Stem Cell Reports
Article

OPEN ACCESS

De Novo DNA Methylation at Imprinted Loci during Reprogramming into Naive and Primed Pluripotency

Masaki Yagi,[1,2,13] Mio Kabata,[2] Tomoyo Ukai,[1,2] Sho Ohta,[1] Akito Tanaka,[2] Yui Shimada,[1,2] Michihiko Sugimoto,[3,14] Kimi Araki,[3] Keisuke Okita,[2] Knut Woltjen,[2,4] Konrad Hochedlinger,[5,6,7,8,9] Takuya Yamamoto,[2,10,11,12,*] and Yasuhiro Yamada[1,2,10,*]

[1]Division of Stem Cell Pathology, Center for Experimental Medicine and Systems Biology, Institute of Medical Science, University of Tokyo, Tokyo 108-8639, Japan
[2]Department of Life Science Frontiers, Center for iPS Cell Research and Application (CiRA), Kyoto University, Kyoto 606-8507, Japan
[3]Institute of Resource Development and Analysis, Kumamoto University, Kumamoto 860-0811, Japan
[4]Hakubi Center for Advanced Research, Kyoto University, Kyoto 606-8501, Japan
[5]Department of Molecular Biology, Massachusetts General Hospital, Boston, MA 02114, USA
[6]Center for Regenerative Medicine, Massachusetts General Hospital, Boston, MA 02114, USA
[7]Cancer Center, Massachusetts General Hospital, Boston, MA 02114, USA
[8]Department of Stem Cell and Regenerative Biology, Harvard University, Cambridge, MA 02138, USA
[9]Harvard Stem Cell Institute, Cambridge, MA 02138, USA
[10]AMED-CREST, AMED 1-7-1 Otemachi, Chiyodaku, Tokyo 100-0004, Japan
[11]Institute for the Advanced Study of Human Biology (WPI-ASHBi), Kyoto University, Yoshida-Konoe-cho, Sakyo-ku, Kyoto 606-8501, Japan
[12]Medical-risk Avoidance Based on iPS Cells Team, RIKEN Center for Advanced Intelligence Project (AIP), Kyoto 606-8507, Japan
[13]Present address: Department of Molecular Biology, Massachusetts General Hospital, Boston, MA 02114, USA
[14]Present address: RIKEN BioResource Research Center, 3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
*Correspondence: takuya@cira.kyoto-u.ac.jp (T.Y.), yasu@ims.u-tokyo.ac.jp (Y.Y.)
https://doi.org/10.1016/j.stemcr.2019.04.008

SUMMARY

CpG islands (CGIs) including those at imprinting control regions (ICRs) are protected from *de novo* methylation in somatic cells. However, many cancers often exhibit CGI hypermethylation, implying that the machinery is impaired in cancer cells. Here, we conducted a comprehensive analysis of CGI methylation during somatic cell reprogramming. Although most CGIs remain hypomethylated, a small subset of CGIs, particularly at several ICRs, was often *de novo* methylated in reprogrammed pluripotent stem cells (PSCs). Such *de novo* ICR methylation was linked with the silencing of reprogramming factors, which occurs at a late stage of reprogramming. The ICR-preferred CGI hypermethylation was similarly observed in human PSCs. Mechanistically, ablation of *Dnmt3a* prevented PSCs from *de novo* ICR methylation. Notably, the ICR-preferred CGI hypermethylation was observed in pediatric cancers, while adult cancers exhibit genome-wide CGI hypermethylation. These results may have important implications in the pathogenesis of pediatric cancers and the application of PSCs.

INTRODUCTION

Pluripotent stem cells (PSCs) have self-renewing activity and are capable of differentiating into various types of cells, making them invaluable tools for regenerative medicine and disease modeling (Stadtfeld and Hochedlinger, 2010; Yamanaka, 2012). In mice there are two types of pluripotent states, naive and primed. Mouse embryonic stem cells (ESCs) have naive pluripotency and are derived from inner cell mass (ICM) of a blastocyst, while mouse epiblast stem cells (EpiSCs) have primed pluripotency and are derived from post-implantation epiblast (Nichols and Smith, 2009). Naive and primed PSCs display distinct transcriptional and epigenetic profiles with different developmental potential (Nichols and Smith, 2009). Notably, naive and primed PSCs can be established from somatic cells by the enforced expression of defined transcription factors, such as *Oct4*, *Sox2*, *Klf4*, and *c-Myc* (OSKM), under appropriate culture conditions (induced pluripotent stem cells [iPSCs] and induced epiblast stem cells [iEpiSCs], respectively) (Han et al., 2011; Takahashi and Yamanaka, 2006). Although a number of studies have revealed that iPSCs and iEpiSCs display shared molecular characteristics with ESCs and EpiSCs, respectively (Choi et al., 2015; Han et al., 2011; Maherali et al., 2008; Mikkelsen et al., 2008), whether induced naive and primed PSCs faithfully recapitulate *in vivo* pluripotency is unknown.

Genomic imprinting is an essential epigenetic mechanism that controls the monoallelic expression of genes and is mediated by gamete-derived allele specific DNA methylation (Ferguson-Smith, 2011). Imprints are established exclusively in the male or female germline through *de novo* DNA methylation at imprinting control regions (ICRs) (Bourc'his et al., 2001; Kaneda et al., 2004). Established ICR methylation together with the concomitant unmethylated state at the other allele is strictly maintained in somatic cells throughout life (Ferguson-Smith, 2011). Previous studies demonstrated that Dnmt1, a maintenance DNA methyltransferase, together with Uhrf1 is responsible for the preservation of ICR methylation (Branco et al., 2008). In contrast, it is not fully understood how unmethylated allele at ICRs are maintained in the unmethylated

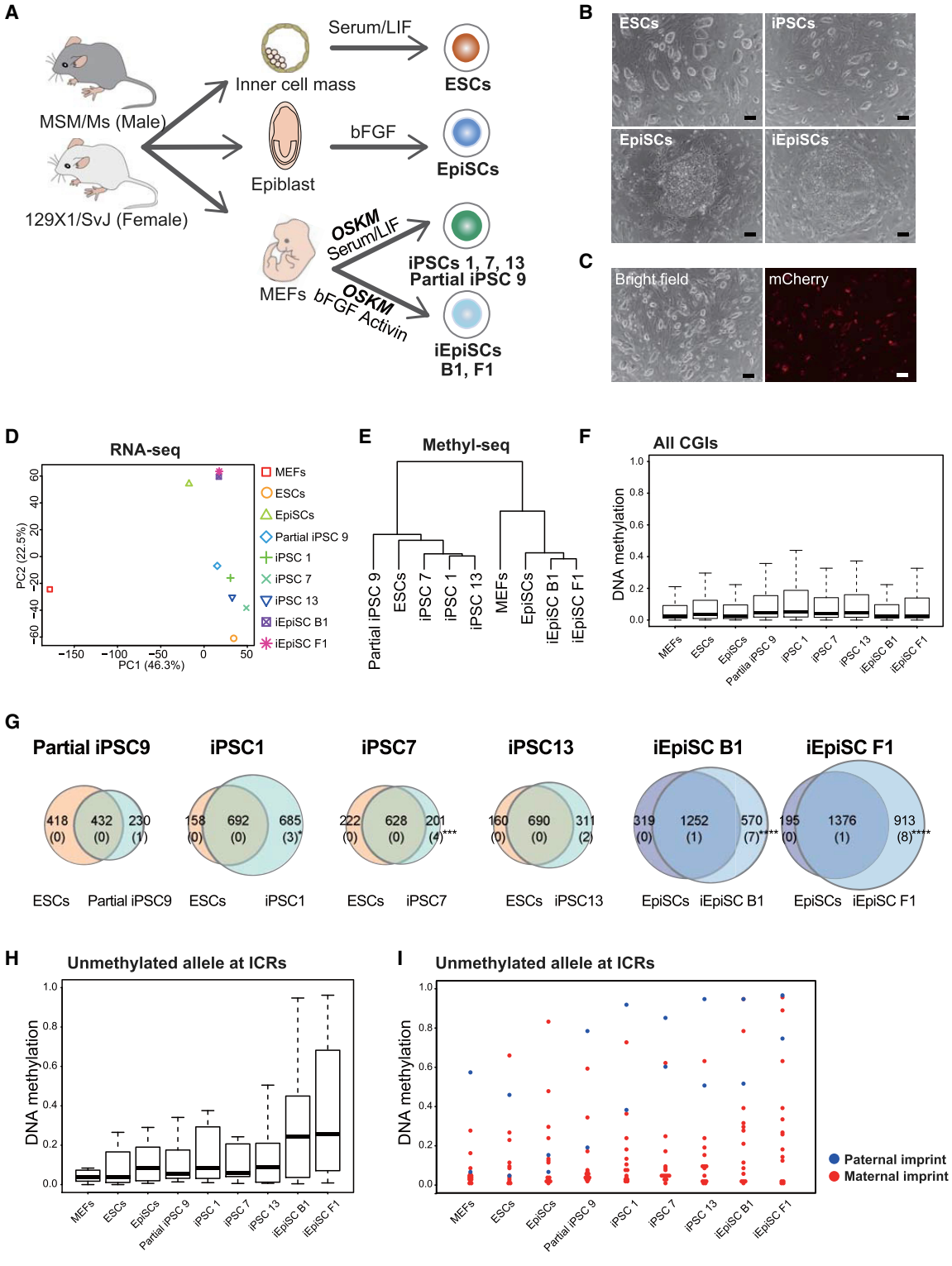


Figure 1. Establishment of Naive and Primed Mouse PSCs and CGI Methylation Analysis

(A) Schematic diagram of the experimental design. Generation of naive and primed PSCs derived directly from embryos or from somatic cells by reprogramming. Parental alleles can be distinguished by SNPs in (129X1/SvJ × MSM/Ms) F1 cells.

(B) Representative images of naive and primed PSCs (ESCs, EpiSCs, iPSCs, iEpiSCs). Scale bars, 100 μm.

(legend continued on next page)

state. It is known that CpG islands (CGIs) are generally protected from *de novo* methylation in somatic cells. Considering that ICRs often consist of CpG-rich regions including CGIs, protection from *de novo* ICR methylation could be mediated through mechanisms whereby CGIs are protected from *de novo* DNA methylation. Notably, cancer cells often display abnormal DNA hypermethylation at both CGIs and ICRs (Sharma et al., 2010), indicating that the machinery for avoiding *de novo* CGI methylation is impaired in cancer cells.

The dysregulation of imprinted genes is implicated in developmental defects and tumorigenesis (Kato et al., 1999; Steenman et al., 1994). Indeed, it has been reported that the dysregulation of imprints compromises the developmental potential of PSCs (Choi et al., 2017b; Yagi et al., 2017a). In addition, CGI hypermethylation in cancer cells are often detectable in tumor-suppressor genes with concomitant transcriptional silencing, which supports the notion that *de novo* CGI methylation plays a role in cancer development. It is therefore important to evaluate the stability of CGI/ICR methylation in PSCs. Several studies have previously examined the status of imprints in mouse and human iPSCs (Bar et al., 2017; Johannesson et al., 2014; Ma et al., 2014; Nazor et al., 2012; Pick et al., 2009). Differentially methylated regions (DMRs) at the *Dlk1-Dio3* gene cluster are often hypermethylated in mouse iPSCs, which is linked with impaired developmental potential (Stadtfeld et al., 2010). A large-scale analysis of allele-specific RNA sequencing (RNA-seq) data revealed that primed human iPSCs display a higher incidence of biallelic expression of imprinted genes (Bar et al., 2017). However, the genome-wide stability of CGI methylation during the reprogramming process of naive and primed pluripotency remains to be fully elucidated.

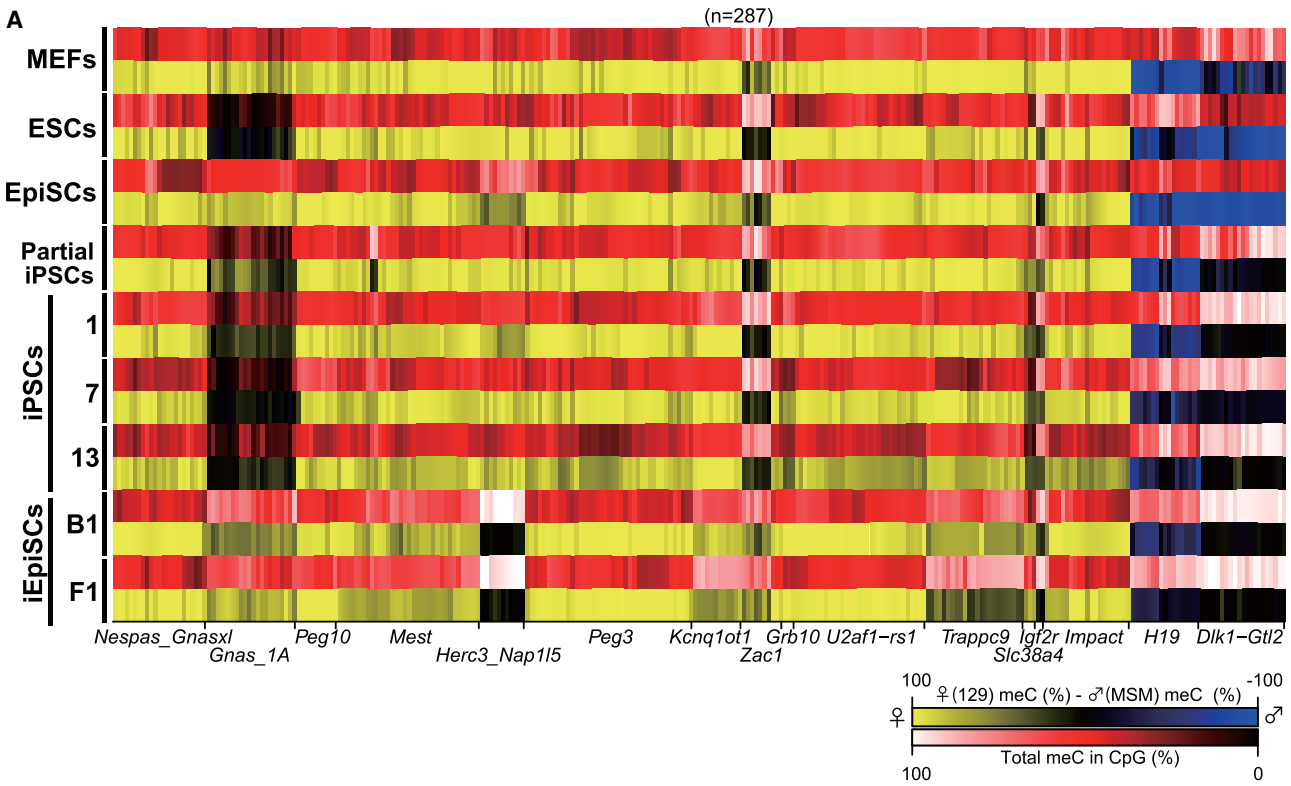
Here we conducted comprehensive methylation analysis for CGIs and ICRs to understand the epigenetic stability in naive and primed PSCs. Given that ICR methylation is highly affected by culture conditions and gender in mouse PSCs (Choi et al., 2017a; Pasque et al., 2018; Yagi et al., 2017a, 2017b), in this study we focus on the male reprogramming process under conventional serum-containing culture conditions. PSCs derived from cells with (129X1/SvJ × MSM/Ms) F1 genetic background allowed us to investigate allele-specific DNA methylation at ICRs by single-nucleotide polymorphisms (SNPs) (Takada et al., 2013; Yagi et al., 2017a). This effort revealed aberrant methylation at several ICRs during the reprogramming of somatic cells into naive and primed pluripotency. Furthermore, our data unveiled similar epigenetic aberrations in pediatric cancers with iPSCs, providing an unappreciated link between reprogramming and childhood cancer development.

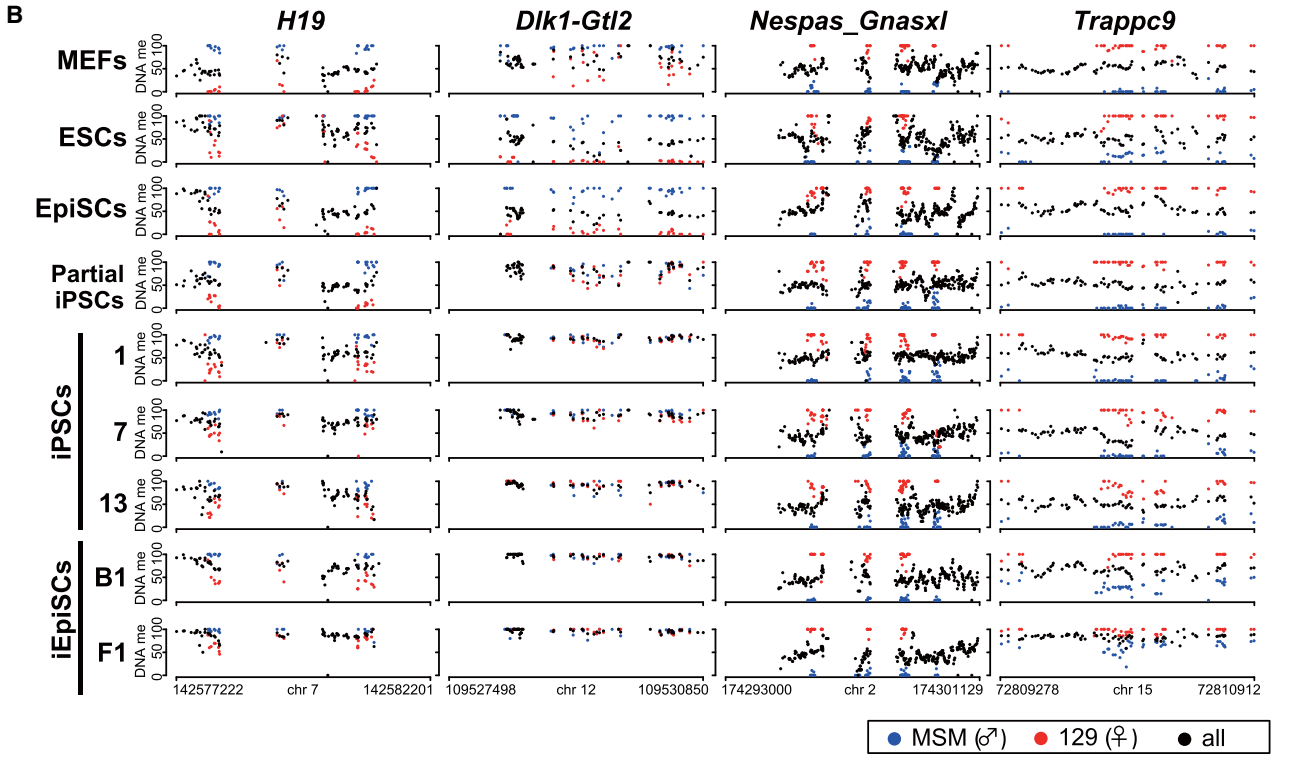
RESULTS

Generation of Naive and Primed Mouse PSCs in which Parental Alleles Are Distinguishable

To elucidate the stability of CGI methylation during reprogramming into naive and primed PSCs, we generated iPSCs and iEpiSCs from male mouse embryonic fibroblasts (MEFs) with *piggyBac* (*PB*) vector containing a doxycycline (Dox)-inducible polycistronic transgene encoding *OSKM* (Kim et al., 2016) (Figures 1A–1C, S1A, and S1B). For their control, we established male ESCs and EpiSCs derived from embryonic day 3.5 (E3.5) blastocyst and E6.5 post-implantation epiblast, respectively (Figures 1A and 1B). These cell lines were derived from (129X1/SvJ × MSM/Ms) F1 MEFs or embryos in which the parental alleles are distinguishable by a large number of SNPs. Early-passage PSCs (passage 3 [p3] to p4) were used to analyze gene expression and DNA methylation in this study. RNA-seq analysis confirmed that the iPSCs and iEpiSCs expressed general pluripotency-associated genes at levels comparable with those of control ESCs and EpiSCs (Figure S1C). iPSCs expressed naive pluripotency-associated genes and iEpiSCs expressed primed pluripotency-associated genes (Figures S1D and S1E). The established iPSC and iEpiSC clones exhibited silencing of transgenes even in the presence of Dox, except for iPSCs 9, 21, and 37, which continuously expressed mCherry, indicating transgene expression (Figures

(C) An image of mouse partial iPSCs. The mCherry signal represents the expression of the *OSKM* transgene. The transgene is not silenced in iPSC 9. Scale bars, 100 μm.
(D) Principal component (PC1 and PC2) analysis of transcriptional profiles by RNA-seq.
(E) Hierarchical clustering analysis of the global DNA methylation status by MethylC-seq.
(F) Box plot of DNA methylation levels at all CGIs in MEFs, ESCs, EpiSCs, iPSCs, and iEpiSCs. Solid lines in each box indicate the median. The bottom and top of the boxes are lower and upper quartiles, respectively. Whiskers extend to ±1.5 interquartile range (IQR).
(G) Venn diagram of CGIs with increased DNA methylation in PSCs compared with MEFs (DNA methylation difference >0.2). Number in parentheses indicates the number of CGIs linked to ICR. Note that ICR-linked CGIs are enriched in reprogrammed PSC-specific methylated CGIs. *$p < 0.05$, ***$p < 0.001$, and ****$p < 0.0001$ (Fisher's test).
(H) Box plot of DNA methylation levels at unmethylated alleles in ICRs in MEFs, ESCs, EpiSCs, iPSCs, and iEpiSCs. Solid lines in each box indicate the median. The bottom and top of the boxes are lower and upper quartiles, respectively. Whiskers extend to ±1.5 IQR.
(I) DNA methylation levels at unmethylated alleles in paternal and maternal ICRs in MEFs, ESCs, EpiSCs, iPSCs, and iEpiSCs.





(legend on next page)

1C and S1F). The silencing of transgenes is critical for achieving complete reprogramming to a stable pluripotent state (Jaenisch and Young, 2008). Consistent with this, iPSC 9 displayed distinct global transcriptional and methylome patterns compared with other ESC/iPSC clones and exhibited a decreased expression level of naive pluripotency-associated genes, which reflects the partial reprogramming state (Figures 1D and 1E). Consistent with this, there were DMRs in partially reprogrammed (partial) iPSC 9 compared with control iPSCs and ESCs (Figure S1G). Naive PSC lines and primed PSC lines were separately clustered in both transcriptome and DNA methylation profiling (Figures 1D and 1E). Overall, the established naive and primed PSC clones by somatic cell reprogramming harbored shared molecular signatures with naive and primed control PSC lines, which were directly derived from embryos.

De Novo DNA Methylation at ICRs in Reprogrammed PSCs

We next investigated genome-wide CGI methylation patterns by conducting target-captured MethylC sequencing (MethylC-seq) analysis of various PSC lines. We first confirmed that MethylC-seq analysis is suitable for the comprehensive analysis of CGI and ICR methylation since the probes capture 94.2% CGIs among all mouse CGIs (21,648 out of 22,948 CGIs). In fact, MethylC-seq analysis had higher sequencing coverage for CGIs and ICRs than whole-genome bisulfite sequencing (WGBS), with comparable numbers of sequencing reads (Figure S2A).

MethylC-seq analysis revealed that most CGIs remained hypomethylated in all PSC lines examined (Figure 1F). However, a small subset of CGIs exhibited increased methylation in reprogrammed PSCs compared with MEFs, which is the origin of the reprogrammed cells (DNA methylation difference >0.2) (Figure 1G). The majority of CGIs with increased methylation in iPSCs and iEpiSCs were similarly methylated in ESCs and EpiSCs, respectively (Figure 1G), which suggests that methylation at these CGIs is cell-type-related methylation in PSCs. Notably, we also observed iPSC/iEpiSC-specific CGI methylation (Figure 1G), suggesting that such CGI methylation is associated with the reprogramming. The reprogramming-associated CGI methylation was observed in various genetic elements (Figure S2B). Of particular interest, CGIs linked to ICRs (n = 27) were significantly enriched within CGIs with reprogramming-associated methylation in most reprogrammed PSC clones ($p < 0.05$ in iPSC 1, $p < 0.001$ in iPSC 7, $p = 0.05466$ in iPSC 13, $p < 0.0001$ in iEpiSC B1, and $p < 0.0001$ in iEpiSC F1, Fisher's test) (Figure 1G). Indeed, 9 out of 27 ICR-linked CGIs exhibited methylation in iEpiSC F1, and 8 of the 9 ICR-linked CGIs were not methylated in embryo-derived EpiSCs (Figure 1G), indicating that ICR methylation is closely related to reprogramming. The comprehensive allele-specific analysis for ICR methylation further confirmed that unmethylated alleles at ICRs were frequently *de novo* methylated in reprogrammed PSCs, a feature especially pronounced in iEpiSCs (Figure 1H). The unmethylated alleles of paternally imprinted ICRs (*H19* DMR and *Dlk1-Gtl2* DMR) were heavily methylated in both iPSCs and iEpiSCs, whereas maternally imprinted ICRs (e.g., *Nap1l5* DMR, *Kcnq1ot1* DMR, *Trappc9* DMR) were often *de novo* methylated in iEpiSCs (Figure 1I).

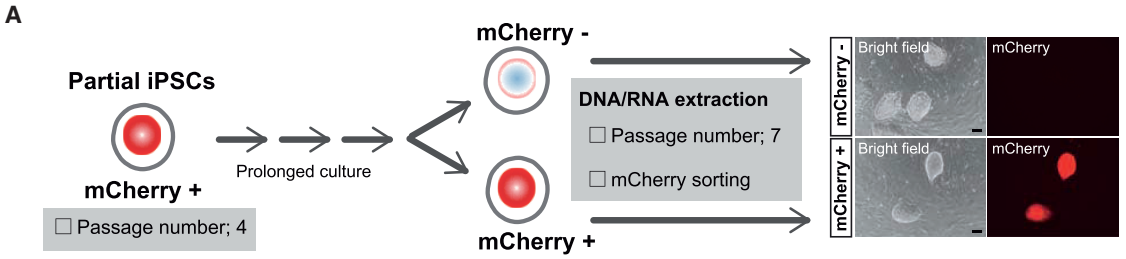
De Novo DNA Methylation at Paternal ICRs during Reprogramming

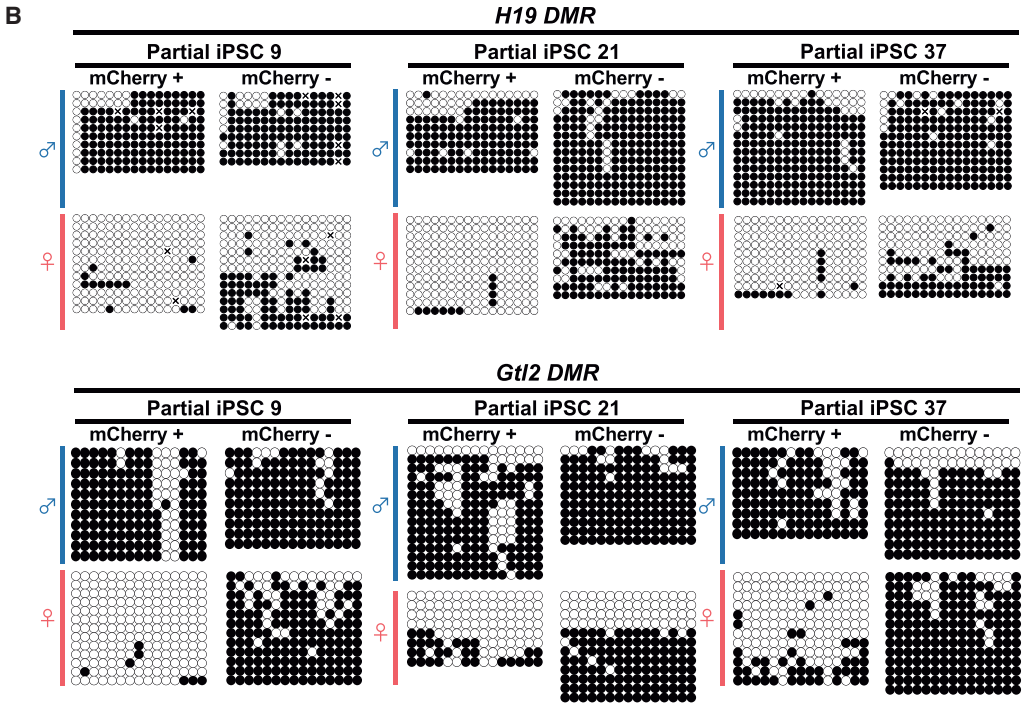
We next generated heatmaps and dot plots showing methylation levels and allelic methylation patterns at ICRs (Figures 2A and 2B). We confirmed that control MEFs retain monoallelic ICR methylation patterns. Consistent with this, the total ICR methylation level in MEFs exhibited approximately 50% as *in vivo* tissues (Figures 2A, 2B, and S2C), except for partial gain of methylation at a part of *Zac1* DMR and *Dlk1-Gtl2* DMR (Figures 2A, 2B, and S2D). In mice, there are three paternal imprinted loci (*Dlk1-Gtl2* DMR, *H19* DMR, and *Rasgrf1* DMR). *De novo* methylation at *Dlk1-Gtl2* DMR was observed not only in iPSC clones but also in iEpiSC clones, but the loci remained unmethylated in both ESCs and EpiSCs (Figures 2A and 2B). Consistent with this observation, *Meg3* and *Rian*, which are regulated by DNA methylation at *Dlk1-Dio3* loci, were repressed in iPSCs and iEpiSCs (Figure S2E). *H19* DMR also acquired *de novo* methylation in reprogrammed PSCs, which was further confirmed with multiple independent PSC clones (Figures 2A, 2B, and S3A). The unmethylated allele of *H19* DMR was similarly methylated in ESCs to some extent, but not in EpiSCs, indicating that *de novo* methylation at *H19* DMR takes place during both the reprogramming and maintenance of naive PSCs (Figures 2A, 2B, and S3A). We next examined the DNA methylation status at *Rasgrf1* DMR, which was not captured by MethylC-seq analysis. *De novo* methylation

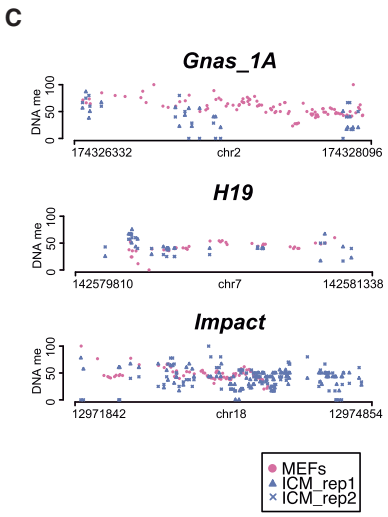
Figure 2. DNA Methylation of ICRs during Reprogramming to Naive and Primed Pluripotency in Mice

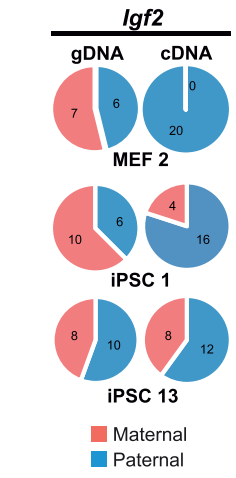
(A) Heatmap for DNA methylation levels and allelic balance at ICRs in MEFs, ESCs, EpiSCs, iPSCs, and iEpiSCs. The heatmap depicts the methylation status at CpG sites in which parental alleles have been distinguished. CpG methylation levels and allelic balance for the methylation are shown for each CpG site. Color scale is shown for DNA methylation levels and allelic balance.

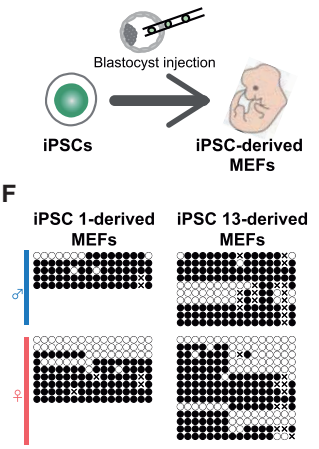
(B) CpG methylation at representative ICRs of MEFs, ESCs, EpiSCs, iPSCs, and iEpiSCs. Each black dot represents a methylation percentage for each CpG site. Red and blue dots indicate methylation levels at maternal 129 allele and paternal MSM allele, respectively.

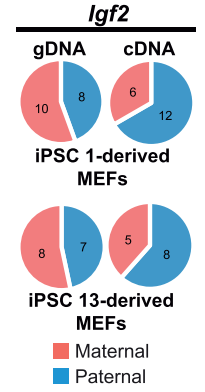












(legend on next page)

at the unmethylated allele of *Rasgrf1* DMR was observed exclusively in iEpiSCs (Figure S3B), providing additional evidence that paternally imprinted DMRs are epigenetically unstable during reprogramming.

De Novo ICR Methylation Is Linked with Silencing of Exogenous Reprogramming Factors

We found that partial iPSC 9 harbors a smaller number of methylated CGIs compared with iPSCs ($p < 0.0001$, Fisher's test) (Figure 1G). Notably, iPSC 9 tended to preserve the monoallelic methylation pattern at ICRs (Figures 1H, 2A, 2B and S3A), which raised the possibility that aberrant ICR *de novo* methylation occurred at a late stage of the reprogramming. To examine this possibility, we extended the culture of iPSC 9 (Figure 3A). We found that a subset of iPSC 9 cells turned into mCherry-negative cells during the prolonged culture. Consistent with the fact that the silencing of reprogramming factors is a critical event for complete reprogramming, mCherry-negative iPSC 9 cells expressed higher levels of naive pluripotency genes (Figure S3C), suggesting that a subset of partial iPSCs converted into fully reprogrammed iPSCs. Most notably, mCherry-negative iPSC 9 cells acquired *de novo* methylation at the unmethylated allele of *H19* DMR and *Gtl2* DMR, while the allele remained unmethylated in mCherry-positive cells (Figure 3B). This transgene silencing-linked *de novo* ICR methylation was similarly observed in partial iPSC clones 21 and 37 (Figure 3B). These data indicate that *de novo* DMR methylation at *H19* and *Gtl2* is coupled with the silencing of exogenous reprogramming factors.

De novo DNA Methylation at a Subset of Maternal ICRs in Primed PSCs

In contrast to the monoallelic DNA methylation pattern at maternally imprinted loci in iPSCs, iEpiSCs exhibited biallelic ICR methylation at a subset of maternal ICRs (e.g., *Nap1l5*, *Trappc9*), whereas embryo-derived EpiSCs showed only a modest increase of methylation at these ICRs (Figures 1I, 2A, and 2B). Hypermethylation at *Nap1l5* DMR was confirmed with multiple iEpiSC clones (Figure S3D). We also found that *Gnas1A* ICR is hypomethylated in naive PSCs (both ESCs and iPSCs) but maintained in primed PSCs (both EpiSCs and iEpiSCs) (Figure 2A). Given that *Gnas1A* ICR is hypomethylated in ICM of preimplantation embryos (Figure 3C), the decreased *Gnas1A* ICR methylation in naive PSCs may reflect the reduced methylation in ICM. Apart from primary ICRs, secondary DMRs (sDMRs) acquire parent-of-origin-dependent DNA methylation patterns after implantation. We found that naive PSCs exhibit reduced methylation levels at *Nespas* sDMR and *Cdkn1c* sDMR, whereas primed PSCs retain monoallelic methylation (Figure S4A), presumably reflecting the DNA methylation status in the *in vivo* counterpart. Collectively, these results suggest that the epigenetic integrity of imprinted DMRs in PSCs is variable and depends on the imprinted loci and two types of pluripotent states (naive and primed).

Inheritance of *De Novo* ICR Methylation and Biallelic Expression of Imprinted Genes after Differentiation

We next examined whether an altered DNA methylation status at ICRs affects the allelic expression pattern of imprinted genes. Among maternally imprinted genes, expressed genes retained monoallelic expression in iEpiSC lines (Figure S4B), which is consistent with our methylome data showing that the corresponding ICRs were stably methylated/unmethylated in reprogrammed PSC lines (Figure 2A). We noticed that maternally imprinted genes that acquire *de novo* ICR methylation (e.g., *Nap1l5*, *Airn*) showed lower or no detectable expression levels in EpiSCs (data not shown). These results may suggest that imprinted

Figure 3. *De Novo* ICR Methylation Occurs at the Late Stage of Somatic Cell Reprogramming

(A) Schematic diagram of the experimental design. Partial iPSC 9 cells were passaged three times, and mCherry-negative/-positive cells (p7) were sorted by fluorescence-activated cell sorting for expression analysis and DNA methylation analysis. Successful sorting was confirmed after expansion of sorted cells. Scale bars, 100 μm.

(B) DNA methylation analysis at *H19* DMR and *Gtl2* DMR in mCherry-positive and mCherry-negative iPSCs (clones 9, 21, 37) by conventional bisulfite sequencing. Open circles represent unmethylated CpGs and closed circles represent methylated CpGs. Crosses indicate undermined methylation status.

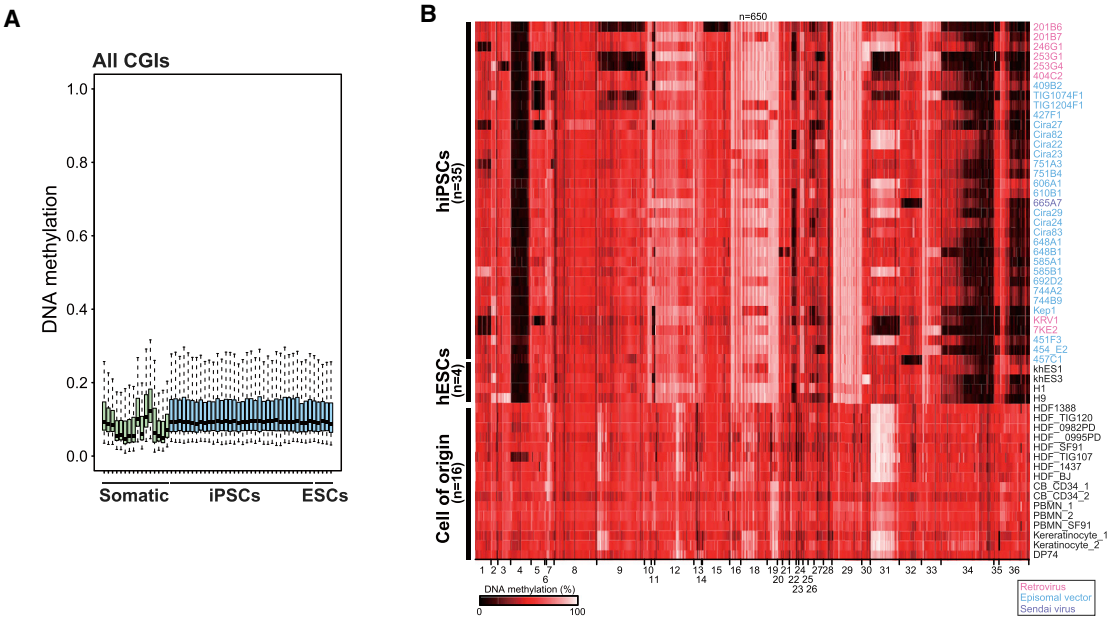
(C) DNA methylation at ICRs in preimplantation embryos. Note that ICMs exhibit reduced methylation levels at *Gnas*_1A ICR, while methylation levels are not altered at *H19* or *Impact* ICRs. WGBS data of ICMs were obtained from GEO: GSE84236. MethylC-seq data of MEFs from our previous study (Yagi et al., 2017a) were used (GEO: GSE84165).

(D) Allelic expression analysis of *Igf2* in MEFs and iPSCs. iPSCs exhibit biallelic expression of *Igf2*, which is consistent with biallelic methylation at *H19* DMR in these cells. Red and blue indicate maternal and paternal alleles, respectively. Numbers in the pie chart display numbers of the subcloned allele.

(E) Establishment of iPSC-derived secondary MEFs by blastocyst injection. iPSC-derived MEFs were selected by neomycin treatment for 7 days.

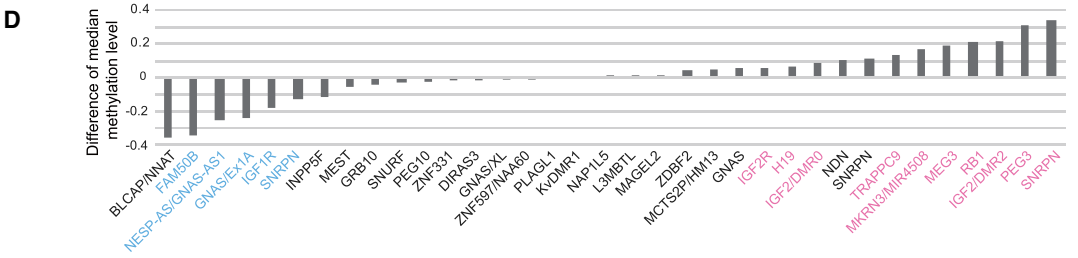
(F) DNA methylation analysis of *H19* DMR in iPSC-derived MEFs by conventional bisulfite sequencing.

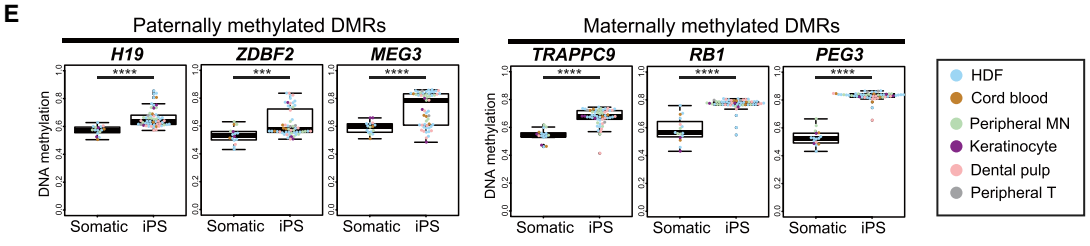
(G) Allelic expression analysis of *Igf2* in iPSC-derived MEFs. Biallelic expression of *Igf2* is detectable in iPSC-derived MEFs. Red and blue indicate maternal and paternal alleles, respectively. Numbers in the pie chart display numbers of the subcloned allele.



C

Number in heatmap	Locus	Methylation origin	Methylation timing	Number in heatmap	Locus	Methylation origin	Methylation timing
1	***DIRAS3***	M	Germline	19	***MKRN3/ miR 4508***	M	Somatic
2	***ZBDF2***	P	Somatic	20	***MAGEL2***	M	Somatic
3	***NAP1L5***	M	Germline	21	***NDN***	M	Somatic
4	***FAM50B***	M	Germline	22	***SNRPN***	M	Somatic
5	***PLAGL1***	M	Germline	23	***SNRPN***	M	Somatic
6	***IGF2R***	M	Germline	24	***SNRPN***	M	Somatic
7	***GRB10***	M	Germline	25	***SNURF***	M	Germline
8	***PEG10***	M	Germline	26	***IGF1R***	M	Germline
9	***MEST***	M	Germline	27	***ZNF597/NAT15***	P	Somatic
10	***PEG13/TRAPPC9***	M	Germline	28	***ZNF331***	M	Germline
11	***INPP5Fv2***	M	Germline	29	***PEG3***	M	Germline
12	***H19***	P	Germline	30	***MCTS2/HM13***	M	Germline
13	***IGF2 DMR2***	P	Somatic	31	***BLCAP/NNAT***	M	Germline
14	***IGF2 DMR0***	P	Somatic	32	***L3MBTL***	M	Germline
15	***KvDMR1***	M	Germline	33	***NESP***	P	Somatic
16	***RB1***	M	Germline	34	***NESP-AS/GNAS-AS1***	M	Germline
17	***IG-DMR***	P	Germline	35	***GNAS XL***	M	Germline
18	***MEG3***	P	Somatic	36	***GNAS Ex1A***	M	Somatic





(legend on next page)

genes with lower expression levels are targets of *de novo* methylation at maternally imprinted ICRs in iEpiSCs.

It is well known that paternally imprinted *H19* DMR regulates the monoallelic expression of *Igf2* from the paternal allele (Steenman et al., 1994). Consistent with our observation that *H19* DMR is biallelically methylated in reprogrammed PSCs, *Igf2* was expressed from both paternal and maternal alleles in iPSC clones (Figure 3D). A previous report suggests that ICR methylation is not recovered in somatic cell lineages once it is lost in PSCs (Holm et al., 2005). Therefore, we next investigated whether the acquired *de novo* ICR methylation in PSCs is inherited after differentiation. For this purpose, we injected iPSCs, which harbor *de novo* methylation at *H19* DMR, into blastocyst and established iPSC-derived secondary MEFs by neomycin selection (Figure 3E). Notably, aberrant DNA methylation patterns at *H19* DMR and biallelic expression of *Igf2* were detected in differentiated MEFs (Figures 3F and 3G), indicating that aberrant *de novo* ICR methylation was sustained even after the differentiation of PSCs. This result is consistent with the silencing of *H19* in ESC nuclei being sustained in differentiated cells after nuclear cloning (Humpherys et al., 2001).

Variable ICR Methylation Status in Human PSCs

To investigate any overlap of aberrant CGI methylation patterns between mouse and human PSCs, we next analyzed the genome-wide CGI methylation status in human primed PSC lines (Nishizawa et al., 2016). Consistent with our results in mouse primed iEpiSCs, most CGIs were hypomethylated in human iPSCs (Figure 4A). However, a subset of imprinted loci was aberrantly methylated in human iPSCs (hypermethylated DMR, n = 10 loci; hypomethylated DMR, n = 5 loci), whereas cells of origin for human iPSCs exhibited normal methylation levels (approximately 50%) at the imprinted DMRs (Figures 4B–4D). The aberrant DMR methylation was not associated with the method of reprogramming (Figure 4B). Notably, paternally imprinted genes such as *H19*, *MEG3*, and *ZDBF2* were frequently hypermethylated in multiple human PSCs, which was similarly observed in mouse PSCs (Figure 4E). These observations are consistent with a recent study demonstrating that the biallelic expression of imprinted genes is observed more frequently at paternally imprinted genes than maternally imprinted genes in human iPSCs (Bar et al., 2017). As observed in mouse iEpiSCs, several maternally methylated DMRs (e.g., *TRAPPC9*, *SNRPN*, *RB1*, and *PEG3*) were hypermethylated in human PSCs (Figure 4E). Conversely, a subset of maternally methylated DMRs (*FAM50B* and *GNAS*) were hypomethylated in human PSCs (Figure 4D). Importantly, *FAM50B* and *GNAS* are hypomethylated in human preimplantation embryos (Hanna et al., 2016), which supports the notion that hypomethylation at imprinted DMRs in PSCs may reflect the decreased methylation level of preimplantation embryos.

Dnmt3a Contributes to *De Novo* ICR Methylation during Reprogramming in Mice and Humans

During germ cell development, imprinted DMR is *de novo* methylated by *Dnmt3a* in conjunction with *Dnmt3l* (Bourc'his et al., 2001; Kaneda et al., 2004). A previous study revealed that *Dnmt3a* but not *Dnmt3l* is responsible for the increased methylation at *Dlk1-Dio3* imprinted loci during reprogramming (Stadtfeld et al., 2012). We therefore asked whether aberrant hypermethylation at other ICRs during reprogramming is also dependent on *Dnmt3a* activity. To test this hypothesis, we analyzed iPSCs derived from C57/BL6 MEFs lacking *Dnmt3a* with lentivirus-mediated Dox-inducible *OKSM* (Stadtfeld et al., 2012) for ICR methylation. We observed hypermethylation at *H19* DMR in control iPSCs, affirming that ICR hypermethylation

Figure 4. Variable ICR Methylation Aberrations in Human PSCs

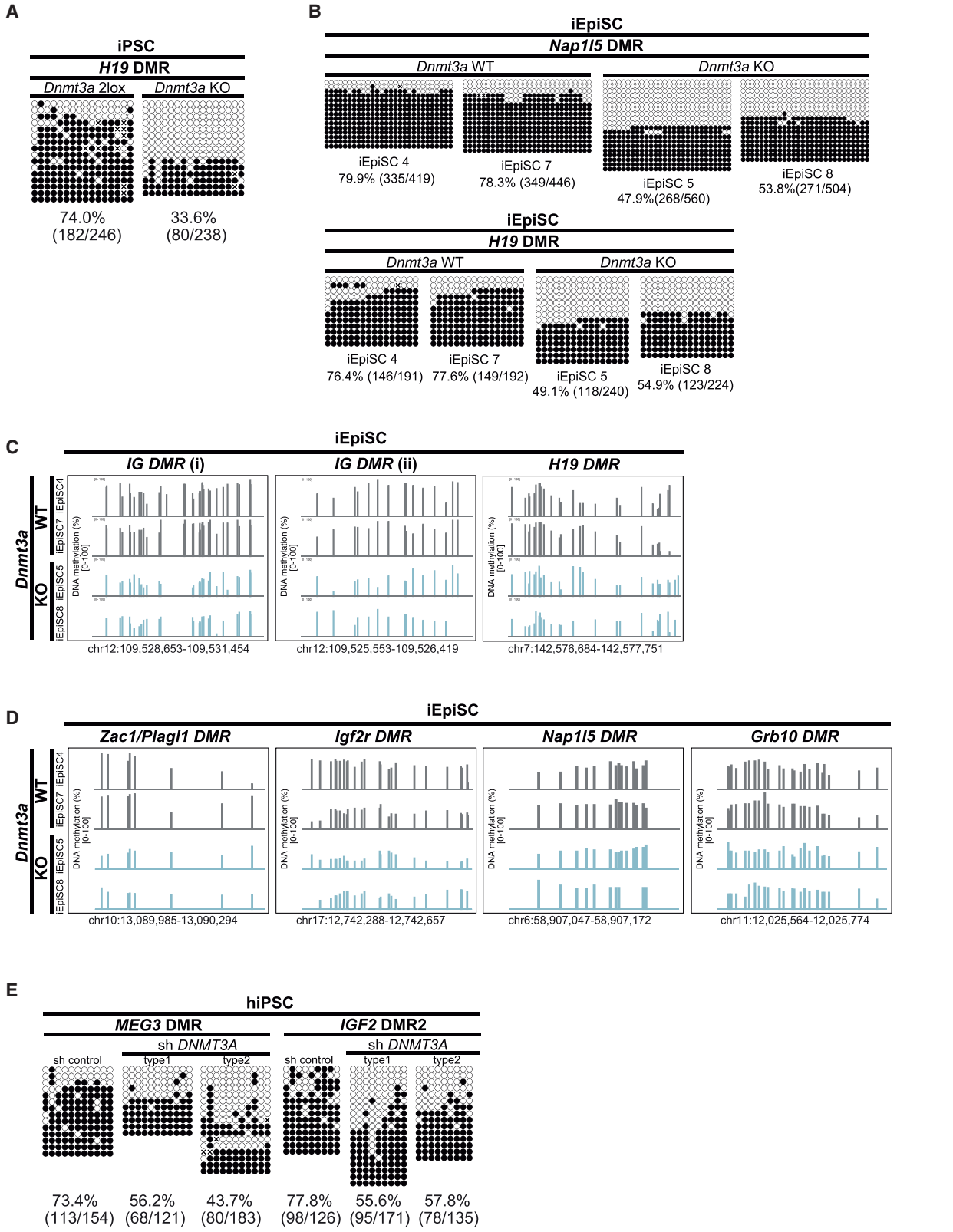
(A) Box plot of DNA methylation levels at all CGIs in human somatic cells and PSCs. Solid lines in each box indicate the median. The bottom and top of the boxes are lower and upper quartiles, respectively. Whiskers extend to ±1.5 IQR. Infinium 450K data of human somatic cells and PSCs were obtained from GEO: GSE60821 and GSE60923.

(B) Heatmap for DNA methylation levels at imprinted DMRs in human iPSCs (hiPSCs), human ESCs (hESCs), and various somatic cells of origin for hiPSCs. Color scale is shown for DNA methylation levels. Infinium 450K data of 35 hiPSC lines (20 male lines and 15 female lines), 4 hESCs, and 16 somatic cells were obtained from GEO: GSE60821 and GSE60923. Names of the hPSC lines are shown at the right of the panel. Colors depict the methods of reprogramming. HDF, human dermal fibroblasts; CB, cord blood cells; PBMN, peripheral blood mononuclear cells; DP, dental pulp cells.

(C) List of 36 human imprinted DMRs shown in (A) (Court et al., 2014). Origin (maternal or paternal allele) and timing (germline or somatic) of methylation are shown for each DMR. *IG*-DMR (17) is not considered in further analyses because the probes of Infinium 450K are not designed at *IG*-DMR.

(D) Difference of median methylation levels at the 35 imprinted DMRs between hiPSCs and somatic cells. The median methylation levels of the 35 iPSC lines and 16 somatic cells in (B) are compared. Hypermethylated DMRs (false discovery rate [FDR] <0.01, median methylation in iPSCs >60%) and hypomethylated DMRs (FDR <0.01, median methylation in iPSCs <40%) are shown in pink and blue, respectively.

(E) Box plots of DNA methylation levels at representative paternal and maternal imprinted DMRs in human iPSCs. Solid lines in each box indicate the median. The bottom and top of the boxes are lower and upper quartiles, respectively. Whiskers extend to ±1.5 IQR. Each color in the boxes represents a cell of origin. ***$p < 0.001$ and ****$p < 0.0001$ (Mann-Whitney U test).



(legend on next page)

occurred irrespective of the mouse genetic background or the *PB* system. Notably, *Dnmt3a*-deficient iPSCs displayed decreased DNA methylation levels at *H19* DMR compared with control cells (Figure 5A). We further established iEpiSCs from *Dnmt3a* KO MEFs and examined ICR methylation (Figure S4C). *Dnmt3a* knockout (KO) iEpiSCs exhibited reduced methylation levels at multiple ICRs compared with control *Dnmt3a* wild-type (WT) iEpiSCs (Figures 5B–5D). Of note, methylation levels at multiple ICRs were close to 50% in *Dnmt3a*-KO iEpiSCs (Figures 5A–5D). These findings suggest that *Dnmt3a* plays a dominant role in *de novo* ICR methylation, although our study does not exclude the possibility of the contribution of *Dnmt3b*. Collectively, *Dnmt3a* contributes to aberrant *de novo* DNA methylation at ICRs during somatic cell reprogramming in mice.

We further tested whether the suppression of *DNMT3A* can prevent human iPSCs from hypermethylation at imprinted DMRs during reprogramming. We generated human iPSCs from human peripheral blood mononuclear cells by inducing reprogramming factors with simultaneous knockdown of *DNMT3A* (Figures S4D and S4E). Of note, DNA methylation levels at *MEG3* DMR and *IGF2* DMR2 were reduced in *DNMT3A* KO human iPSCs compared with control iPSCs (Figure 5E), indicating that *DNMT3A* is responsible for hypermethylation at imprinted DMRs during reprogramming in both mice and humans.

Increased Methylation Levels at a Subset of Imprinted DMRs in Pediatric Cancers

Increased methylation levels at imprinted DMRs have been observed in a subset of cancers. Particularly, *H19* DMR hypermethylation and the concomitant biallelic expression of *IGF2* are frequently detectable in Wilms' tumors, the most common human pediatric kidney cancer (Hubertus et al., 2011; Steenman et al., 1994). We previously demonstrated that premature termination of *in vivo* reprogramming in mice leads to the development of cancers (Ohnishi et al., 2014). Notably, reprogramming-associated kidney cancers resembled Wilms' tumor and often harbored imprinting aberrations, including *H19* DMR hypermethylation (Ohnishi et al., 2014). These results may provide a link between reprogramming-associated ICR methylation and epigenetic aberrations in pediatric cancers. Therefore, to investigate a possible association of aberrant DNA methylation between Wilms' tumors and iPSCs, we performed comprehensive DMR methylation analysis in Wilms' tumors as well as normal kidney tissues and renal cell carcinomas (RCCs), a representative of adult kidney cancer, using public datasets. Wilms' tumors often exhibited increased methylation levels at imprinted DMRs not only at *H19* DMR but also other DMRs, whereas normal kidney samples and RCCs displayed relatively stable DMR methylation levels (Figures 6A and 6B). Increased DMR methylation levels at imprinted loci including *H19* DMR were also detectable in neuroblastomas, another type of pediatric cancer (Figure S4F). Together, these results suggest that *de novo* methylation at particular imprinted DMRs is a shared aberration in reprogrammed PSCs and pediatric cancers.

Previous studies demonstrated that CGI hypermethylation, particularly at CGIs linked to polycomb (PcG) target genes, is a general feature of epigenetic abnormalities found in cancers (Feinberg et al., 2006; Schlesinger et al., 2007; Widschwendter et al., 2007). In contrast, we found that most CGIs, except for imprinted DMRs, remain hypomethylated in iPSCs, which suggests that aberrant CGI methylation patterns in cancers are distinct from those in iPSCs. Consistent with previous reports, we detected CGI hypermethylation in a wide variety of adult cancers and found it was more prominent at CGIs close to PcG target genes (Figures 6C–6E). Notably, however, the same analysis revealed that pediatric cancers as well as iPSCs displayed no evidence of hypermethylation at global CGIs or even at PcG target gene-linked CGIs (Figures 6C–6E). Collectively, our data demonstrated that pediatric cancers harbor similar patterns of aberrant CGI methylation with iPSCs and highlighted the distinct features of epigenetic abnormalities in pediatric cancers and adult cancers.

DISCUSSION

In the present study, we conducted comprehensive methylation analysis for CGIs to elucidate the epigenetic stability

Figure 5. *Dnmt3a* Mediates Reprogramming-Associated *De Novo* ICR Methylation in Mice and Humans

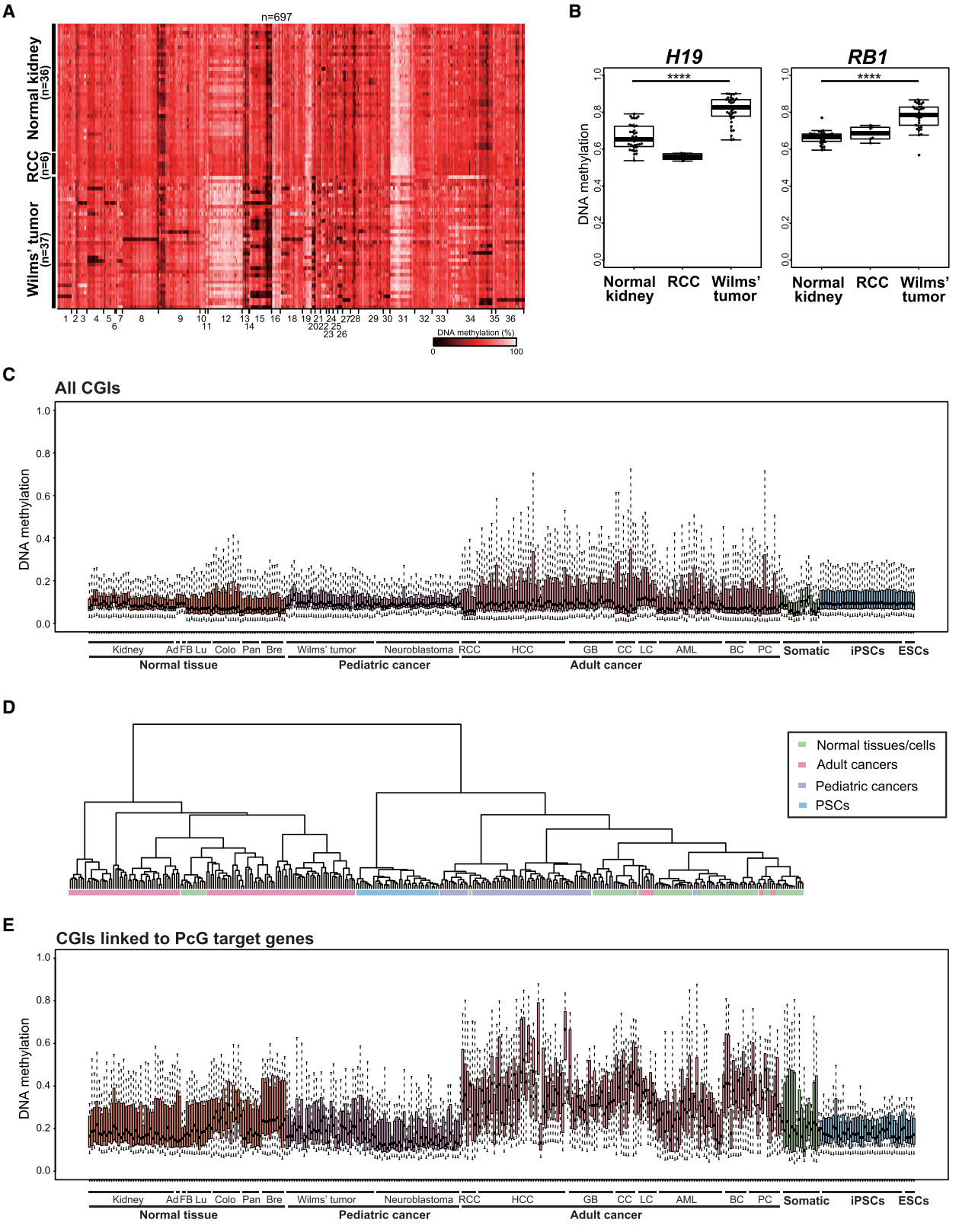
(A) DNA methylation status at *H19* DMR by conventional bisulfite sequencing in *Dnmt3a* control (2lox) and KO iPSCs.

(B) DNA methylation status at *Nap1l5* DMR and *H19* DMR by conventional bisulfite sequencing in *Dnmt3a* wild-type (WT) and null (KO) iEpiSCs.

(C) DNA methylation status at paternally methylated DMRs in *Dnmt3a* wild-type (WT) and null (KO) iEpiSCs. Each bar indicates a CpG site, and bar height represents methylation percentage (0%–100%) by MethylC-seq.

(D) DNA methylation status at maternally methylated DMRs in *Dnmt3a* wild-type (WT) and null (KO) iEpiSCs. Each bar indicates a CpG site, and bar height represents methylation percentage (0%–100%) by MethylC-seq.

(E) DNA methylation status at *MEG3* DMR and *IGF2* DMR2 by conventional bisulfite sequencing in hiPSCs established with *DNMT3A* short hairpin RNA (shRNA) and control shRNA treatment.



(legend on next page)

during somatic cell reprogramming into naive and primed PSCs. This analysis unveiled that most CGIs are precluded from being methylated, but CGIs linked to ICRs preferentially undergo *de novo* methylation during reprogramming. Consistent with this, paternally methylated ICRs often gained aberrant *de novo* methylation in both iPSCs and iEpiSCs. A subset of maternally methylated ICRs (e.g., *Nap1l5* and *Trappc9*) was also hypermethylated in primed iEpiSCs. We also show that human iPSCs exhibit aberrant CGI methylation at several imprinted DMRs. Notably, the biallelic expression or silencing of imprinted genes in reprogrammed PSCs are sustained in PSC-derived differentiated cells. Since the proper establishment and maintenance of genomic imprints, particularly at *H19* DMR (Kono et al., 2004), are important for normal development in mammals, our results may have important implications in various applications of PSCs, including regenerative medicine, drug screening, and the study of early developmental biology.

Interestingly, *de novo* ICR methylation was not prominent in partial iPSCs, suggesting that the aberrant ICR methylation in PSCs was not related to incomplete reprogramming. Notably, *de novo* ICR methylation occurred in accordance with the silencing of reprogramming factors. A previous study demonstrated that *de novo* DNA methylation plays a role in *PB* silencing (Troyanovsky et al., 2016). It is thus possible that the same machinery is involved in *de novo* ICR methylation and *PB* silencing. Nevertheless, considering that aberrant DMR hypermethylation is detectable in human iPSCs established by various methods, we propose that the *de novo* methylation at particular ICRs is a general phenomenon during somatic cell reprogramming. However, it should be noted that there exists clonal variation in ICR methylation patterns. The clonal variation may reflect the stochastic nature of the *de novo* methylation.

Although we demonstrated that *Dnmt3a* contributes to the ICR methylation in both mice and humans, the precise mechanism for *de novo* ICR methylation remains unclear. A previous study demonstrated that unmethylated alleles of imprinted genes are marked with both H3K4me2/3 and H3K27me3 and exhibit monoallelic bivalent chromatin when they are transcriptionally inactive (Maupetit-Mehouas et al., 2016). Of note, a recent study demonstrated that Dnmt3a preferentially binds to bivalent regions (Manzo et al., 2017). Collectively, it is possible that bivalent modifications at the unmethylated allele of imprinted genes might be a target of Dnmt3a binding, thus acting as a possible cause of *de novo* methylation during reprogramming.

We found that CGIs of multiple ICRs are aberrantly methylated in both human PSCs and pediatric cancers. However, we also identified that adult cancers exhibit distinct patterns of CGI hypermethylation from iPSCs; adult cancers display genome-wide global CGI hypermethylation while iPSCs exhibit ICR-preferred CGI hypermethylation. Notably, in sharp contrast to adult cancers, pediatric cancers did not show global CGI hypermethylation but exhibited ICR-preferred CGI hypermethylation, indicating that pediatric cancers harbor shared aberrant epigenetic signatures with human PSCs. A recent study demonstrated that CGI hypermethylation in adult cancers is associated with activated fibroblast growth factor (FGF) signaling (Smith et al., 2017). The fact that pediatric cancers infrequently harbor genetic aberrations in components of the FGF pathway might explain the absence of global CGI in pediatric cancers. Collectively, these findings highlighted the unique patterns of epigenetic aberrations in pediatric cancers, which exhibit differences from adult cancers but similarities with iPSCs. Considering that the premature termination of *in vivo* reprogramming causes pediatric cancer-like tumors in mice (Ohnishi et al., 2014) and that pediatric cancers harbor PSC-like transcriptional signatures (Terada et al., 2019), these results raised the possibility that some aspects of reprogramming to PSCs may drive the development of pediatric cancers with concomitant aberrations in ICR methylation.

Figure 6. Pediatric Cancers Exhibit Hypermethylation at ICRs but Not at Global CGIs

(A) Heatmap for DNA methylation levels at imprinted DMRs in human normal kidney samples, renal cell carcinomas (RCC), and Wilms' tumors. Color scale is shown for DNA methylation levels. Infinium 450K data of normal kidney tissues and Wilms' tumors were obtained from GEO: GSE59157. Those of RCCs were obtained from GEO: GSE70303.

(B) Box plots of DNA methylation levels at representative hypermethylated imprinted DMRs in Wilms' tumors. Solid lines in each box indicate the median. The bottom and top of the boxes are lower and upper quartiles, respectively. Whiskers extend to ±1.5 IQR. Note that *H19* DMR and *RB1* DMR are hypermethylated in Wilms' tumor but not in normal kidney or RCC. ****$p < 0.0001$ (Mann-Whitney U test).

(C) Box plot of DNA methylation levels at all CGIs in normal tissues, pediatric cancers, adult cancers, somatic cells, and PSCs. Solid lines in each box indicate the median. The bottom and top of the boxes are lower and upper quartiles, respectively. Whiskers extend to ±1.5 IQR. Note that increased CGI methylation is observed in adult cancers but not in pediatric cancers. Data of somatic cells and PSCs are the same as in Figure 4A.

(D) Hierarchical clustering analysis based on the methylation status at all CGIs.

(E) Box plot of DNA methylation levels at CGIs linked to PcG target genes (Lee et al., 2006) in normal tissues, pediatric cancers, adult cancers, somatic cells, and PSCs. Solid lines in each box indicate the median. The bottom and top of the boxes are lower and upper quartiles, respectively. Whiskers extend to ±1.5 IQR.

To conclude, our findings about *de novo* CGI methylation may provide important insights into the faithful recapitulation of *in vivo* pluripotent cells *in vitro*. Our findings may also underscore the significant relevance of reprogramming-associated epigenetic aberrations in the development of pediatric cancers.

EXPERIMENTAL PROCEDURES

Detailed descriptions of experimental procedures can be found in Supplemental Information.

Establishment and Culture of Male ESCs and EpiSCs

Zygotes with an (129X1/SvJ × MSM/Ms) F1 genetic background were obtained by *in vitro* fertilization (IVF). ESCs established in a previous study were used in this study (Yagi et al., 2017a). Epiblasts were divided from the extraembryonic regions and transferred into culture plates to derive EpiSCs.

Generation and Culture of Male iPSCs and iEpiSCs

Dox-inducible PB vector containing tetO-*Oct4-Sox2-Klf4-cMyc*-IRES-mCherry-EF1-rtTA-IRES-Neo (PB-*OSKM*) was used for reprogramming. After transfection, cultured medium was switched to ESC medium containing 2 μg/mL Dox (Sigma) for iPSC derivation and iEpiSC medium for iEpiSC derivation.

Generation of iPSCs and iEpiSCs from *Dnmt3a*-Deficient MEFs

Dnmt3a WT and KO MEFs were obtained by crossing *Dnmt3a* hetero KO (B6; 129S4-Dnmt3a <tm1Enl>) mice (Okano et al., 1999). *Dnmt3a* WT and KO iEpiSCs were generated by Dox-inducible PB-*OSKM*. *Dnmt3a* WT and KO iPSCs were generated in a previous study (Stadtfeld et al., 2012).

Animals

All experiments using animals were performed under the ethical guidelines of Kyoto University, University of Tokyo, and Kumamoto University. MSM/Ms were obtained from RIKEN Bio Resource Center (Takada et al., 2013, 2015).

Generation and Culture of hiPSCs

For generation of human iPSCs (hiPSCs), human peripheral blood mononuclear cells (Cellular Technology) were cultured. After transduction of the plasmid mixture (pCXLE-hOCT3/4-shp53-F, pCXLE-hUL, pCXWB-EBNA1, and pCXLE-hSK encoding short hairpin RNA for DNMT3A), the cells were seeded on a 6-well plate coated with iMatrix-511 (Takara) and cultured until hiPSC colony formation. The selected hiPSCs were expanded in StemFit AK02N (Takara).

Library Preparation

Library preparation was performed with SureSelect Mouse Methyl-Seq Reagent Kit (Agilent Technologies). DNA was bisulfite-treated using the EZ DNA Methylation-Gold Kit. The libraries were then sequenced on HiSeq 2500 (2 × 100-bp or 2 × 101-bp paired-end reads, Illumina). RNA-seq libraries were generated using the Truseq Stranded mRNA LT sample prep kit (Illumina). RNA-seq libraries were sequenced on NextSeq500 (75-bp single read, Illumina).

DNA Methylation Analyses

For allelic methylation analyses, the SNP data for MSM/Ms were obtained from NIG Mouse Genome Database (MSMv4HQ, http://molossinus.lab.nig.ac.jp/msmdb/index.jsp). The B6-derived and MSM/Ms-derived sequenced reads were determined based on the MSM/Ms SNP data. Previously described mouse CGIs (Illingworth et al., 2010) and mouse ICRs (Court et al., 2014; Tomizawa et al., 2011) were used for CGI and imprinting analyses, respectively. Infinium array data were obtained from publicly available datasets. Previously described human CGIs (Illingworth et al., 2010), human ICRs (Court et al., 2014; Tomizawa et al., 2011), and PcG target genes (Lee et al., 2006) were used for methylation analyses. The average methylation signals of each ICR were used for comparison among samples.

ACCESSION NUMBERS

The accession numbers for the MethylC-seq and RNA-seq data reported in this paper are GEO: GSE111173 and GSE84165.

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.stemcr.2019.04.008.

AUTHOR CONTRIBUTIONS

M.Y. and Y.Y. designed and conceived the study and wrote the manuscript. M.Y., T.U., Y.S., M.S., and K.A. generated the cell lines, performed the experiments, and generated the MethylC-seq and RNA-seq libraries. A.T. performed the blastocyst injections. K.O. established the human iPSCs with short hairpin RNA. K.W. provided the OSKM *piggyBac* vector. K.H. provided *Dnmt3a* 2flox and null iPSCs. M.K., S.O., and T.Y. analyzed the RNA-seq, MethylC-seq, and Infinium 450K data.

ACKNOWLEDGMENTS

We are grateful to K. Nakabayashi for providing details on human imprinted DMRs. Y.Y. was supported in part by cancer research grant P-CREATE, Japan Agency for Medical Research and Development (AMED); SICORP, AMED; JSPS KAKENHI 18H04026; the Takeda Science Foundation; and the Naito Foundation. Y.Y. and T.Y. were supported by AMED-CREST under grant numbers JP18gm1110004 (Y.Y and T.Y.) and JP18gm0610017 (T.Y.); and Core Center for iPS Cell Research, Research Center Network for Realization of Regenerative Medicine, AMED. T.Y. was supported by the iPS Cell Research Fund. M.Y. was supported by JSPS KAKENHI 15J05792. The ASHBi is supported by the World Premier International Research Center Initiative, MEXT, Japan.

Received: November 29, 2018
Revised: April 5, 2019
Accepted: April 8, 2019
Published: May 2, 2019

REFERENCES

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.

Bourc'his, D., Xu, G.L., Lin, C.S., Bollman, B., and Bestor, T.H. (2001). Dnmt3L and the establishment of maternal genomic imprints. Science *294*, 2536–2539.

Branco, M.R., Oda, M., and Reik, W. (2008). Safeguarding parental identity: Dnmt1 maintains imprints during epigenetic reprogramming in early embryogenesis. Genes Dev. *22*, 1567–1571.

Choi, J., Clement, K., Huebner, A.J., Webster, J., Rose, C.M., Brumbaugh, J., Walsh, R.M., Lee, S., Savol, A., Etchegaray, J.P., et al. (2017a). DUSP9 modulates DNA hypomethylation in female mouse pluripotent stem cells. Cell Stem Cell *20*, 706–719.e7.

Choi, J., Huebner, A.J., Clement, K., Walsh, R.M., Savol, A., Lin, K., Gu, H., Di Stefano, B., Brumbaugh, J., Kim, S.Y., et al. (2017b). Prolonged Mek1/2 suppression impairs the developmental potential of embryonic stem cells. Nature *548*, 219–223.

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., Simon, 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.

Feinberg, A.P., Ohlsson, R., and Henikoff, S. (2006). The epigenetic progenitor origin of human cancer. Nat. Rev. Genet. *7*, 21–33.

Ferguson-Smith, A.C. (2011). Genomic imprinting: the emergence of an epigenetic paradigm. Nat. Rev. Genet. *12*, 565–575.

Han, D.W., Greber, B., Wu, G., Tapia, N., Arauzo-Bravo, M.J., Ko, K., Bernemann, C., Stehling, M., and Scholer, H.R. (2011). Direct reprogramming of fibroblasts into epiblast stem cells. Nat. Cell Biol. *13*, 66–71.

Hanna, C.W., Penaherrera, M.S., Saadeh, H., Andrews, S., McFadden, D.E., Kelsey, G., and Robinson, W.P. (2016). Pervasive polymorphic imprinted methylation in the human placenta. Genome Res. *26*, 756–767.

Holm, T.M., Jackson-Grusby, L., Brambrink, T., Yamada, Y., Rideout, W.M., 3rd, and Jaenisch, R. (2005). Global loss of imprinting leads to widespread tumorigenesis in adult mice. Cancer Cell *8*, 275–285.

Hubertus, J., Lacher, M., Rottenkolber, M., Muller-Hocker, J., Berger, M., Stehr, M., von Schweinitz, D., and Kappler, R. (2011). Altered expression of imprinted genes in Wilms tumors. Oncol. Rep. *25*, 817–823.

Humpherys, D., Eggan, K., Akutsu, H., Hochedlinger, K., Rideout, W.M., 3rd, Biniszkiewicz, D., Yanagimachi, R., and Jaenisch, R. (2001). Epigenetic instability in ES cells and cloned mice. Science *293*, 95–97.

Illingworth, R.S., Gruenewald-Schneider, U., Webb, S., Kerr, A.R., James, K.D., Turner, D.J., Smith, C., Harrison, D.J., Andrews, R., and Bird, A.P. (2010). Orphan CpG islands identify numerous conserved promoters in the mammalian genome. PLoS Genet. *6*, e1001134.

Jaenisch, R., and Young, R. (2008). Stem cells, the molecular circuitry of pluripotency and nuclear reprogramming. Cell *132*, 567–582.

Johannesson, B., Sagi, I., Gore, A., Paull, D., Yamada, M., Golan-Lev, T., Li, Z., LeDuc, C., Shen, Y., Stern, S., et al. (2014). Comparable frequencies of coding mutations and loss of imprinting in human pluripotent cells derived by nuclear transfer and defined factors. Cell Stem Cell *15*, 634–642.

Kaneda, M., Okano, M., Hata, K., Sado, T., Tsujimoto, N., Li, E., and Sasaki, H. (2004). Essential role for de novo DNA methyltransferase Dnmt3a in paternal and maternal imprinting. Nature *429*, 900–903.

Kato, Y., Rideout, W.M., 3rd, Hilton, K., Barton, S.C., Tsunoda, Y., and Surani, M.A. (1999). Developmental potential of mouse primordial germ cells. Development *126*, 1823–1832.

Kim, S.I., Oceguera-Yanez, F., Sakurai, C., Nakagawa, M., Yamanaka, S., and Woltjen, K. (2016). Inducible transgene expression in human iPS cells using versatile all-in-one piggybac transposons. Methods Mol. Biol. *1357*, 111–131.

Kono, T., Obata, Y., Wu, Q.L., Niwa, K., Ono, Y., Yamamoto, Y., Park, E.S., Seo, J.S., and Ogawa, H. (2004). Birth of parthenogenetic mice that can develop to adulthood. Nature *428*, 860–864.

Lee, T.I., Jenner, R.G., Boyer, L.A., Guenther, M.G., Levine, S.S., Kumar, R.M., Chevalier, B., Johnstone, S.E., Cole, M.F., Isono, K., et al. (2006). Control of developmental regulators by Polycomb in human embryonic stem cells. Cell *125*, 301–313.

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.

Maherali, N., Ahfeldt, T., Rigamonti, A., Utikal, J., Cowan, C., and Hochedlinger, K. (2008). A high-efficiency system for the generation and study of human induced pluripotent stem cells. Cell Stem Cell *3*, 340–345.

Manzo, M., Wirz, J., Ambrosi, C., Villasenor, R., Roschitzki, B., and Baubec, T. (2017). Isoform-specific localization of DNMT3A regulates DNA methylation fidelity at bivalent CpG islands. EMBO J. *36*, 3421–3434.

Maupetit-Mehouas, S., Montibus, B., Nury, D., Tayama, C., Wassef, M., Kota, S.K., Fogli, A., Cerqueira Campos, F., Hata, K., Feil, R., et al. (2016). Imprinting control regions (ICRs) are marked by mono-allelic bivalent chromatin when transcriptionally inactive. Nucleic Acids Res. *44*, 621–635.

Mikkelsen, T.S., Hanna, J., Zhang, X., Ku, M., Wernig, M., Schorderet, P., Bernstein, B.E., Jaenisch, R., Lander, E.S., and Meissner, A. (2008). Dissecting direct reprogramming through integrative genomic analysis. Nature *454*, 49–55.

Nazor, K.L., Altun, G., Lynch, C., Tran, H., Harness, J.V., Slavin, I., Garitaonandia, I., Muller, 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.

Nichols, J., and Smith, A. (2009). Naive and primed pluripotent states. Cell Stem Cell *4*, 487–492.

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.

Ohnishi, K., Semi, K., Yamamoto, T., Shimizu, M., Tanaka, A., Mitsunaga, K., Okita, K., Osafune, K., Arioka, Y., Maeda, T., et al. (2014). Premature termination of reprogramming in vivo leads to cancer development through altered epigenetic regulation. Cell *156*, 663–677.

Okano, M., Bell, D.W., Haber, D.A., and Li, E. (1999). DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. Cell *99*, 247–257.

Pasque, V., Karnik, R., Chronis, C., Petrella, P., Langerman, J., Bonora, G., Song, J., Vanheer, L., Sadhu Dimashkie, A., Meissner, A., et al. (2018). X chromosome dosage influences DNA methylation dynamics during reprogramming to mouse iPSCs. Stem Cell Reports *10*, 1537–1550.

Pick, M., Stelzer, Y., Bar-Nur, O., Mayshar, Y., Eden, A., and Benvenisty, N. (2009). Clone- and gene-specific aberrations of parental imprinting in human induced pluripotent stem cells. Stem Cells *27*, 2686–2690.

Schlesinger, Y., Straussman, R., Keshet, I., Farkash, S., Hecht, M., Zimmerman, J., Eden, E., Yakhini, Z., Ben-Shushan, E., Reubinoff, B.E., et al. (2007). Polycomb-mediated methylation on Lys27 of histone H3 pre-marks genes for de novo methylation in cancer. Nat. Genet. *39*, 232–236.

Sharma, S., Kelly, T.K., and Jones, P.A. (2010). Epigenetics in cancer. Carcinogenesis *31*, 27–36.

Smith, Z.D., Shi, J., Gu, H., Donaghey, J., Clement, K., Cacchiarelli, D., Gnirke, A., Michor, F., and Meissner, A. (2017). Epigenetic restriction of extraembryonic lineages mirrors the somatic transition to cancer. Nature *549*, 543–547.

Stadtfeld, M., Apostolou, E., Akutsu, H., Fukuda, A., Follett, P., Natesan, S., Kono, T., Shioda, T., and Hochedlinger, K. (2010). Aberrant silencing of imprinted genes on chromosome 12qF1 in mouse induced pluripotent stem cells. Nature *465*, 175–181.

Stadtfeld, M., Apostolou, E., Ferrari, F., Choi, J., Walsh, R.M., Chen, T., Ooi, S.S., Kim, S.Y., Bestor, T.H., Shioda, T., et al. (2012). Ascorbic acid prevents loss of Dlk1-Dio3 imprinting and facilitates generation of all-iPS cell mice from terminally differentiated B cells. Nat. Genet. *44*, 398–405, S391-392.

Stadtfeld, M., and Hochedlinger, K. (2010). Induced pluripotency: history, mechanisms, and applications. Genes Dev. *24*, 2239–2263.

Steenman, M.J., Rainier, S., Dobry, C.J., Grundy, P., Horon, I.L., and Feinberg, A.P. (1994). Loss of imprinting of IGF2 is linked to reduced expression and abnormal methylation of H19 in Wilms' tumour. Nat. Genet. *7*, 433–439.

Takada, T., Ebata, T., Noguchi, H., Keane, T.M., Adams, D.J., Narita, T., Shin, I.T., Fujisawa, H., Toyoda, A., Abe, K., et al. (2013). The ancestor of extant Japanese fancy mice contributed to the mosaic genomes of classical inbred strains. Genome Res. *23*, 1329–1338.

Takada, T., Yoshiki, A., Obata, Y., Yamazaki, Y., and Shiroishi, T. (2015). NIG_MoG: a mouse genome navigator for exploring intersubspecific genetic polymorphisms. Mamm. Genome *26*, 331–337.

Takahashi, K., and Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell *126*, 663–676.

Terada, Y., Jo, N., Arakawa, Y., Sakakura, M., Yamada, Y., Ukai, T., Kabata, M., Mitsunaga, K., Mineharu, Y., Ohta, S., et al. (2019). Human pluripotent stem cell-derived tumor model uncovers the embryonic stem cell signature as a key driver in atypical teratoid/rhabdoid tumor. Cell Rep. *26*, 2608–2621.e6.

Tomizawa, S., Kobayashi, H., Watanabe, T., Andrews, S., Hata, K., Kelsey, G., and Sasaki, H. (2011). Dynamic stage-specific changes in imprinted differentially methylated regions during early mammalian development and prevalence of non-CpG methylation in oocytes. Development *138*, 811–820.

Troyanovsky, B., Bitko, V., Pastukh, V., Fouty, B., and Solodushko, V. (2016). The functionality of minimal piggybac transposons in mammalian cells. Mol. Ther. Nucleic Acids *5*, e369.

Widschwendter, M., Fiegl, H., Egle, D., Mueller-Holzner, E., Spizzo, G., Marth, C., Weisenberger, D.J., Campan, M., Young, J., Jacobs, I., et al. (2007). Epigenetic stem cell signature in cancer. Nat. Genet. *39*, 157–158.

Yagi, M., Kishigami, S., Tanaka, A., Semi, K., Mizutani, E., Wakayama, S., Wakayama, T., Yamamoto, T., and Yamada, Y. (2017a). Derivation of ground-state female ES cells maintaining gamete-derived DNA methylation. Nature *548*, 224–227.

Yagi, M., Yamanaka, S., and Yamada, Y. (2017b). Epigenetic foundations of pluripotent stem cells that recapitulate in vivo pluripotency. Lab. Invest. *97*, 1133–1141.

Yamanaka, S. (2012). Induced pluripotent stem cells: past, present, and future. Cell Stem Cell *10*, 678–684.

Stem Cell Reports, Volume *12*

Supplemental Information

De Novo DNA Methylation at Imprinted Loci during Reprogramming into Naive and Primed Pluripotency

Masaki Yagi, Mio Kabata, Tomoyo Ukai, Sho Ohta, Akito Tanaka, Yui Shimada, Michihiko Sugimoto, Kimi Araki, Keisuke Okita, Knut Woltjen, Konrad Hochedlinger, Takuya Yamamoto, and Yasuhiro Yamada

Supplemental Information

De novo DNA methylation at imprinted loci during reprogramming into naïve and primed pluripotency

Masaki Yagi, Mio Kabata, Tomoyo Ukai, Sho Ohta, Akito Tanaka, Yui Shimada, Michihiko Sugimoto, Kimi Araki, Keisuke Okita, Knut Woltjen, Konrad Hochedlinger, Takuya Yamamoto, Yasuhiro Yamada

Inventory of Supplemental Information

Figure S1

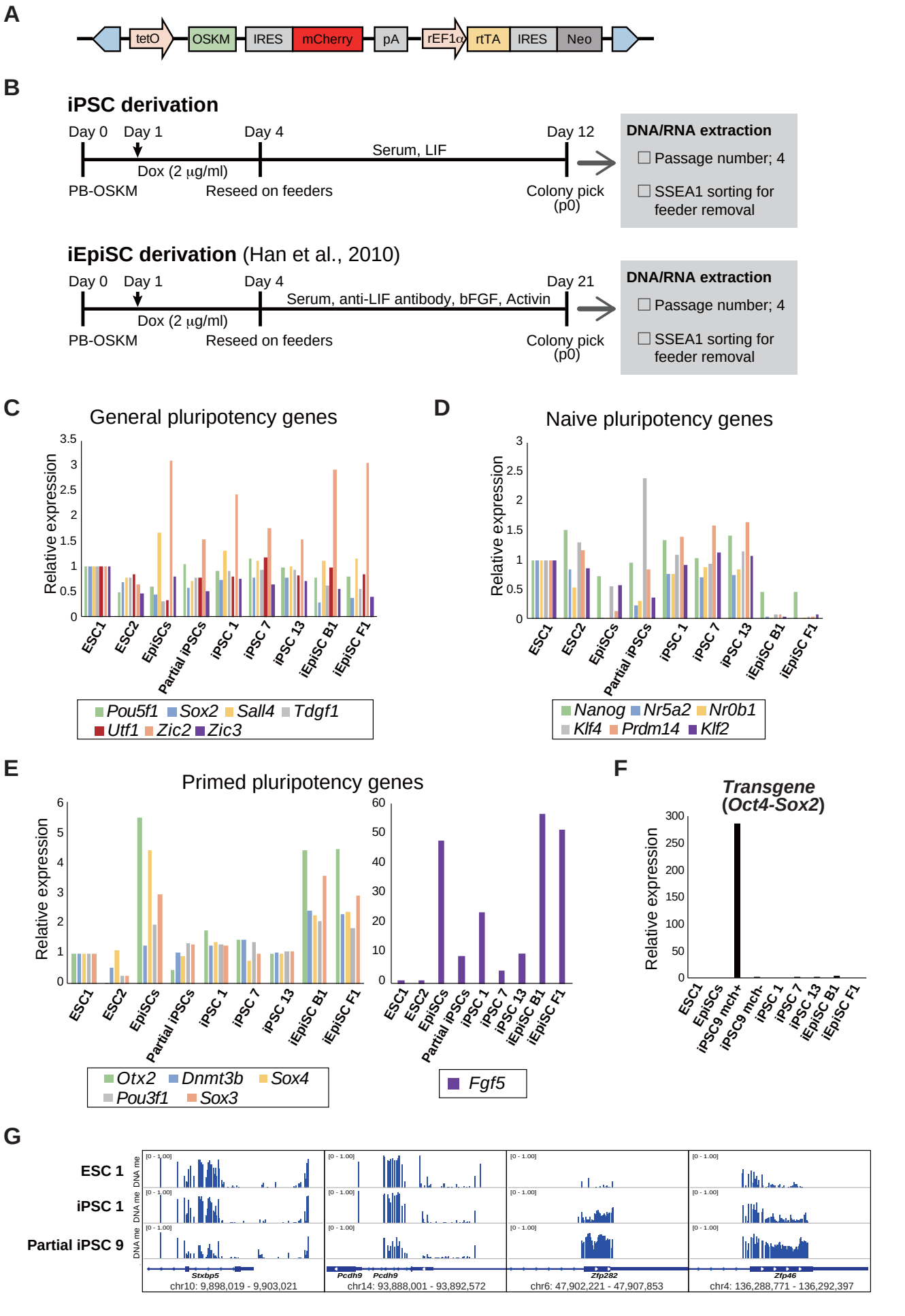

A
tetO
OSKM
IRES
mCherry
pA
rEF1α
rtTA
IRES
Neo
B
iPSC derivation
Day 0
Day 1
Day 4
Day 12
Serum, LIF
Dox (2 μg/ml)
PB-OSKM
Reseed on feeders
Colony pick (p0)
DNA/RNA extraction
Passage number; 4
SSEA1 sorting for feeder removal
iEpiSC derivation (Han et al., 2010)
Day 0
Day 1
Day 4
Day 21
Serum, anti-LIF antibody, bFGF, Activin
Dox (2 μg/ml)
PB-OSKM
Reseed on feeders
Colony pick (p0)
DNA/RNA extraction
Passage number; 4
SSEA1 sorting for feeder removal
C
General pluripotency genes
Relative expression
ESC1
ESC2
EpiSCs
Partial iPSCs
iPSC 1
iPSC 7
iPSC 13
iEpiSC B1
iEpiSC F1
Pou5f1
Sox2
Sall4
Tdgf1
Utf1
Zic2
Zic3
D
Naive pluripotency genes
Relative expression
Nanog
Nr5a2
Nr0b1
Klf4
Prdm14
Klf2
E
Primed pluripotency genes
Relative expression
Otx2
Dnmt3b
Sox4
Pou3f1
Sox3
Fgf5
F
Transgene (Oct4-Sox2)
Relative expression
iPSC9 mch+
iPSC9 mch-
G
ESC 1
iPSC 1
Partial iPSC 9
DNA me
Stxbp5
Pcdh9
Zfp282
Zfp46
chr10: 9,898,019 - 9,903,021
chr14: 93,888,001 - 93,892,572
chr6: 47,902,221 - 47,907,853
chr4: 136,288,771 - 136,292,397

Figure S1, related to Figure 1: Derivation and characterization of mouse naïve and primed PSCs

A: Doxycycline (Dox)-inducible *piggyBac* vector containing *tetO-OSKM-IRES-mCherry-EF1-rtTA-IRES-Neo* (*PB-OSKM* all-in-one) was used for reprogramming of MEFs into iPSCs and iEpiSCs.

B: Protocol for derivation of iPSC and iEpiSC from MEFs.

C: Relative expression (FPKM) levels of general pluripotency genes in ESCs, EpiSCs, iPSCs and iEpiSCs by RNA-seq. All PSC lines express general pluripotency genes. The expression levels of ESC 1 was set to 1.

D: Relative expression (FPKM) levels of naive pluripotency genes in ESCs, EpiSCs, iPSCs and iEpiSCs by RNA-seq. Naïve PSC lines (ESCs and iPSCs) highly express naive pluripotency genes. Note that the expression of naive pluripotency genes is lower in partial iPSCs compared to iPSCs. The expression levels of ESC 1 was set to 1.

E: Relative expression (FPKM) levels of primed pluripotency genes in ESCs, EpiSCs, iPSCs and iEpiSCs by RNA-seq. Primed PSC lines (EpiSCs and iEpiSCs) express higher levels of primed pluripotency genes. The expression levels of ESC 1 was set to 1.

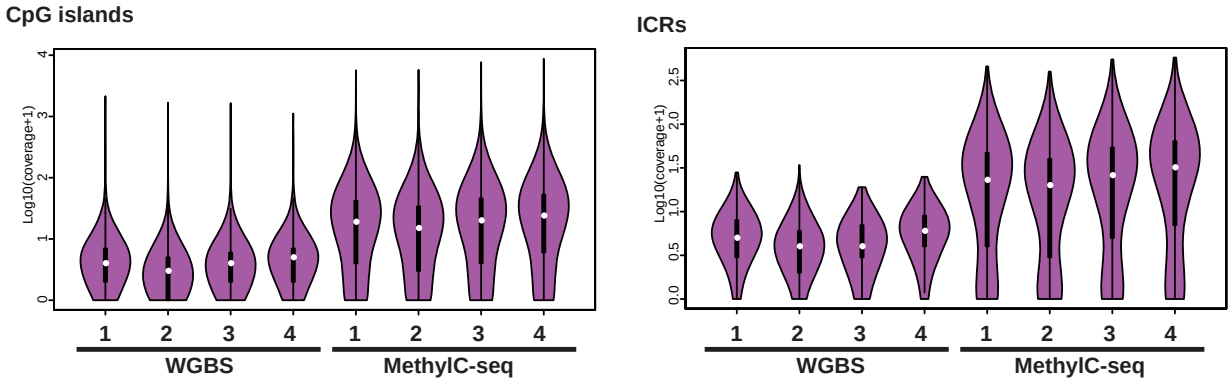
F: A qRT-PCR analysis confirmed transgene silencing in PSCs. Primers were set in *Oct4* and *Sox2* within polycistronic gene and thus can only amplify the transgenes. Data are presented as the mean of technical triplicates. The mean expression level of transgene-silenced iPS9 mch- was set to 1.

G: DNA methylation status at representative differentially methylated regions (hyper or hypomethylated regions) in partial iPSC 9 compared to PSCs (ESCs and iPSC 1). Each bar indicates a CG site, and the height of the bar represents the methylation percentage. Location of RefSeq genes and gene symbols for representative genes are indicated below.

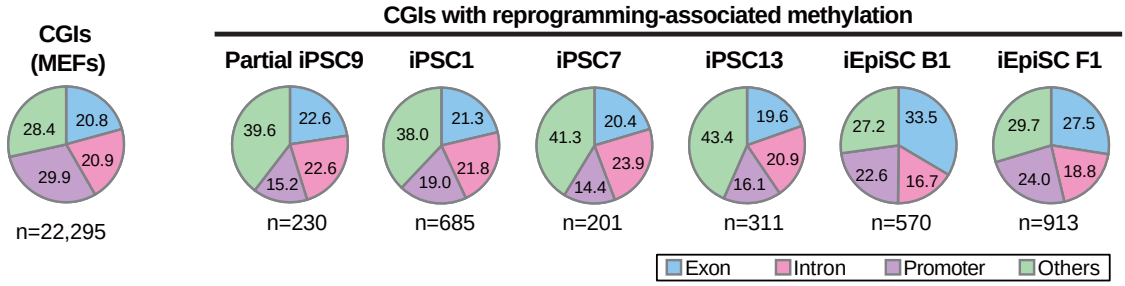
Figure S2

A

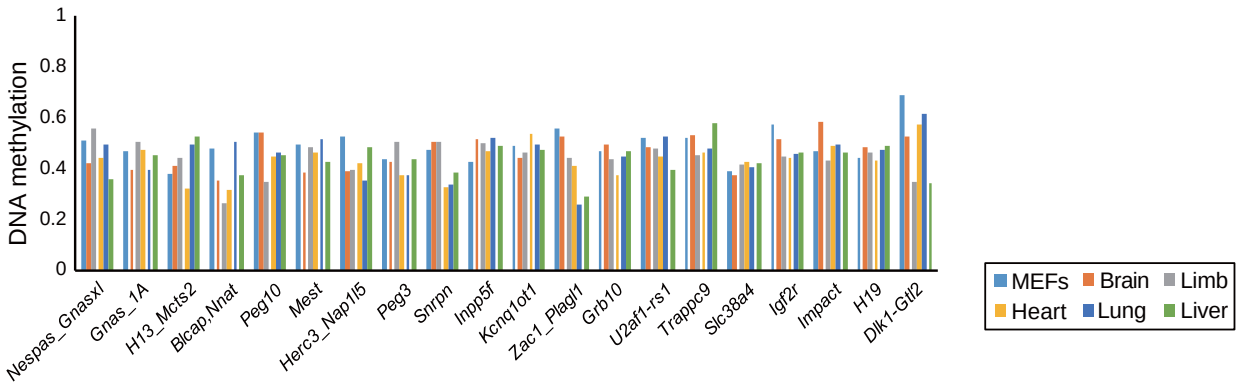
Sample	Method	Read Number (Mb)	Seq type
Sample 1	WGBS	117.6 (×2)	Hiseq (pair)
Sample 2	WGBS	117.3 (×2)	Hiseq (pair)
Sample 3	WGBS	107.9 (×2)	Hiseq (pair)
Sample 4	WGBS	116.9 (×2)	Hiseq (pair)
Sample 1	Methylseq	77.3 (×2)	Hiseq (pair)
Sample 2	Methylseq	70.7 (×2)	Hiseq (pair)
Sample 3	Methylseq	91.8 (×2)	Hiseq (pair)
Sample 4	Methylseq	96.0 (×2)	Hiseq (pair)



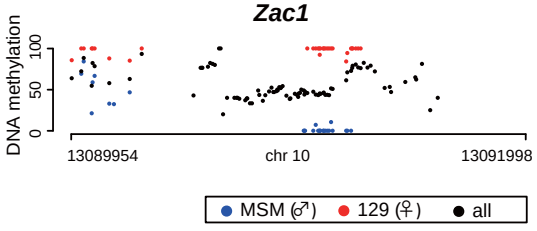
B



C



D



E

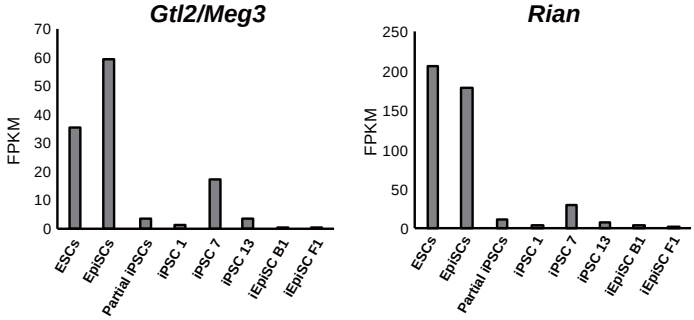


Figure S2, related to Figure 1 and Figure 2: DNA methylation analyses in mouse naïve and primed PSCs

A: Numbers of sequencing read and sequencing coverage at CGIs and ICRs by WGBS and MethylC-seq. WGBS and MethylC-seq data for ESC sample 1-4 were obtained from GSE84165.

B: Composition of each cluster of reprogrammed PSC-specific methylated CGIs are shown as pie charts. Composition at MEFs is shown as a control.

C: DNA methylation levels at ICRs in MEFs and various tissues obtained from an embryo at E14.5. Note that MEFs and tissues displayed relatively normal methylation levels.

D: CpG methylation at *Zac1* DMR in MEFs. Each black dot represents a methylation percentage for each CpG site. Red and blue dots indicate methylation levels at maternal 129 allele and paternal MSM allele, respectively. Note that some CpG sites show increased methylation at unmethylated allele in MEFs.

E: Gene expression (FPKM) levels of *Meg3* and *Rian* in ESCs, EpiSCs, iPSCs and iEpiSCs by RNA-seq. Note that both *Meg3* and *Rian* are repressed in reprogrammed PSCs (iPSCs and iEpiSCs).

Figure S3

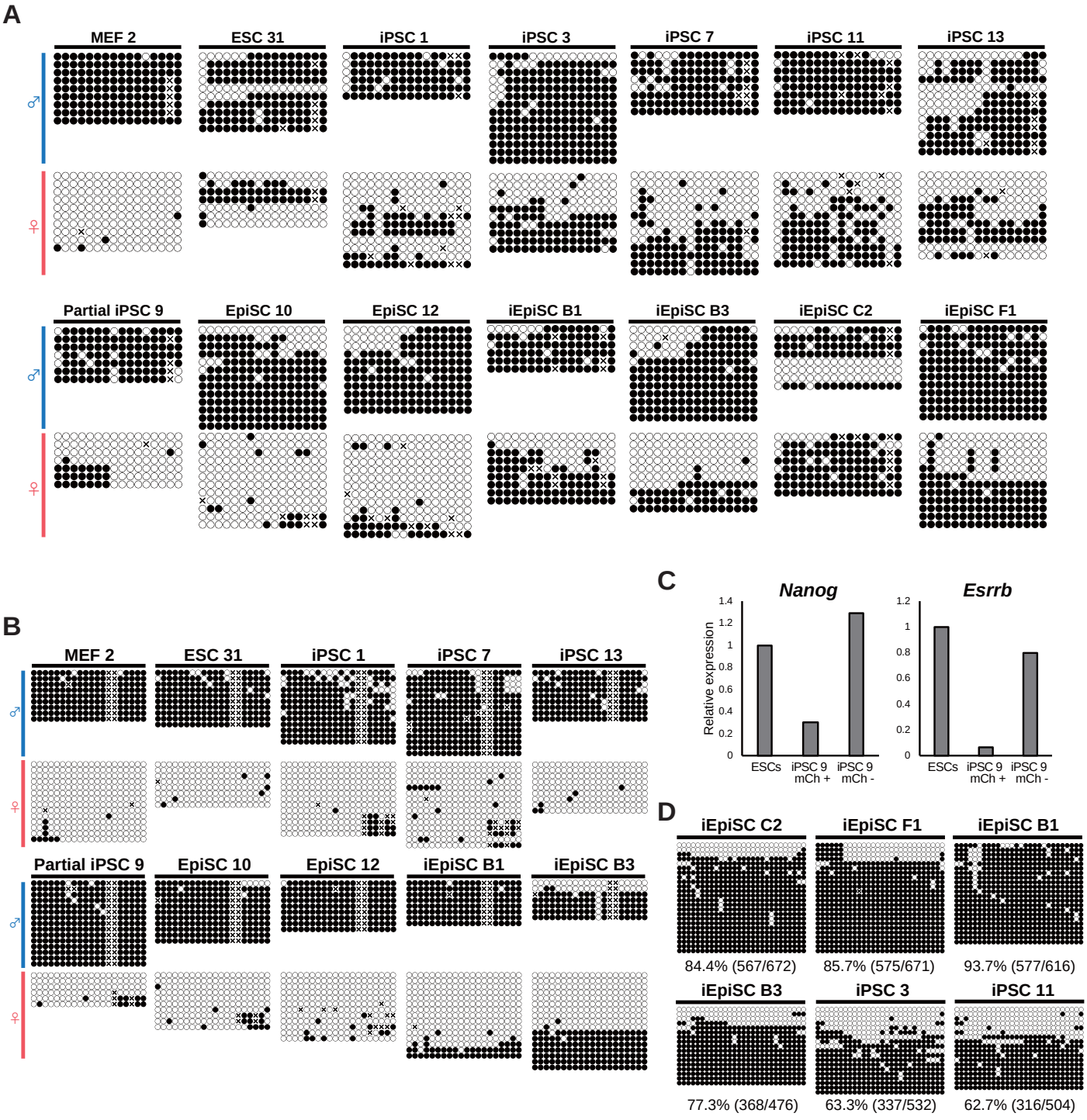

A
MEF 2
ESC 31
iPSC 1
iPSC 3
iPSC 7
iPSC 11
iPSC 13
Partial iPSC 9
EpiSC 10
EpiSC 12
iEpiSC B1
iEpiSC B3
iEpiSC C2
iEpiSC F1
♂
♀
B
MEF 2
ESC 31
iPSC 1
iPSC 7
iPSC 13
Partial iPSC 9
EpiSC 10
EpiSC 12
iEpiSC B1
iEpiSC B3
C
Nanog
Esrrb
Relative expression
ESCs
iPSC 9 mCh +
iPSC 9 mCh -
D
iEpiSC C2
iEpiSC F1
iEpiSC B1
84.4% (567/672)
85.7% (575/671)
93.7% (577/616)
iEpiSC B3
iPSC 3
iPSC 11
77.3% (368/476)
63.3% (337/532)
62.7% (316/504)

Figure S3, related to Figure 2 and Figure 3: Aberrant *de novo* methylation at ICRs in PSCs
A: DNA methylation analysis at *H19* DMR in MEFs, ESCs, EpiSCs, iPSCs and iEpiSCs by conventional bisulfite sequencing. Open circles represent unmethylated CpGs, and closed circles represent methylated CpGs. Crosses indicate undermined methylation status.
B: DNA methylation analysis at *Rasgrf1* DMR in MEFs, ESCs, EpiSCs, iPSCs and iEpiSCs by conventional bisulfite sequencing. *Rasgrf1* DMR is *de novo* methylated at the maternal allele in iEpiSCs. Open circles represent unmethylated CpGs, and closed circles represent methylated CpGs. Crosses indicate undermined methylation status.
C: A qRT-PCR analysis for *Nanog* and *Esrrb* in ESCs, mCherry-positive iPSC 9 cells and mCherry-negative iPSC 9 cells. mCherry-negative iPSC 9 cells express comparative levels of *Nanog* and *Esrrb* to ESCs. Data are presented as the mean of technical triplicates. The mean expression level of ESCs was set to 1.
D: DNA methylation analysis at *Nap1l5* DMR in multiple iEpiSC and iPSC clones by conventional bisulfite sequencing. Note that iEpiSC clones exhibit hypermethylation at *Nap1l5* DMR. Open circles represent unmethylated CpGs, and closed circles represent methylated CpGs. Crosses indicate undermined methylation status.

Figure S4

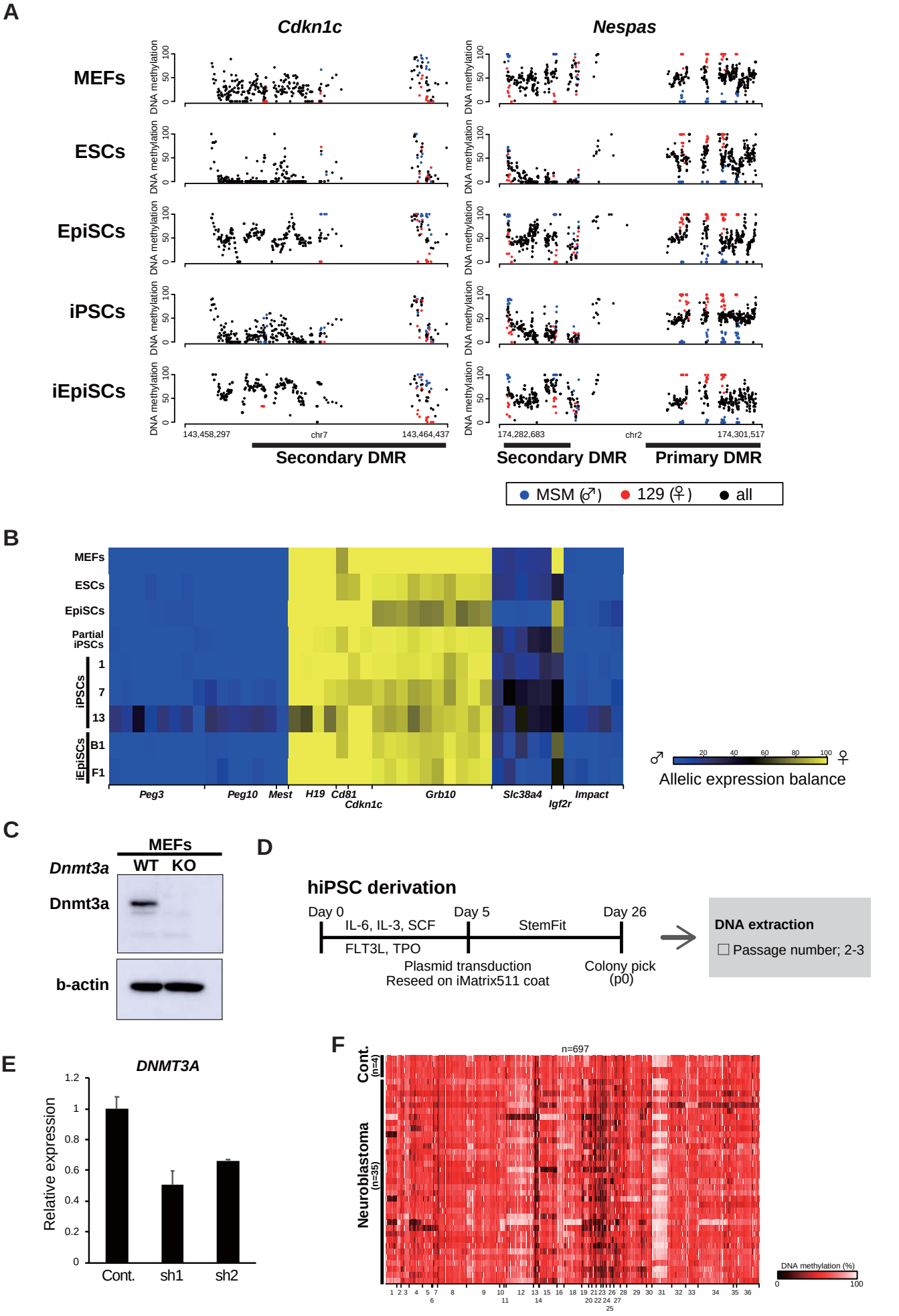

A
Cdkn1c
Nespas
MEFs
ESCs
EpiSCs
iPSCs
iEpiSCs
DNA methylation
0
50
100
143,458,297
chr7
143,464,437
Secondary DMR
174,282,683
chr2
174,301,517
Secondary DMR
Primary DMR
MSM (♂)
129 (♀)
all
B
MEFs
ESCs
EpiSCs
Partial iPSCs
iPSCs
1
7
13
iEpiSCs
B1
F1
Peg3
Peg10
Mest
H19
Cd81
Cdkn1c
Grb10
Slc38a4
Igf2r
Impact
♂
20
40
60
80
100
♀
Allelic expression balance
C
MEFs
Dnmt3a
WT
KO
Dnmt3a
b-actin
D
hiPSC derivation
Day 0
Day 5
Day 26
IL-6, IL-3, SCF
FLT3L, TPO
StemFit
Plasmid transduction
Reseed on iMatrix511 coat
Colony pick
(p0)
DNA extraction
☐ Passage number; 2-3
E
DNMT3A
Relative expression
1.2
1
0.8
0.6
0.4
0.2
0
Cont.
sh1
sh2
F
n=697
Cont.
(n=4)
Neuroblastoma
(n=35)
DNA methylation (%)
0
100

Figure S4, related to Figure 3, Figure 5 and Figure 6: DNA methylation analyses in normal tissues and cancers

A: CpG methylation at representative sDMRs of MEFs, ESCs, EpiSCs, iPSCs and iEpiSCs. Each black dot represents a methylation percentage for each CpG site. Red and blue dots indicate methylation levels at maternal 129 allele and paternal MSM allele, respectively. Note that DNA methylation levels at sDMRs are decreased in naïve PSCs (ESCs and iPSCs) while they are maintained in primed PSCs (EpiSCs and iEpiSCs).

B: Allelic expression analysis for imprinted genes of MEFs and PSCs. Genes containing >10 read counts in all samples are listed. Note that most "expressed" imprinted genes are monoallelically expressed.

C: Western blotting analysis for Dnmt3a confirmed a lack of Dnmt3a protein in *Dnmt3a* KO MEFs.

D: Protocol for hiPSC derivation with sh*DNMT3A* transduction.

E: A qRT-PCR analysis for evaluating knockdown efficiency by treatment with shRNA targeting *DNMT3A*. Hela cells were used to evaluate the efficiency. Data are presented as the mean ± SD of triplicates. The mean expression level of control shRNA-treated cells was set to 1.

F: Heat map for DNA methylation levels at imprinted DMRs in adrenal gland tissues, fetal brain tissues and neuroblastomas. Color scale is shown for DNA methylation levels. Infinium450K data were obtained from GSE54719.

Supplemental Experimental Procedures

Data reporting

No statistical methods were used to predetermine sample size. None of the experiments were randomized. The investigators were not blinded to allocation during the experiments.

Establishment and culture of ESCs

Zygotes with an (129X1/SvJ × MSM/Ms) F1 genetic background were obtained by *in vitro* fertilization (IVF). ESCs established in a previous study were used in this study (Yagi et al., 2017). Cells were maintained at 37 °C with 5% CO_2. Briefly, ESCs were established and maintained on feeders (mitomycin C (Kyowa Hakko Kirin Co., Ltd.)-treated MEFs) in knockout DMEM (GIBCO) with 2 mM L-glutamine (Nacalai tesque), 100×NEAA (Nacalai tesque), 100 U/ml penicillin, 100 μg/ml streptomycin (P/S) (Nacalai tesque), 15% Fetal Bovine Serum (FBS) (GIBCO), 0.11 mM β-mercaptoethanol (GIBCO) and 1000 U/ml human recombinant leukemia inhibitory factor (LIF) (Wako). For the establishment of ESCs, a blastocyst at E3.5 was placed on mitomycin C treated MEFs (feeders) in 24 well plates (passage number 0: p0). After the expansion of ICM, the cells were passaged into 24 well plates (p1), followed by a second passage into 6 well plates (p2) and then a third passage into 6 cm dishes (p3). ESCs were maintained on feeders. Male ESCs were used in this study.

Establishment and culture of EpiSCs

Male MSM/Ms mice were mated with female 129X1/SvJ mice. Noon of the day when the plug was observed was designated as embryonic day (E) 0.5. The day before isolating the embryos, feeders were harvested on 0.2% gelatin-coated 24-well culture plates using MEF medium. At E6.5, the uterus was removed by cutting across the cervix and two uterotubal junctions, and the embryos were placed in HEPES (GIBCO). Then, the muscle layer was removed, and the deciduum was dissected by using needles. Epiblasts were divided from the extra-embryonic regions and transferred into prepared 24-well culture plates to derive EpiSCs. After the outgrowth of epiblasts, cells were passaged on feeders in 24-well culture plates (p1), 6-well culture plates (p2), and 6 cm culture dishes (p3). EpiSCs were cultured in DMEM/F12+L-Glu (GIBCO), NEAA, P/S, 0.11 mM β-mercaptoethanol, 20% KnockOut Serum Replacement (KSR) (GIBCO), which was supplemented with 10 ng/ml Human Recombinant Fibroblast Growth Factor (basic) (rh bFGF) (wako), and 10 μM Y-27632 (wako). Cells were maintained at 37 °C

with 5% CO_2. Male EpiSCs were used in this study.

Establishment and culture of MEFs

MEFs were isolated from E14.5 embryos (129X1/SvJ × MSM/Ms) and cultured in DMEM (Nacalai tesque) with 2 mM L-glutamine, NEAA, P/S and 10% FBS. Cells were maintained at 37 °C with 5% CO_2.

Generation and culture of iPSCs

Male MEFs (129X1/SvJ × MSM/Ms) at p1-2 were seeded onto 6-well plates (1-2×10^5 cells/well) pre-coated with 0.2% gelatin using MEF medium. Doxycycline (Dox)-inducible PB vector containing tetO-*Oct4-Sox2-Klf4-cMyc*-IRES-mCherry-EF1-rtTA-IRES-Neo (PB-*OSKM*) was used for the reprogramming. 2.5 μg of PB-*OSKM* and 2.5 μg of Transposase were transfected using Fugene HD Transfection Reagent according to the supplier's instruction. After the transfection, cultured medium was switched to ESC medium containing 2 μg/ml Dox (Sigma). At day 10-12, emerged iPSC-like colonies were picked up and harvested onto 96-well plates (p1). Each iPSC clone was passaged into a 24-well plate (p2), 6-well plate (p3) and finally a 6 cm dish (p4). iPSCs at p4 were established as individual clones and used in this study.

Generation and culture of iEpiSCs

Male MEFs (129X1/SvJ × MSM/Ms) at p1-2 were seeded onto 6-well plates (1-2×10^5 cells/well) pre-coated with 0.2% gelatin using MEF medium. Doxycycline (Dox) inducible PB vector containing tetO-*Oct4-Sox2-Klf4-cMyc*-IRES-mCherry-EF1-rtTA-IRES-Neo (PB-*OSKM*) was used for reprogramming. 2.5 μg of PB-*OSKM* and 2.5 μg of Transposase were transfected using Fugene HD Transfection Reagent according to the supplier's instruction. After the transfection, cultured medium was switched to iEpiSC medium: DMEM/F12+L-Glu supplemented with B27 (GIBCO) and N2 cell-supplements (GIBCO), NEAA, P/S, 0.11 mM β-mercaptoethanol, 15% FBS, which was supplemented with 10 ng/ml rh bFGF, 20 ng/ml Recombinant Activin A (R&D), 200 ng/ml anti-LIF antibody (Funakoshi), and 2 μg/ml Dox (Sigma). At around day 21, emerged iEpiSC-like colonies were picked up and harvested onto 96-well plates (p1). Each iEpiSC clone was passaged into a 24-well plate (p2), 6-well plate (p3) and finally a 6 cm dish (p4). During the expansion of iEpiSCs, cells were maintained in iEpiSCs medium with 10 μM Y-27632. iEpiSCs at p4 were established as individual clones and used in this study.

Generation of iPSCs and iEpiSCs from *Dnmt3a* deficient MEFs

Dnmt3a WT and KO MEFs were obtained by crossing *Dnmt3a* hetero knockout (B6;129S4-Dnmt3a<tm1Enl>) mice (Okano et al., 1999). *Dnmt3a* WT and KO iEpiSCs were generated by Dox inducible PB-OSKM. *Dnmt3a* WT and KO iPSCs were generated in a previous study (Stadtfeld et al., 2012).

Removal of feeders by FACS

All PSC lines (ESCs, EpiSCs, iPSCs, iEpiSCs) cultured on feeders were sorted by fluorescence activated cell sorting (FACS) (Aria II, BD) to remove feeders. Alexa Flour 647 conjugated anti-SSEA1 antibody (Santa Cruz, sc-21702 AF647 clone MC480 or BD Pharmingen, Clone MC480, cat: 562277) was used for the sorting. Sorted cells were used for DNA and RNA isolation.

Animals

All experiments using animals were performed under the ethical guidelines of Kyoto University and University of Tokyo. MSM/Ms were obtained from RIKEN Bio Resource Center (Takada et al., 2013; Takada et al., 2015). 129X1/SvJ, ICR mice and pseudopregnant ICR mice were purchased from SLC. Noon of the day when the plug was observed was designated as embryonic day (E) 0.5.

Generation and culture of hiPSCs

Human peripheral blood mononuclear cells (Cellular Technology Ltd.) isolated from a 29 year-old female were cultured in StemSpan ACF medium (StemCell Technologies) supplemented with 100 ng/mL IL-6, 300 ng/mL SCF, 300 ng/mL TPO, 300 ng/mL FLT3L and 10 ng/mL IL-3 for 5 days. After transduction of the plasmid mixture (pCXLE-hOCT3/4-shp53-F, pCXLE-hUL, pCXWB-EBNA1 and pCXLE-hSK encoding shRNA for DNMT3A) by Amaxa Human CD34+ Cell Nucleofector Kit (Lonza), the cells were seeded on a 6-well plate coated with iMatrix-511 (Takara) and cultured until hiPSC colony formation. The selected hiPSCs were expanded in StemFit AK02N (Takara). To construct shRNA, the following oligos were hybridized, inserted into pSilencer 1.0-U6 (Thermo Fisher Scientific) and cloned into *BamH*I site of pCXLE-hSK: type-1F (5'-GCATCCACTGTGAATGATAAGTTCAAGAGACTTATCATTCACAGTGGATGCTTTTTTGG-3'), type-1R (5'-AATTCCAAAAAAGCATCCACTGTGAATGATAAGTCTCTTGAACTTATCATTCACAGTGGATGCGGCC-3') (Rinaldi et al., 2016), type-2F (5'-

GCCCAAGGTCAAGGAGATTATTTCAAGAGAATAATCTCCTTGACCTTGGGCTT TTTTGG-3') and type-2R (5'-AATTCCAAAAAAGCCCAAGGTCAAGGAGATTATTCTCTTGAAATAATCTCCTT GACCTTGGGCGGCC-3') (Onder et al., 2012).

Cell line availability

Cell lines used for this study were derived in the Yamada laboratory or Hochedlinger laboratory and tested for mycoplasma contamination.

Gender determination

The gender of the cell lines and mouse embryos was determined by PCR using primers for *Sry* and *Uty* located on the Y chromosome. The primers used were *Sry*-Fw: CGTGGTGAGAGGCACAAGTT; *Sry*-Rv: AGGCAACTGCAGGCTGTAAA; *Uty*-Fw: GAGTTCTTCTTGCGTTCACCATCTG; and *Uty*-Rv: CTATCTAATCCACAAAGCGCCTTCTTC.

DNA/RNA extraction and cDNA synthesis

DNA was extracted by PureLink Genomic DNA Mini Kit (Invitrogen), and RNA was extracted by RNAeasy plus Mini Kit (QIAGEN) or NucleoSpin RNA Plus (MACHEREY-NAGEL). DNA and RNA were quantified by Nanodrop 2000c (Thermo Scientific) and Qubit (Thermo Scientific). Reverse transcription was performed with PrimeScript^{TM}II 1st strand cDNA Synthesis Kit (Takara) according to the supplier's instructions.

Quantitative RT-PCR

Quantitative PCR was performed using GoTaq qPCR Mater Mix and CXR Reference Dye according to the supplier's instruction (Promega). StepOnePlus Real-Time PCR system (Applied Biosystems) was used. Transcript levels were normalized by β-actin. The experiments were performed in technically triplicate. The primers used were: *Oct4* (*Oct4-Sox2*)-Fw: TGGGCTCTCCCATGCATTCA; *Sox2* (*Oct4-Sox2*)-Rv: GGCTTCAGCTCCGTCTCCAT; *β-Actin*-Fw: GCCAACCGTGAAAAGATGAC; *β-Actin*-Rv: TCCGGAGTCCATCACAATG; *DNMT3A*-Fw: GGCATCCACTGT GAATGATAAG; *DNMT3A*-Rv: GGAGTTTGACCTCGTAGTAATG; *Nanog*-Fw: TGCTTACAAGGGTCTGCTACTG; and *Nanog*-Rv: TAGAAGAATCAGGGCTGCCTTG; *Esrrb*-Fw: GCACCTGGGCTCTAGTTGC; and *Esrrb*-Rv: TACAGTCCTCGTAGCTCTTGC; *GAPDH*-Fw:

AAGGTCATCCCAGAGCTGAA; and *GAPDH*-Rv: CTGCTTCACCACCTTCTTGA.

Western blot

Cells were dissolved in 500 μl of RIPA lysis buffer containing protease inhibitor cocktail (Nacalai Tesque), dithiothreitol (Wako), and phosphatase inhibitor cocktail (Nacalai Tesque). Proteins were denatured with 2× SDS at 95°C for 5 min. A total of 30 μg of denatured protein was run on a 10% SDS/PAGE gel and transferred to Amersham Hybond-P PVDF Membrane (GE Healthcare) using PowerPac HC (Bio-Rad). Membranes were blocked in blocking buffer [1× TBS with 0.05% Tween-20 (TBST) containing 4% nonfat milk (Nacalai Tesque)], and then incubated overnight at 4°C with primary antibodies diluted in blocking buffer. Primary antibodies were as follows: anti-Dnmt3a (Santa Cruz Biotechnology, sc-20703; dilution, 1:200) and anti-β-actin (Santa Cruz Biotechnology, sc-47778; dilution, 1:1000). Blots were rinsed with TBST. Membranes were incubated for 90 min at room temperature with HRP-conjugated secondary antibodies. Secondary antibodies used were ECL Anti-mouse IgG, HRP-linked whole antibody from sheep (NA931V, GE Healthcare), and ECL Anti-rabbit IgG, HRP-linked whole antibody from donkey (NA934V, GE Healthcare). After rinsing with TBST, Pierce ECL plus Western Blotting Substrate (Thermo Scientific) was used for visualization, and bands were detected on an LAS4000 (GE Healthcare).

Allelic expression analysis

A SNP of *Igf2* was used to distinguish parental alleles. cDNA was amplified using KOD-Plus-Neo (TOYOBO). Amplified products were cloned into pCR4-TOPO-Blunt vector (Invitrogen). Subcloned colonies were sequenced with a forward primer by ABI 3500xL. The primers used were *Igf2*-Fw: CATCGTGGAAGAGTGCTGCT; and *Igf2*-Rv: TTGAGCTCTTTGGCAAGCAT.

Bisulfite sequencing

200 ng of DNA was bisulfite-treated using EZ DNA Methylation-Gold Kit™ (ZYMO RESEARCH) according to the supplier's instruction. PCR was performed with EX Taq HS (Takara). PCR products were cloned into pCR4-TOPO vector (Invitrogen). Subcloned colonies were sequenced with a M13 reverse primer by ABI 3500xL (Applied Biosystems). The primers used were *H19*-Fw out: GAGTATTTAGGAGGTATAAGAATT and *H19*-Rv out: ATCAAAAACTAACATAAACCCCT; *H19*-Fw in: GTAAGGAGATTATGTTTATTTTTGG and *H19*-Rv in:

CCTCATTAATCCCATAACTAT; *Rasgrf1*-Fw out: GAGAGTATGTAAAGTTAGAGTTGTGTTG and *Rasgrf1*-Rv out: ATAATACAACAACAACAATAACAATC; *Rasgrf1*-Fw in: TAAAGATAGTTTAGATATGGAATTTTGGG and *Rasgrf1*-Rv in (same as *Rasgrf1*-Rv out): ATAATACAACAACAACAATAACAATC; and *Gtl2*-Fw: GGAAGGAAAAGATAAAATGTAGAAA and *Gtl2*-Rv CATAAATAAATAAACCCATAATCCC; *Nap1l5*-Fw: AGAGTTTGGAATTTTTTGTTAAATTTGG and *Nap1l5*-Rv TCTCTAAACCAACTCTATTACAAACT; *IGF2* DMR2-Fw: GGGATTGGGTTAGGAGAAGTTT and *IGF2* DMR2-Rv CCCCCAAAAATAACCAACAAT; and *MEG3*-Fw: GTAAGTTTTATAGGTTGTAAAGGGGGTGTT and *MEG3*-Rv: CCACAACTAATAACTAAAAAAATAAACATT.

Chimera formation

Female ICR mice were treated with pregnant mare serum gonadotropin (PMSG; 7.5 IU) and human chorionic gonadotropin (hCG; 7.5 IU) by intraperitoneal (i.p.) injection. Embryos were rinsed with M2 medium (Sigma) and cultured in KSOM medium until development into blastocysts. ESCs were trypsinized, and 7-8 cells per embryo were injected into the blastocoels of E3.5 blastocysts. Injected blastocysts were transferred into the uteri of pseudopregnant ICR mice.

Library preparation for target-captured bisulfite sequencing (MethylC-seq)

3 μg of DNA calculated by Qubit was fragmented by sonication. Subsequently, library preparation was performed with SureSelect^{XT} Mouse Methyl-Seq Reagent Kit (Agilent Technologies). The Methyl-Seq Kit could enrich 109 Mb of mouse genomic regions including CpG islands, Gencode promoters, DNase I hypersensitive sites and tissue specific DMRs. DNA was bisulfite-treated using the EZ DNA Methylation-Gold Kit^{TM}. Sequencing libraries were assessed by Bioanalyzer and quantified with KAPA Library Quantification Kits. The libraries were then sequenced on HiSeq 2500 (2x 100 bp or 2x 101 bp paired-end reads, Illumina).

Library preparation for RNA sequencing

200 ng of total RNA was prepared for library construction. High-quality RNA (RNA Integrity Number value >=8) assessed by Bioanalyzer was used for the library preparation. RNA-seq libraries were generated using the Truseq Stranded mRNA LT

sample prep kit (Illumina). PolyA-containing mRNA was purified by poly-T oligo-attached magnetic beads, and the RNA was fragmented and primed for cDNA synthesis. Cleaved RNA fragments were reverse transcribed into first strand cDNA using transcriptase and random primers. Second strand cDNA was synthesized by the incorporation of dUTP, and ds cDNA was separated using AMPure XP beads (BECKMAN COULTER). A single 'A' nucleotide was added to the 3' ends of the blunt fragments, and then adapters with index were ligated to the ends of the ds cDNA. ds cDNA fragments were amplified by PCR with PCR primer Cocktail. The number of PCR cycles was minimized (11-15 cycles) to avoid skewing the representation of the libraries. RNA-seq libraries were sequenced on NextSeq500 (75 bp single, Illumina).

DNA methylation analyses for target captured MethylC-seq data

For allelic methylation analyses, the SNPs data for MSM/Ms were obtained from NIG Mouse Genome Database (MSMv4HQ, http://molossinus.lab.nig.ac.jp/msmdb/index.jsp), and MSM/Ms mouse genome was reconstructed from mm10 using the SNPs. The information about in/dels was not used in this study. Y chromosome and M chromosome were omitted from the MSM/Ms reference genome because of the lack of SNP data. The bases with low quality scores and the adapters in all the sequenced reads were trimmed with cutadapt-1.14 (Martin, 2011). The filtered reads were mapped to both B6 mouse genome (mm10) and MSM/Ms mouse genome independently by Bismark software-v0.18.2 (Krueger and Andrews, 2011) with bowtie2 (version 2.3.2) (Langmead and Salzberg, 2012). The reads uniquely mapped to the same chromosome and positions of both B6 and MSM/Ms genomes were used for further analyses. The B6-derived and MSM/Ms-derived sequenced reads were determined based on the MSM/Ms SNPs data. SNPs in CpG sites were excluded, and the patterns of bisulfite conversion were taken into account in the determination of parental alleles. Methylated cytosines were extracted from the reads by the Bismark methylation extractor with --ignore 10 --ignore_r2 10 --ignore_3prime 5 --ignore_3prime_r2 5 options. Because probe sequences of the SureSelect^{XT} were designed for the original top strand, only the original bottom (OB) data were used for the analysis. The R software (v3.3.2) was used to visualize the CpG methylated status, create heat maps and perform clustering analysis. The methylation percentages for a CpG site at ≥ 5x coverage in all samples were used for clustering analysis. In the dot plots, each dot represents the methylation percentage for a CpG site at ≥ 5x coverage. CpG sites at ≥ 5x coverage in all samples were captured by heat maps. Previously described mouse CGIs (Illingworth et al., 2010) and mouse ICRs (Court et al., 2014; Tomizawa et al., 2011) were used for CGI and imprinting analyses, respectively.

To reduce effects of artifact genomic regions, the ENCODE blacklisted regions (Consortium, 2012) were excluded from the CGIs for our methylation analyses with NGS. The UCSC LiftOver tools (http://genome.ucsc.edu/,(Rosenbloom et al., 2015) were used to convert the coordinates of mm9 assembly into those of mm10 assembly. After the conversion of the CGI regions, overlap regions were merged into a single region. The CGI regions which include >= 10 CpG coverages of sequenced reads were used for comparison of CGI methylation levels between MEFs and PSCs.

DNA methylation analyses for infinium methylation array data

Infinium array data were obtained from publically available data sets (GEO) to represent somatic cells (GSE60923), PSC cells (GSE60821), normal tissues (GSE53051, GSE59157, GSE54719) and tumor samples (GSE53051, GSE54719, GSE58298, GSE59157, GSE62298, GSE70303, GSE75041, and GSE85845). Previously described human CGIs (Illingworth et al., 2010), human ICRs (Court et al., 2014; Tomizawa et al., 2011) and PcG target genes (Lee et al., 2006) were used for methylation analyses. The UCSC LiftOver tools (http://genome.ucsc.edu/,(Rosenbloom et al., 2015) were used to convert the coordinates of hg18 assembly into those of hg19 assembly. After the conversion of the CGI regions, overlap regions were merged into a single region. The probe data lacking methylation signals in any of the samples in each heat map of ICRs were excluded for visualizing the methylation levels. The average methylation signals of each ICR were used for comparison among samples. The CGI regions which include >= 10 probe data (infinium methylation signals) were used for representation of boxplots and clustering analyses.

RNA-seq data analyses

The sequenced reads were mapped to the mouse reference genome (mm10) using STAR version 2.5.3 (Dobin et al., 2013) with the GENCODE version M15 annotation gtf file and the aligner Bowtie2-2.3.2 (Langmead and Salzberg, 2012) after trimming adaptor sequences and low-quality bases by cutadapt-1.14 (Martin, 2011). For allelic expression analysis, the filtered reads were also mapped to MSM genomes used in the DNA methylation data analyses section, and the reads mapped to the same chromosome and positions of both B6 and MSM/Ms genomes with high mapping quality (MAPQ>=20) were used for further analyses. The expression level of each gene was calculated as Fragments Per Kilobase of exon model per Million mapped fragments (FPKM) by cufflinks-2.2.1 (Trapnell et al., 2010). For PCA analysis, low expressed genes (FPKM < 5) in all samples were excluded, and log2 (FPKM + 1) values were used.

Data availability

All data analyzed by MethylC-seq and RNA-seq have been deposited in the Gene Expression Omnibus (GEO) under accession number GSE111173. MethylC-seq and RNA-seq data of MEFs have been deposited in GSE84165. MethylC-seq data, WGBS data for ESC sample1-4 and RNA-seq data of ESCs have been deposited in GSE84165.

References

Consortium, E.P. (2012). An integrated encyclopedia of DNA elements in the human genome. Nature *489*, 57-74.

Court, F., Tayama, C., Romanelli, V., Martin-Trujillo, A., Iglesias-Platas, I., Okamura, K., Sugahara, N., Simon, 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.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15-21.

Illingworth, R.S., Gruenewald-Schneider, U., Webb, S., Kerr, A.R., James, K.D., Turner, D.J., Smith, C., Harrison, D.J., Andrews, R., and Bird, A.P. (2010). Orphan CpG islands identify numerous conserved promoters in the mammalian genome. PLoS Genet *6*, e1001134.

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.

Lee, T.I., Jenner, R.G., Boyer, L.A., Guenther, M.G., Levine, S.S., Kumar, R.M., Chevalier, B., Johnstone, S.E., Cole, M.F., Isono, K.*, et al.* (2006). Control of developmental regulators by Polycomb in human embryonic stem cells. Cell *125*, 301-313.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. 2011 *17*.

Okano, M., Bell, D.W., Haber, D.A., and Li, E. (1999). DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. Cell *99*, 247-257.

Onder, T.T., Kara, N., Cherry, A., Sinha, A.U., Zhu, N., Bernt, K.M., Cahan, P., Marcarci, B.O., Unternaehrer, J., Gupta, P.B.*, et al.* (2012). Chromatin-modifying enzymes as modulators of reprogramming. Nature *483*, 598-602.

Rinaldi, L., Datta, D., Serrat, J., Morey, L., Solanas, G., Avgustinova, A., Blanco, E., Pons, J.I.,

Matallanas, D., Von Kriegsheim, A.*, et al.* (2016). Dnmt3a and Dnmt3b Associate with Enhancers to Regulate Human Epidermal Stem Cell Homeostasis. Cell Stem Cell *19*, 491-501.

Rosenbloom, K.R., Armstrong, J., Barber, G.P., Casper, J., Clawson, H., Diekhans, M., Dreszer, T.R., Fujita, P.A., Guruvadoo, L., Haeussler, M.*, et al.* (2015). The UCSC Genome Browser database: 2015 update. Nucleic Acids Res *43*, D670-681.

Stadtfeld, M., Apostolou, E., Ferrari, F., Choi, J., Walsh, R.M., Chen, T., Ooi, S.S., Kim, S.Y., Bestor, T.H., Shioda, T.*, et al.* (2012). Ascorbic acid prevents loss of Dlk1-Dio3 imprinting and facilitates generation of all-iPS cell mice from terminally differentiated B cells. Nat Genet *44*, 398-405, S391-392.

Takada, T., Ebata, T., Noguchi, H., Keane, T.M., Adams, D.J., Narita, T., Shin, I.T., Fujisawa, H., Toyoda, A., Abe, K.*, et al.* (2013). The ancestor of extant Japanese fancy mice contributed to the mosaic genomes of classical inbred strains. Genome Res *23*, 1329-1338.

Takada, T., Yoshiki, A., Obata, Y., Yamazaki, Y., and Shiroishi, T. (2015). NIG_MoG: a mouse genome navigator for exploring intersubspecific genetic polymorphisms. Mamm Genome *26*, 331-337.

Tomizawa, S., Kobayashi, H., Watanabe, T., Andrews, S., Hata, K., Kelsey, G., and Sasaki, H. (2011). Dynamic stage-specific changes in imprinted differentially methylated regions during early mammalian development and prevalence of non-CpG methylation in oocytes. Development *138*, 811-820.

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.

Yagi, M., Kishigami, S., Tanaka, A., Semi, K., Mizutani, E., Wakayama, S., Wakayama, T., Yamamoto, T., and Yamada, Y. (2017). Derivation of ground-state female ES cells maintaining gamete-derived DNA methylation. Nature *548*, 224-227.