 cells exhibited mono-allelic expression in the majority of X-linked genes. This finding suggests the possibility that both of the differential epigenetic states—i.e., the H3K27me3/H3K9me3-partitioned state (F3) and the H3K9me3-propagated state (F4/F5)—maintain XCI in a comparable manner (Figures 2A, 2B, and 2C).

It has been reported that X chromosome erosion progresses during the passages of hPSCs (Mekhoubad et al., 2012; Nazor et al., 2012). To verify this point, we analyzed the gene expression status in early- and late-passage cells of F3, F4, and F5 clones. XIST expression was maintained at a high level in F3 cells, whereas it was downregulated in F4 and F5 cells during the passages (Figure 2D). However, in F4 and F5 cells, the disappearance of XIST RNA was not accompanied by gene reactivation; rather, the number of mono-allelically expressed genes was slightly increased (Figure 2E), indicating that the erosion of XCI did not occur in these cells. We further collated these data with the data of Balaton et al. (2015), which aggregated the inactivation statuses of X-linked genes in several human studies (Figure 2F; Balaton et al., 2015). In F1 and F2 cells, about one-half of the genes reported as subject to XCI ("S") were converted to a bi-allelically expressed state. In contrast, a large fraction of genes reported as escaped from XCI ("E") were in a mono-allelically expressed state in F3 and F4 cells, and conversion from the bi- to the mono-allelic expression during the passages was observed in F5 cells. These data indicate that the Xi state is stably maintained during the passages in the F3, F4, and F5 hiPSC lines, regardless of the presence or absence of XIST RNA.

By using X chromosome painting combined with immunodetection, we verified the allelic accumulation of H3K27me3 or H3K9me3 in the female hiPSC lines (Figure 2G; Figures S2C

---

(D) XIST expression dynamics in early (E), middle (M), and late (L) passage of the hiPSC lines. "M" samples correspond to the samples shown in (B) and (C).

(E) Allelic expression of X-linked genes as in (B). Genes expressed through early (E) to late (L) passage cells were sorted by their chromosome positions.

(F) Classification of X-linked genes according to the consensus inactivation status (Balaton et al., 2015). S, subject to XCI; E, escape from XCI; PAR, pseudo-autosomal region. S or E includes the S/mostly S, or E/mostly E/variably E/mostly VE categories in Balaton et al. (2015), respectively.

(G) (a) Immunofluorescence (IF) of H3K27me3 or H3K9me3 (green), combined with DNA-fluorescence in situ hybridization (FISH) for chromosome X (red) and DAPI staining (gray), on middle (M) passage cells. Surfaces: 3D-area presentation of the DNA-FISH signals generated by Imaris 9.0 software. Representative nuclei for each mark are shown. Scale bar, 3 μm. (b) Dot plots for the accumulation levels of H3K27me3 or H3K9me3 signals on chromosome X. The chromosome surface showing higher accumulation of the mark within a nucleus was assigned to "allele 2." AU, arbitrary unit. Numbers of analyzed nuclei are shown.

(H) Violin plots for the H3K27me3 or H3K9me3 accumulation on the allele 2 as in (G), in E and M passage cells. Numbers of analyzed nuclei are shown in parentheses. *p < 1 × 10$^{-4}$, **p < 1 × 10$^{-5}$; n.s., not significant (Welch two sample t test).

(I) Schematic of X chromosome states in female hiPSCs. The "Xi," "Xe," or "Xa" was used based on the allelic expression status of X-linked genes.
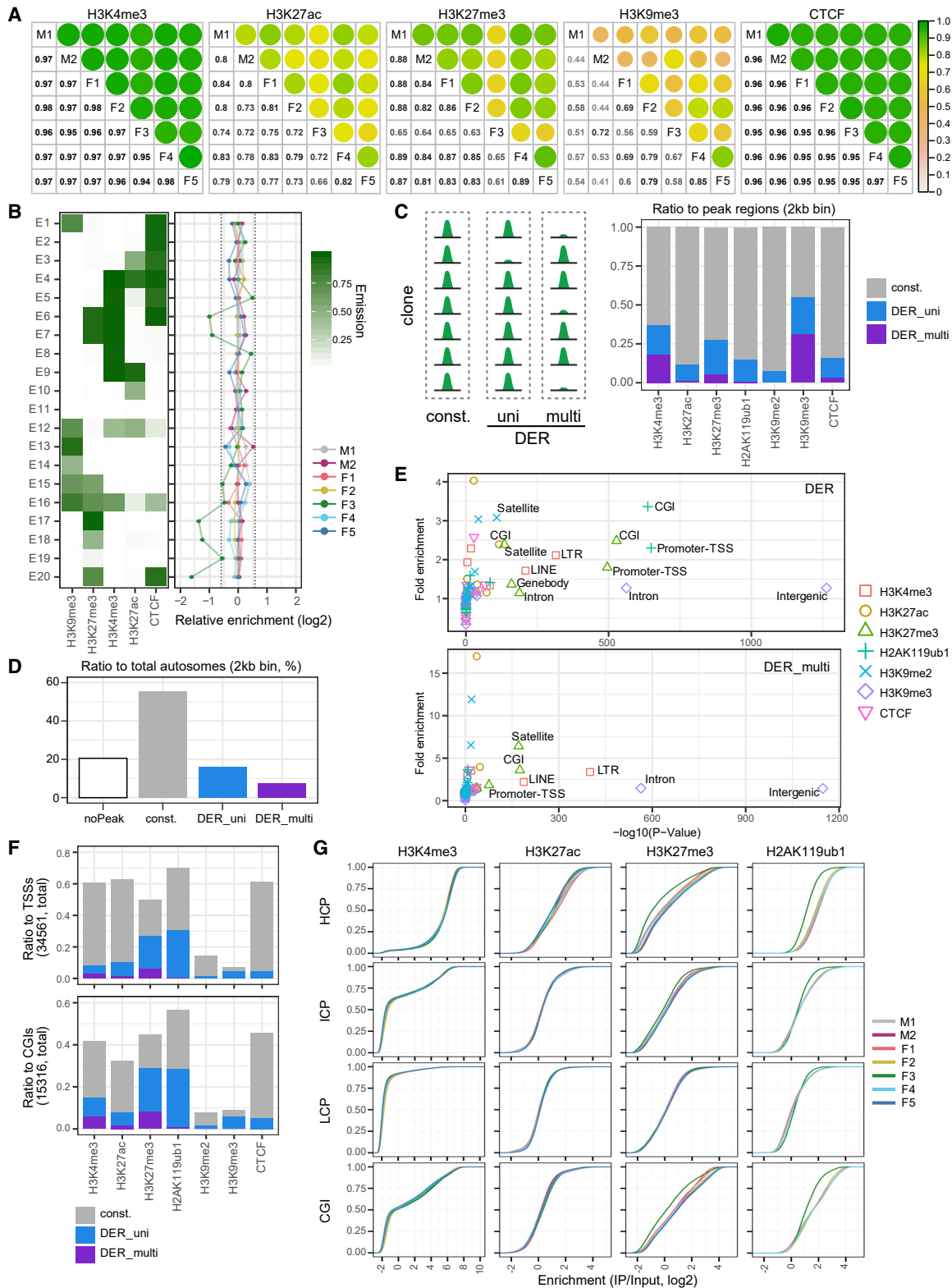
See also Figure S2.

CellPress
OPEN ACCESS

Cell Reports
Article



**Figure 3. Epigenetic heterogeneity in the autosomes of hiPSC lines**

(A) Correlation coefficient (Pearson's R) matrices of the $\log_2$-transformed enrichment values (IP/input, autosomal 2-kb bins) of the indicated epigenetic marks. Reads mapped to the autosomes were used as total reads for normalization.

*(legend continued on next page)*

and S2D). Moreover, we found that the extent of H3K27me3 accumulation in F4 cells decreased during the passages, coupled with the downregulation of *XIST* (Figures 2D and 2H). Collectively, our data revealed the presence of multiple distinct epigenetic states of the X chromosome in female hiPSCs, namely, (1) XaXi$_e$$^{XIST-}$, (2) XaXi$^{XIST+}$, or (iii) XaXi$^{XIST-}$ (Figure 2I), with (iii) XaXi$^{XIST-}$ as a unique state that maintains XCI through chromosome-wide propagation of H3K9me3 in the absence of *XIST* (Figures S2A and S2B).

## Epigenetic heterogeneity in the autosomes of hiPSC lines

We next examined epigenetic heterogeneity in the autosomes among these hiPSC lines. To minimize the influences of their variations in the X chromosomes (Figures 1 and 2), autosomal reads (chromosomes 1 to 22) were used as total reads for normalization. The comparison of enrichment values (IP/input) among the hiPSC clones exhibited the highest correlation in H3K4me3 and CTCF samples (Pearson's R, ≥0.94 and ≥0.95, respectively) and the lowest correlation in H3K9me3 samples (R, ≥0.41) (Figure 3A; Figure S3A). We assessed the genome-wide association of these marks along the autosomes by using a 20-state ChromHMM model (Ernst and Kellis, 2017). Although the enrichment of regions annotated to each state was globally similar among the hiPSC lines with less than 1.5-fold differences, an underrepresentation of the H3K27me3-associated states was noticed in F3 cells (Figure 3B). Next, we identified differentially enriched regions (DERs) for each epigenetic mark, based on the enrichment values per 2-kb bin overlapped with its merged peak regions (≥2-fold difference in ≥1 pair of clones) (Figure 3C; see STAR Methods). Additionally, DERs were marked as "multi" if the variations were observed in more than one clone; otherwise they were marked as "uni" (Figure 3C). The ratio of DERs and multi-DERs was the highest in H3K9me3 peaks (54.9% and 31.5%, respectively) (Figure 3C). In total, the sum of defined DERs of seven epigenetic marks reached 23.9% of the autosomal genome, and one-third of them were assigned as multi-DERs (Figure 3D). Interestingly, the ratio of DERs was the second-highest in H3K4me3 peaks (Figure 3C); most of the DERs were observed among the bins with lower H3K4me3 levels, whereas the bins showing higher enrichment levels were mostly "constant" among samples, contributing to the high correlations of H3K4me3 in the entire bin set (Figure S3B; Figure 3A). In contrast, DERs were biased toward higher enrich-

ment levels in H3K9me3 peaks (Figure S3B). The inspection of genomic features associated with the DERs by use of the HOMER tool (Heinz et al., 2010) revealed that the promoter regions or CpG islands (CGIs) were overrepresented in DERs of the H3K27me3 or H2AK119ub1 peaks, whereas intronic/intergenic regions or long terminal repeats (LTRs)/long interspersed nuclear elements (LINEs) were overrepresented in the H3K9me3 or H3K4me3 DERs, respectively (Figure 3E; Figure S3C). On the other hand, CTCF DERs were found to constitute only a small fraction, and the enrichment of particular genomic features was not prominent among the DERs (Figures 3C and 3E), indicating that the distribution of CTCF is highly conserved among the hiPSC lines.

We further analyzed the regions around TSSs and non-promoter CGIs by using the same criteria and found DERs in a large proportion of the regions carrying H3K27me3 (9,298 TSSs; 4,452 CGIs) or H2AK119ub1 (10,607 TSSs; 4,375 CGIs) (Figure 3F). Indeed, a decrease in the enrichment levels of H3K27me3 or H2AK119ub1 was observed in F3 cells with respect to TSSs of high-CpG density promoters (HCPs) as well as non-promoter CGIs (Figure 3G; Figures S3D and S3E), which is in agreement with the results of the ChromHMM analysis (Figure 3B). Of note, a decrease in the enrichment of H2AK119ub1 in F3 cells was also observed in the HCPs that showed little or no occupancy of H3K27me3 (Figure S3E). A majority of these regions were also enriched with H3K4me3, H3K27ac, or CTCF; the enrichment levels of those active marks were similar in all hiPSC lines, with little effect on the expression levels of associated genes (Figure 3G; Figures S3D–S3F). Collectively, these data indicate that the enrichment of PRC-mediated epigenetic marks at CpG-dense regions is compromised in F3 cells. Differences in such a "bivalent state" may affect the developmental potential of the hiPSC lines (Voigt et al., 2013).

## Variable H3K4me3 enrichment at evolutionarily recent transposable elements (TEs)

In contrast to the high consistency in H3K4me3 enrichment at the gene-coding regions (Figures 3F and 3G; Figures S3C–S3E), H3K4me3 peaks associated with LTRs or LINEs exhibited a large fraction of DERs (LTRs, 78.5%; LINEs, 63.8%), and multi-DERs (LTRs, 60.5%; LINEs, 39.6%) (Figure S3C). A comparison of the enrichment levels of 1-kb regions around H3K4me3-peak summits also revealed low correlation coefficients and high variances in LTR- or LINE-associated peaks (Figure 4A), confirming

(B) ChromHMM analysis using a 20-state model on the basis of the enrichment of indicated 5 epigenetic marks in the 7 hiPSC lines. A higher emission value represents a greater probability of observation (left). The enrichment of regions annotated to each state in each cell line, relative to the median of the seven hiPSC lines is shown (right). Vertical dotted lines, 1.5-fold differences from the median (0). The values of replicate 1 were used (Table S3).

(C) Schematic diagram of differentially enriched regions (DERs) (see the Results and STAR Methods). The right panel shows bar plots of the proportion of autosomal 2-kb bins assigned to constant regions, uni-DERs, and multi-DERs in each epigenetic mark.

(D) Bar plots of the proportion of autosomal 2-kb bins as in (C) in at least one epigenetic mark. The 2-kb bins that were not overlapped with any of the examined marks are labeled as "noPeak."

(E) 2D plots for the enrichment of genomic features in DERs (top) or multi-DERs (bottom) by using the annotation of the HOMER tool. p values (hypergeometric test) and fold enrichment values are shown on the x and y axis, respectively. The annotations "5′UTR," "3′UTR," "exon," and "TTS" were combined as "Genebody."

(F) Bar plots of the proportion of TSSs (top) or non-promoter CGIs (bottom) as in (C).

(G) CDFs of enrichment values of the indicated epigenetic marks at the TSSs or non-promoter CGIs. TSSs were classified as high, intermediate, or low CpG promoters (HCP, ICP, and LCP, respectively).
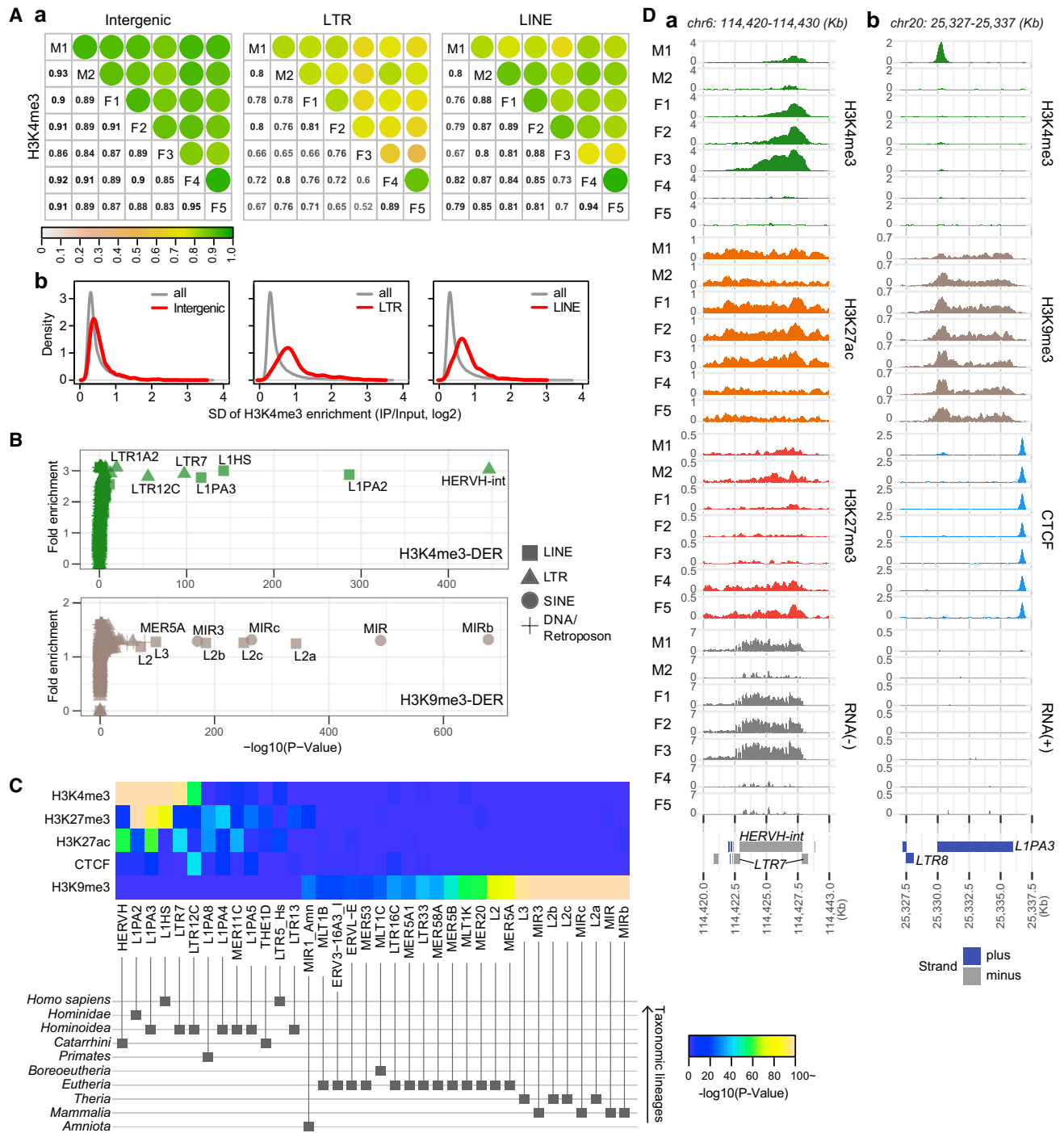
See also Figure S3.

京都大学
KYOTO UNIVERSITY

京都大学学術情報リポジトリ
KURENAI 紅
Kyoto University Research Information Repository

CellPress
OPEN ACCESS

Cell Reports
Article

**Figure 4. Variable H3K4me3 enrichment at evolutionarily recent TEs**

(A) (a) Correlation coefficient (Pearson's R) matrices of the $\log_2$-transformed enrichment values (IP/input) of H3K4me3 at the peak summits (±500 bp) detected in ≥1 sample, classified by the HOMER tool annotation. (b) Density curves of the standard deviation (SD) values of H3K4me3 enrichment as in (a) (red), compared to whole peaks in the autosomes (gray).

(B) 2D plots for the enrichment of TE subfamilies in DERs of H3K4me3 (top) or H3K9me3 (bottom) by using RepeatMasker annotation. p values (hypergeometric test) and fold enrichment values are shown in the x and y axis, respectively.

*(legend continued on next page)*

that the H3K4me3 enrichment at the repeat sequences is variable among these hiPSC lines.

We next investigated which subfamily of TEs was enriched in DERs by using RepeatMasker annotation (http://genome.ucsc.edu/). Human endogenous retrovirus subfamily H (HERVH), long terminal repeat 7 (LTR7), and LTR12C, as well as human- or primate-specific LINE-1 elements (L1HS or L1PA2/L1PA3, respectively), were significantly enriched in H3K4me3 DERs ($p < 1 \times 10^{-50}$) (Figures 4B and 4C). On the other hand, ancient TEs such as LINE-2 (L2) or mammalian-wide interspersed repeat (MIR) subfamilies were significantly enriched in the DERs of H3K9me3 ($p < 1 \times 10^{-100}$) (Figures 4B and 4C), which also share a substantial fraction of the peak regions (LTRs, 46.5%; LINEs, 44.8%) (Figure S3C). These data indicate that the associations of H3K4me3 and other active epigenetic marks at the newly evolved TEs tend to be variable among hiPSC lines (Figure 4C), relating to the differential transcriptional activities of those elements in some cases (Figure 4D). In addition, we observed a positive tendency in the correlation between the TE-associated H3K4me3 enrichment and the adjacent gene expression, suggesting the influence of TE-associated epigenetic variations on the neighbor chromatin states (Figures S4A and S4B).

## Heterogeneity in the enrichment of H3K9me3 over large domains

The accumulation of H3K9me3 was often spread over a large region, and the pattern was variable among the hiPSC lines (Figure 5A). Inspection of H3K9me3-enriched peaks by combining all samples revealed a bimodal distribution of their lengths (Figure 5B; Figure S4C). Indeed, smaller peaks ($\leq 10$ kb) included DERs at a low ratio (34.1%), whereas larger peaks ($\geq 100$ kb) included a large fraction of DERs (83.6%), of which most were annotated as multi-DERs (61.8%). We defined 254 H3K9me3 large regions ($\geq 100$ kb) among the autosomes, including centromeres, based on the peak domains called by the epic2 software (Stovner and Saetrom, 2019; Table S6). We noticed that in some regions the levels of input reads were high and variable among samples (Figure S4D), possibly reflecting the copy number variation (CNV) or variable number tandem repeat (VNTR) in the genomes. The regions with higher variance in input reads (class 1) exhibited a strong bias in their positions toward centromeres and telomeres, whereas this tendency was much weaker in the regions with lesser variance (class 2), and even megascale domains were located in the middle of chromosome arms (Figure 5C). We redefined the class-2 regions (n = 145) as H3K9me3-large domains in the following analysis.

The level of H3K9me3 enrichment in these large domains was highly variable among hiPSC lines (Figure 5D; Table S6). A total of 53 domains bore a high accumulation of H3K9me3 in a majority of cell lines ($\geq 5$ out of 7), whereas about half the domains bore a high H3K9me3 accumulation only in the outlier cell lines (Figure 5E). The average levels of H3K9me3 accumulation were

not correlated with domain lengths (Figure S4E). Remarkably, most of the H3K9me3-large domains were within lamina-associated domains (LADs) in differentiated somatic cells; in contrast, only a subset of the H3K9me3-large domains was overlapped with the LADs of H1 ESCs (Figure 5E) (The 4DN Data Portal) (Dekker et al., 2017; van Schaik et al., 2020). Of note, the domains with high H3K9me3 accumulation in a majority of cell lines were more frequently observed within inter-LAD regions in H1 ESCs (Figure 5E), suggesting that the nuclear architectures might differ between hiPSCs and hESCs.

The H3K9me3-large domains were enriched with H3K9me2; in contrast, the other epigenetic marks accumulated at their boundaries (Figure S4F). We next extracted the boundary regions of each H3K9me3 peak called by epic2 software in all samples and examined the association of H3K4me3 or CTCF; the results showed that, indeed, H3K4me3 or CTCF peaks were observed in those regions more frequently than expected (Figure 5F). In agreement with this finding, a large number of transcription factor (TF)-binding motifs were significantly enriched in the boundary regions of H3K9me3 peaks, and the motifs for CTCF or CTCFL binding were prominently enriched when H3K9me3 peaks were over 100 kb (Figure 5G). Indeed, by exploring the published TF-ChIP-seq data on hESCs (Sloan et al., 2016; Tsankov et al., 2015), we found that the binding of TFs, such as YY1, SP2, and SMAD2/3, was frequently observed in the boundary regions of H3K9me3 peaks (Figures 5G and S4G). These data suggest that the recruitment of these chromatin-binding factors may determine the boundary of H3K9me3 propagation and vice versa, and their variations likely cause the heterogeneity among the hiPSC lines.

We next examined the distribution of TEs within the H3K9me3 peaks and observed that smaller H3K9me3 peaks ($\leq 10$ kb) were significantly enriched with newly evolved TE subfamilies (Figure 5H). Significant enrichment of many of these subfamilies was observed in H3K4me3 DERs (Figure 4B), and in fact, such H3K4me3 DERs were located within or proximal to smaller H3K9me3 peaks (Figure S4H). In contrast, ancient subfamilies such as L2 or MIR were present within larger H3K9me3 peaks ($\geq 100$ kb) in a more frequent manner (Figure 5H), consistent with the significant enrichment of those subfamilies in H3K9me3 DERs (Figures 5B, 4B, and 4C). However, the presence of such ancient subfamilies in the H3K9me3 peaks was still less frequent if compared to the randomly chosen regions (Figure 5H), implying that TEs are unlikely to be a key factor in the formation of large H3K9me3 domains, while they actively contribute to the formation of smaller H3K9me3 peaks.

To assess the effects of passage on H3K9me3 enrichment, we examined early passage cells of the F4 and F5 lines (termed F4e and F5e, respectively) (Table S1). The comparison of the H3K9me3 enrichment between early and middle passage cells exhibited high correlations in both cell lines, indicating that the heterogeneity in H3K9me3 is well maintained during cell propagation

---

(C) Heatmap for the enrichment of TE subfamilies (p values, hypergeometric test) in DERs and their taxa (Storer et al., 2021) indicated with gray rectangles (bottom). The subfamilies that are overrepresented in $\geq 1$ epigenetic mark-DERs ($p < 1 \times 10^{-20}$) are shown.

(D) Distribution of the indicated epigenetic marks and stranded RNA-seq reads at the indicated *HERVH* and *L1PA3* loci. For epigenetic marks, the y axis represents $\log_2$-transformed IP reads normalized by autosomal total reads.
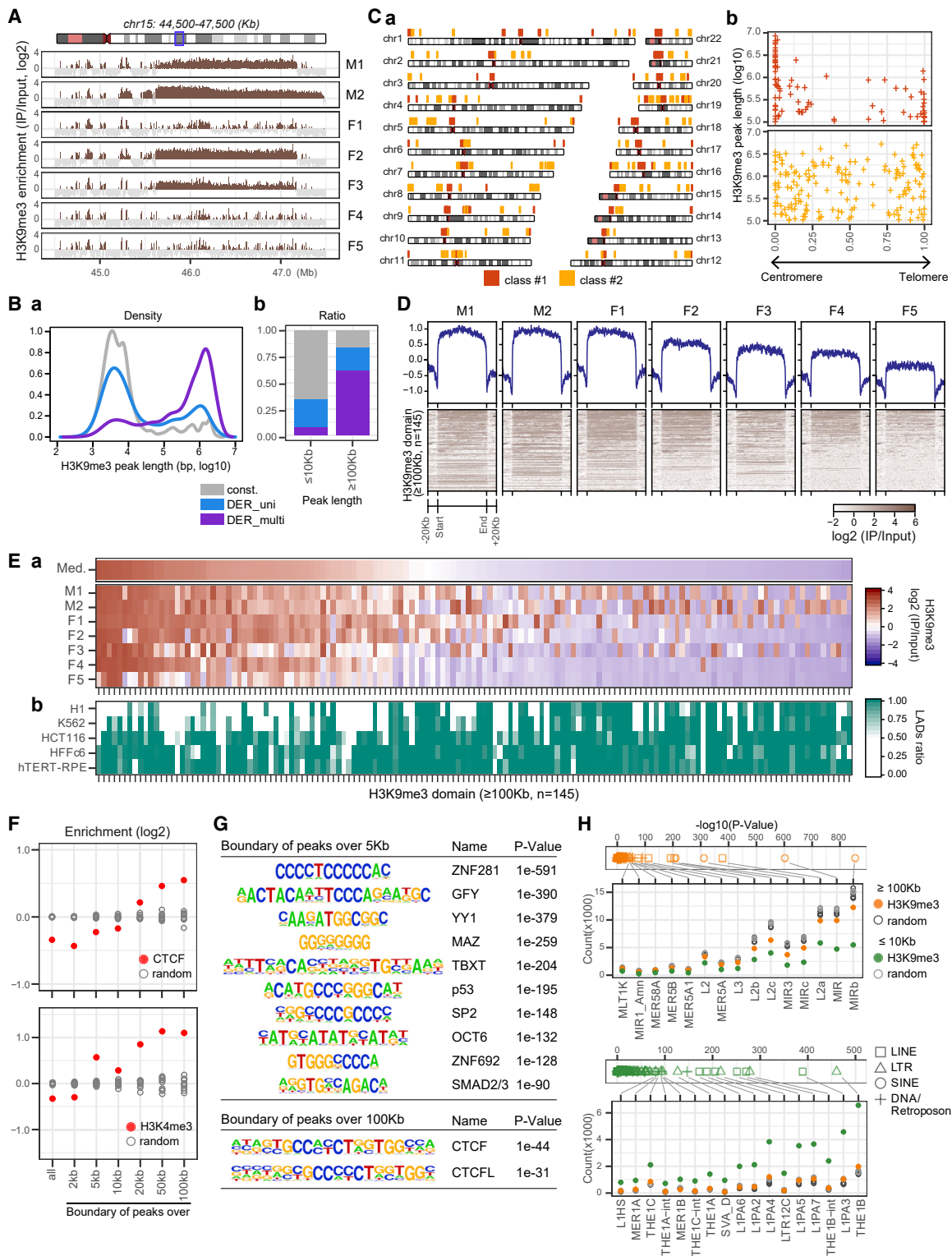
See also Figure S4.

**Figure 5. Heterogeneity in the enrichment of H3K9me3 over large domains**

(A) Distribution of H3K9me3 on chromosome 15 in the indicated hiPSC lines.

(B) The frequency plots (a) and bar plots (b) of autosomal 2-kb bins assigned to constant regions, uni-DERs, and multi-DERs of H3K9me3, against the length of the peaks to which they belong (log$_{10}$ transformed). Peak regions called by epic2 software in each sample were merged.

# Cell Reports
## Article

CellPress
OPEN ACCESS

(Figure 6A). Accordingly, differences between F4 and F5 cells were already observed in early passage cells, especially in large domains (≥100 kb). In the X chromosomes, a decrease in the H3K27me3 enrichment during the passages was observed in megabase-scale regions including the *XIST*, *DXZ4*, or *FIRRE* locus, likely in association with the downregulation of *XIST* (Figures S5A and S5B; Figures 2D and 2H); the enrichment of H3K9me3 was well preserved in the vast majority of the X chromosomes in F4 and F5 cells (Figures S5A and S5B). Note that the heterogeneity in large H3K9me3 domains was also observed in hESC lines in certain regions (Figure S6), suggesting that the occurrence of such epigenetic variability may not be restricted to the somatic cell reprogramming process for the hiPSC derivation.

## Inverse correlation between H3K9me3 and DNA methylation status

To explore the interplay between DNA methylation and the epigenomic variations we have described, we performed whole-genome bisulfite sequencing (WGBS) on F1, F4e, and F5e cells (Table S4). In contrast to the heterogeneity observed in the H3K9me3 enrichment levels (Figure 6A), the genome-wide methylated CpG (mCG) levels were in general highly correlated among these samples (Figure 6B). On the other hand, in agreement with the previous reports (Ma et al., 2014; Nazor et al., 2012), differences in the mCG levels were observed in a subset of imprinted domains (Court et al., 2014; Okae et al., 2014; Yamashiro et al., 2018; Figure S5C). Previous studies have demonstrated aberrant non-CG methylation in megabase-scale regions of hiPSCs and significant enrichment of H3K9me3 in such regions (Lister et al., 2011). Inspection of the methylated C (mC) levels in the H3K9me3-large domains revealed that not only mCA but also mCG levels were inversely correlated with the H3K9me3 enrichment levels (Figure 6C). Nevertheless, the differences in mCG levels were only partial, with more than 75% of CG remaining methylated in the majority of regions. These data suggest the possibility that the accumulation of H3K9me3 may impair the efficiency of the DNA methylation machineries in both the maintenance and *de novo* pathways. A similar inverse correlation between H3K9me3 and mC was observed in the X chromosomes (Figure S5D).

## Influences of genetic variations on epigenomic heterogeneities

To address the influences of genetic differences among the hiPSCs on their epigenomic states, we performed whole-genome sequencing of M2, F2, F3, F4, and F5 cells (Tables S1 and S4). When we examined the genome-wide frequency of CNVs, small insertions/deletions (INDELs), and SNPs, we observed a slight increase in the frequencies of CNVs and SNPs in the autosomal DERs, but no such difference for INDELs (Figure 6D). Notably, SNPs were frequently found in H3K4me3-associated LTRs and LINEs (peak summits ± 500 bp), implying a link between the genetic variations and the differences in H3K4me3 enrichment discussed in Figure 4 (Figure S7A). We next assessed the overlap of CNVs with the H3K9me3 domains (≥100 kb) determined in Figure 5. Among the 145 domains in class 2 (Figures 5C and S4D), only 15 contained CNVs within the region, indicating that CNVs made only a small contribution, if any, to engendering the heterogeneity in the H3K9me3 enrichment over large domains (Figures S7B and S7C). Together, these findings demonstrate that a majority of epigenomic heterogeneity in hiPSCs can arise independently of their genetic differences.

## Genes associated with epigenomic variations implicated in human germ cell differentiation

We next examined the relationships between the epigenetic differences and the efficiency for PGCLC induction among the hiPSC lines. We calculated the correlation between the enrichment levels of epigenetic marks and the median values of induction rates that were analyzed in our previous study (Figure 7A; Yokobayashi et al., 2017). In autosomal regions, only a small fraction of DERs exhibited high correlation coefficients (Pearson's R) in any of the epigenetic marks (0.8%–2.6%, ≥0.75; 0.6%–2.7%, ≤−0.75). In contrast, the DERs in X chromosomes showed strong biases, namely, a positive correlation with the enrichment levels of H3K9me3 (62.6%, ≥0.75) and negative correlation with those of H3K4me3, H3K27ac, or CTCF (30.0%, 44.3%, or 41.5%, respectively, ≤−0.75) (Figure 7A). These data indicate that an increase in the enrichment of active marks, i.e., reactivation of XCI, has a negative effect on the PGCLC induction (Figures 2B and 2C).

Next, by examining the overlaps between the DERs and gene regulatory elements annotated in the GeneHancer database (Fishilevich et al., 2017), we extracted genes that bear highly correlated DERs (R ≥ 0.75 or ≤ −0.75) in their regulatory regions. We obtained 1,325 genes in total, of which 1,012 genes were located in the autosomal regions (Figure 7B). Gene Ontology (GO) analysis revealed that genes for transcriptional regulation, cell growth regulation, or canonical WNT signaling pathways were significantly enriched among these 1,325 genes (Figure 7C; Table S7).

(C) (a) Chromosome ideograms of the 254 large H3K9me3 domains (≥100 kb). The 145 domains showing lower heterogeneity at the level of input samples were classified as "class 2" (see Figure S4D; Table S6). (b) Chromosomal positions of the large H3K9me3 domains. The x axis represents domain positions relative to centromeres and telomeres; the y axis represents the $\log_{10}$-transformed peak lengths.

(D) Heatmap representation and profile plots (top) of the H3K9me3 enrichment levels over H3K9me3 domains and flanking genomic regions. The values of replicate 1 were used (Table S3).

(E) (a) Heatmap of the level of H3K9me3 abundance over domains (vertical bars in the x axis) in each hiPSC line. The median values of the seven hiPSCs are shown on the top. (b) Overlap of LADs with H3K9me3 domains in the indicated cell types (the 4DN Data Portal).

(F) The frequency of the observation of CTCF (top) or H3K4me3 (bottom) peaks at the boundary of H3K9me3 peaks grouped by the lengths, relative to those in 10 randomly chosen regions (gray).

(G) The enrichment of known motifs at the boundary of H3K9me3 peaks as in (F) analyzed by the HOMER tool.

(H) The enrichment of TE subfamilies in H3K9me3 peak groups (≥100 kb [top]; ≤ 10 kb [bottom]). p values (hypergeometric test) (top panel) and the number of observations of the indicated subfamilies in the H3K9me3 peaks and 5 randomly chosen regions (light or dark gray) (bottom panel) are shown.
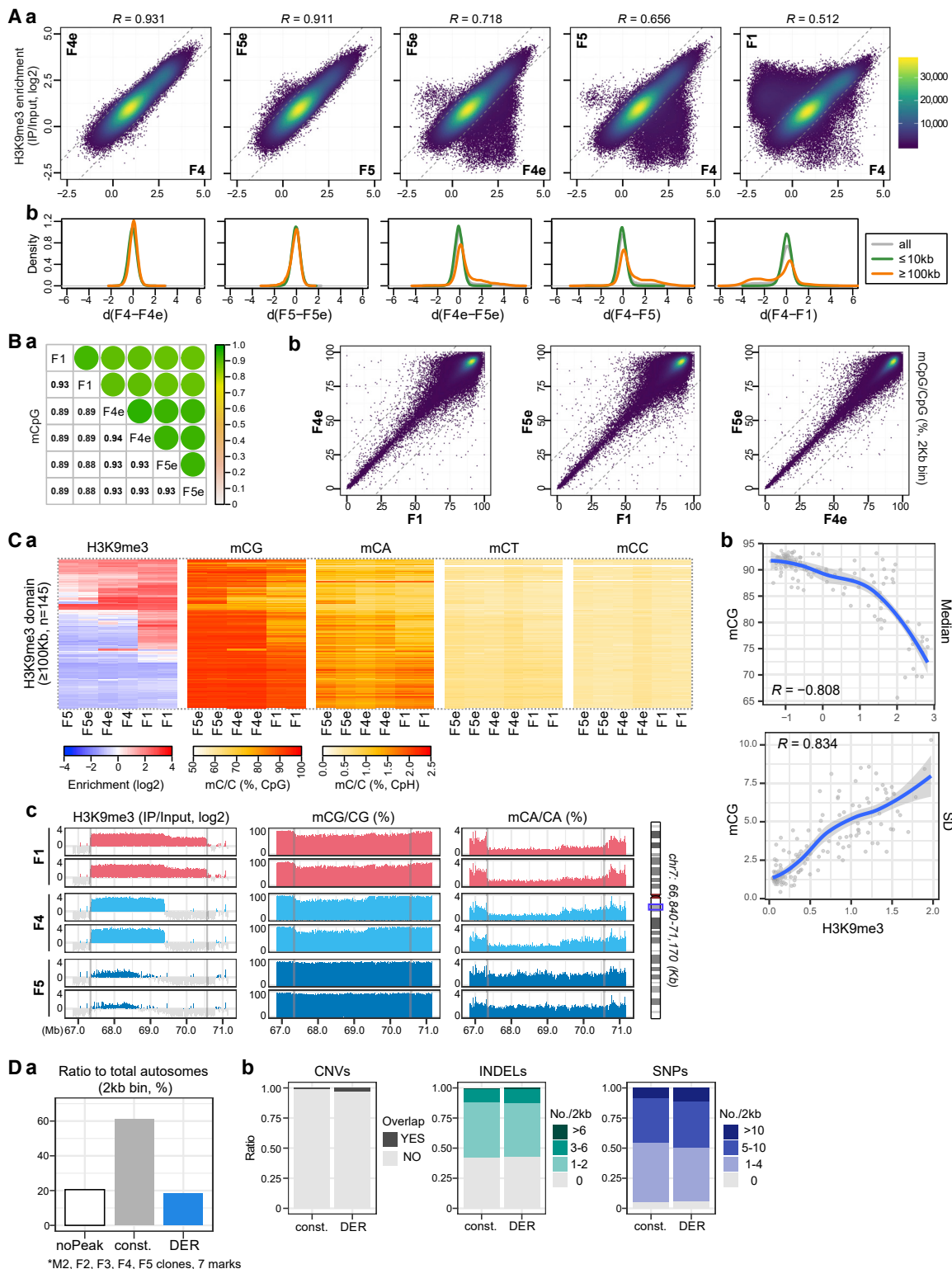
See also Figure S4.

CellPress
OPEN ACCESS

Cell Reports
Article



**Figure 6. Inverse correlation between H3K9me3 and DNA methylation status**

(A) (a) Scatterplots of the log₂-transformed enrichment values of H3K9me3 in the indicated hiPSC samples. Each point represents an autosomal 2-kb bin overlapped with merged H3K9me3 peak regions. The values of replicate 1 or 2 were used in the F1 or F4 sample, respectively. (b) Density curves of the H3K9me3 enrichment differences per 2-kb bin between the indicated samples.

*(legend continued on next page)*

# Cell Reports
## Article

Of note, WNT signaling is one of the major pathways that are activated at the initial step of PGCLC induction (Sasaki et al., 2015). Moreover, the expression levels of many of these genes were dynamically regulated during human germ cell differentiation, in which hiPSCs were induced into early PGCLCs and further differentiated into oogonia/gonocyte-like cells (Figure 7D; Kojima et al., 2017; Yamashiro et al., 2018). Interestingly, *CHCHD2*, which has been reported to play a role in the suppression of the transforming growth factor β (TGF-β) pathway (Zhu et al., 2016), exhibited a dynamic expression pattern during human germ cell differentiation and was associated with the clonal heterogeneity in the propagation of H3K9me3 domains in these hiPSC lines (Figure 7E). Additionally, we found autosomal regions that showed differential enrichment of epigenetic marks between males and females, and these were found in the regulatory elements of 18 germline-associated genes (Figure S7D). Taken together, these data suggest that epigenetic differences in the autosomes also have effects on the modulation of the gene network during the human germ cell differentiation process.

## DISCUSSION

Our study clarified the following distinct states of the X chromosome in female hiPSC lines: XaXi$_e$$^{XIST-}$ (F1 and F2), XaXi$^{XIST+}$ (F3), or XaXi$^{XIST-}$ (F4 and F5) (Figure 2I). In F3 cells where *XIST* expression was maintained, H3K27me3 and H3K9me3 were distributed along the X chromosomes in a largely mutually exclusive manner (Figure 1B), as described previously in H9 ESCs (Figure 1D; Vallot et al., 2015). In the other female hiPSCs, the enrichment of H3K27me3 was mostly lost; the enrichment of H3K9me3 was confined to narrower regions in XaXi$_e$$^{XIST-}$ cells (F1/F2), whereas it was extended to almost the entire chromosome in XaXi$^{XIST-}$ cells (F4/F5) (Figures 1B and 1C). In the XaXi$_e$$^{XIST-}$ state, approximately one-half of X-linked genes exhibited bi-allelic expression and the increase in gene dosage, whereas the rest of the genes remained silenced (Figures 2B and 2C). Such erosion of XCI (increase in the number of bi-allelically expressed genes) was not observed even in the late passage of F4 and F5 cells (Figures 2E and 2F), implying that hiPSCs establish two divergent states as the outcome of the downregulation of *XIST*. Although F4/F5 cells were generated under the feeder-free condition, F1/F2 cells were generated and passaged on SNL-feeder cells and subsequently transferred to and maintained under the feeder-free condition; such differences in the hiPSC-derivation culture conditions could be a cause for the different XCI states, a possibility that warrants investigation (Nakagawa et al., 2014; Okita et al., 2013; Takahashi et al., 2007; Tomoda et al., 2012; Table S1).

It has been well established that *XIST* recruits PRC2, and hence H3K27me3, to the Xi (Galupa and Heard, 2018); however, it remains unclear how the H3K9me3-enriched domains are established on the Xi and what determines the boundary of H3K27me3/H3K9me3 domains. In fact, such boundaries were abolished in F4 and F5 cells (Figure 1B). In the early passage of F4 cells, H3K27me3 was preferentially retained at around *XIST*, *DXZ4*, and *FIRRE* loci (Figure S5B). These loci form long-range chromatin loops in the Xi (Darrow et al., 2016; Rao et al., 2014), suggesting a close relationship between the distribution of repressive histone marks and the formation of a higher-order chromatin structure. The erosion of XCI has also been observed in cancer cells, which often exhibit dispersed or no *XIST* and H3K27me3 accumulation on the Xi (Chaligné et al., 2015). Further investigation, probably including comparative studies with other cell types, would be required to understand the molecular mechanisms causing the different XCI states in female hPSCs.

Consistent with the highly similar gene expression profiles among the hiPSC lines, in the autosomes, epigenetic variations were enriched not in genic but in non-genic regions, such as intergenic or intronic regions and TEs (Figures 3 and S1A). We did not observe a stark enrichment of genetic variations in these epigenetically variable regions (Figure 6D; Figure S7E). It has been reported that several loci of TEs are associated with differential DNA methylation in some hiPSC lines (Koyanagi-Aoi et al., 2013). In this study, using a more systematic approach, we identified differential enrichment of H3K4me3 in newly evolved TE subfamilies, including HERV-H/LTR7, LTR12C, or human- or primate-specific LINE1 elements (Figures 4B and 4C). These elements can have autonomous promoter activities (e.g., LTR12C) (van Arensbergen et al., 2017) and affect the formation of higher-order chromatin structures (e.g., HERV-H/LTR7) (Zhang et al., 2019). Therefore, the variegated epigenetic states of these elements may influence the property of hiPSCs by regulating the expression of neighboring genes (Figures S4A and S4B). Interestingly, we found that SNPs tended to be observed at a high frequency in the H3K4me3 peak regions associated with LINEs and LTRs (Figure S7A), suggesting that the SNPs may cause the differential epigenetic states, possibly by influencing the accessibility of chromatin binding factors (Carcamo-Orive et al., 2017; DeBoever et al., 2017; Skelly et al., 2020). It remains to be investigated whether these TE-associated H3K4me3 variations are unique to hiPSCs.

Previous studies have shown that, unlike hESCs, hiPSCs bear megascale regions with the hypo-methylation at non-CG dinucleotides and the enrichment of H3K9me3, which are preferentially located near centromeres or telomeres (Lister et al., 2011;

---

(B) (a) Correlation coefficient (Pearson's R) matrix of the mCG levels (autosomal 2-kb bins) of the indicated hiPSC samples. (b) Scatterplots of the mCG levels (autosomal 2-kb bins).

(C) (a) Heatmap of the levels of H3K9me3, mCG, mCA, mCT, or mCC in the 145 H3K9me3 domains (Figures 5D and 5E) in the indicated hiPSC lines as in (A). (b) Comparison of the median (top) or SD (bottom) values between the H3K9me3 and mCG levels in the hiPSC samples indicated in (a). Each point represents a H3K9me3 domain. Correlation coefficient (Pearson's R) and loess regression curves (blue) with 95% confidence intervals (gray shadow) are shown. (c) The levels of H3K9me3, mCG, and mCA on chromosome 7.

(D) (a) Bar plots of the proportion of autosomal 2-kb bins as shown in Figure 3D, analyzed on the indicated five hiPSC lines whose WGS data were examined. (b) Bar plots of the proportion of 2-kb bins classified into "const." or "DER," as shown in (a). Overlap with CNVs (left) and frequency of INDELs (center) or SNPs (right) are shown.
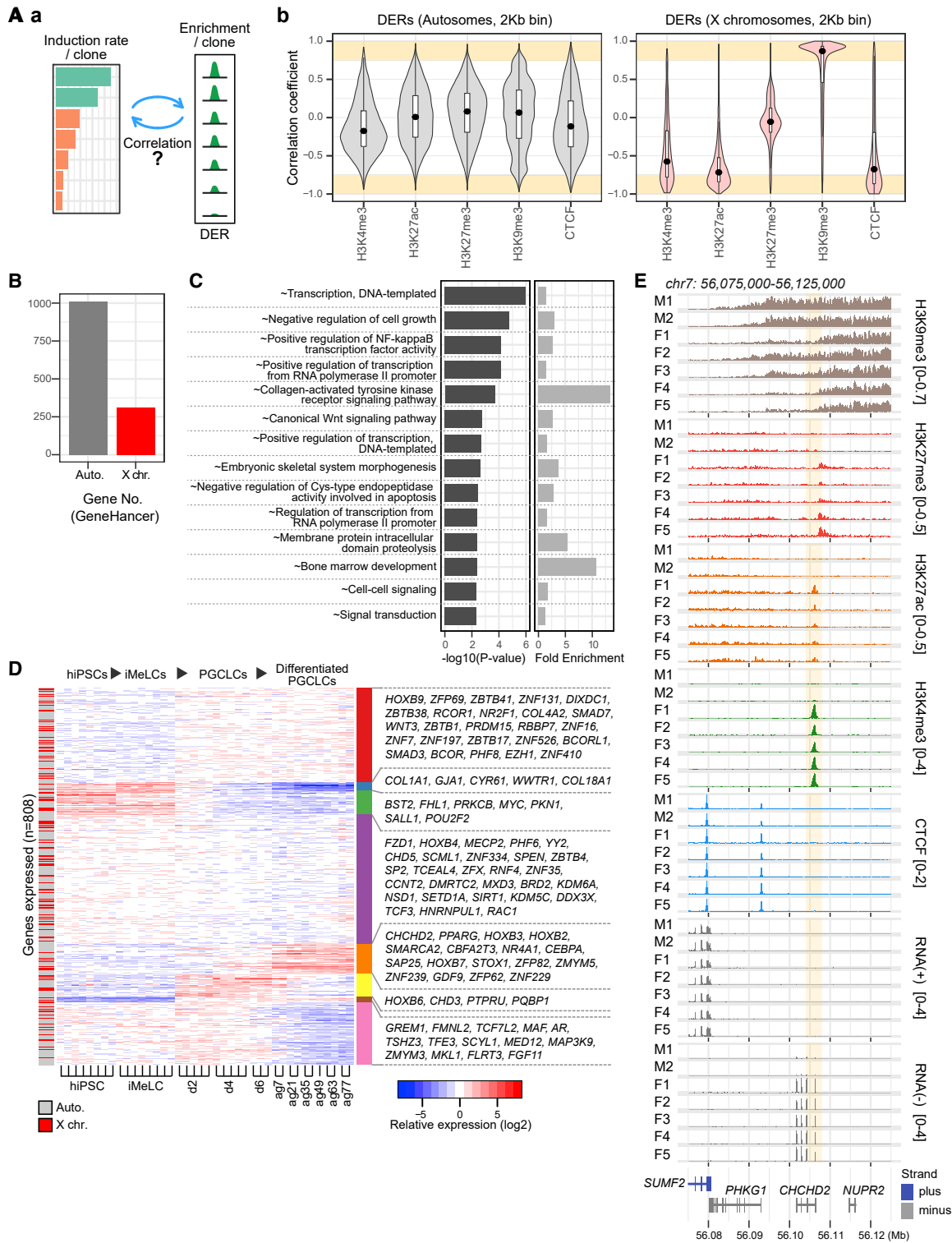
See also Figures S5–S7.

CellPress
OPEN ACCESS

Cell Reports
Article



**Figure 7. Genes associated with epigenomic variations implicated in human germ cell differentiation**

(A) (a) Schematic for the correlation analysis between the PGCLC induction rates (Yokobayashi et al., 2017) and the enrichment levels of epigenetic marks. (b) Violin plots of correlation coefficient (Pearson's R) as in (a), calculated for each 2-kb bin overlapped with the DERs of the indicated epigenetic marks. The median values and 25th and 75th percentiles are shown in overlaid boxplots.

京都大学
KYOTO UNIVERSITY

KURENAI 紅
Kyoto University Research Information Repository

# Cell Reports
## Article

Cell Press
OPEN ACCESS

Ma et al., 2014). In the present study, we identified 145 large domains where the distribution of H3K9me3 is heterogeneous among hiPSC lines. Such domains were in the range of 100 kb to >5 Mb, were distributed broadly across the genome (Figure 5C), and were rarely overlapped with CNVs (Figures S7B and S7C). WGBS analysis revealed that the level of H3K9me3 enrichment in the large domains was inversely correlated with the level of DNA methylation at not only non-CG but also CG sites, although the differences in CG-methylation levels were only partial (Figure 6C). This result likely indicates an interference with the access of the DNA methylation machinery (de novo and/or replication coupled) to the regions within the H3K9me3-enriched large domain, possibly due to the phase-separated chromatin structures formed by the recruitment of H3K9me3 and heterochromatin protein 1 (HP1) (Larson et al., 2017; Strom et al., 2017). A subset of these H3K9me3 domains exhibited dynamic changes in the interaction to nuclear lamina between differentiated somatic cells and hESCs (Figure 5E), suggesting that the heterogeneity in H3K9me3-large domains might be linked to the heterogeneity in higher-order chromatin organizations. Collectively, these results indicated that H3K9me3 would be a good landmark for evaluating the domain-scale epigenetic heterogeneity among hiPSCs in future studies. Moreover, it will be important to investigate H3K9me3 states in additional hESC lines that also show some degree of such heterogeneity (Figure S6).

Among the female hiPSC lines, the efficiency of hPGCLC induction decreases in the order of XaXi$^{XIST-}$ (F4, F5) > XaXi$^{XIST+}$ (F3) > XaXi$_i$/e$^{XIST-}$ (F1, F2) states (Yokobayashi et al., 2017; Figure S7F), suggesting that having the XCI state is an advantage for hPSCs to differentiate into the germ cell lineage. The dosage differences between XaXi/e and XaXi were less than 2-fold but have an effect on hundreds of genes, including transcriptional/epigenetic regulators or factors involved in cytokine signaling pathways (Figures 7C and 7D; Table S7). Given the dynamic regulation of the XCI state during mammalian development (Patrat et al., 2020), it will be important to clarify in future studies how the orchestration of these genes affects cell fate decisions in the germ cell lineage as well as in somatic cell lineages. Our analysis also identified epigenetic variations in autosomal regions that are highly correlated with the efficiency of hPGCLC induction (Figure 7), and some of these variations exhibited significant differences between male and female hiPSCs (Figure S7D). We and others have shown that female hPSCs tended to be less efficient for hPGCLC induction (Chen et al., 2017; Yokobayashi et al., 2017). It warrants further investigation whether such epigenetic differences in autosomes could be attributed to the differences in the XCI or sex chromosome dosages, as pointed out in previous studies (Brenes et al., 2021; Milagre et al., 2017; Pasque et al., 2018; Song et al., 2019).

In summary, the present study revealed the basis of epigenomic variations existing in hiPSCs, through the in-depth analysis of seven relevant hiPSC lines. In addition to the female X chromosomes, the properties of genomic regions, such as the CpG density, TE insertion, or interaction with the nuclear lamina, likely contribute to the emergence and/or retention of the heterogeneity of epigenomic states. These variations appear to be neither beneficial nor detrimental for cells in the undifferentiated state but may determine the latent propensity for the differentiation capacity of hiPSCs. Decoding the epigenomic information would enable us to prospectively assess the fate of hiPSCs, which requires future studies on a large scale.

### Limitations of the study
A limitation of our study is the use of seven lines and a limited number of derivation/culture conditions. A larger number of cell lines, including both hESCs and hiPSCs, should be examined in future studies for further clarification of the potential effects of epigenomic variations on the PGCLC induction proposed in this study. The examination of a wider variety of female cell lines with different derivation/culture conditions would be especially important for understanding the effects of the XCI state on the in vitro specification of human germ cell fate and subsequent differentiation processes.

### STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
  - Stranded RNA-seq library preparation and sequencing
  - ChIP-seq library preparation and sequencing
  - WGBS library preparation and sequencing
  - Whole genome sequencing
  - Immuno-DNA FISH
  - RNA-seq data processing
  - ChIP-seq data processing
  - Identification of DERs
  - WGBS data processing
  - Analysis of H3K9me3 domains
  - SNP analysis for X-linked genes
  - Analysis of CNVs, INDELs, and SNPs in the autosomes

(B) Number of genes whose regulatory elements are overlapped with the highly correlated DERs (Pearson's R, ≥0.75 or ≤−0.75; shadowed in B), defined by GeneHancer database.

(C) Gene Ontology analysis of the genes shown in (B) (Table S7).

(D) K-means clustering (K = 8) and heatmap of the relative expression values of the genes shown in (B) during the PGCLC induction process, examined on 585B1-868 reporter hiPSCs (Kojima et al., 2017; Yamashiro et al., 2018). Genes whose $\log_2$(RPM+1) values were <4 in all samples were excluded. Representative genes are shown on the right.

(E) The levels of the indicated epigenetic marks and stranded RNA-seq reads on chromosome 7.

See also Figure S7.

**Cell Reports**
Article

**SUPPLEMENTAL INFORMATION**

Supplemental information can be found online at https://doi.org/10.1016/j.celrep.2021.109909.

**AUTHOR CONTRIBUTIONS**

Conceptualization and design, S.Y.; methodology, S.Y., Y.Y., B.H., Y.M., T.N., G.B., and T.Y.; analysis, S.Y., Y.Y., and T.Y.; investigation, S.Y. and Y.M.; resources, M.N. and K.O.; data interpretation, S.Y., G.B., and M.S.; writing, S.Y. and M.S.; supervision, J.M. and M.S.

**DECLARATION OF INTERESTS**

The authors declare no competing interests.

**INCLUSION AND DIVERSITY STATEMENT**

We worked to ensure diversity in experimental samples through the selection of the cell lines.

**REFERENCES**

Adewumi, O., Aflatoonian, B., Ahrlund-Richter, L., Amit, M., Andrews, P.W., Beighton, G., Bello, P.A., Benvenisty, N., Berry, L.S., Bevan, S., et al.; International Stem Cell Initiative (2007). Characterization of human embryonic stem cell lines by the International Stem Cell Initiative. Nat. Biotechnol. *25*, 803–816.

Balaton, B.P., Cotton, A.M., and Brown, C.J. (2015). Derivation of consensus inactivation status for X-linked genes from genome-wide studies. Biol. Sex Differ. *6*, 35.

Bansal, P., Ahern, D.T., Kondaveeti, Y., Qiu, C.W., and Pinter, S.F. (2021). Contiguous erosion of the inactive X in human pluripotency concludes with global DNA hypomethylation. Cell Rep. *35*, 109215.

Bar, S., Schachter, M., Eldar-Geva, T., and Benvenisty, N. (2017). Large-Scale Analysis of Loss of Imprinting in Human Pluripotent Stem Cells. Cell Rep. *19*, 957–968.

Bock, C., Kiskinis, E., Verstappen, G., Gu, H., Boulting, G., Smith, Z.D., Ziller, M., Croft, G.F., Amoroso, M.W., Oakley, D.H., et al. (2011). Reference Maps of human ES and iPS cell variation enable high-throughput characterization of pluripotent cell lines. Cell *144*, 439–452.

Boeva, V., Popova, T., Bleakley, K., Chiche, P., Cappo, J., Schleiermacher, G., Janoueix-Lerosey, I., Delattre, O., and Barillot, E. (2012). Control-FREEC: a tool for assessing copy number and allelic content using next-generation sequencing data. Bioinformatics *28*, 423–425.

Borgel, J., Guibert, S., Li, Y., Chiba, H., Schübeler, D., Sasaki, H., Forné, T., and Weber, M. (2010). Targets and dynamics of promoter DNA methylation during early mouse development. Nat. Genet. *42*, 1093–1100.

Brenes, A.J., Yoshikawa, H., Bensaddek, D., Mirauta, B., Seaton, D., Hukelmann, J.L., Jiang, H., Stegle, O., and Lamond, A.I. (2021). Erosion of human X chromosome inactivation causes major remodeling of the iPSC proteome. Cell Rep. *35*, 109032.

Carcamo-Orive, I., Hoffman, G.E., Cundiff, P., Beckmann, N.D., D'Souza, S.L., Knowles, J.W., Patel, A., Papatsenko, D., Abbasi, F., Reaven, G.M., et al. (2017). Analysis of Transcriptional Variability in a Large Human iPSC Library Reveals Genetic and Non-genetic Determinants of Heterogeneity. Cell Stem Cell *20*, 518–532.e9.

Chaligné, R., Popova, T., Mendoza-Parra, M.A., Saleem, M.A., Gentien, D., Ban, K., Piolot, T., Leroy, O., Mariani, O., Gronemeyer, H., et al. (2015). The inactive X chromosome is epigenetically unstable and transcriptionally labile in breast cancer. Genome Res. *25*, 488–503.

Chen, D., Liu, W., Lukianchikov, A., Hancock, G.V., Zimmerman, J., Lowe, M.G., Kim, R., Galic, Z., Irie, N., Surani, M.A., et al. (2017). Germline competency of human embryonic stem cells depends on eomesodermin. Biol. Reprod. *97*, 850–861.

Choi, J., Lee, S., Mallard, W., Clement, K., Tagliazucchi, G.M., Lim, H., Choi, I.Y., Ferrari, F., Tsankov, A.M., Pop, R., et al. (2015). A comparison of genetically matched cell lines reveals the equivalence of human iPSCs and ESCs. Nat. Biotechnol. *33*, 1173–1181.

Court, F., Tayama, C., Romanelli, V., Martin-Trujillo, A., Iglesias-Platas, I., Okamura, K., Sugahara, N., Simón, C., Moore, H., Harness, J.V., et al. (2014). Genome-wide parent-of-origin DNA methylation analysis reveals the intricacies of human imprinting and suggests a germline methylation-independent mechanism of establishment. Genome Res. *24*, 554–569.

Darrow, E.M., Huntley, M.H., Dudchenko, O., Stamenova, E.K., Durand, N.C., Sun, Z., Huang, S.C., Sanborn, A.L., Machol, I., Shamim, M., et al. (2016). Deletion of DXZ4 on the human inactive X chromosome alters higher-order genome architecture. Proc. Natl. Acad. Sci. USA *113*, E4504–E4512.

DeBoever, C., Li, H., Jakubosky, D., Benaglio, P., Reyna, J., Olson, K.M., Huang, H., Biggs, W., Sandoval, E., D'Antonio, M., et al. (2017). Large-Scale Profiling Reveals the Influence of Genetic Variation on Gene Expression in Human Induced Pluripotent Stem Cells. Cell Stem Cell *20*, 533–546.e7.

Dekker, J., Belmont, A.S., Guttman, M., Leshyk, V.O., Lis, J.T., Lomvardas, S., Mirny, L.A., O'Shea, C.C., Park, P.J., Ren, B., et al.; 4D Nucleome Network (2017). The 4D nucleome project. Nature *549*, 219–226.

Ernst, J., and Kellis, M. (2017). Chromatin-state discovery and genome annotation with ChromHMM. Nat. Protoc. *12*, 2478–2492.

Fishilevich, S., Nudel, R., Rappaport, N., Hadar, R., Plaschkes, I., Iny Stein, T., Rosen, N., Kohn, A., Twik, M., Safran, M., et al. (2017). GeneHancer: genome-wide integration of enhancers and target genes in GeneCards. Database (Oxford) *2017*, bax028.

Galupa, R., and Heard, E. (2018). X-Chromosome Inactivation: A Crossroads Between Chromosome Architecture and Gene Regulation. Annu. Rev. Genet. *52*, 535–566.

Guenther, M.G., Frampton, G.M., Soldner, F., Hockemeyer, D., Mitalipova, M., Jaenisch, R., and Young, R.A. (2010). Chromatin structure and gene expression programs of human embryonic and induced pluripotent stem cells. Cell Stem Cell *7*, 249–257.

Hayashi-Takanaka, Y., Yamagata, K., Wakayama, T., Stasevich, T.J., Kainuma, T., Tsurimoto, T., Tachibana, M., Shinkai, Y., Kurumizaka, H., Nozaki,

N., and Kimura, H. (2011). Tracking epigenetic histone modifications in single cells using Fab-based live endogenous modification labeling. Nucleic Acids Res. *39*, 6475–6488.

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

Huang, W., Sherman, B.T., and Lempicki, R.A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat. Protoc. *4*, 44–57.

Kajiwara, M., Aoi, T., Okita, K., Takahashi, R., Inoue, H., Takayama, N., Endo, H., Eto, K., Toguchida, J., Uemoto, S., and Yamanaka, S. (2012). Donor-dependent variations in hepatic differentiation from human-induced pluripotent stem cells. Proc. Natl. Acad. Sci. USA *109*, 12538–12543.

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

Kimura, H., Hayashi-Takanaka, Y., Goto, Y., Takizawa, N., and Nozaki, N. (2008). The organization of histone H3 modifications as revealed by a panel of specific monoclonal antibodies. Cell Struct. Funct. *33*, 61–73.

Kojima, Y., Sasaki, K., Yokobayashi, S., Sakai, Y., Nakamura, T., Yabuta, Y., Nakaki, F., Nagaoka, S., Woltjen, K., Hotta, A., et al. (2017). Evolutionarily Distinctive Transcriptional and Signaling Programs Drive Human Germ Cell Lineage Specification from Pluripotent Stem Cells. Cell Stem Cell *21*, 517–532.e5.

Koyanagi-Aoi, M., Ohnuki, M., Takahashi, K., Okita, K., Noma, H., Sawamura, Y., Teramoto, I., Narita, M., Sato, Y., Ichisaka, T., et al. (2013). Differentiation-defective phenotypes revealed by large-scale analyses of human pluripotent stem cells. Proc. Natl. Acad. Sci. USA *110*, 20569–20574.

Krueger, F., and Andrews, S.R. (2011). Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. Bioinformatics *27*, 1571–1572.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359.

Larson, A.G., Elnatan, D., Keenen, M.M., Trnka, M.J., Johnston, J.B., Burlingame, A.L., Agard, D.A., Redding, S., and Narlikar, G.J. (2017). Liquid droplet formation by HP1α suggests a role for phase separation in heterochromatin. Nature *547*, 236–240.

Lawrence, M., Huber, W., Pagès, H., Aboyoun, P., Carlson, M., Gentleman, R., Morgan, M.T., and Carey, V.J. (2013). Software for computing and annotating genomic ranges. PLoS Comput. Biol. *9*, e1003118.

Lee, T.I., Johnstone, S.E., and Young, R.A. (2006). Chromatin immunoprecipitation and microarray-based analysis of protein location. Nat. Protoc. *1*, 729–748.

Li, H., and Durbin, R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics *26*, 589–595.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079.

Lister, R., Pelizzola, M., Kida, Y.S., Hawkins, R.D., Nery, J.R., Hon, G., Antosiewicz-Bourget, J., O'Malley, R., Castanon, R., Klugman, S., et al. (2011). Hotspots of aberrant epigenomic reprogramming in human induced pluripotent stem cells. Nature *471*, 68–73.

Ma, H., Morey, R., O'Neil, R.C., He, Y., Daughtry, B., Schultz, M.D., Hariharan, M., Nery, J.R., Castanon, R., Sabatini, K., et al. (2014). Abnormalities in human pluripotent cells due to reprogramming mechanisms. Nature *511*, 177–183.

Mekhoubad, S., Bock, C., de Boer, A.S., Kiskinis, E., Meissner, A., and Eggan, K. (2012). Erosion of dosage compensation impacts human iPSC disease modeling. Cell Stem Cell *10*, 595–609.

Milagre, I., Stubbs, T.M., King, M.R., Spindel, J., Santos, F., Krueger, F., Bachman, M., Segonds-Pichon, A., Balasubramanian, S., Andrews, S.R., et al. (2017). Gender Differences in Global but Not Targeted Demethylation in iPSC Reprogramming. Cell Rep. *18*, 1079–1089.

Miura, F., Enomoto, Y., Dairiki, R., and Ito, T. (2012). Amplification-free whole-genome bisulfite sequencing by post-bisulfite adaptor tagging. Nucleic Acids Res. *40*, e136.

Nakagawa, M., Taniguchi, Y., Senda, S., Takizawa, N., Ichisaka, T., Asano, K., Morizane, A., Doi, D., Takahashi, J., Nishizawa, M., et al. (2014). A novel efficient feeder-free culture system for the derivation of human induced pluripotent stem cells. Sci. Rep. *4*, 3594.

Nazor, K.L., Altun, G., Lynch, C., Tran, H., Harness, J.V., Slavin, I., Garitaonandia, I., Müller, F.J., Wang, Y.C., Boscolo, F.S., et al. (2012). Recurrent variations in DNA methylation in human pluripotent stem cells and their differentiated derivatives. Cell Stem Cell *10*, 620–634.

Nishizawa, M., Chonabayashi, K., Nomura, M., Tanaka, A., Nakamura, M., Inagaki, A., Nishikawa, M., Takei, I., Oishi, A., Tanabe, K., et al. (2016). Epigenetic Variation between Human Induced Pluripotent Stem Cell Lines Is an Indicator of Differentiation Capacity. Cell Stem Cell *19*, 341–354.

Nozawa, R.S., Nagao, K., Igami, K.T., Shibata, S., Shirai, N., Nozaki, N., Sado, T., Kimura, H., and Obuse, C. (2013). Human inactive X chromosome is compacted through a PRC2-independent SMCHD1-HBiX1 pathway. Nat. Struct. Mol. Biol. *20*, 566–573.

Ohi, Y., Qin, H., Hong, C., Blouin, L., Polo, J.M., Guo, T., Qi, Z., Downey, S.L., Manos, P.D., Rossi, D.J., et al. (2011). Incomplete DNA methylation underlies a transcriptional memory of somatic cells in human iPS cells. Nat. Cell Biol. *13*, 541–549.

Okae, H., Chiba, H., Hiura, H., Hamada, H., Sato, A., Utsunomiya, T., Kikuchi, H., Yoshida, H., Tanaka, A., Suyama, M., and Arima, T. (2014). Genome-wide analysis of DNA methylation dynamics during early human development. PLoS Genet. *10*, e1004868.

Okita, K., Yamakawa, T., Matsumura, Y., Sato, Y., Amano, N., Watanabe, A., Goshima, N., and Yamanaka, S. (2013). An efficient nonviral method to generate integration-free human-induced pluripotent stem cells from cord blood and peripheral blood cells. Stem Cells *31*, 458–466.

Ortmann, D., Brown, S., Czechanski, A., Aydin, S., Muraro, D., Huang, Y., Tomaz, R.A., Osnato, A., Canu, G., Wesley, B.T., et al. (2020). Naive Pluripotent Stem Cells Exhibit Phenotypic Variability that Is Driven by Genetic Variation. Cell Stem Cell *27*, 470–481.e6.

Pasque, V., Karnik, R., Chronis, C., Petrella, P., Langerman, J., Bonora, G., Song, J., Vanheer, L., Sadhu Dimashkie, A., Meissner, A., and Plath, K. (2018). X Chromosome Dosage Influences DNA Methylation Dynamics during Reprogramming to Mouse iPSCs. Stem Cell Reports *10*, 1537–1550.

Patel, S., Bonora, G., Sahakyan, A., Kim, R., Chronis, C., Langerman, J., Fitz-Gibbon, S., Rubbi, L., Skelton, R.J.P., Ardehali, R., et al. (2017). Human Embryonic Stem Cells Do Not Change Their X Inactivation Status during Differentiation. Cell Rep. *18*, 54–67.

Patrat, C., Ouimette, J.F., and Rougeulle, C. (2020). X chromosome inactivation in human development. Development *147*, dev183095.

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841–842.

Ramírez, F., Ryan, D.P., Grüning, B., Bhardwaj, V., Kilpert, F., Richter, A.S., Heyne, S., Dündar, F., and Manke, T. (2016). deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Res. *44*, W160–W165.

Rao, S.S., Huntley, M.H., Durand, N.C., Stamenova, E.K., Bochkov, I.D., Robinson, J.T., Sanborn, A.L., Machol, I., Omer, A.D., Lander, E.S., and Aiden, E.L. (2014). A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell *159*, 1665–1680.

Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011). Integrative genomics viewer. Nat. Biotechnol. *29*, 24–26.

Roost, M.S., Slieker, R.C., Bialecka, M., van Iperen, L., Gomes Fernandes, M.M., He, N., Suchiman, H.E.D., Szuhai, K., Carlotti, F., de Koning, E.J.P., et al. (2017). DNA methylation and transcriptional trajectories during human development and reprogramming of isogenic pluripotent stem cells. Nat. Commun. *8*, 908.

Ruiz, S., Diep, D., Gore, A., Panopoulos, A.D., Montserrat, N., Plongthongkum, N., Kumar, S., Fung, H.L., Giorgetti, A., Bilic, J., et al. (2012). Identification of a specific reprogramming-associated epigenetic signature in human induced pluripotent stem cells. Proc. Natl. Acad. Sci. USA *109*, 16196–16201.

Salomonis, N., Dexheimer, P.J., Omberg, L., Schroll, R., Bush, S., Huo, J., Schriml, L., Ho Sui, S., Keddache, M., Mayhew, C., et al. (2016). Integrated Genomic Analysis of Diverse Induced Pluripotent Stem Cells from the Progenitor Cell Biology Consortium. Stem Cell Reports *7*, 110–125.

Sasaki, K., Yokobayashi, S., Nakamura, T., Okamoto, I., Yabuta, Y., Kurimoto, K., Ohta, H., Moritoki, Y., Iwatani, C., Tsuchiya, H., et al. (2015). Robust In Vitro Induction of Human Germ Cell Fate from Pluripotent Stem Cells. Cell Stem Cell *17*, 178–194.

Schulz, E.G., Meisig, J., Nakamura, T., Okamoto, I., Sieber, A., Picard, C., Borensztein, M., Saitou, M., Blüthgen, N., and Heard, E. (2014). The two active X chromosomes in female ESCs block exit from the pluripotent state by modulating the ESC signaling network. Cell Stem Cell *14*, 203–216.

Shirane, K., Kurimoto, K., Yabuta, Y., Yamaji, M., Satoh, J., Ito, S., Watanabe, A., Hayashi, K., Saitou, M., and Sasaki, H. (2016). Global Landscape and Regulatory Principles of DNA Methylation Reprogramming for Germ Cell Specification by Mouse Pluripotent Stem Cells. Dev. Cell *39*, 87–103.

Skelly, D.A., Czechanski, A., Byers, C., Aydin, S., Spruce, C., Olivier, C., Choi, K., Gatti, D.M., Raghupathy, N., Keele, G.R., et al. (2020). Mapping the Effects of Genetic Variation on Chromatin State and Gene Expression Reveals Loci That Control Ground State Pluripotency. Cell Stem Cell *27*, 459–469.e8.

Sloan, C.A., Chan, E.T., Davidson, J.M., Malladi, V.S., Strattan, J.S., Hitz, B.C., Gabdank, I., Narayanan, A.K., Ho, M., Lee, B.T., et al. (2016). ENCODE data at the ENCODE portal. Nucleic Acids Res. *44*, D726–D732.

Song, J., Janiszewski, A., De Geest, N., Vanheer, L., Talon, I., El Bakkali, M., Oh, T., and Pasque, V. (2019). X-Chromosome Dosage Modulates Multiple Molecular and Cellular Properties of Mouse Pluripotent Stem Cells Independently of Global DNA Methylation Levels. Stem Cell Reports *12*, 333–350.

Storer, J., Hubley, R., Rosen, J., Wheeler, T.J., and Smit, A.F. (2021). The Dfam community resource of transposable element families, sequence models, and genome annotations. Mob. DNA *12*, 2.

Stovner, E.B., and Sætrom, P. (2019). epic2 efficiently finds diffuse domains in ChIP-seq data. Bioinformatics *35*, 4392–4393.

Strom, A.R., Emelyanov, A.V., Mir, M., Fyodorov, D.V., Darzacq, X., and Karpen, G.H. (2017). Phase separation drives heterochromatin domain formation. Nature *547*, 241–245.

Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T., Tomoda, K., and Yamanaka, S. (2007). Induction of pluripotent stem cells from adult human fibroblasts by defined factors. Cell *131*, 861–872.

Terranova, R., Yokobayashi, S., Stadler, M.B., Otte, A.P., van Lohuizen, M., Orkin, S.H., and Peters, A.H. (2008). Polycomb group proteins Ezh2 and Rnf2 direct genomic contraction and imprinted repression in early mouse embryos. Dev. Cell *15*, 668–679.

Thomson, J.A., Itskovitz-Eldor, J., Shapiro, S.S., Waknitz, M.A., Swiergiel, J.J., Marshall, V.S., and Jones, J.M. (1998). Embryonic stem cell lines derived from human blastocysts. Science *282*, 1145–1147.

Tomoda, K., Takahashi, K., Leung, K., Okada, A., Narita, M., Yamada, N.A., Eilertson, K.E., Tsang, P., Baba, S., White, M.P., et al. (2012). Derivation conditions impact X-inactivation status in female human induced pluripotent stem cells. Cell Stem Cell *11*, 91–99.

Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M.J., Salzberg, S.L., Wold, B.J., and Pachter, L. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat. Biotechnol. *28*, 511–515.

Tsankov, A.M., Gu, H., Akopian, V., Ziller, M.J., Donaghey, J., Amit, I., Gnirke, A., and Meissner, A. (2015). Transcription factor binding dynamics during human ES cell differentiation. Nature *518*, 344–349.

Vallot, C., Ouimette, J.F., Makhlouf, M., Féraud, O., Pontis, J., Côme, J., Martinat, C., Bennaceur-Griscelli, A., Lalande, M., and Rougeulle, C. (2015). Erosion of X Chromosome Inactivation in Human Pluripotent Cells Initiates with XACT Coating and Depends on a Specific Heterochromatin Landscape. Cell Stem Cell *16*, 533–546.

van Arensbergen, J., FitzPatrick, V.D., de Haas, M., Pagie, L., Sluimer, J., Bussemaker, H.J., and van Steensel, B. (2017). Genome-wide mapping of autonomous promoter activity in human cells. Nat. Biotechnol. *35*, 145–153.

van Schaik, T., Vos, M., Peric-Hupkes, D., Hn Celie, P., and van Steensel, B. (2020). Cell cycle dynamics of lamina-associated DNA. EMBO Rep. *21*, e50636.

Voigt, P., Tee, W.W., and Reinberg, D. (2013). A double take on bivalent promoters. Genes Dev. *27*, 1318–1338.

Yamanaka, S. (2020). Pluripotent Stem Cell-Based Cell Therapy-Promise and Challenges. Cell Stem Cell *27*, 523–531.

Yamashiro, C., Sasaki, K., Yabuta, Y., Kojima, Y., Nakamura, T., Okamoto, I., Yokobayashi, S., Murase, Y., Ishikura, Y., Shirane, K., et al. (2018). Generation of human oogonia from induced pluripotent stem cells in vitro. Science *362*, 356–360.

Yin, T., Cook, D., and Lawrence, M. (2012). ggbio: an R package for extending the grammar of graphics for genomic data. Genome Biol. *13*, R77.

Yokobayashi, S., and Saitou, M. (2018). Reconstitution of germ cell development in vitro. In Cell Biology of the Ovary (Springer).

Yokobayashi, S., Okita, K., Nakagawa, M., Nakamura, T., Yabuta, Y., Yamamoto, T., and Saitou, M. (2017). Clonal variation of human induced pluripotent stem cells for induction into the germ cell fate. Biol. Reprod. *96*, 1154–1166.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., and Liu, X.S. (2008). Model-based analysis of ChIP-Seq (MACS). Genome Biol. *9*, R137.

Zhang, Y., Li, T., Preissl, S., Amaral, M.L., Grinstein, J.D., Farah, E.N., Destici, E., Qiu, Y., Hu, R., Lee, A.Y., et al. (2019). Transcriptionally active HERV-H retrotransposons demarcate topologically associating domains in human pluripotent stem cells. Nat. Genet. *51*, 1380–1388.

Zhu, L., Gomez-Duran, A., Saretzki, G., Jin, S., Tilgner, K., Melguizo-Sanchis, D., Anyfantis, G., Al-Aama, J., Vallier, L., Chinnery, P., et al. (2016). The mitochondrial protein CHCHD2 primes the differentiation potential of human induced pluripotent stem cells to neuroectodermal lineages. J. Cell Biol. *215*, 187–202.

# Cell Reports
## Article

**CellPress**
OPEN ACCESS

## STAR★METHODS

### KEY RESOURCES TABLE

| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
|---|---|---|
| **Antibodies** | | |
| anti-H3K4me3 | Gift from Dr. H. Kimura (Kimura et al., 2008) | Clone CMA304 |
| anti-H3K27ac | Gift from Dr. H. Kimura (Kimura et al., 2008) | Clone CMA309 |
| anti-H3K9me3 | Gift from Dr. H. Kimura (Hayashi-Takanaka et al., 2011) | Clone CMA318 |
| anti-H3K27me3 | Gift from Dr. H. Kimura (Hayashi-Takanaka et al., 2011) | Clone CMA323 |
| anti-H3K9me2 | MBL | MABI0307; RRID: AB_11124951 |
| anti-H2AK119ub1 | CST | D27C4; RRID: AB_10891618 |
| anti-CTCF | CST | D31H2; RRID: AB_2086791 |
| anti-H3K27me3 | MBL | MABI0323; RRID: AB_11123929 |
| anti-H3K9me3 | MBL | MABI0318 |
| Donkey anti-mouse IgG-Alexa 488 | Thermo Fisher Scientific | A32766; RRID: AB_2762823 |
| XCyting Chromosome Paints Orange | Metasystems | D-0323-100-OR |
| **Chemicals, peptides, and recombinant proteins** | | |
| StemFit AK03N | Ajinomoto | N/A |
| iMatrix-511 | Nippi | 892012 |
| TryPLE Select | Thermo Fisher Scientific | 12563011 |
| Y-27632 | Wako Pure Chemical Industries | 253-00511 |
| 16% Formaldehyde (w/v), Methanol-free | Thermo Fisher Scientific | 28906 |
| cOmplete, EDTA-free | SIGMA | 4693132001 |
| Dynabeads M-280 Sheep anti-Mouse IgG beads | Invitrogen | 11202D |
| Dynabeads Protein A beads | Invitrogen | 10001D |
| BSA Fraction V | GIBCO | 15260-037 |
| RNase A | Thermo Fisher Scientific | EN0531 |
| Proteinase K | Thermo Fisher Scientific | AM2546 |
| Unmethylated Lambda phage DNA | Promega | D1521 |
| Paraformaldehyde | Nacalai Tesque | 26126-25 |
| BSA | SIGMA | A7030 |
| ibidi mounting medium | ibidi | ib50001 |
| Poly-d-Lysine | SIGMA | P6407 |
| **Critical commercial assays** | | |
| RNeasy Micro Kit | QIAGEN | 74004 |
| TruSeq Stranded mRNA LT Sample Prep Kit | Illumina | RS-122-2101 |
| QIA quick PCR purification kit | QIAGEN | 28104 |
| KAPA Hyper Prep Kit | Roche | KK8502 |
| Adaptor kit | FastGene | FG-NGSAD24 |
| KAPA Library Quantification Kit | Roche | KK4824 |
| All Prep DNA/RNA mini kit | QIAGEN | 80204 |

*(Continued on next page)*

***Continued***

| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
|---|---|---|
| **Deposited data** | | |
| RNA-seq data | This paper | GEO: GSE165869 |
| ChIP-seq data | This paper | GEO: GSE165869 |
| Whole genome bisulfite sequence data | This paper | GEO: GSE165869 |
| Whole genome sequence data | This paper | JGA: JGAS000310 |
| ChIP-seq data | Sloan et al., 2016 | https://www.encodeproject.org/ |
| ChIP-seq data | Vallot et al., 2015 | GSE62562 |
| ChIP-seq data | Nozawa et al., 2013 | DRA000894 |
| ChIP-seq data | Tsankov et al., 2015 | GSE61475 |
| DamID-seq data | Dekker et al., 2017 | https://data.4dnucleome.org/ |
| RNA-seq data | Kojima et al., 2017 | GSE99350 |
| RNA-seq data | Yamashiro et al., 2018 | GSE117101 |
| Whole genome bisulfite sequence data | Okae et al., 2014 | JGAS00000000006 |
| **Experimental models: Cell lines** | | |
| Human iPSC clones | Listed in Table S1 | Table S1 |
| **Software and algorithms** | | |
| Imaris 9.0 software | Bitplane | N/A |
| Tophat v2.1.0 | Kim et al., 2013 | https://ccb.jhu.edu/software/tophat/index.shtml |
| Bowtie2 v2.2.7 (RNA-seq) Bowtie2 v2.2.6 (ChIP-seq) | Langmead and Salzberg, 2012 | http://bowtie-bio.sourceforge.net/bowtie2/index.shtml |
| Cufflinks v2.2.1 | Trapnell et al., 2010 | http://cole-trapnell-lab.github.io/cufflinks/ |
| SAMtools v1.3 | Li et al., 2009 | http://www.htslib.org/ |
| IGVTools v2.3.52 | Robinson et al., 2011 | https://software.broadinstitute.org/software/igv/igvtools |
| wigToBigWig tool | http://hgdownload.cse.ucsc.edu/admin/exe/ | N/A |
| Cutadapt v1.9.1 (ChIP-seq) Cutadapt v1.8.1 (WGS) | https://cutadapt.readthedocs.io/en/stable/index.html | N/A |
| FastQC v0.11.4 | https://www.bioinformatics.babraham.ac.uk/projects/fastqc/ | N/A |
| Picard-tools v2.1.0 (ChIP-seq) Picard-tools v1.134 (WGS) | https://broadinstitute.github.io/picard/ | N/A |
| BEDtools v2.25.0 | Quinlan and Hall, 2010 | https://bedtools.readthedocs.io/en/latest/ |
| MACS2 v2.1.1 | Zhang et al., 2008 | https://github.com/macs3-project/MACS |
| epic2 v0.0.41 | Stovner and Saetrom, 2019 | https://github.com/biocore-ntnu/epic2 |
| ChromHMM v1.12 | Ernst and Kellis, 2017 | http://compbio.mit.edu/ChromHMM/ |
| Trim_galore v0.4.1 | https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/ | N/A |
| Bismark v0.17.0 | Krueger and Andrews, 2011 | https://www.bioinformatics.babraham.ac.uk/projects/bismark/ |
| HOMER v4.10.4 | Heinz et al., 2010 | http://homer.ucsd.edu/homer/ |
| BWA-MEM v0.7.12 | Li and Durbin, 2010 | http://bio-bwa.sourceforge.net/ |

# Cell Reports
## Article

**Continued**

| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
|---|---|---|
| GATK tools v3.4-0 | https://gatk.broadinstitute.org/hc/en-us | N/A |
| iSAAC-04.18.11.09 | https://github.com/Illumina/Isaac4 | N/A |
| Strelka v2.9.10 | https://github.com/Illumina/strelka/ | N/A |
| Control-FREEC v11.5 | Boeva et al., 2012 | https://github.com/BoevaLab/FREEC |
| DAVID 6.8 | Huang et al., 2009 | https://david.ncifcrf.gov/summary.jsp |
| R software v3.6.2 | https://www.r-project.org/ | N/A |
| gplots | https://github.com/talgalili/gplots | N/A |
| corrplot | https://github.com/taiyun/corrplot | N/A |
| ggbio | Yin et al., 2012 | https://www.bioconductor.org/packages/release/bioc/html/ggbio.html |
| GenomicRanges | Lawrence et al., 2013 | https://bioconductor.org/packages/release/bioc/html/GenomicRanges.html |
| ggplot2 | https://github.com/tidyverse/ggplot2 | N/A |
| ggpointdensity | https://cran.r-project.org/web/packages/ggpointdensity/index | N/A |
| deepTools v3.4.3 | Ramírez et al., 2016 | https://deeptools.readthedocs.io/en/develop/ |
| **Other** | | |
| Picoruptor | Diagenode | N/A |
| LabChIP GX | PerkinElmer | N/A |
| LSM780 confocal microscope | Zeiss | N/A |

## RESOURCE AVAILABILITY

### Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Shihori Yokobayashi (yokobayashi@anat2.med.kyoto-u.ac.jp).

### Materials availability

This study did not generate new unique reagents.

### Data and code availability

- The RNA-seq, ChIP-seq, and WGBS data generated in this study were deposited in the NCBI database (GEO: GSE165869). The WGS data generated in this study were deposited in the JGA database (JGA: JGAS000310).
- This paper does not report original code. Published analysis code used in this study are available on request from the lead author.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

## EXPERIMENTAL MODEL AND SUBJECT DETAILS

The hiPSC lines used in this study with their detailed description, including sex and passage numbers, were provided in Table S1. hiPSCs were maintained under a feeder-free condition (StemFit AK03 or AK03N [Ajinomoto] medium on recombinant LAMININ511 [iMatrix-511; Nippi]-coated cell culture plates) (Nakagawa et al., 2014). For single-cell dissociation for the passage, the cells were treated with 0.5x TryPLE Select (Life Technologies) solution in 0.5 mM EDTA/PBS for 8-10 min at 37°C and harvested with cell scrapers. 10 μM of Rho-associated coiled coil-forming kinase (ROCK) inhibitor (Y-27632; Wako Pure Chemical Industries) was added and the treatment

was continued for 24 h after seeding. hiPSCs were cultured at 37°C under an atmosphere of 5% CO2 in air. All the experiments were performed under the ethical guidelines of Kyoto University. The experiments using human iPSCs were approved by the Institutional Review Board of Kyoto University and the Ministry of Education, Culture, Sports, Science and Technology (MEXT) of Japan.

## METHOD DETAILS

### Stranded RNA-seq library preparation and sequencing
Total RNA was extracted from harvested hiPSCs ($\sim$4 × 10^5) by using RNeasy Micro Kit columns (QIAGEN) following the manufacturer's guidelines. RNA-seq libraries were prepared using a TruSeq Stranded mRNA LT Sample Prep Kit (Illumina). Libraries were sequenced as paired-end reads (2 × 71 bp or 72 bp) on an Illumina HiSeq 2500 platform (rapid mode).

### ChIP-seq library preparation and sequencing
The ChIP experiment was performed as previously described (Lee et al., 2006) with minor modifications. Harvested hiPSCs were crosslinked with 1% formaldehyde (methanol-free; Thermo Fisher Scientific) in PBS for 10 min at room temperature and quenched with 125 mM glycine. Crosslinked cells were incubated in lysis buffer 1 (50 mM HEPES-KOH pH 7.5, 1 mM EDTA, 140 mM NaCl, 10% glycerol, 0.5% NP-40, 0.25% Triton X-100, proteinase inhibitors) for 10 min at 4°C. After centrifugation, the pellets were resuspended and incubated in lysis buffer 2 (10 mM Tris-HCl pH 8, 1 mM EDTA, 0.5 mM EGTA, 200 mM NaCl, proteinase inhibitors) for 10 min at 4°C. After centrifugation, the pellets were resuspended and incubated in lysis buffer 3 (50 mM Tris-HCl pH 8, 1 mM EDTA, 0.5 mM EGTA, 100 mM NaCl, 0.1% Na-Deoxycholate, 0.5% N-lauroylsarcosine, proteinase inhibitors) for 10 min at 4°C. Chromatin was sonicated using a Picoruptor (Diagenode) to achieve a mean DNA fragment size of around 200$\sim$400 base pairs. 1/10 volume of Triton X-100 was added and the supernatants (sonicated chromatin) after centrifugation were used for immunoprecipitation.

Sonicated chromatin was diluted in lysis buffer 3 to an optimal volume for antibody reaction and pre-cleared for 35 min at 4°C using Dynabeads M-280 Sheep anti-Mouse IgG beads (Invitrogen) or Dynabeads Protein A beads (Invitrogen) blocked with 0.5% BSA in PBS. Pre-cleared chromatin was then incubated with antibody that was pre-incubated with the appropriate Dynabeads in 0.5% BSA in PBS overnight at 4°C, as follows: chromatin equivalent of 5 × 10^5 cells with anti-H3K4me3 (mouse, gift from H. Kimura, 2.5 μl) (Kimura et al., 2008), anti-H3K27ac (mouse, gift from H. Kimura, 2.5 μl) (Kimura et al., 2008), anti-H3K9me3 (mouse, gift from H. Kimura, 2.5 μl) (Hayashi-Takanaka et al., 2011), anti-H3K27me3 (mouse, gift from H. Kimura, 2.5 μl) (Hayashi-Takanaka et al., 2011), and anti-H3K9me2 (mouse, MBL, MABI0307, 5 μl); 2 × 10^6 cells with anti-H2AK119ub1 (rabbit monoclonal, CST, D27C4, 10 μl); and 1 × 10^6 cells with anti-CTCF (rabbit monoclonal, CST, D31H2, 5 μl). After the incubation for 6 h at 4°C, the beads were washed 4 times in wash buffer 1 (20 mM Tris-HCl pH 8, 2 mM EDTA, 150 mM NaCl, 1% Triton X-100, 0.1% SDS), twice in wash buffer 2 (20 mM Tris-HCl pH 8, 2 mM EDTA, 500 mM NaCl, 1% Triton X-100, 0.1% SDS), and twice in wash buffer 3 (10 mM Tris-HCl pH 8, 1 mM EDTA, 250 mM LiCl, 1% Na-Deoxycholate, 1% NP-40). The washed beads were eluted in 10 mM Tris-HCl pH 8, 5 mM EDTA, 300 mM NaCl, and 1% SDS, and crosslinks were reversed overnight at 65°C. Input samples (chromatin equivalent of 2 × 10^5 cells) were treated in a similar manner. The following day, IP and Input samples were incubated with RNase A (Thermo Fisher Scientific) for 25 min at 37°C, and proteinase K (Thermo Fisher Scientific) for 40 min at 55°C. IP or Input DNA was purified using the QIA quick PCR purification kit (QIAGEN).

ChIP-seq libraries were prepared using a KAPA Hyper Prep Kit (Roche) following the manufacturer's guidelines. An adaptor kit (FastGene) was used for the sample indexes. The average size and concentration of library samples were analyzed using a LabChIP GX (PerkinElmer) and KAPA Library Quantification Kit (Roche), respectively. Libraries were sequenced as 50 or 75 bp paired-end reads on an Illumina HiSeq 2500 platform.

### WGBS library preparation and sequencing
WGBS libraries were generated with a post-bisulfite adaptor tagging (PBAT) method (Miura et al., 2012; Shirane et al., 2016). Genomic DNA of hiPSCs ($\sim$4 × 10^5) was purified using an All Prep DNA/RNA mini kit (QIAGEN) following the manufacturer's guidelines. Unmethylated lambda phage DNA (1/200 of the estimated gDNA mass) (Promega) was spiked into the samples, and then the samples were subjected to bisulfite conversions and library constructions for amplification-free WGBS analyses (Miura et al., 2012). Libraries were sequenced as 104 bp single-end reads on an Illumina HiSeq 2500 platform.

### Whole genome sequencing
Genomic DNA extraction was performed manually. The harvested hiPSCs ($\sim$1 × 10^6) were lysed in 10 mM Tris-HCl pH 8.0, 10 mM EDTA, 150 mM NaCl, 0.1% SDS, and proteinase K overnight at 55°C, and then treated with RNase A for 1 h at 37°C. gDNA was purified by a series of phenol, phenol/chloroform/isoamyl alcohol, and chloroform extractions, followed by isopropanol precipitation. gDNA was dissolved in TE buffer (10 mM Tris-HCl pH 8.0, 1 mM EDTA). WGS library preparation and sequencing were performed at Macrogen Inc. (https://www.macrogen.com).

### Immuno-DNA FISH
Immuno-DNA FISH experiments were performed as described previously (Terranova et al., 2008). Human iPSCs were plated on coverslips pre-treated with poly-d-lysine (SIGMA) and cultured for several days. Cells were fixed in 3% paraformaldehyde in PBS pH 7.4 for 10 min, permeabilized with 0.5% Triton X-100 in PBS for 5 min on ice, blocked in PBS containing 1% BSA for 30 min,

# Cell Reports
## Article

and then incubated with the primary antibody, anti-H3K9me3 (MBL, MABI0318, 1:250), and anti-H3K27me3 (MBL, MABI0323, 1:250) overnight at 4°C, followed by secondary antibody anti-mouse IgG-Alexa 488 for 1 h. All immunostaining steps were performed at room temperature unless otherwise stated, and excess antibodies were washed in PBS three times for 10 min each. The stained cells were post-fixed in 3% PFA for 10 min at RT, permeabilized with 0.7% Triton X-100/0.1M HCl in PBS for 10 min on ice, rinsed in 2 × SSC, and denatured for 30 min in 2 × SSC/70% formamide at 80°C. The cells were then washed in cold 2 × SSC, and hybridized with denatured XCyting Chromosome Paints Orange (Metasystems) overnight at 37°C in a humid chamber. Excess probe was eliminated through two washes in 2 × SSC (37°C for 30 min each), followed by one wash in 1 × SSC (RT for 30 min) and one wash in 0.5x SSC (RT for 30 min). After staining with 250 ng/ml DAPI in 0.5x SSC, samples were mounted in ibidi mounting medium (ibidi) prior to microscopy analysis. Images were acquired using a Zeiss LSM780 confocal microscope and analyzed using Imaris 9.0 software (Bitplane).

### RNA-seq data processing

The human genome sequence (GRCh38.p2, chromosomes 1 to 22 and X) was obtained from the NCBI ftp site (ftp://ftp.ncbi.nlm.nih.gov/). The transcript annotation (gencode.v25.annotation.gff3) was obtained from the GENCODE site (https://www.gencodegenes.org/). Paired-end reads were aligned using Tophat v2.1.0/Bowtie2 v2.2.7 with the "-no-coverage-search" option (Kim et al., 2013; Langmead and Salzberg, 2012), followed by cufflinks v2.2.1 with the "-compatible-hits-norm," "-library-type fr-firststrand," and "-max-mle-iterations 50000" options (Trapnell et al., 2010) to estimate fragments per kbp of transcript per million mapped reads (FPKM) per gene. BAM files including R1 or R2 reads were created separately using SAMtools v1.3 (Li et al., 2009); bigWig files were created from the same BAM files using IGVTools v2.3.52 (Robinson et al., 2011) with the "-strands read" option and the wigToBigWig tool (http://hgdownload.cse.ucsc.edu/admin/exe/), with normalization by million mapped reads.

### ChIP-seq data processing

Paired-end reads were processed using Cutadapt v1.9.1 (https://cutadapt.readthedocs.io/en/stable/) to obtain 50 bp reads uniformly. The GC content per read was analyzed using FastQC v0.11.4 (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). The truncated reads were aligned to the human genome (GRCh38.p2, chromosome 1 to 22 and X) using Bowtie2 v2.2.6 with the "-very-sensitive" and "-X 2000" options (Langmead and Salzberg, 2012). Duplicate reads were removed using the MarkDuplicates command of Picard-tools v2.1.0 (https://broadinstitute.github.io/picard/) with the "REMOVE_DUPLICATES = true" option, and proper paired reads were extracted using SAMtools v1.3 (Li et al., 2009) with the "-f 0x2" option. The BAM files were converted to BEDPE format using the BEDtools v2.25.0 (Quinlan & Hall, 2010).

Identification of regions enriched by epigenetic marks was performed using peak calling tools. For H3K4me3 and CTCF, macs2 v2.1.1 (Zhang et al., 2008) was used with the "BAMPE" option. For H3K27ac, H3K27me3, H3K9me3, H3K9me2 and H2AK119ub1, the macs2 with the "BAMPE," "-q 0.01," "-broad," and "-broad-cutoff 0.05" options, and epic2 v0.0.41 (Stovner and Saetrom, 2019) were used. The scores of the fraction of reads in peaks (FRiP) were calculated for each IP sample with each peak caller. The merged peak regions for each epigenetic mark were generated by using the BEDtools merge command on peak regions of each sample called by the macs2, unless otherwise stated.

The levels of IP or Input reads over genome-wide 2 kb bins were counted, and normalized by the window size (2 kb) and total million mapped reads (FPKM). For the analysis of autosomes, the window size (2 kb) and million mapped reads to chromosome 1 to 22 were used for normalization. The 2 kb bins whose $\log_2$ (FPKM + 0.01) $\leq -3$ in all Input samples were excluded. The 2 kb bins overlapping with the centromeres were also excluded. For the calculation of enrichment values (IP/Input), the higher of $\log_2$ (FPKM + 0.01) of the corresponding region or the surrounding 10 kb region was used as an Input value. The 2 kb bins overlapping the merged peak regions were defined for each epigenetic mark using the BEDtools intersect command with the "-f 0.10" option ($\geq$ 200 bp).

The annotations for transcription start sites (TSSs) and CpG islands were obtained from the UCSC browser (http://genome.ucsc.edu/). Promoter regions were defined as regions between 900 bp upstream and 400 bp downstream of transcription start sites (TSSs). High-, intermediate-, and low-CpG promoters (HCP, ICP, and LCP, respectively) were calculated as described previously (Borgel et al., 2010). The summit positions of H3K4me3 peaks of each sample were aggregated and merged if within 200 bp. The levels of IP or Input reads over TSSs (±1 kb), CGIs, and H3K4me3 peaks (±500 bp from the center of merged summits), and the enrichment values (IP/Input) were calculated as described above. The RepeatMasker annotation (http://genome.ucsc.edu/) was used for the analysis of transposable elements, and the Dfam database (https://dfam.org/) was used to obtain the taxon of each TE subfamily. The genome-wide association of the combination of epigenetic marks was analyzed using ChromHMM v1.12 (Ernst and Kellis, 2017).

### Identification of DERs

Differentially enriched regions (DERs) were determined per 2 kb bin overlapping the merged peak regions of each epigenetic mark. For the enrichment value $E$ on bin $x$ in sample $i$ (hiPSC line A) and $j$ (hiPSC line B) of a certain epigenetic mark, we scored the difference between those two samples as follows.

$$E = IP_{\log_2(FPKM + 0.1)} - Input_{\log_2(FPKM + 0.1)}$$

$$if\ E(s_{ix}) - E(s_{jx}) \le -1\ or \ge 1,\ Pair(s_{ix},\ s_{jx}) = 1, else\ Pair(s_{ix},\ s_{jx}) = 0$$

We defined bin $x$ as "DER" if the following equation was satisfied.

$$D_x = \sum_{i,j} Pair(s_{ix},\ s_{jx}) * WF,\ D_x \ge 1$$

Here, $WF$ is a weighting factor to compensate for the difference in the presence or absence of biological replicates (Tables S1 and S3): 1 if neither sample $i$ nor $j$ has another biological replicate of the epigenetic mark; 0.5 if either sample $i$ or $j$ does; 0.25 if both sample $i$ and $j$ do. The differences contributed by a given hiPSC line $i$, $D_{(x,\ clone\ i)}$, were calculated in a similar manner, but with the sample pairs including the hiPSC line $i$. Among DERs, we defined bin $x$ as "multi" if the following equation was satisfied for all hiPSC lines.

$$D_x - D_{(x,\ clone\ i)} \ge 1$$

DERs were analyzed using all hiPSC lines: seven hiPSC lines for H3K4me3, H3K27ac, H3K27me3, H3K9me3, and CTCF; five hiPSC lines (M1/F1/F2/F3/F4) for H3K9me2 and H2AK119ub1. Exceptionally, in Figure 6D, DERs were analyzed using five hiPSC lines (M2/F2/F3/F4/F5) for H3K4me3, H3K27ac, H3K27me3, H3K9me3, and CTCF; three hiPSC lines (F2/F3/F4) for H3K9me2 and H2AK119ub1.

The DERs in TSSs (±1 kb), CGIs, and H3K4me3 peak summits (±500 bp) were determined in a similar manner. The genomic features of peak regions, DERs, or multi-DERs in autosomal 2 kb bins were examined using the annotatePeaks.pl program of HOMER v4.10.4 (Heinz et al., 2010).

### WGBS data processing

The read processing, mapping, and estimation of the methylated C levels were performed as described previously (Shirane et al., 2016). Single-end reads were processed with the Trim_galore program (https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/) to remove four bases from the 5′-ends, one base from the 3′-ends, adaptor sequences, and low-quality reads (quality score < 20). The processed reads were aligned to the human genome (GRCh38.p2, chromosome 1 to 22 and X) with the bismark program from Bismark v0.17.0 with the "–pbat" option (Krueger and Andrews, 2011). Mapped data (BAM format) were converted to methylation levels using the bismark_methylation_extractor program from Bismark.

Only CpGs in which the read depth is between 4 and 200 are used for % methylated C (mC) calculations. Only regions in which CpG (read depth $\ge$ 4) counts were four or more within genome-wide 2 kb bins were used for calculating averages. For the analysis of CpA, CpC, and CpT methylations, percent methylations were calculated as % mC per kb as follows: The number of methylation calls over genome-wide 1 kb bins was summed and divided by the sum of the number of unmethylation calls in the same region. The methylation levels per kb for CpA, CpC, and CpT were calculated separately throughout the whole genome.

### Analysis of H3K9me3 domains

To define large H3K9me3 domains, H3K9me3 peaks of each hiPSC sample called by epic2 software were merged using BEDtools; and then the merged peaks over 100 kb were extracted and further merged if within 100 kb using BEDtools. Next, the levels of Input reads on the defined H3K9me3 domains (n = 254) were examined; the 145 domains with lower variations among Input samples were determined as "class2" as described in Figure S4, and used for downstream analysis. The median of the enrichment values of 2 kb bins overlapping with each domain was used as the level of H3K9me3 abundance in each hiPSC line. The median of the mCG/CG values of 2 kb bins overlapping with each domain, or the median of the mCH/CH values of 1 kb bins overlapping with each domain, was used as the level of methylated DNA over H3K9me3 domains in each hiPSC line.

The BED files for lamina associated domains (LADs) of five human cell types (H1 ESCs, HFFc6, HCT116, K562, and hTERT-RPE1 cells) examined by DamID-seq experiments were downloaded from the 4DN Data Portal (Dekker et al., 2017; Table S5). The ratio of the region covered by LADs in each H3K9me3 domain was calculated and plotted in Figure 5E.

The start or end positions of H3K9me3 peaks were used for the analysis of the boundary regions of H3K9me3. H3K9me3 peaks called by epic2 software in individual samples were grouped by the length of the merged peaks to which they belonged. Randomly chosen regions were obtained from the autosomal genomic regions using the shuffle command of BEDtools with the "-noOverlapping" option, and used as a control. The frequencies of the observation of H3K4me3 peak summits (±500 bp) or CTCF peak summits (±500 bp) at the H3K9me3 boundary regions (±1 kb) were counted using BEDtools. The enrichment of known motifs was examined using the findMotifsGenome.pl program of HOMER v4.10.4 (Heinz et al., 2010). For this analysis, the regions between −1000 bp and +200 bp of the start positions and −200 bp and +1000 bp of the end positions were used as the H3K9me3 boundary regions. Regarding the public TF ChIP-seq data, the bed files of peak regions were obtained and processed as described above.

### SNP analysis for X-linked genes

The WGS raw data on 201B7 (shared by the Genome Analysis team in CiRA Foundation) were processed using the cutadapt v1.8.1 (https://cutadapt.readthedocs.io/en/stable/) to remove low quality bases and adaptor sequences. The trimmed reads were mapped

to the human reference genome (hg 38, chromosome 1 to 22, X, and M) using bwa mem v0.7.12 (Li and Durbin, 2010). The uniquely and properly mapped reads were filtered with samtools-1.2 and the duplicate reads were removed with picard-tools-1.134. The remaining reads were subjected to reprocessing steps including local realignment and base quality recalibration using GATK tools (v3.4-0). Finally, the VCF files were generated by the HaplotypeCaller tool in GATK (v3.4-0). The SNPs were first extracted from the VCF files with GT = "0/1" and GQ $\geq$ 20 settings. Then the SNPs located inside of exons (gencode.v25.annotation.gff3) were extracted.

The WGS data acquired in this study (1205A, 1381B1, 1390G3, and 1390C1) were processed according to the analysis pipeline of Macrogen Inc. (https://www.macrogen-japan.co.jp/next_seq_0602.php). First, after the removal of low quality bases and adaptor sequences, paired-end sequences were mapped to the human reference genome (original GRCh38 from NCBI, Dec. 2013) using iSAAC aligner (iSAAC-04.18.11.09). Strelka (2.9.10) was performed to identify single-nucleotide variants (SNVs) and short insertions and deletions (Indels). Through read processing, low quality reads and PCR duplicates were filtered out, and read realignment was carried out to increase accuracy. The SNPs were extracted from the generated VCF files with VARTYPE = SNP, HET, and GQ $\geq$ 20 settings. Then the SNPs located inside of exons (gencode.v25.annotation.gff3) were extracted.

RNA-seq reads covering each exon-SNP position were counted using the mpileup function of SAMtools v1.3. The SNPs having five or more valid reads were used for downstream analysis. The ratio of the reference or alternate base to the total base counts was calculated for each SNP, and the base showing a higher ratio was assigned to "Allele 1." For genes having more than one SNP, the median and minimum values of "Allele-1-ratio" were calculated per gene. The median values of "Allele-1-ratio" were used for the plots in Figures 2B and 2E. The minimum values of "Allele-1-ratio" were used for calling the allelic expression status ("mono" or "bi") in Figure 2F: if the value was greater than 0.9, the gene was classified as mono-allelic.

### Analysis of CNVs, INDELs, and SNPs in the autosomes

The lists of CNVs, INDELs, and SNPs of five hiPSC lines (1383D6, 1205A, 1381B1, 1390G3, and 1390C1) were generated according to the analysis pipeline of Macrogen Inc. (https://www.macrogen-japan.co.jp/next_seq_0602.php). Briefly, Strelka (2.9.10) was performed to identify single-nucleotide variants (SNVs) and short insertions and deletions (Indels). Control-FREEC (11.5) (Boeva et al., 2012) was performed to identify copy number variant with 10,000 bp window size and no additional options. Control-FREEC also used GC-content bias to normalize read counts and XY for sex. The regions of CNVs of each hiPSC line were merged and examined for the overlap with respect to autosomal 2kb bins. The positions of SNPs or INDELs found in each hiPSC line were pooled together and counted per autosomal 2 kb bin, 1 kb bin, or the region of interest.

### Association analysis with the clonal PGCLC induction efficiency

The correlation coefficient (Pearson's R) between the $\log_2$-transformed median values of PGCLC induction rates (Yokobayashi et al., 2017) and ChIP enrichment values (IP/Input, $\log_2$ transformed) was calculated for each 2 kb bin overlapped with the DERs of H3K4me3, H3K27ac, H3K27me3, H3K9me3, and CTCF. In the case of bins on chromosome X, the data of female hiPSC lines were used for the calculation. The annotations for gene regulatory elements were obtained from the GeneHancer database (Fishilevich et al., 2017). Gene ontology (GO) analysis was performed using DAVID 6.8 (Huang et al., 2009). The expression profiles of hiPSCs, iMeLCs, hPGCLCs, and hPGCLCs in xenogeneic reconstituted ovaries, using the 585B1 reporter hiPSC line bearing the *BLIMP1/PRDM1-2A-tdTomato* (BT) and *TFAP2C-2A-EGFP* (AG) alleles (Sasaki et al., 2015), were acquired from previous studies and processed accordingly (Kojima et al., 2017; Yamashiro et al., 2018).

### Public data analysis

All the public data used in this study are listed in Table S5 and Key resources table. All the read data of ChIP-seq experiments were processed using the same procedure as described above.

### QUANTIFICATION AND STATISTICAL ANALYSIS

All statistical details can be found in the Figure legends, Figures, Results, and relevant STAR Methods sections, including the statistical test used, number of genes/nuclei/regions analyzed, and dispersion and precision measures (mean, median, SD, and confidence intervals).

The following analyses were performed using R software version 3.6.2 (https://www.r-project.org/). The mean values of replicate #1 and #2 were used in the samples with biological replicates (listed in Table S3) unless otherwise stated. Heatmaps were created using the heatmap.2 function in the gplots package together with the RColorBrewer package. Unsupervised hierarchical clustering was performed with the dist function with the method = "euclidean" setting and with the hclust function with the method = "ward.D2" setting. Correlation matrices were created using the corrplot package. The cor function was used with the method = "pearson" and use = "pairwise.complete.obs" settings. Chromosome ideograms were created using the ggbio and GenomicRanges packages (Lawrence et al., 2013; Yin et al., 2012). The plot function, the ggplot2 package, and the ggpointdensity package were used for the creation of plots. The sample function was applied for creating the plots in Figure 6B-b (n = 150,000) and Figure S1E (n = 100,000). Loess regression curves were drawn using the stat_smooth function in the ggplot2 package. A hypergeometric test was carried out using the phyper function with the lower.tail = FALSE setting for testing over-representation. The CDF (cumulative

CellPress
OPEN ACCESS

Cell Reports
Article

distribution function) plots were created using the stat_ecdf function with the geom = "step" setting in the ggplot2 package. Welch two sample t test was carried out using the t.test function with the default setting. Mann-Whitney $U$ test was carried out using the wilcox.test function with the default setting.

Heatmap representation and profile plots of the enrichment levels of epigenetic marks over genomic regions were created using deepTools (v3.4.3) (Ramírez et al., 2016). Normalized bigWig files were generated using the bamCompare command, with the "-ignoreForNormalization chrX" option for the analysis of autosomal regions. The computeMatrix and plotHeatmap commands were used for creating plots.

ChIP-seq tracks were created as follows. For chromosome-wide representation, the median values of the enrichment levels (IP/Input) per 100 kb bin were calculated based on the enrichment of the overlapping 2 kb bins, and visualized using the ggplot2 package. For megabase-scale representation, 10 kb bins were used instead of 100 kb bins and processed in a similar manner. For sub-megabase-scale representation, the IP levels on bins of an appropriate size were calculated using the computeMatrix command of deepTools, and visualized using the ggplot2 package. RNA-seq tracks were created in a similar manner. In the samples with biological replicates, the mean values of two replicates were used in Figure 1B; the values of replicate #1 were used in Figures 1E, 4D, 5A, 7E, S4B, and S7C (Table S3).