Differentiation-defective phenotypes revealed by large-scale analyses of human pluripotent stem cells

Mikiyoh Koyanagi-Aoi,1,2 Mari Ohnuki,1,2 Kazutoshi Takahashi, Keisuke Okita, Hisashi Noma, Yuka Sawamura, Ito Teramoto, Megumi Narita, Yoshiko Sato, Tomoko Ichisaka, Naoki Amano, Akira Watanabe, Yasuhiro Yamada, Toshiyuki Sato, Jun Takahashi, and Shinya Yamanaka

Abstract

We examined the gene expression and DNA methylation of 49 human induced pluripotent stem cells (hiPSCs) and 10 human embryonic stem cells (hESCs) that had been cultured under the same conditions. Comparisons of the in vitro neural differentiation of 10 hESCs and 10 human embryonic stem cells showed that seven hiPSC clones retained a significant number of undifferentiated cells even after neural differentiation culture and formed teratoma when transplanted into mouse brains. These differentiation-defective hiPSC clones were marked by higher expression levels of several genes, including those expressed from long terminal repeats of specific human endogenous retroviruses. These data demonstrated a subset of hiPSC lines that have aberrant gene expression and defective potential in neural differentiation, which need to be identified and eliminated before applications in regenerative medicine.

Significance

In the past few years, findings have been controversial in regard to whether hESCs and hiPSCs are distinct cell types. Some researchers have reported that their differences in their gene expression (7–10), DNA methylation (10–13), and capacity for differentiation (14). In the latter papers, relatively small numbers of cell lines were generally compared. In addition, most comparisons used pluripotent cell lines from various laboratories, so the observed differences may be attributable to laboratory-specific variations owing to technical differences (15).

H

Human pluripotent stem cells possess a robust potential for proliferation and provide useful sources of cells for regenerative medicine and drug discovery. Two types of human pluripotent stem cells have been generated: human embryonic stem cells (hESCs) derived from blastocysts (1) and induced pluripotent stem cells (hiPSCs), which are generated from somatic cells by factor-mediated reprogramming (2, 3).

The in vitro directed neural differentiation of these pluripotent stem cells.

Results

Overlapped Variations of mRNA Expression and DNA Methylation in hiPSCs and hESCs.

We analyzed a total of 49 hiPSCs derived from four types of somatic cells, including human dermal fibroblasts (HDFs), dental-pulp stem cells (DP), cord blood cells (CB), and peripheral blood mononuclear cells (PBMC), generated using three gene delivery methods, including those using retroviruses, nonintegration episomal plasmids, and Sendai viruses (Table 1 and Dataset S1). Most clones were generated in our own laboratory, except for three clones that were established in another laboratory (16). Before the analyses of gene/miRNA expression (Fig. 1A and B) and DNA methylation (Fig. 1C) we cultured these hiPSCs, as well as 10 hESCs, under the same culture conditions for at least three passages. In addition, we analyzed the original somatic cells, two human embryonic carcinoma cell (hECC) lines (Ntera2 cloneD1 and 2102EpbD3), and three cancer cell lines (HepG2, MCF7, and Jurkat).

The miRNA microarray analyses (Fig. 1A) identified 61 probes with significant differences in expression between hESCs and hiPSCs [FDR < 0.05]. Each of the 61 probes showed variable expression among both the hESCs and hiPSCs, and the distributions of the expression levels in the two groups overlapped (Fig. 1D). Of note, hESCs established at Kyoto University (Kyoto hESCs) were more similar to hiPSCs than to the remaining hESCs (other hESCs) in their expression of 15 probes that were differentially expressed between hESCs and hiPSCs [FDR < 0.05 and fold change (FC) > 3] (Fig. S1A). In contrast, hierarchical clustering using all probes showed no clear-cut separation among Kyoto hESCs, other ESCs, and iPSCs, indicating that the similarities between Kyoto ESCs and iPSCs are confined to a small set of genes (Fig. S1B). In addition, the miRNA array analyses (Fig. 1A) did not find any significant differences between hESCs and hiPSCs (t test, FDR < 0.05). The expressions of hsa-miR-886-3p and hsa-miR-142-3p tended to be higher in hiPSCs, but the expression levels of these probes have been deposited in the Gene Expression Omnibus (GEO) database, www.ncbi.nlm.nih.gov/geo (accession no. GSE49053).


Conflict of interest statement: S.Y. is a member without salary of the scientific advisory boards of iPierian, IPS Academia Japan, Megakaryon Corporation, and HEALIOS K. K. Japan.

Data deposition: Gene expression, miRNA expression, DNA methylation, and exon array data have been deposited in the Gene Expression Omnibus (GEO) database, www.ncbi.nlm.nih.gov/geo (accession no. GSE49053).

†M.K.-A. and M.O. contributed equally to this work.

‡To whom correspondence should be addressed. E-mail: yamanaka@cira.kyoto-u.ac.jp.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1319061110/-/DCSupplemental.

www.pnas.org/cgi/doi/10.1073/pnas.1319061110

PNAS | December 17, 2013 | vol. 110 | no. 51 | 20569–20574

PNAS | December 17, 2013 | vol. 110 | no. 51 | 20569–20574

PNAS | December 17, 2013 | vol. 110 | no. 51 | 20569–20574
miRNAs showed overlapped variations among hiPSCs and hESCs (Fig. S1).

We next compared the global DNA methylation status between hiPSCs and hESCs by the Illumina Infinium Human Methylation27 BeadChip assay. Among 27,445 CpG dinucleotides examined, we did not identify significantly differentially methylated CpG regions (CG-DMR) between the hESCs and hiPSCs (Mann–Whitney U test, FDR < 0.05) (Fig. 1C).

We then validated the CG-DMRs reported in previous studies. Three studies identified a total of 205 regions as CG-DMRs, including 130, 71, and 4 regions identified by comparing five hiPSCs and two hESCs (13), three versus three (12), and nine versus three (10) cell lines, respectively. Of the 205 regions, 46 regions containing 66 CpG dinucleotides were covered by the Infinium platform used in our study (Table S1). Based on the methylation levels in our hiPSCs and hESCs, these CpGs were clustered into three groups (Fig. 1E). Two-thirds of these CpGs belonged to group A: They tended to be highly methylated in hESCs, ECCs, and cancer cell lines and to be hypomethylated in hiPSCs, as well as somatic cells. However, they were also hypo-methylated in Kyoto ESCs (17). The methylation status of the upstream region of the paraoxonase 3 (PON3), a representative example of CpGs in group A, was confirmed by pyrosequencing (Fig. 1F). Thus, the CpG methylation status in group A may
distinguish some, but not all, hESCs from hiPSCs. Seventeen CpGs belonged to group B, which showed similar methylation levels in both hESCs and hiPSCs. Five CpGs, representing three genes, belonged to group C and showed higher methylation levels in some, but not all, hiPSCs compared with hESCs. The remaining hiPSCs showed low methylation levels, comparable to those in hESCs. We confirmed the methylation status of a representative example of CpGs in group C, the upstream region of the transcription elongation regulator 1-like (TCERG1L), by pyrosequencing; the methylation levels were low in 21 of the 49 hiPSCs (Fig. 1G). Therefore, the CpGs in group C may distinguish some, but not all, iPSCs from hESCs.

A previous report (12) showed that many CG-DMRs were located in CpG shores, rather than CpG islands. Because only some, but not all, iPSCs from hESCs.

Levels in hESCs and hiPSCs are overlapped. They may distinguish some, but not all, hiPSCs from hESCs.

A Subset of hiPSC Clones Retain Undifferentiated Cells After Neural Differentiation. To examine whether hESCs and hiPSCs have comparable differentiation potential, we performed in vitro directed differentiation into neural stem and progenitor cells using the modified serum-free floating culture of embryoid body-like aggregates (SFEBq) method (Fig. 2) (18). We initially performed the neural induction of two hESCs and 21 hiPSCs. Fourteen days after induction, the differentiation efficiency was evaluated based on the expression of an early neural marker, polysialylated neural cell adhesion molecule (PSA-NCAM). We found that all hESCs and hiPSCs differentiated into PSA-NCAM+ cells with more than 80% efficiency (Fig. 2B). We also quantified the expression levels of the early neural marker paired box 6 (PAX6) and the late neural marker microtubule-associated protein 2 (MAP2) in neurospheres by quantitative RT-PCR (qRT-PCR) (Fig. S2A). All of the examined hES/iPSCs expressed PAX6 at >100-fold higher levels and MAP2 at >20-fold higher levels in comparison with undifferentiated H9 hESC. However, in some hiPSCs, we noticed slightly lower differentiation efficiency than in the remaining hiPSCs and hESCs (Fig. 2B). This lower efficiency in neural differentiation was inversely correlated with a higher proportion of POU class 5 homeobox 1 (POUSF1, also known as OCT3/4)+ and TRA1-60+ undifferentiated cells (Fig. 2C). We also detected residual undifferentiated cells after a different neural differentiation protocol using adhesion culture (19) (Fig. S2B).

We then increased the number of clones and examined the proportions of OCT3/4+ undifferentiated cells after neural induction from 10 hESCs and 40 hiPSCs. The 50 clones were ranked according to their proportions of OCT3/4+ cells on day 14 (Dataset S1and Fig. 2D). The proportions of OCT3/4+ cells varied from 0 to ~20%. Thirty-eight clones, including nine hESCs and 29 hiPSCs, showed less than 1% OCT3/4+ cells in all experiments. We designated these clones as “good” clones. However, seven hiPSCs contained more than 10% OCT3/4+ cells after neural differentiation in at least one experiment. We designated these clones as “differentiation-defective” clones. Clones that were not good or defective were categorized as “intermediate.”

Activation of Specific LTR7 Elements in Differentiation-Defective Clones. To identify molecular signatures that can predict differentiation-defective clones, we compared the global gene expression patterns of 38 good clones and seven differentiation-defective clones under the culture conditions used for the undifferentiated state. We identified 19 probes (13 putative transcripts) that showed greater than fivefold differences in expression, with an FDR <0.05, shown by magenta dots in Fig. 3A and listed in Table S1B.

Of the 19 probes identified, five probes recognized HHLA1 (human endogenous retrovirus-H LTR-associating 1). Previous reports have shown that HHLA1 is regulated by a long terminal repeat (LTR) of a human endogenous retrovirus-H (HERV-H) (20). The LTR in HHLA1 is classified as LTR7. Moreover, among the genes recognized by the 19 probes, we found that at least two others, αβ-hydroxylase domain containing 12B (ABHD12B) and chromosome 4 open reading frame 51 (C4orf51), also contained LTR7 sequences in their gene bodies. According to a microarray analysis, we confirmed that these three LTR7-containing genes were up-regulated in the differentiation-defective hiPSCs, as well as the nullipotent hECC line 2102Ep 4D3 (21), but they were expressed at lower levels in the good hiPSCs, hESCs, and pluripotent hECC line NTer2 cloneD1. They were almost not expressed in the original somatic cells (Fig. 3B).

The Agilent Technologies microarray platform has 12 probes, including two reverse probes [d(r), f(r)], for HHLA1 and its neighboring gene, otoconin 90 (OC90), which is reported to make a fusion transcript with HHLA1 (20) (Fig. 3C). Among them, seven probes located downstream of LTR7 showed higher expression levels in differentiation-defective clones than in good clones (Fig. 3 C and D). Similarly, there are two probes for ABHD12B, designed for exons 4 and 13 (Fig. 3C). Only the exon
13 probe, located downstream of LTR7, showed a higher expression in differentiation-defective clones than in good clones (Fig. 3D). We also performed an exon array (Affymetrix) of ABHD12B and C4orf51 and found that exons downstream of LTR7 were preferentially up-regulated in differentiation-defective hiPSC clones (Fig. 3E). We also found that the methylation status of LTR7s in these three genes were lower in differentiation-defective hiPSC clones than in good clones (Fig. 3F). These results indicate that the three genes are transcribed from activated LTR7.

**DNA Hypomethylation Exists in Some, but Not All, LTR7s in Differentiation-Defective Clones.** According to the Repeatmasker software program, there are 3,523 LTR7 elements in the human genome. To extract microarray probes that are potentially affected by LTR7s, we first selected genes containing LTR7s in their gene bodies or regions 2 kb upstream from their transcription start sites. We then retrieved the microarray probes located between each LTR7 and the 3’ end of the corresponding gene body. As a result, we selected 763 probes as LTR7-related probes (Fig. S3A and Table S1C). We found that most of these probes showed comparable expression levels in good and defective lines (Fig. S3B), with the exception of some probes, such as those corresponding to arrenalin, beta 1 (ARRB1), fatty acid amide hydrolase 2 (FAAH2), and TBC1 domain family, member 23 (TBC1D23), that were differentially expressed between good and defective clones (FDR <0.05 and FC >2) and showed slightly higher expression in defective lines.

We then checked the DNA methylation status of the LTR7 regions in these three genes and three other genes DNA (cytosine-5-)methyltransferase 3 beta (DNMT3B), ATP-binding cassette, sub-family A (ABCA1), member 1 (ABCA1), and amyloid beta (A4) precursor protein (APP) whose expression levels were not significantly different between the good and defective clones. By pyrosequencing and clonal bisulfite sequencing, we found that the LTR7 regions in four genes (ARRB1, FAAH2, TBC1D23, and APP) were hypomethylated in defective clones compared with good clones. In contrast, the LTR7 regions in two genes (DNMT3B and ABCA1) did not show such hypomethylation (Fig. S3C). Therefore, the activation of LTR7 is not confined to HHLA1, ABHD12B, and C4orf51; DNA hypomethylation exists in some, but not all, LTR7s in defective hiPSCs.

**Differentiation-Defective hiPSC Clones Form Teratomas in Mouse Brains.** To further evaluate the defective hiPSCs, we induced their differentiation into dopaminergic neurons, which were then transplanted into the striata of nonobese diabetic/severe combined immune-deficient (NOD/SCID) mouse brains (Fig. 4A). Thirty and 60 d after transplantation, we obtained T2-weighted images of the mouse brains with an MRI scanner to observe the graft sizes at the transplanted sites (Fig. 4B). The quantification of the MRI images showed that defective hiPSC clones resulted in significantly larger graft sizes than good clones (Fig. 4C). Notably, some mice that had received defective clones died or developed symptoms that required euthanasia before day 60 (Table S2). Therefore, we could not obtain the graft size data on day 60 in these mice.

To identify the composition of the surviving grafts, we performed a histological analysis of the brains of animals that died or that became moribund after transplantation. The remaining healthy mice were euthanized 14–41 wk after transplantation. Sections were stained with H&E. Thirty-six of the 42 grafts were positive for human neural cell adhesion molecule (NCAM) (Fig. S4A). The remaining healthy mice were euthanized 14–41 wk after transplantation. Sections were stained with H&E. Thirty-six of the 42 grafts were positive for human neural cell adhesion molecule (NCAM) (Fig. S4A). The remaining healthy mice were euthanized 14–41 wk after transplantation. Sections were stained with H&E. Thirty-six of the 42 grafts were positive for human neural cell adhesion molecule (NCAM) (Fig. S4A). The remaining healthy mice were euthanized 14–41 wk after transplantation. Sections were stained with H&E. Thirty-six of the 42 grafts were positive for human neural cell adhesion molecule (NCAM) (Fig. S4A). The remaining healthy mice were euthanized 14–41 wk after transplantation. Sections were stained with H&E. Thirty-six of the 42 grafts were positive for human neural cell adhesion molecule (NCAM) (Fig. S4A). The remaining healthy mice were euthanized 14–41 wk after transplantation. Sections were stained with H&E. Thirty-six of the 42 grafts were positive for human neural cell adhesion molecule (NCAM) (Fig. S4A). The remaining healthy mice were euthanized 14–41 wk after transplantation. Sections were stained with H&E. Thirty-six of the 42 grafts were positive for human neural cell adhesion molecule (NCAM) (Fig. S4A). The remaining healthy mice were euthanized 14–41 wk after transplantation. Sections were stained with H&E. Thirty-six of the 42 grafts were positive for human neural cell adhesion molecule (NCAM) (Fig. S4A). The remaining healthy mice were euthanized 14–41 wk after transplantation. Sections were stained with H&E. Thirty-six of the 42 grafts were positive for human neural cell adhesion molecule (NCAM) (Fig. S4A).
We also observed that 14 out of the 63 (22.2%) grafts from good clones, including those from hESCs, contained a nonneural component in the graft tissue after transplantation (Fig. 4D and Table S2), although these clones did not show high expression levels of OCT3/4 in the pretransplantation samples (Fig. 4E). We referred to these clones as “type-2 defective” clones, which were distinct from “type-1 defective” clones that contained OCT3/4+ undifferentiated cells in the pretransplantation samples. We observed higher expression levels of SRY (sex determining region Y)-box 17 (SOX17, an endoderm marker) and goosecoid homebox (GSC, an endoderm and mesoderm marker) in the pretransplantation samples of type-2 defective clones (Fig. 4F and G), demonstrating the presence of other lineages in these pretransplantation samples.

**Discussion**

We identified two types of defective pluripotent stem cell lines in this study. The first type consisted of hiPSCs that retained a substantial number of undifferentiated cells after in vitro directed neural differentiation. Seven out of the 40 iPSCs (17.5%) examined in this study fell into this category. In contrast, we did not observe such defects in any of the 10 hESCs. More clones should be analyzed to confirm that hESCs are free from this deficiency. Nevertheless, it is likely that type-1 defectiveness is more common in hiPSCs than in hESCs. The type-1 defective hiPSCs are accompanied by an aberrant epigenetic status. Among the 13 putative transcripts that were highly expressed in these defective clones, at least three were expressed from the LTR of endogenous retroviruses. Normally, theseLTRs are silenced by various epigenetic modifications, including DNA methylation (22–24). In type-1 defective iPSC clones, the LTR locus in the three genes showed lower DNA methylation levels than in good clones and original somatic cells. Notably, the same regions were hypomethylated in the nullipotent hESC line, 2102Ep4D3, suggesting that the loss of DNA methylation in these LTR loci is correlated with the lower ability to differentiate. At present, the biological significance and relationship between activation of specific LTRs and the defective phenotype is unclear. Recent reports showed that endogenous retrovirus may play roles in the establishment and maintenance of transcription network in pluripotent stem cells (25, 26). Furthermore, updated annotations revealed that one of the differentially expressed probes in type-1 defective hiPSCs (A_19_P00325604) encoded large intergenic noncoding RNA regulator of reprogramming (Line-ROR), which contained LTR7 in its 5’ region. Line-ROR is reported to have multiple roles in the induction and maintenance of pluripotency (27, 28).

Future studies should be undertaken to clarify why these epigenetic abnormalities occur and how they are related to the defective differentiation.

Kim et al. (29) showed that there is an inverse correlation between the hsa-mir-371–373 expression and the efficiency of neural differentiation. They also showed that Kruppel-like factor 4 (KLF4) may induce the expression of hsa-mir-371–373. In our analyses, the hsa-mir-371–373 cluster was highly expressed in all of the seven type-1 defective hiPSC clones (Fig. S5A). However, the cluster was also highly expressed in many good clones. KLF4 was highly expressed in some defective clones (Fig. S5B), and four out of six retroviral defective clones failed to silence KLF4 retrogene transgenes (Fig. S5C). Together, these findings indicate that high expression levels of the hsa-mir-371–373 cluster, KLF4, and transgenes cannot function as absolute markers for type-1 defectiveness.

We previously reported that the origin of mouse iPSCs was a major determinant of defectiveness in directed neural differentiation; mouse iPSCs from adult tail tip fibroblasts showed the highest incidence of resistance to differentiation (30). In the present study using human iPSCs, five out of seven type-1 defective clones were derived from fibroblasts of donors of various ages, and six out of the seven clones were generated using retroviruses (Table 1). This may suggest that type-1 defectiveness is associated with fibroblast origin and retroviral induction. However, in this study, most of the fibroblast-derived iPSCs were generated by retroviruses, and most of the nonfibroblast iPSCs were generated by nonretroviral methods. Future studies will need to be undertaken to determine whether the origin or the generation method (or both) has a significant impact on the frequency of type-1 differentiation-defective iPSCs.

The second type of defective group includes hiPSCs and hESCs that contained differentiated cells of nonneural lineages after in vitro directed differentiation into dopaminergic neurons. We have previously shown that the optimal conditions for hepatic
differentials were different for each clone (16). By optimizing the protocols, it may be possible to induce complete neural differentiation to avoid type-2 defective clones. Alternatively, purification of neural cells using a cell sorter may work to avert type-2 defective clones.

Several studies have reported sets of genes whose DNA methylation status is different between hiPSCs and hESCs. We validated these CG-DMRs and found that many of them can distinguish some hESCs from hiPSCs (group A in Fig. 1E). They are highly methylated in some hESCs, but not in hiPSCs or original somatic cells. Thus, these CG-DMRs may represent epigenetic memories of somatic cells in iPSCs. However, we found a set of hESCs that showed low methylation status of these CG-DMRs, which were comparable to hiPSCs. We also found another set of the reported CG-DMRs that showed high methylation status in some hiPSCs but not in original somatic cells or hESCs (group C in Fig. 1E). These likely represent aberrant methylation associated with reprogramming. However, we also found many hiPSCs showed normal methylation patterns of these CG-DMRs. A recent study identified nine genes that can segregate hiPSCs from hESCs in DNA methylation and gene expression (31). However, we did not observe such a clear distinction in gene expression of these genes between our hiPSCs and hESCs (Fig S6). Two of these genes, TCF7RGL and FAM19A5, may distinguish some but not all, hiPSCs from hESCs.

In our analyses, 35 hiPSCs had records of the donor's genetic background; 14 were derived from Caucasians and 21 were from Japanese subjects (Dataset S1). Thus, the similarity of some signatures between the Kyoto hESCs and our hiPSCs cannot be attributed to the racial or ethnic backgrounds of the donors. Another possible cause of the differences is the method used to establish the hESCs and the subsequent culture conditions. The Kyoto hESCs were generated on feeders consisting of a 1:1 mixture of mouse embryonic fibroblasts and SL1 cells (17, 32), which were subcloned from STO cells. Most of our hiPSCs were established on SNL feeders, which were also derived from STO cells. A recent report showed that the feeders have profound effects on established hiPSCs (33). To confirm the importance of the culture conditions, more studies comparing hESC/hiPSCs established under different conditions will be needed.

In conclusion, our results revealed that a subset of hiPSCs is deficient in neural differentiation and marked with activation of endogenous retroviruses. We also confirmed that some hiPSCs are different from hESCs in molecular signatures, including CG-DMRs, which has been previously reported. It remains to be determined whether these molecular signatures specific for some hiPSCs have functional consequences.

Materials and Methods

Gene expression profiling was carried out using the SurePrint G3 human GE microarray (Agilent Technologies). Most of the data were analyzed using the GeneSpring GX 11.5.1 software program (Agilent Technologies). Neuronal induction was performed as described previously (18). Detailed descriptions of methods are available in SI Materials and Methods.

Acknowledgments. We thank Drs. Takashi Aoi, Yoshinori Yoshida, Masato Nakagawa, and other members of the S.Y. research group for their valuable scientific discussions; Drs. Takuya Yamamoto, Toshiki Taya, and Tomaoki Tsubota for advice about bioinformatics; Dr. Hirofumi Suemori for hiPSC cell lines; Drs. Koji Eto and Naoya Takayama for TKCB hiPSC clones, and Shoko Otsuka for RNA and genomic DNA samples; Dr. Takafumi Kimura and Takaomi Futakami for HLA analysis; and Aki Sasaki, Tomomi Ito, Midori Yokura, and Ayumi Ichikawa for technical assistance. We also thank Naoki Nagata, Yoko Miyake, Rie Kato, Eri Minamitani, Sayaka Takehashima, and Ryoko Fujiwara for their valuable administrative support. This work was supported in part by Grants-in-Aid for Scientific Research from the Japanese Society for the Promotion of Science; from the Ministry of Education, Culture, Sports, Science, and Technology; by a grant from the Leading Project of the Ministry of Education, Culture, Sports, Science, and Technology; and by a grant from the Funding Program for World-Leading Innovative Research and Development on Science and Technology (First Program) of the Japan Society for the Promotion of Science. M.K.-A. and M.D. were Research Fellows of the Japanese Society for the Promotion of Science.

Koyanagi-Aoi et al. 10.1073/pnas.1319061110

Supporting Information

SI Materials and Methods

Cell Culture. The human embryonic stem cells (hESCs) and human induced pluripotent stem cells (hiPSCs) were maintained in Primate ES cell medium (ReproCELL) supplemented with 4 ng/mL of human recombinant basic fibroblast growth factor (bFGF) (Wako) on SNL feeders (1–3). Human dermal fibroblasts (HDFs) were obtained from the Japanese Collection of Research Bio-resources or were purchased from Cell Applications, Inc. Dental pulp (DP) cells were kindly provided by Dr. Ken-ichi Tszuka (Gifu University Graduate School of Medicine, Gifu Prefecture, Japan). HDFs and human embryonic carcinoma cell (hECC) lines were maintained in DMEM (Nacalai Tesque) containing 10% (vol/vol) FBS (Thermo) and 0.5% penicillin/streptomycin (Life Technologies). DP cells were cultured in MSBGM medium (Lonza), CD34+ cord blood cells were obtained from the Stem Cell Resource Network in Japan (banks at Miyagi, Tokyo, Kanagawa, Aichi, and Hyogo) through the RIKEN BioResource Center (Tsukuba, Ibaraki, Japan). The peripheral blood was harvested from healthy donors whose written informed consent was obtained in accordance with the ethical committee of the Department of Medicine and Graduate School of medicine, Kyoto University. The mononuclear cells were then isolated by density gradient centrifugation with Ficoll-Paque PLUS (GE Healthcare).

Generation of Human iPSCs. The generation of hiPSCs from HDFs, DP cells, and blood samples using a retroviral system or episomal vectors was performed as described previously (1, 4–6). TKCBV4-2, 5-6, and TKCB7-2 iPSCs were kindly provided by Drs. Koji Eto and Naoya Takayama (Center for iPS Cell Research and Application, Kyoto University, Kyoto, Japan) (7).

During the generation of hiPSCs from blood using Sendai viral vectors, vectors encoding POU class 5 homeobox 1 (OCT3/4), KROPP-like factor 4 (KLF4), and v-myv avian myelocytomatosis virus oncogene homolog (MYC, also known as c-MYC) (CytoTune-iPS; DNASPEC) were infected into CD34+ cells at a multiplicity of infection of 3 or 10 in αMEM medium supplemented with 10% (vol/vol) FBS, 50 ng/mL IL-6, 50 ng/mL IL-4R, 50 ng/mL stem-cell factor, 10 ng/mL thrombopoietin, 20 ng/mL Flt3 ligand, and 20 ng/mL IL-3. The next day, the infected cells were centrifuged to remove residual virus, plated onto six-well plates covered with MEF feeder cells, and cultured in Primate ES cell medium supplemented with 4 ng/mL of bFGF until colonies were formed. Sendai virus infection and the generation of iPSCs from αβT cells were carried out as described previously (8).

RNA Extraction. We lysed the cells at subconfluent density using TRIzol reagent (Life Technologies) and total RNA was purified by a standard protocol. The RNA concentration and purity were determined through measurement of the A260/280 ratios with a Nanodrop instrument (Thermo Scientific). For microarrays, confirmation of the RNA quality was performed using a 2100 Bioanalyzer (Agilent Technologies).

mRNA Expression Analysis. The gene expression profiling was carried out using the SurePrint G3 human GE microarray (Agilent Technologies) according to the manufacturer’s protocol. The data were analyzed using the GeneSpring GX 11.5.1 software program (Agilent Technologies). The data processing was performed as follows: (i) Threshold raw signals were set to 1.0, (ii) log base 2 transformation was performed, and (iii) 75th percentile normalization was chosen as the normalized algorithm (http://genespring-support.com/faq/normalization). The flag setting was performed as follows: Feature is not positive and significant (not detected), not uniform (compromised), not above background (not detected), saturated (promised), or is a population outlier (compromised). Control probes were removed and only the “detected” probes that were present in at least one sample in all hES/hiPS cell samples were used for the further analysis. The number of probes used in the analysis was 36,757 (Fig. 1A) and 36,083 (Fig. 1A).

MicroRNA Microarray Analysis. The miRNA expression profiling was carried out using the Agilent Technologies human miRNA microarray release 12.0 according to the manufacturer’s protocol. The data were analyzed using the GeneSpring GX 11.5.1 software program (Agilent Technologies), and data processing was performed in the same way as for the mRNA expression analysis, except that 90th percentile normalization was chosen as the normalized algorithm. The number of probes used in the analysis was 476 (Fig. 1B).

Genomic DNA Extraction and Bisulfite Treatment. Genomic DNA extraction and purification from cultured cells was carried out using a Gentra Puregene kit (Qiagen). Extracted DNA was quantitated by using the Nanodrop instrument, and the quality was assessed by gel electrophoresis. A total of 500 ng of genomic DNA was treated with bisulfite using the EZ DNA MethylationGold Kit (Zymo Research Corp.) according to the manufacturer’s protocol.

DNA Methylation Analysis with a Beads Array. Genome-wide DNA methylation profiling was performed using the Illumina Infinium Human Methylation27 BeadChip (Illumina). Bisulfite-converted DNA was used, and the remaining assay steps were performed using the reagents supplied by Illumina and their specified conditions. The readout from the array was a β-value, which was defined as the ratio between the fluorescent signal from the methylated allele to the sum of both methylated and unmethylated alleles, and thus correlated with the level of DNA methylation. A β-value of 0.5 corresponds to complete methylation and 0 is equal to no DNA methylation. To exclude potential sources of technical bias, we only used CpG sites with detection P values <0.05 in at least 56 out of 59 samples. Normalization was not performed. The number of probes used in the analysis was 27,445 (Fig. 1E).

Generation of Heat Maps. We used Microsoft Excel to visualize the values as heat maps. The color spectrum expands from green (lower value) to magenta (higher value) through black in the gene/miRNA expression analysis, and from white (hypomethylation) to magenta (hypermethylation) in the DNA methylation analysis.

Bioinformatic Analysis. A hierarchical clustering analysis was performed using the GeneSpring GX 11.5.1 software program.

DNA Methylation Analysis with Pyrosequencing and Clonal Sequencing. Pyrosequencing was carried out with primers designed using the Pyromark Assay Design software program, version 2.0 (Qiagen). The primer sequences are shown in Table S1. PCR was performed in a 25-μL reaction mixture containing 25 ng bisulfite-converted DNA, 1× Pyromark PCR Master Mix (Qiagen), 1× Coral Load Concentrate, and 0.2 μM forward and 5′ biotinylated reverse primers. The PCR conditions were 45 cycles of 95 °C for 30 s, 56 °C for 30 s, and 72 °C for 30 s. The PCR product was

Koyanagi-Aoi et al. www.pnas.org/cgi/content/short/1319061110
bound to streptavidin Sepharose beads (GE Healthcare) and was purified, washed, denatured, and washed again. Then, 16 pmol of the sequencing primer was annealed to the purified PCR product. Pyrosequencing reactions were performed using the PSQ HS 96 Pyrosequencing System. The degree of methylation is shown as the percentage of methylated cytosines divided by the sum of methylated and unmethylated cytosines (percentage of mC).

We performed the exon array of the abhydrolase 3, 201B7, H1, and 253G1), and somatic cells, cells were prepared without contaminating cell-free embryoid body formation (SFEBq) as described previously (12). The animal experiments were performed in accordance with the Guidelines for Animal Experiments of Japan (2017). The animal experiments were performed in accordance with the Guidelines for Animal Experiments of Japan (2017).

To prepare samples for injection, we scraped dissociated into single cells and transferred at 9,000 cells per well to 96-well cell-adhesion plates (Lipidure-Coat Plate A-U96; NOF Corp.). The cells were cultured for 14 d in DFK5 medium consisting of DMEM/F-12 (Life Technologies), 5% (vol/vol) Knockout Serum Replacement (Life Technologies), 1% (vol/vol) MEM-nonsential amino acids (Life Technologies), 2 mM L-glutamine (Life Technologies), 0.1 mM 2-mercaptoethanol (Life Technologies), and 0.5% penicillin/streptomycin. We used DFK5 medium supplemented with 10 μM Y-27632, 2 μM dorsomorphin (Sigma), and 10 μM SB431542 (Sigma) for the first 4 d.

Adhering neural differentiation of dual SMAD inhibition was described as previously (11). Briefly, the cells were plated on matrigel-coated plate and after reaching 90% confluency they were cultured with 100 nM LDN193189 (Stemgent), 10 μM SB431542 in 15% (vol/vol) knockout serum replacement, 2 mM L-glutamine, and 10 μM β-mercaptoethanol–containing DMEM for 12 d.

For dopaminergic differentiation, we first transfected ESCs or iPSCs onto 96-well low-cell-adhesion plates with Y-27632, dorsomorphin, and SB431542 in the same way as indicated for SFEBq. We supplemented the cultures with 100 ng/mL FGF8 (Peprotech) and 20 ng/mL WNT1 (Peprotech) from days 5 to 12, and with 200 ng/mL Sonic Hedgehog (SHH) (R&D) from days 8 to 12. Twelve days after induction, aggregates were transferred onto six-well plates coated with laminin (Becton–Dickinson) and poly-L-ornithine (Sigma) and were cultured with neurobasal medium (Life Technologies) containing 2% (vol/vol) B27 supplement (Life Technologies), 2 mM L-glutamine, and 0.5% penicillin/streptomycin. We added 200 ng/mL SHH to the medium from days 12–15 and 1 ng/mL FGF20 and 12.5 ng/mL BFGF from days 15–22. On day 22, the cells were dissected into clumps and plated on new six-well plates coated with laminin and poly-L-ornithine and were then cultured with neurobasal medium supplemented with 2 ng/mL GDNF (R&D), 20 ng/mL BDNF (R&D), 400 mM dbcAMP (Sigma), and 200 mM ascorbic acid (Sigma) until day 29.

Flow Cytometric Analysis. Neural aggregates were dissociated and treated with the flow cytometric analysis by a FACS Aria II instrument (Becton–Dickinson). To analyze the proportion of OCT3/4+ cells, the cells were fixed with 3.7% (wt/vol) formaldehyde, permeabized with 0.2% TritonX-100, and stained with the appropriate antibody. To count the number of PSA-NCAM+ cells or TRA-1-60+ cells, cells were prepared without fixation. To eliminate the number of dead cells from the total cell population, we stained the cells with propidium iodide after labeling them with the anti–TRA-1-60 or anti-PSA-NCAM antibody, or with red fluorescent reactive dye from the LIVE/DEAD Fixable Dead Cell Stain Kits (Invitrogen) before fixing the cell suspension.

Transplantation of ES/iPS Cell-Derived Dopaminergic Neuron Cultures into the Brains of Nonobese Diabetic/Severe Combined Immune-Deficient Mice. To prepare samples for injection, we scraped and mechanically dissected cells by gently pipetting them up and down a few times, suspended them in culture medium (1 × 10^6 cells/μL), and injected 2 μL of the cell suspension into the right striatum (2 mm lateral, 1 mm rostral to the bregma; depth, 3 mm from the dura) of nonobese diabetic/severe combined immune-deficient mice (6 wk old, female) using a glass micropipette, as described previously (12). The animal experiments were performed in accordance with the Guidelines for Animal Experiments of Kyoto University, the Institutional Animal Care and Use Committee of Kobe Institute in RIKEN, and the Guide for the Care and Use of Laboratory Animals of the Institute of Laboratory Animal Resources (Washington, DC).
**Magnetic Resonance Imaging.** Graft imaging was performed with an MRmini SA instrument (DS Pharma Biomedical) by using a cylindrical slotted holder with a 20-mm radio frequency coil constructed for mice. T2-weighted images (repetition time, 2,000 ms; echo time, 69 ms) were recorded. The brains were imaged coronally in a single section through the graft center by using an image matrix of 256 × 128, a filed of view of 2 × 4 cm², and two excitations. Parametric images were generated by using the Sampler XPNI software program (DS Pharma Biomedical). Graft section surfaces were measured by using the INTAGE Realia Professional imaging software program (Cybernet).

**Immunostaining.** Anti-NCAM (ERIC1) antibody (Santa Cruz) was used as a primary antibody and anti-Mouse Ig biotin (Dako) was used as a secondary antibody for immunocytochemistry.

**Statistical Analysis. Gene/miRNA expression.** We conducted the t test (variances assumed equal) for the normalized, filtered data and controlled the false discovery rate (FDR) at 0.05 using the Benjamini–Hochberg method to identify probes that differed significantly between hESCs and hiPSCs, or for neural differentiation good and defective clones.

**DNA methylation determined using the beads array.** We conducted a Mann–Whitney U test on the filtered data controlling the FDR at 0.05 using the Benjamini–Hochberg method to identify probes that differed significantly between hESCs and hiPSCs.

**DNA methylation as determined by pyrosequencing and clonal sequencing.** A Mann–Whitney U test was used to compare the quantitative methylation values between hESCs and hiPSCs or defective and good groups. Calculations were carried out with the Statview software program.

**Graft size after transplantation.** A t test was used to compare the graft sizes derived from defective and good clones (Fig. 4C). In the case of comparisons between graft sizes between unsorted and depleted cell cultures (Fig. S4C), we performed a paired t test.

Fig. S1. Gene expression patterns among Kyoto hESCs, other hESCs, and hiPSCs and DNA methylation levels of previously reported ES-iPS differentially methylated regions (DMRs) in our cell lines. (A) Heat maps showing the expression levels of 15 probes that were differentially expressed between hESCs and hiPSCs (FDR < 0.05 and absolute fold change (FC) > 3) and hsa-miR-142-3p and hsa-miR-886-3p in various cell lines. A hierarchical clustering analysis for 15 probes was performed using the Euclidean distance and average linkage algorithm. (B) A hierarchical clustering analysis of the global gene expression patterns in various cell lines was performed using the Euclidean distance and average linkage algorithm. (C) Previously reported hES-iPS DMRs, A2BP1, IGF1R, ZNF184, POU3F4, and PTPRT, were examined by pyrosequencing in 10 hESCs and 49 hiPSCs in our laboratory. Each CpG dinucleotide position was assayed in triplicate, and average values were plotted. A Mann-Whitney U test was used to compare the quantitative methylation values between hESCs and hiPSCs. (n.s., not significant; *P < 0.05, **P < 0.01).
Fig. S2. Most of the cells can differentiate into neural cells although some clones retain undifferentiated cells after neural differentiation. (A) The expression levels of PAX6 (Left) and MAP2 (Right) in neurospheres of differentiated hESC line H9 and hiPSC lines (253G1, TIG108-4f3, and TKCBV5-6) were determined by quantitative RT-PCR. The expression levels in hESC line H9 before differentiation were set to 1, and relative expression levels were presented in log scale. (B) A comparison of the proportion of TRA-1-60–positive cells after neural induction between the SFEBq method (white) and the adhesion culture method (black).
**Fig. S3.** Activation of LTR7 is not confined to HHLA1, ABHD12B, and C4orf51; DNA hypomethylation exists in some, but not all, LTR7s in defective hiPSC clones. (A) The extraction of 763 probes corresponding to 435 genes as LTR7-related probes from the Agilent Technologies human G3 microarray (design ID 028004). (B) A comparison of the expression levels of 763 LTR7-related probes between good and defective clones. Magenta-colored genes (ABHD12B, HHLA1, and C4orf51) and yellow-colored genes (ARRB1, FAAH2, and TBC1D23) are differentially expressed between the good and defective clones (FDR < 0.05 and FC > 2, respectively), and green-colored genes (DNMT3B, ABCA1, and APP) did not show any differences. (C) The DNA methylation status of LTR7 and its neighboring regions of ARRB1, FAAH2, and TBC1D23 were examined by pyrosequencing and those of DNMT3B, ABCA1, and APP were examined by clonal sequencing. A Mann–Whitney *U* test was used to compare the quantitative methylation values between defective and good clones (n.s., not significant; *P* < 0.05).
Fig. S4. The histology of grafts derived from hESCs and hiPSCs. (A) Transplanted mouse brains were fixed with formaldehyde, embedded in paraffin, sectioned, and stained with H&E. The ratios indicate neural cells and nonneural cells as determined by a microscopic observation. (Scale bar, 500 μm.) (B) H&E-stained sections (Left) and human NCAM-stained sections (Center and Right) of mouse brains transplanted with 29-d differentiated hESC/iPSCs. (C) The maximum surface size of graft sections 45 or 60 d after transplantation. Transplanted cells were prepared with or without depletion of TRA-1-60+ cells 22 d after differentiation (*t test, paired P < 0.05).
Fig. S5. High expression levels of the hsa-mir-371–373 cluster, KLF4, and transgenes are not absolute markers for type-1 defectiveness. The expression levels of hsa-mir-371–373 (A) and KLF4 (B) were examined by a microarray analysis in seven defective clones, five intermediate clones, 38 good clones, two hECC lines (NTera2 and 2102Ep), six somatic cell lines (HDF, DP, CB1, CB2, PBMN1, and PBMN2), and three cancer cell lines (HepG2, MCF7, and Jurkat). The total and retroviral transgene expression levels of KLF4 (C) and OCT3/4 (D) were measured by qPCR in 18 hiPS clones established by a retroviral method, the defective clone 451F3, which was generated using an episomal plasmid vector, four hESCs, two hECCs, and two somatic cell lines (HDF and DP).
Previously reported iPS-specific aberrantly methylated genes’ expressions. A heat map for 10 hESCs and 49 hiPSCs examined in our laboratory based on the gene expression levels of reported aberrantly methylated genes that can distinguish hiPSCs and hESCs. Of nine previously reported genes, probes for C22orf34 were not detected in all of the samples in our microarray platform, so we only evaluated the other eight genes (eight probes, detected in at least two clones of our samples).

Other Supporting Information Files

Table S1 (DOCX)
Table S2 (DOCX)
Dataset S1 (XLSX)