An improved pipeline for reprogramming human induced pluripotent stem cells with TET1

Makoto Motono[1,2,4], Keiko Hiraki-Kamon[1,4], Masayoshi Kamon[1], Hidenori Kiyosawa[3], Yoichi Kondo[2], and Hidemasa Kato[1,*]

[1]Department of Developmental Biology and Functional Genomics, Ehime University Graduate School of Medicine, Toon, Ehime, 791-0295, Japan. [2]Department of Anatomy and Cell Biology, Osaka Medical and Pharmaceutical University, 2-7, Daigakumachi, Takatsuki, Osaka, 569-8686, Japan. [3]Department of Life Science, Chiba Institute of Technology, 2-17-1, Tsudanuma, Narashino, Chiba, 275-0016, Japan. [4]These authors contributed equally to this work: Makoto Motono, Keiko Hiraki-Kamon.
*Correspondence: e-mail: hidemasa.kato.zz@ehime-u.ac.jp

ABSTRACT

Despite offering insights into regenerative medicine, induced pluripotent stem cells (iPSCs) exhibit inconsistent differentiation potential, which may negatively impact their downstream applications. Here, we potentiated the reprogramming process of iPSCs by adding ten-eleven translocation 1 (TET1), a DNA demethylase, to a conventional reprogramming cocktail to produce TET1-iPSCs (T-iPSCs). We selected differentiation-elite iPSC clones through testing using a default differentiation procedure that challenged the evolutionarily conserved differentiation propensity toward neuroectoderm formation. By comparing the differentiation efficiencies of 46 in-house-generated iPSC clones, we identified extraembryonic markers that may indicate differentiation defects. Elite T-iPSCs showed an epithelialized morphology and differentiated into midbrain dopaminergic neurons with unprecedented fidelity and efficiency. We propose the T-iPSC production pipeline described here as a *de facto* standard for obtaining human pluripotent stem cells with enhanced differentiation potential.

INTRODUCTION

Human pluripotent stem cells (hPSCs) exhibit heterogeneous differentiation capabilities, which are usually more pronounced in human induced pluripotent stem cells (hiPSCs)[1,2] than in human embryonic stem cells (hESCs)[3]. Previous data have shown that the expression levels of miR-371/2/3, stage-specific embryonic antigen (SSEA) 3/4, and TRA1-60 in hPSCs are correlated negatively with their differentiation capabilities, although SSEA3/4 and TRA1-60 are considered pluripotent markers for hPSCs[4]. Superfluous LTR7-retrovirus-driven transcription has been reported as a risk factor for differentiation defectiveness[1,2]. Furthermore, hiPSCs have hotspots of aberrant epigenomic reprogramming that are maintained after differentiation[5]. The authors of these studies concluded that the expression of *Krüppel-like factor 4* (*KLF4*) appeared to be partly responsible for differentiation defects or epigenetic abnormalities[2,4,5].

The mesenchymal-epithelial transition (MET) is a key event during the early phase of iPSC reprogramming from fibroblasts[6]. Ten-eleven translocation (TET) dioxygenases are crucial to MET[7]. Although Tet1 expression begins in the early stages of mouse reprogramming where E-cadherin-positive epithelialized cells appear within the first week[6,8,9], human TET1 expression occurs later and to a moderate extent, which may

lead to retarded epitheliazation[10]. Moreover, early Tet1 induction during iPSC reprogramming is essential for unleashing the differentiation potential of mouse iPSCs[9]. Therefore, it is arguable that Yamanaka factors alone may be insufficient to demethylate human mesenchymal cells.

Parkinson's disease has attracted much research on ventral midbrain dopaminergic (mDA) neurons[11,12]. HiPSCs are posited to be more refractory to neural induction than hESCs[13]. Many laboratories employ dual SMAD inhibition[14], which improves the differentiation efficiency of a wide variety of hiPSC clones into the neuroectoderm. However, it is unclear whether or how early differentiation defects affect later differentiation stages, such as the development of mDA neurons.

This study developed a practical pipeline for producing hiPSCs with high differentiation potential. We hypothesized that the inconsistent differentiation potentials of various hiPSCs are caused by insufficient TET-mediated DNA demethylation activity during reprogramming. The conventional reprogramming cocktail was therefore reinforced with *TET1*. We noticed novel differentiation-resistant markers, representing bias toward extraembryonic differentiation, by correlating candidate gene expression before differentiation to the differentiation outputs. We linked the degree of

extraembryonic differentiation propensity to the differentiation and refractoriness of mDA progenitors.

RESULTS

TET1-aided human cell reprogramming. Given the benefits of TET1 in reprogramming and pluripotency[9,15], we generated episomal hiPSCs from male neonatal and female embryonic fibroblasts using the conventional expression vectors (including *POU class 5 homeobox 1* [*POU5F*1], *SRY-box transcription factor 2* [*SOX2*], *KLF4*, *MYCL proto-oncogene* [*MYCL*], *lin-28 homolog A* [*LIN28A*], and an RNA interference against p53[16]) with or without *TET1*-overexpression vector (Fig. 1a). We denoted TET1-aided iPSCs as T-iPSCs or conventional iPSCs as C-iPSCs. To reprogram the cultures, we used the protocol established by Thomson's group[17], which enabled the passage of cells after trypsin digestion and trituration onto a non-feeder surface and better discrimination of the morphological aspects of clones (see Methods; Fig. 1b). T-iPSCs and C-iPSCs were induced in parallel by the same experimenter(s), using an equal number of starting fibroblasts for each induction. Colonies were selected for further expansion after approximately 28 days. Most of the emerging T-iPSC colonies exhibited an epithelialized morphology (Fig. 1c, d). These epithelialized colonies were rare in the

C-iPSCs (Fig. 1c, d). Retrospective analyses revealed that only some of the selected epithelialized colonies had good differentiation capabilities; therefore, we discarded the mesenchymal-like colonies with fuzzy borders (Fig. 1c). Upon expansion in culture, we eliminated revertant clones with mesenchymal-like morphologies (Supplemental data 1) as a dimorphic morphology reappeared, even if the seemingly epithelialized part of the colony had been lifted and subcultured. Coincidentally, a significant proportion of clones lost their vigorous cell-proliferation capacities. Therefore, we considered clones that reached more than 90% confluency in a 100 mm dish to be "established." In this manner, we established 121 T-iPSC clones (out of 454 colonies) and 13 C-iPSC clones (out of 72 colonies), which indicated a clear over-representation of T-iPSC clones.

Residual reprogramming vectors can negatively influence cell differentiation[1] or potentiate carcinogenesis upon transplantation[18]. We detected residual episomal plasmids using a conventional polymerase chain reaction (PCR) analysis (data not shown)[19]. When we switched to quantitative PCR (qPCR), we found that conventional PCR only detected clones with > 0.16 plasmid copies/genome as positive. The mean copy number of "vector-free" clones (as determined by conventional PCR) was 0.0062 ± 0.0067 (mean ± standard deviation; Fig. 1e). Among these clones, we correlated the residual vector copy number with neural markers after default differentiation on day 8

(Fig. 2b, c) and found no significant trend (data not shown). We deduced that a residual episomal copy number of < 0.03, the maximal value among vector-free clones, should not affect the differentiation capability, and tentatively used 0.03 as the vector-free status threshold.

We obtained 22 male isogenic vector-free clones (3 C-iPSCs and 19 T-iPSCs out of 526 colonies) and 24 female clones (3 C-iPSCs and 21 T-iPSCs out of 566 colonies) from five independent inductions.

Transcriptomic features affecting differentiation efficiency. It is common to check the expression of "pluripotency-associated" genes in established iPSC clones[20]. We analyzed the expression of *POU5F1*, *SOX2*, *Nanog homeobox* (*NANOG*), *KLF4*, *MYC proto-oncogene* (*MYC*), and *MYCN proto-oncogene* (*MYCN*) in all clones in the undifferentiated state on day 0 (Fig. 2b). No variance over 2-fold from the average was observed for *POU5F1, SOX2, NANOG,* and *MYC* (except for the 2T1 clone; Fig. 2a, Supplemental data 2a). This analysis was performed in low-passage iPSC lines (passages 5 to 6), which are known to be prone to transcriptional noise[21]. The average expression of the pluripotency-associated factors demonstrated the overall uniformity of the obtained hiPSC clones. In contrast, the expression of *MYCN* and *KLF4* exhibited

considerable fluctuations (Fig. 2a, middle and bottom panels). The data collectively suggest that all the analyzed clones conformed to the qualitative criteria of pluripotency. Notably, we found a strong correlation between the transcription of *POU5F1* and *NANOG*, which was consistent with a previous report[22] (Fig. 3a, Supplemental data 2b). In contrast, the expression of *SOX2* showed only a minor correlation with that of *POU5F1* (Supplemental data 2c).

We independently investigated the differentiation capabilities of integration-free hiPSCs (Fig. 2b). We chose the default model of neural induction as the test method, which represents an evolutionarily conserved developmental trajectory among vertebrates. This model revealed that dissociated early embryonic cells defaulted to neural induction when incubated in a developmentally neutral culture medium[23]. We, in addition to other researchers, have previously shown that mouse, monkey, and human ESCs robustly default to neurogenesis under the conditions used herein[24-26].

This survey revealed striking differences in the differentiation capabilities of isogenic cell lines. The neural progenitor master regulator *paired box 6* (*PAX6*)[27] and radial glial cell marker *Hes family bHLH transcription factor 5* (*HES5*)[13] were expressed at levels comparable to those in our standard T-iPSC clone (1T47) in only one-fourth of the clones analyzed (Fig. 2c). Uniform-sized dense spheres or spheres

with sac-like structures showed high or low differentiation capabilities to the neural lineage, respectively (Supplemental data 3).

NANOG is a direct inhibitor of neural induction[28], and T-box transcription factor T (TBXT) is expressed in the mesendoderm and non-neuronal ectoderm[29] in humans; therefore, we examined these markers after differentiation. Some cell lines retained relatively high expression of *NANOG* or *TBXT*, which provided evidence of deviation from default neurogenesis (Fig. 2d).

Our results show that approximately three-fourths of the hiPSC clones, which were derived in a controlled manner, still had inherent differentiation defects. Given that our hiPSC clones were derived from isogenic fibroblasts, inter-clonal variations in their differentiation capabilities were barely attributable to genotype variations[30]. We excluded cell lines exhibiting high expression levels of *NANOG* and *TBXT* after differentiation and ranked clones based on the expression levels of five neural genes (*SOX21*, *SOX3*, *PAX6*, *HES5*, and *SOX1*) (Fig. 2c, d). We named the hiPSC clones with overall good differentiation propensities as “differentiation-elite” clones and investigated their karyotypes (Supplemental data 4).

We next investigated the correlations between gene expression levels before and after default differentiation by employing our hiPSC clone battery (Fig. 3a). As expected, the

classical pluripotency-associated genes (*POU5F1*, *SOX2*, and *NANOG*) showed no significant correlations with the neural differentiation markers (Fig. 3a). Moreover, the expression of undifferentiated endogenous *KLF4* and *MYCN* correlated negatively with the neural markers upon differentiation (Fig. 3a, b).

We further analyzed candidate KLF4-activated gene expression as KLF4 expression appeared to be partly responsible for the epigenetic abnormalities or differentiation defects in hPSCs[2,4,5]. Based on a previously reported secondary hiPSC derivation system[10], we explored the *NLR family pyrin domain containing 7* (*NLRP7*), *developmental pluripotency-associated 3* (*DPPA3*), and *DNA methyltransferase 3 like* (*DNMT3L*) genes as candidate KLF4 targets. The expression of *NLRP7* and *DNMT3L* before differentiation showed a negative correlation with the neural markers after differentiation (Fig. 3a, b). *HERV-H LTR-associating 1* (*HHLA1*), *arrestin beta 1* (*ARRB1*), *frizzled class receptor 10* (*FZD10*), and *gastrulation brain homeobox 2* (*GBX2*)[2], which have previously been associated with differentiation defects, did not correlate with the neural markers (Fig. 3a). We also examined *heart and neural crest derivatives expressed 1* (*HAND1*) as a candidate marker of partial reprogramming[10]. *HAND1* was most strongly negatively correlated with the neural markers (Fig. 3a, b).

We analyzed the expression patterns of several differentiation-resistant markers (including *HAND1*, *KLF4*, *MYCN*, *NLRP7*, and *DNMT3L*) within the human developmental context using published single-cell RNA-sequencing data obtained from cultured human blastocysts[31] or a gastrulating embryo[32]. Surprisingly, most of the differentiation-resistant markers (*MYCN*, *KLF4*, *NLRP7*, and *HAND1*) were expressed in the extraembryonic compartments (amniotic epithelium [AME] or intermediate state cells [intermediate]) but only at low levels in the late epiblasts of the cultured embryos (Fig. 4a, b). The negative correlation between the extraembryonic markers and differentiation potentials suggests an intriguing possibility that human pluripotency may exist in a bistable state between late epiblasts and extraembryonic lineages.

To directly assess this possibility, we analyzed the single-cell transcriptome of our representative elite hiPSCs. We integrated data obtained from three elite hiPSC clones (1T47, 6T9, and 6C6), which were displayed based on uniform manifold approximation and projection (UMAP) dimension reduction (Fig. 4c), and were grouped into nine clusters (Fig. 4d). The 6T9 and 6C6 clones (main clusters 1 and 6) versus 1T47 (the remaining clusters) were characterized into two distinct patterns (Fig. 4d). Surprisingly, cluster 1, which was mainly represented in the 6T9 and 6C6 cells, exhibited an exclusive expression of representative amnion markers (including *BMP4*, *ACTC1*, and

GATA3) (Fig. 4e)[31,33]. Furthermore, an epithelial-mesenchymal transition related marker, *TWIST1*, which is expressed in the human extraembryonic compartment (Fig. 4b), was expressed in the cluster (Fig. 4e). This indicated that, even after being epithelialized once during the reprogramming process, some of the cells failed to adequately suppress *BMP4* expression. They may have acquired a pro-mesenchymal trait[31,34] reminiscent of the mesenchymal revertants observed during the reprogramming process (Supplemental data 1). Conversely, the expression levels of *SOX2* and *TET1* were low in the cells, which was consistent with the downregulation of SOX2 in the nascent amnions of monkeys[29]. Cluster 6 comprised genes related to the mitotic metaphase (data not shown). Notably, the expression levels of *POU5F1* and *NANOG* were relatively uniform among the clusters, suggesting that all cells retained a certain degree of pluripotency. Collectively, our data suggest that a propensity for extraembryonic-like differentiation within the *SOX2*/*TET1*-low cluster can be observed among hiPSC clones (Fig. 4e, f).

T-iPSCs exert enhanced directed differentiation

Next, we explored the differentiation potential of the clones beyond the early neural fate by default differentiation (Fig. 2c). We then tested the degree of mDA specification for the selected clones.

During development, the regionalization of the central nervous system occurs through the successive specification of the anteroposterior (AP) and dorsoventral axes[35]. Mouse midbrain AP coordinates are set by isthmus-derived fibroblast growth factor 8 (FGF8)- and autocrine Wnt family member (WNT)-signaling cues[36]. The ventralization of midbrain precursors is mediated by sonic hedgehog (SHH) signaling, which emanates initially from the notochord. In this study, we employed a rationale-based protocol developed in Studer's lab[12] with minor modifications to derive mDA neurons (see Methods; Supplemental data 5a).

We began by investigating the differences in the differentiation capabilities between our standard differentiation-elite clone (1T47) and randomly selected non-elite clones (2C10, 4T17, 6C7, and 6T8). We carefully prepared to avoid counting dead cells by studying pyknotic nuclei combined with beta tubulin 3 (TuJ1) immunostaining (Supplemental data 6). On day 19 of the differentiation process, we analyzed the expression of forkhead box A2 (FOXA2), which is a floor plate-associated transcription factor and known precursor marker for mDA neurons[37]. Non-differentiation-elite clones

differentiated less efficiently toward FOXA2-positive cells than did 1T47 cells (Supplemental data 5b). FOXA2-ventral progenitors acquire LIM homeobox transcription factor 1 alpha (LMX1A) in the ventral midbrain[36]. The proportion of the LMX1A-inducibility of 1T47 cells (88.5% ± 1.47%, Supplemental data 5c) was as good as the best results reported previously using standard hESCs (H9)[12].

To explore the identities of the LMX1A-negative cells, we analyzed NK6 homeobox 1 (NKX6-1) and PAX6, which are expressed in the adjacent dorsal regions of the mouse midbrain[38,39]. The proportions of the NKX6-1- and PAX6-positive cells were 8.78% ± 2.98% and 3.80% ± 1.19%, respectively (Supplemental data 5c, d). NKX6-1/PAX6 double-positive cells were rare (< 0.07%) (Supplemental data 5d). In most cases, the sum of the percentages of LMX1A-, NKX6-1-, and PAX6-positive cells was < 100%. This result suggests that some cells were differentiated outside of the midbrain. Therefore, we sought to improve the differentiation efficiency by modifying the standard protocol to better mimic developmental processes.

Inherent WNT activity in hESCs antagonizes neuroectoderm differentiation and induces other lineages[40]. Because the timing of WNT signaling activation influences the AP axis coordinates of induced cells[41], we added a WNT inhibitor only during the first 2 days of differentiation, concurrently with dual SMAD inhibition (Fig. 5a). The transient

upregulation of *neuronal pentraxin 1* (*NPTX1*), an essential factor for human neural induction[42], was more pronounced in the presence of the WNT inhibitor (Fig. 5b). Moreover, differentiation was enhanced in epiblast-like/early neuroectodermal-like cells, as indicated by *POU class 3 homeobox 1* (*POU3F1*) expression and the increased expression of the forebrain marker *SIX homeobox 3* (*SIX3*)[43] (Fig. 5b). Therefore, initial WNT inhibition increased the efficiency of neural induction, which is important for midbrain differentiation.

Next, we investigated the optimal timing for activating ventralizing signals. Previously, ventralizing signals have been provided starting on day 1, preceding AP patterning conferred by WNT activation (Supplemental data 5a)[12]. Therefore, we added ventralizing factors at later time points (Fig. 5a). Although FOXA2 was equally expressed (Supplemental data 7a), the LMX1A induction was more thorough, accompanied by decreases in NKX6-1 and PAX6 cells. This resulted in approximately 100% differentiation (Fig. 5c, Supplemental table 1).

We further differentiated the other elite clones using the improved protocol. The differentiation efficiencies in the midbrain region for the clones 6T9 and 6T20 were exceedingly high, but more variable than for 1T47. The clones 6C4 and 6C6 showed subtotal differentiation efficiencies (Fig. 5d, e, Supplemental data 7b, Supplemental

table 1). These variations in differentiation efficiencies beyond the initial default neural induction of the differentiation-elite clones signify that the differentiation defectiveness of hPSCs was revealed more unambiguously at the later neuronal differentiation stage than at the early neuroectodermal stage. Our data showed that the early differentiation propensity for extraembryonic lineages (Fig. 4f) also affected the proper spatiotemporal control of the coordinates of differentiation (LMX1A) and neural-maturation (microtubule-associated protein 2, MAP2) targets (Fig. 5d, 6c).

The high differentiation efficiencies achieved in this study encouraged us to determine whether these progenitors could mature into mDA progenitors/neurons. We investigated the differentiation efficiencies of cells positive for nuclear receptor subfamily 4 group A member 2 (NR4A2; also known as NURR1), which is a definitive marker of mDA progenitors/neurons[36], using the elite clone, 1T47 (Fig. 6a). The percentage of NURR1-positive cells was 56.5% ± 5.22% on day 29 but progressed to an exceedingly high efficiency of 85.0% ± 5.23% on day 42. Furthermore, we noticed that the dendrites of the neurons derived from T-iPSCs were more elaborate than those from C-iPSCs (Fig. 6b, c). These results suggest that T-iPSC clones can form more mature and functional neurons than C-iPSCs.

DISCUSSION

We exhibited our novel pipeline for the production and quality-controlled selection of T-iPSCs with enhanced differentiation capabilities. Based on two findings, we propose that the lack of TET1 activity is a significant cause of the insufficient differentiation potentials of hiPSCs. With the conventional reprogramming regime, we observed fewer epithelialized cells than T-iPSCs (Fig. 1d), as acknowledged by the role of TET1 in MET during mouse reprogramming[6]. Our single-cell transcriptome analyses further revealed a subpopulation of cells with extraembryonic and pro-mesenchymal propensities accompanied by relatively low *TET1* expression (Fig. 4e, f). Strikingly, the 1T47 T-iPSC clone with minimum extraembryonic-like cells not only expressed *TET1* uniformly and exhibited epithelialized colony growth, but also exhibited near-complete differentiation toward mDA neurons (Fig. 4e, f, 5). We attributed the heterogeneous differentiation capabilities among hiPSCs to the metastabilities between the epiblast cells (as somatic precursors) and the extraembryonic-like cells (non-somatic cells; Fig. 4f). The idea that somatic versus extraembryonic biases may be the source of differentiation inconsistencies is in sharp contrast to the prevailing idea that differentiation propensities are exhibited among the three germ lineages of somatic differentiation[3,44], a point that warrants further scrutiny. Retrospectively, default

differentiation may reveal successive differentiation potentials; first, to the late anterior epiblast versus the late posterior epiblast/extraembryonic cells[45], and then to the neuroectoderm in humans. This study also vindicates that bonafide pluripotency-associated genes would be difficult to discern through bulk expression screens as the uniformity of such marker gene expression is cardinal for exerting full pluripotency, as shown here for *TET1* (Fig. 4e).

A confounding issue in reprogramming is the lack of a direct link between reprogramming factors and the expressivity of their endogenous counterparts. With regard to the episomal delivery of reprogramming factors, such factors are only forcefully expressed at earlier stages of reprogramming. The collective effects of their ectopic expression lead to the induction of the same factors, but from the endogenous gene loci. Therefore, the expressivity of *TET1* observed among various hiPSC lines, for example, is not a direct output of the expression of the reprogramming vector. Consequently, we could not decipher how TET1 conferred differentiation advantages to T-iPSCs. As complete reprogramming in the 1T47 clone is still a rare event, a direct assessment of the demethylation events responsible for attaining full pluripotency in the established iPSC clones before differentiation would be highly beneficial. Currently, we are investigating the genome-wide DNA methylation statuses of the C/T-iPSC clones.

We achieved the quasi-total differentiation of ventral midbrain cells using hiPSCs. The differentiation protocol was finely tuned by identifying the non-targeted cells and adjusting developmental coordinates accordingly. This procedure may merit wide applicability in revising differentiation protocols once a battery of standardized hiPSC clones becomes available.

In summary, we developed a new system of hiPSC induction, as determined by cell-autonomous differentiation testing, and showed the quasi-total induction of mDA neurons from T-iPSCs. Differentiation-elite T-iPSCs provided differentiation results that were comparable to, if not better, than those of hESCs. No C-iPSCs could differentiate as efficiently as differentiation-elite T-iPSCs, especially during long-term differentiation. We therefore propose the demonstrated T-iPSC-production pipeline as a *de facto* standard for obtaining hPSCs with improved differentiation potentials.

METHODS

hiPSC induction and culture. Human fibroblasts (from male neonatal foreskin, 106-05N; female fetal fibroblasts, 106-05F) were purchased from Cell Applications Inc. and cultured in Dulbecco's modified Eagle's medium (DMEM; Nacalai Tesque) supplemented with 10% fetal bovine serum (FBS; Hyclone), 100 μg/mL L-ascorbic acid 2-phosphate trisodium salt (FUJIFILM Wako), and 1× penicillin/streptomycin (P/S; Gibco).

HiPSCs were induced by applying a previously reported method using episomal vectors[19]. Episomal vectors (pCXLE-hOCT3/4-shp53-F: 27077, pCXLE-hSK: 27078, and pCXLE-hUL: 27080) were obtained from Addgene. To establish the T-iPSCs, a TET1 episomal vector (comprising complementary DNA encoding the human *TET1* open-reading frame, cloned into the pCXLE vector) was used to top the three episomal vectors upon reprogramming. The cells were transfected using the Human Dermal Fibroblast Nucleofection® Kit (Amaxa Biosystems [Lonza]) and Nucleofector™ 2B apparatus (Amaxa Biosystems [Lonza]) according to the manufacturer's instructions. Approximately 1×10^6 low-passage fibroblasts and 1 μg of each vector were used per transfection, and the cells were plated immediately in a human fibroblast medium after electroporation. On the following day, the culture medium was replaced with Essential

8™ Medium (E8M; Gibco) supplemented with 100 nM hydrocortisone (Sigma-Aldrich) and 1× P/S. The medium was changed every other day. E8M, which contains TGF-β, was chosen due to its potentiating effect of activating SMAD2/3 during reprogramming[46]. The medium was changed every other day. From day 10 of iPSC induction, E8M without hydrocortisone was used. The iPSC colonies with sharp contours were picked up, triturated into E8M supplemented with 10 μM Y-27632 (FUJIFILM Wako), and plated onto STO-feeder cells treated with mitomycin C (FUJIFILM Wako). The culture medium was replaced the next day with DMEM/F12 (Kohjin Bio) supplemented with 20% KnockOut™ Serum Replacement (KSR; Gibco), 2 mM GlutaMAX™-I (Gibco), 1× MEM non-essential amino acids (NEAA, Gibco), 0.1 mM β-mercaptoethanol (2-ME, FUJIFILM Wako), 5 ng/mL basic fibroblast growth factor (bFGF, Peprotech), 5 mM NaOH (FUJIFILM Wako), and 1× P/S (KF medium).

After the colonies had been selected, the medium was changed daily. When the iPSC colonies grew, they were dissociated with 0.01% trypsin/0.1 mM EDTA buffer (trypsin/EDTA; Nacalai Tesque) and plated onto 5 ng/mL vitronectin (Invitrogen)-coated 6-well plates in KFA medium supplemented with 10 μM Y-27632. The KFA medium was supplemented with 10 ng/mL activin A (R&D Systems), and the bFGF concentration was increased to 15 ng/mL. The culture medium was replaced the

following day with the KFA medium without Y-27632. The clones were further subcultured in a 100 mm dish (Becton Dickinson) until a confluent state had been reached. Half of the cells were cryopreserved in cell stock solution composed of 10% dimethylsulfoxide (Nacalai Tesque) and 90% FBS. Genomic DNA was obtained from the remaining cells to analyze the genomic integration of the plasmid vectors. HiPSCs were dissociated in trypsin/EDTA for feeder-free culture and replated in KFA medium containing 10 μM Y-27632. Twenty-four hours after each passage, the medium was replaced with KFA medium lacking Y-27632. The hiPSCs were passaged at 80-90% confluence every 3 or 4 days.

On average, it took 28 days for the colonies to form and a further 17 days for the iPSC clones to be expanded sufficiently to prepare frozen stocks.

PCR and qPCR analyses of integrated episomal vectors. Genomic DNA (gDNA) was extracted using the DNeasy Blood & Tissue Kit (Qiagen). Conventional (semi-quantitative) PCR was performed using Ex Taq DNA polymerase (TaKaRa Bio). qPCR was performed using Thunderbird SYBR qPCR Mix (Toyobo). The cycle threshold (Ct) values for the episomal vector were normalized against gDNA (measured using the CFTR primer set). The following primers were used for gDNA detection: CFTRexon24-

F: 5′-GGCAGTACGATTCCATCCAG-3′ and CFTRexon24-R: 5′-GAAAGAGCTTCACCCTGTCG-3′. The pEP4-SF1 (5′-TTCCACGAGGGTAGTGAACC-3′) and pEP4-SR1 (5′-TCGGGGGTGTTAGAGACAAC-3′) primers were used to detect the episomal vector backbones for semi-quantitative PCR[20]. For qPCR, the pEP4-SR1 primer was replaced with an in-house designed pEP4-SR3 (5′-CAGGCGAAGATTCAGGAGAG-3′). Genomic DNA (50 or 20 ng) was used as a template for conventional PCR or qPCR, respectively. gDNA extracted from male human dermal fibroblasts was used as a negative control.

Evaluating the differentiation capabilities of hiPSCs by default differentiation.

Pilot experiments using hiPSCs and attempts to derive neural cells from monkey ESCs[25] revealed that KSR-containing culture media were incompatible with default differentiation (data not shown). Therefore, we gradually changed the KSR-media to chemically defined medium (CDM) before differentiation to facilitate the efficient derivation of neural cells[25]. KFA-cultured hiPSCs were dissociated with trypsin/EDTA, and 1×10^6 cells were replated onto a vitronectin-coated 100 mm dish in CDM supplemented with 10 ng/mL bFGF, 2 ng/mL activin A, 3% KSR, and 10 μM Y-27632

(CFA). CDM is composed of Iscove's modified Dulbecco's medium/F-12 (Kohjin Bio), 2 mM GlutaMAX^{TM}-I, 7 μg/ml Insulin (Roche), 15 μg/ml transferrin (Roche), 450 μM monothioglycerol (Sigma-Aldrich), 1× lipid concentrate (Gibco), 5 mg/ml bovine serum albumin fraction V (Sigma-Aldrich), 200 μM L-ascorbic acid 2-phosphate trisodium salt, 0.1 mM NEAA, 80 nM sodium selenite (Sigma-Aldrich), and 1× P/S. Half of the media were carefully refreshed with CFA without Y-27632 daily. On the third day of replication, the cells were dissociated with trypsin/EDTA and plated at 500 cells per well in a 96-well low-binding Sphera plate (NUNC, Thermo Fisher Scientific) in 150 μL of CDM supplemented with 10 μM Y-27632. The day the cells were seeded was designated as differentiation day 0. Two days after seeding (day 2), 100 μL of the medium was removed to avoid aspiration of the spheres, and 150 μL of fresh CDM without Y-27632 was added. On differentiation day 5, 100 μL of the medium was removed, and 100 μL of fresh CDM was added. On differentiation day 8, the induced aggregates were collected.

Differentiation of ventral midbrain neural progenitor cells. Three different media were used to induce ventral midbrain neural progenitors. The first medium tested was a differentiation medium (DM), which was composed of DMEM (Nacalai Tesque), 15%

KSR, 2 mM GlutaMAX™-I, 1 mM sodium pyruvate (Sigma-Aldrich), 10 μM 2-ME, and 0.1× P/S. The second medium tested was N2 medium (N2M), which was composed of DMEM/F12 (FUJIFILM Wako), 1× N-2 supplement (Gibco), 2 mM GlutaMAX™-I, and 0.1× P/S. The third medium tested was the neural induction medium (NIM), composed of 50% N2M and 50% NB27. NB27 comprised neurobasal medium (Gibco), 1× B27 (Gibco), 2 mM GlutaMAX™-I, and 0.1× P/S. Differentiation-elite or non-differentiation-elite cell lines were cultured in KFA medium and subcultured two or three times. The cells were dissociated with trypsin/EDTA, and 1.5×10^5 cells/cm^2 were seeded onto plates coated with 0.5 μg/cm^2 Laminin-511 (Nippi) in DM supplemented with 10 μM Y-27632, 100 nM LDN193189 (LDN; Sigma-Aldrich), 10 μM SB431542 (SB; Sigma-Aldrich), and 10 nM WNT-C59 (C59; Cellagen Technology). The day that the cells were seeded was designated as differentiation day 0. On the following day (day 1), the medium was changed to DM supplemented with LDN, SB, and C59. On differentiation days 5, 7, and 9, the medium was changed to 75% DM and 25% N2M, 50% DM and 50% N2M, or 25% DM and 75% N2M, respectively. On differentiation day 11, the medium was changed to N2M supplemented with 0.5× B27. LDN was added to the medium each day from day 0 to 11; SB was added from day 0 to 4; and C59 was added from day 0 to 2. Ventralization was initiated by adding 100 ng/mL SHH

C25II (R&D Systems) and 2 μM purmorphamine (FUJIFILM Wako) from day 3 to 8. CHIR99021 (3 μM; Sigma-Aldrich) was added to the medium from day 3 to 12. Beginning on differentiation day 13, the medium was changed to NIM supplemented with 10 ng/mL BDNF (FUJIFILM Wako), 10 ng/mL GDNF (FUJIFILM Wako), 1 ng/mL TGF-β3 (R&D Systems), 200 μM L-ascorbic acid 2-phosphate trisodium salt, 200 μM N6,2-O-dibutyryladenosine 3, 5-cyclic monophosphate (Sigma-Aldrich), and 5 μM DAPT (Tocris Bioscience). For immunostaining, the cells were dissociated with Accumax (Innovative Cell Technologies, Inc). Then, 4×10^5 cells/cm^2 were seeded onto culture slides (BD Biosciences) coated with 15 μg/mL poly-l-ornithine (Sigma-Aldrich), 2 μg/mL fibronectin (Sigma-Aldrich), and 0.5 μg/cm^2 Laminin-511 in NIM supplemented with the above growth factors and 10 μM Y-27632. The medium was changed every day from day 0 to 16, and half of the medium was changed every 2 or 3 days from day 17 onward.

Immunostaining the induced cells. On days 19, 29, and 42, individual induced cells were fixed in 4% paraformaldehyde (FUJIFILM Wako), washed, and preserved in phosphate-buffered saline (PBS) at 4 °C until use. They were blocked in 5% (v/v) normal donkey serum (NDS; Millipore)/0.3% (v/v) Triton X-100 (Nacalai Tesque)/PBS

for 1 h. Primary antibodies were diluted in 1% NDS/0.3% Triton X-100/PBS and incubated overnight at 4 °C. The samples were then washed in 0.05% (v/v) Tween-20 (Sigma-Aldrich)/PBS and incubated with secondary antibodies in 1% NDS/0.3% Triton X-100/PBS for 2 h, followed by washing with 0.05% Tween-20/PBS and incubation with DAPI (Invitrogen) in 1% NDS/0.3% Triton X-100/PBS for 10 min. Finally, the samples were washed with 0.05% Tween-20/PBS and mounted with anti-photobleaching mounting medium comprising 2.4 g Mowiol (Calbiochem), 6 mL glycerol (FUJIFILM Wako), 6 mL distilled water, 12 mL 200 mM Tris-HCl (Nacalai Tesque), and 0.66 g of 1.4-diazabicycle (2.2.2) octane (FUJIFILM Wako).

For immunostaining, we used primary antibodies against LMX1A (1:1000; Millipore), PAX6 (1:500; BD Biosciences), MAP2 (1:2000; Sigma-Aldrich), beta-tubulin 3 (TUJ1, 1:2000; Covance), NKX6-1 (1:1000; R&D), NR4A2 (NURR1) (1:1000; Santa Cruz Biotechnology), and FOXA2 (1:500; R&D Systems). Alexa Fluor™-conjugated antibodies (1:500, Invitrogen) were used as secondary antibodies. Given the fluorescence filter constraints, we could not simultaneously perform three-color immunostaining and nuclear staining. Therefore, we individually stained cells for LMX1A and NKX6-1, or NKX6-1 and PAX6, and counted approximately 3000 cells in three randomly selected fields to determine the ratios of cells positive for the respective

markers. In serial sections, immunostaining for LMX1A and PAX6 revealed very few double-positive cells for both markers (data not shown). The stained cells were imaged using a Biorevo BZ-9000 fluorescence microscope (Keyence).

Quantitative PCR analysis. Total RNA was purified from the samples using the RNeasy Plus Mini Kit (Qiagen). Over 50 ng of total RNA was reverse-transcribed using ReverTra Ace (Toyobo) and oligo-dT primers (Toyobo). The qPCR analysis was conducted using Thunderbird SYBR qPCR Mix (Toyobo) and a Thermal Cycler Dice Real Time System (Bio-Rad). The data were assessed using the ΔΔCt method and normalized to *glyceraldehyde-3-phosphate dehydrogenase* (*GAPDH*) expression. The sequences of the primers used for qPCR are shown in Supplemental table 2.

Evaluating MAP2-positive induced cells. The lengths of the MAP2-positive neurites were measured using ImageJ software. The relative dendrite region was calculated as the average number of dendrites extending from the cell body times the average dendrite length[2].

Isolation and RNA-sequencing of single cells. Single-cell RNA-sequencing (RNA-seq) was performed using the 10× Chromium platform using the Single Cell 3' reagent kits v3 Chemistry according to the manufacturer's protocol (10× Genomics). Libraries were sequenced on a DNBSEQ-G400 (MGI) with paired-end sequencing. Reads were aligned to the human reference genome (GRCh38/hg38) using the STARsolo package of STAR software (version 2.7.6a)[47]. The integration analysis was performed using the Seurat package in R v4.0.0[48]. The Seurat package was also used for normalization, variable gene identification, anchor identification, scaling, linear dimensional reduction, and clustering analysis (dims = 1:30). Nonlinear dimensional reduction analysis (UMAP) was performed using the RunUMAP function (dims = 1:30) in the Seurat package. Gene expression profiles were visualized on a UMAP plot using the FeaturePlot function or the RNA assay function in the Seurat package.

Reanalysis of published single-cell RNA-seq data. Single-cell RNA-seq data from cultured human early embryos (6 to 14 days post-fertilization) were downloaded from the Gene Expression Omnibus database (GSE136447)[31]. Reads were aligned to the human reference genome (GRCh38/hg38) using the STARsolo package of STAR software[47]. The Seurat package in R software was used for normalization, variable gene

identification, scaling, linear dimensional reduction, and clustering analysis (dims = 1:20 and resolution = 0.5). The 1483 most variable genes in 552 cells were selected. Nonlinear dimensional reduction analysis (t-SNE) was performed using the RunTSNE function (dims = 1:20) in the Seurat package. Gene expression profiles were visualized on a t-SNE plot using the FeaturePlot function in the Seurat package. We adhered the annotations of most cell lineages in the original article with a few modifications. Pre-EPI (Epiblast) and Post-EPI were renamed to early-EPI and late-EPI, respectively, based on the expression of the naïve markers, *Sushi domain containing 2* (*SUSD2*) and *KLF17*, which demarcate early-EPI. PSA-EPI was renamed to PGC as these cells express primordial germ cell markers, such as *PR/SET domain 14* (*PRDM14*).

Statistical analysis. Values were expressed as the mean ± standard deviation. All statistical analyses were performed using the R statistical software package. We used Pearson's test for correlation analyses when the expression values followed a normal distribution; otherwise, Spearman's test was used. The Shapiro–Wilk test was used to assess normal distributions. A two-sample *t*-test was used for comparisons between two groups. Probability values of < 5% were considered to reflect statistically significant differences.

hiPSC karyotyping. Nihon Gene Research Laboratories, Inc. karyotyped the T/C-iPSC lines by performing Giemsa testing.

Data availability

The accession number for the single-cell RNA-seq data generated in the present study is DRA013769 (DNA Data Bank of Japan [DDBJ]).

ACKNOWLEDGMENTS

We thank Y. Moriyama, Y. Tokuzawa, M. Bagheri, S. Funayama and H. Kawakami for earlier contributions and technical assistance. This study was supported by Zeon Corporation, SATAKE MultiMix Corporation, Yokogawa Electric Corporation, SyntheticGestalt, and AMED under Grant Number JP20bk0104090. The procedure for T-iPS cell production was filed under patent no. WO2014-069479.

COMPETING INTERESTS

All authors declare no competing interests.

AUTHOR CONTRIBUTIONS

H. Kato conceived the project. M. M., K. H-K., M. K., H. Kiyosawa, Y. K., and H. Kato wrote the manuscript. M. M., K. H-K., M. K., and H. Kato performed experiments and analyzed data. All authors have reviewed, edited, and approved the final version of the manuscript.

REFERENCES

1. Koyanagi-Aoi, M. et al. Differentiation-defective phenotypes revealed by large-scale analyses of human pluripotent stem cells. *Proc Natl Acad Sci U S A* **110**, 20569–20574 (2013).
2. Ohnuki, M. et al. Dynamic regulation of human endogenous retroviruses mediates factor-induced reprogramming and differentiation potential. *Proc Natl Acad Sci U S A* **111**, 12426–12431 (2014).
3. Osafune, K. et al. Marked differences in differentiation propensity among human embryonic stem cell lines. *Nat. Biotechnol.* **26**, 313–315 (2008).
4. Kim, H. et al. miR-371-3 expression predicts neural differentiation propensity in human pluripotent stem cells. *Cell Stem Cell* **8**, 695–706 (2011).
5. Lister, R. et al. Hotspots of aberrant epigenomic reprogramming in human induced pluripotent stem cells. *Nature* **471**, 68–73 (2011).
6. Li, R. et al. A mesenchymal-to-epithelial transition initiates and is required for the nuclear reprogramming of mouse fibroblasts. *Cell Stem Cell* **7**, 51–63 (2010).
7. Hu, X. et al. Tet and TDG mediate DNA demethylation essential for mesenchymal-to-epithelial transition in somatic cell reprogramming. *Cell Stem Cell* **14**, 512–522 (2014).
8. Samavarchi-Tehrani, P. et al. Functional genomics reveals a BMP-driven mesenchymal-to-epithelial transition in the initiation of somatic cell reprogramming. *Cell Stem Cell* **7**, 64–77 (2010).
9. Bartoccetti, M. et al. Regulatory dynamics of Tet1 and Oct4 resolve stages of global DNA demethylation and transcriptomic changes in reprogramming. *Cell Rep* **30**, 2150–2169 e2159 (2020).
10. Cacchiarelli, D. et al. Integrative analyses of human reprogramming reveal dynamic nature of induced pluripotency. *Cell* **162**, 412–424 (2015).
11. Barker, R. A., Parmar, M., Studer, L. & Takahashi, J. Human trials of stem cell-derived dopamine neurons for Parkinson's disease: dawn of a New Era. *Cell Stem Cell* **21**, 569–573 (2017).
12. Steinbeck, J. A. et al. Optogenetics enables functional analysis of human embryonic stem cell-derived grafts in a Parkinson's disease model. *Nat. Biotechnol.* **33**, 204–209 (2015).
13. Tchieu, J. et al. A modular platform for differentiation of human PSCs into all major ectodermal lineages. *Cell Stem Cell* **21**, 399–410.e397 (2017).
14. Chambers, S. M. et al. Highly efficient neural conversion of human ES and iPS

cells by dual inhibition of SMAD signaling. *Nat. Biotechnol.* **27**, 275–280 (2009).
15. Verma, N. et al. TET proteins safeguard bivalent promoters from de novo methylation in human embryonic stem cells. *Nat. Genet.* **50**, 83–95 (2018).
16. Okita, K. et al. A more efficient method to generate integration-free human iPS cells. *Nat. Methods* **8**, 409–412 (2011).
17. Chen, G. et al. Chemically defined conditions for human iPSC derivation and culture. *Nat. Methods* **8**, 424–429 (2011).
18. Ohnishi, K. et al. Premature termination of reprogramming in vivo leads to cancer development through altered epigenetic regulation. *Cell* **156**, 663–677 (2014).
19. Yu, J. et al. Human induced pluripotent stem cells free of vector and transgene sequences. *Science* **324**, 797–801 (2009).
20. International Stem Cell, I. et al. Characterization of human embryonic stem cell lines by the International Stem Cell Initiative. *Nat. Biotechnol.* **25**, 803–816 (2007).
21. Chin, M. H. et al. Induced pluripotent stem cells and embryonic stem cells are distinguished by gene expression signatures. *Cell Stem Cell* **5**, 111–123 (2009).
22. Fogarty, N. M. E. et al. Genome editing reveals a role for OCT4 in human embryogenesis. *Nature* **550**, 67–73 (2017).
23. Ozair, M. Z., Kintner, C. & Brivanlou, A.H. Neural induction and early patterning in vertebrates. *Wiley Interdiscip Rev Dev Biol* **2**, 479–498 (2013).
24. Bouhon, I. A., Kato, H., Chandran, S. & Allen, N. D. Neural differentiation of mouse embryonic stem cells in chemically defined medium. *Brain Res. Bull.* **68**, 62–75 (2005).
25. Ono, T. et al. A single-cell and feeder-free culture system for monkey embryonic stem cells. *PLoS One* **9**, e88346 (2014).
26. Patani, R. et al. Activin/Nodal inhibition alone accelerates highly efficient neural conversion from human embryonic stem cells and imposes a caudal positional identity. *PLoS One* **4**, e7327 (2009).
27. Zhang, X. et al. Pax6 is a human neuroectoderm cell fate determinant. *Cell Stem Cell* **7**, 90–100 (2010).
28. Loh, K. M. & Lim, B. A precarious balance: pluripotency factors as lineage specifiers. *Cell Stem Cell* **8**, 363–369 (2011).
29. Sasaki, K. et al. The germ cell fate of cynomolgus monkeys is specified in the nascent amnion. *Dev. Cell* **39**, 169–185 (2016).
30. Liang, G. & Zhang, Y. Genetic and epigenetic variations in iPSCs: potential causes and implications for application. *Cell Stem Cell* **13**, 149–159 (2013).

31. Xiang, L. et al. A developmental landscape of 3D-cultured human pre-gastrulation embryos. *Nature* **577**, 537–542 (2020).
32. Tyser, R. C. V. et al. Single-cell transcriptomic characterization of a gastrulating human embryo. *Nature* **600**, 285–289 (2021).
33. Chhabra, S. & Warmflash, A. BMP-treated human embryonic stem cells transcriptionally resemble amnion cells in the monkey embryo. *Biol Open* **10**, bio058617 (2021).
34. Richter, A. et al. BMP4 promotes EMT and mesodermal commitment in human embryonic stem cells via SLUG and MSX2. *Stem Cells* **32**, 636–648 (2014).
35. Tao, Y. & Zhang, S. C. Neural subtype specification from human pluripotent stem cells. *Cell Stem Cell* **19**, 573–586 (2016).
36. Arenas, E., Denham, M. & Villaescusa, J. C. How to make a midbrain dopaminergic neuron. *Development* **142**, 1918–1936 (2015).
37. Ferri, A. L. et al. Foxa1 and Foxa2 regulate multiple phases of midbrain dopaminergic neuron development in a dosage-dependent manner. *Development* **134**, 2761–2769 (2007).
38. Duan, D., Fu, Y., Paxinos, G. & Watson, C. Spatiotemporal expression patterns of Pax6 in the brain of embryonic, newborn, and adult mice. *Brain Struct Funct* **218**, 353–372 (2013).
39. Andersson, E. et al. Identification of intrinsic determinants of midbrain dopamine neurons. *Cell* **124**, 393–405 (2006).
40. Blauwkamp, T. A., Nigam, S., Ardehali, R., Weissman, I. L. & Nusse, R. Endogenous Wnt signalling in human embryonic stem cells generates an equilibrium of distinct lineage-specified progenitors. *Nat Commun* **3**, 1070 (2012).
41. Kim, T. W. et al. Biphasic activation of WNT signaling facilitates the derivation of midbrain dopamine neurons from hESCs for translational use. *Cell Stem Cell* **28**, 343–355 e345 (2021).
42. Boles, N. C. et al. NPTX1 regulates neural lineage specification from human pluripotent stem cells. *Cell Rep* **6**, 724–736 (2014).
43. Metzis, V. et al. Nervous system regionalization entails axial allocation before neural differentiation. *Cell* **175**, 1105–1118 e1117 (2018).
44. Shao, Y. et al. Self-organized amniogenesis by human pluripotent stem cells in a biomimetic implantation-like niche. *Nat Mater* **16**, 419–425 (2017).
45. Avila-Gonzalez, D. et al. Unraveling the spatiotemporal human pluripotency in embryonic development. *Front Cell Dev Biol* **9**, 676998 (2021).

46. Ruetz, T. et al. Constitutively active SMAD2/3 are broad-scope potentiators of transcription-factor-mediated cellular reprogramming. *Cell Stem Cell* **21**, 791–805.e9 (2017).
47. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
48. Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* **36**, 411–420 (2018).

FIGURE LEGENDS

Figure 1 | HiPSC-production pipelines. a, Reprogramming factor cocktail used for T-iPSCs and C-iPSCs. OCT3 is an alias of POU5F1. **b**, Schematic representation of the timeline and procedure used for hiPSC production. E8M, Essential 8^{TM} medium. HC, hydrocortisone. Y, Y-27632, a ROCK inhibitor. VTN, vitronectin. The components of the STO, KF, and KFA are described in the Methods section. **c**, Characteristic morphologies of the emerging hiPSCs. Note that the non-reprogrammed fibroblasts served as feeder cells for the hiPSCs. Scale bars, 50 μm. **d**, Number of epithelialized (sharp contour) vs. mesenchymal (fuzzy contour) colonies for both reprogramming methods. **e**, Copy numbers of integrated reprogramming episomal vectors. (**d**, n = 454 T-iPSCs, n = 72 C-iPSCs; **e**, n = 46)

Figure 2 | System for evaluating the differentiation of hiPSC clones. a, The relative expression levels of pluripotency-related genes are indicated (top graph: white, *NANOG*; grey, *POU5F1*; black, *SOX2*), *MYCN*, and *KLF4*. The x-axis shows the clone names. The red dotted lines indicate the average expression level, and the black lines indicate half or two-times the average expression level. **b**, The default differentiation

procedure used to evaluate the differentiation capabilities of T/C-iPSC clones. **c**, **d**, The relative expression levels of (c) neural markers, or (d) *NANOG* and *TBXT* on day 8 in male (c, top panel) and female (c, bottom panel) hiPSCs were normalized to those of clone 1T47.

Figure 3 | The gene expression profile predicted the differentiation refractoriness of hiPSCs. a, Correlation matrix for marker expression levels and the default differentiation behaviors of all vector-free hiPSC clones. The colored boxes indicate a significant correlation ($P < 0.05$) between two markers; white boxes indicate cases where no significant correlation was observed. **b**, Scatter plots showing the relative gene expression levels of the neural markers on differentiation day 8. The red data points indicate the differentiation-elite clones (**a**, **b**, total clone number = 46). Both axes indicate the *GAPDH*-normalized expression levels of each gene.

Figure 4 | Single-cell RNA-seq analysis of the differentiation-resistance factors in cultured human early embryo cells (reported by Xiang *et al.*[31]) and the representative elite hiPSCs. a, t-SNE plot of the single-cell RNA profiles of early human embryos. Different colors denote different cell types. **b**, The gene expression of

differentiation-resistance markers and the epithelial-mesenchymal transition marker, *TWIST1*, was visualized using feature plots. **c**, UMAP plot for the representative elite hiPSCs. Left panel: the merged UMAP plot for the four cell types. Right panel: individual clones contributions to the merged UMAP plot in the left panel. 1T47-1 and 1T47-2 are technical duplicates. **d**, UMAP plot of the clusters. Left panel: the merged UMAP plot for the four cell types. Right panel: individual clones contributions to the merged UMAP plot in the left panel. Different colors denote different clusters. **e**, The gene expression of AME markers in the integrated analysis was visualized using feature plots. Expression level data for *POU5F1* and *NANOG* were evaluated using RNA assay instead of FeaturePlot used for other genes. **f**, Proposed view of bistability of pluripotency in human.

Figure 5 | Improved mDA differentiation by coordinating the developmental rationale. a, Improved mDA-differentiation protocol. **b**, qPCR analyses for neuroectodermal and forebrain markers at an early stage of differentiation. Control (original protocol, blue line; as shown in Supplemental data 5a) and New (new protocol, red line). n = 3. The values are represented as the mean ± standard deviation. **c**, **d**, Differentiation efficiency of the ventral or lateral midbrain regions on day 19 for clone

1T47 (**c, d**) and other differentiation-elite clones (**d**). n = 3. The values are represented as the mean ± standard deviation. **e**, The sum of all ratios of cells expressing ventral midbrain regional markers on day 19. n = 3.

Figure 6 | Enhanced directed differentiation and maturation of differentiation-elite T-iPSCs to mDA neurons. a, 1T47-derived NURR1-positive mDA neurons on day 29 (arrows, NURR1-positive neurons; arrowheads, NURR1-negative neurons; the cells in sphere-like aggregates encircled in red were not counted). **b**, **c**, Morphological evaluation of MAP2-positive cells on day 29. The values are represented as the mean ± standard deviation. Scale bars, 50 μm. The clone names are indicated on the x-axis. One-way analysis of variance and Tukey's multiple-comparison test were performed to determine significance of differences (**b**). * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$. n = 3.

SUPPLEMENTAL DATA LEGENDS

Supplemental data 1 | Mesenchymal revertants during iPSC expansion. A subconfluent culture during the expansion of a picked colony in KFA medium. Note that the cells in the encircled area show a mesenchymal-like morphology. Scale bar, 50 μm.

Supplemental data 2 | Pluripotency-associated gene expression and correlations among T/C-iPSCs. a, Relative expression levels of *MYC*. The x-axis shows the clone names. The red dotted line shows the average expression levels, and the black line indicates half of the average expression level**. b**, Relative expression levels of *NANOG* versus *POU5F1*. **c**, Relative expression levels of *SOX2* versus *NANOG* and *POU5F1* on day 0. Results of the Pearson's correlation analysis are presented as correlation coefficients (R) and p values in each graph.

Supplemental data 3 | Morphology of spheroids on day 8 of default differentiation. All clones with high differentiation capabilities formed uniform-sized and dense spheres, such as 6T20 (left panel). Spheres with sac-like structures such as 6T4 (right) exhibited low levels of expression of neural markers.

Supplemental data 4 | Karyotyping of differentiated elite clones. The numbers in square brackets represent the number of observed karyotypes among the 20 nuclei counted. +mar, marker chromosome of unknown origin.

Supplemental data 5 | Differentiation capabilities of differentiation elite and non-elite clones. a, Schematic representation of the differentiation protocol used[12], with minor modifications. SB, SB431542; LDN, LDN193189; CHIR, CHIR99021; OFL, ornithine, fibronectin, and Laminin-511. **b**, Immunostaining of induced cells on day 19. 1T47 is a differentiation-elite clone, and the other clones are non-differentiation-elite clones. The percentages shown indicate the ratios of FOXA2-positive cells. **c**, Schematic representation of the developing mouse midbrain region with its marker (adapted from references 38 and 39). Red shading, LMX1A-positive region; yellow shading, LMX1A and NKX6-1-double-positive regions; green, NKX6-1-positive region; white shading, more lateral region. The percentages show the ratios of each cell type in the induced 1T47 clone on day 19. The values are represented as the mean ± standard deviation. **d**, Immunostaining of midbrain-region markers in induced 1T47 cells. Blue, DAPI-positive; green, NKX6-1- or PAX6-positive; red, LMX1A- or NKX6-1-positive. (**b**, n = 1; **c**, **d**, n = 3). Scale bars, 50 μm.

Supplemental data 6 | Preparatory histological examination to avoid counting dead cells. Pyknotic nuclei with a vigorous blue intensity (yellow arrows and dotted circles) were not counted as the cells did not express TuJ1, a pan-ectodermal marker, and appeared to be dead or dying. A few cells appeared to be stained with TuJ1 (white arrowheads). Based on the bundled appearance of the TuJ1-positive fibers, we believe that they represent both neuronal and radial glial projections.

Supplemental data 7 | Developmental rationale improves the differentiation efficiency in the ventral midbrain. a, **b**, Representative immunostaining data with DAPI (blue) and (a) FOXA2 (green) or (b) LMX1A (red) on day 19 in differentiation-elite clones after induction using the new differentiation protocol. The percentages indicate the ratio of each positive cell type. Scale bars, 50 μm.

Figure 1

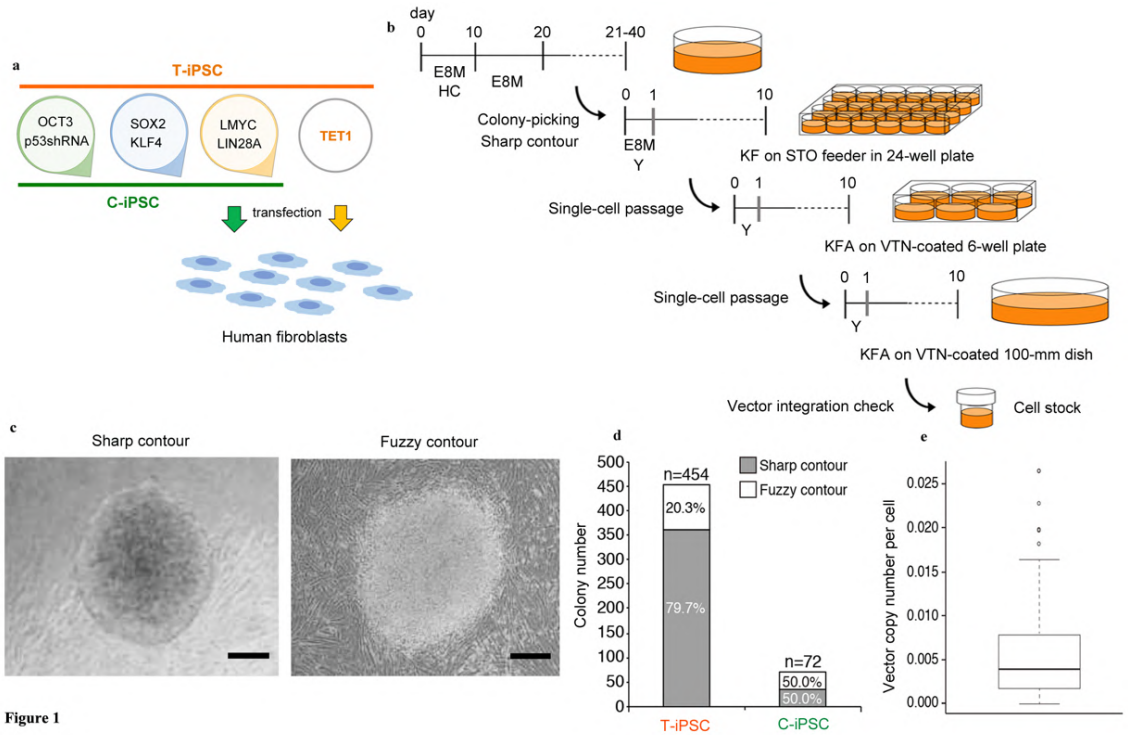


Figure 1

Figure 2

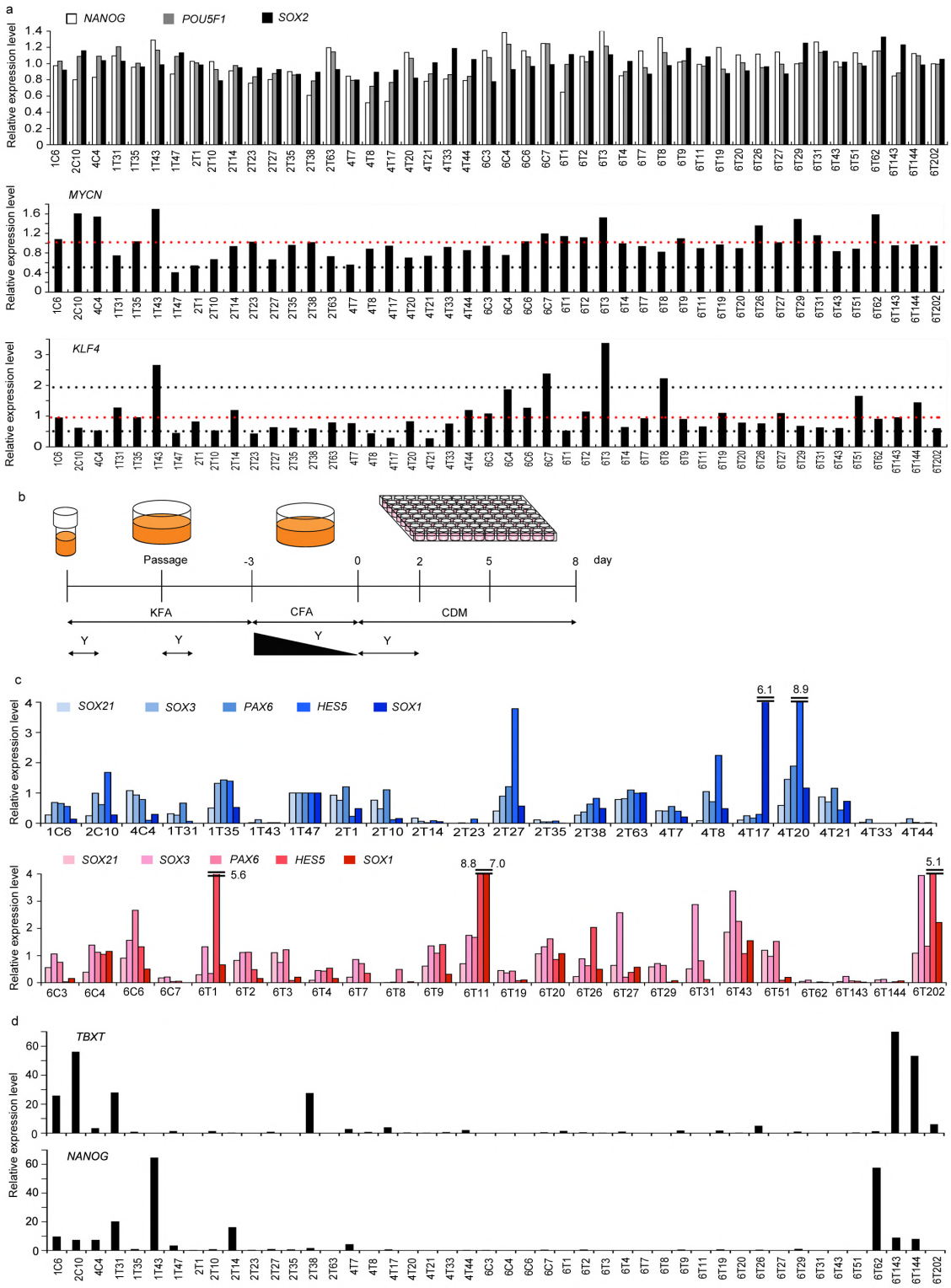


Figure 2

Figure 3

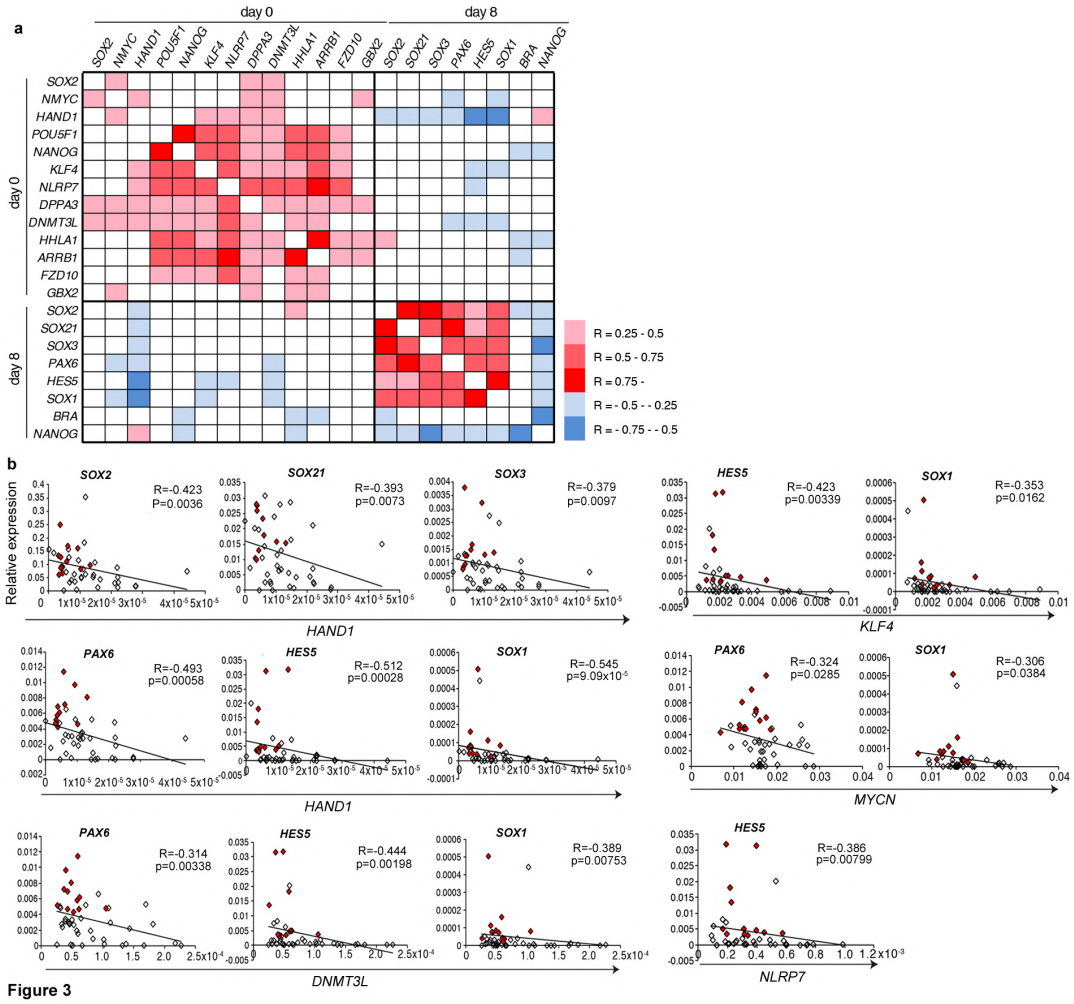


Figure 3

Figure 4

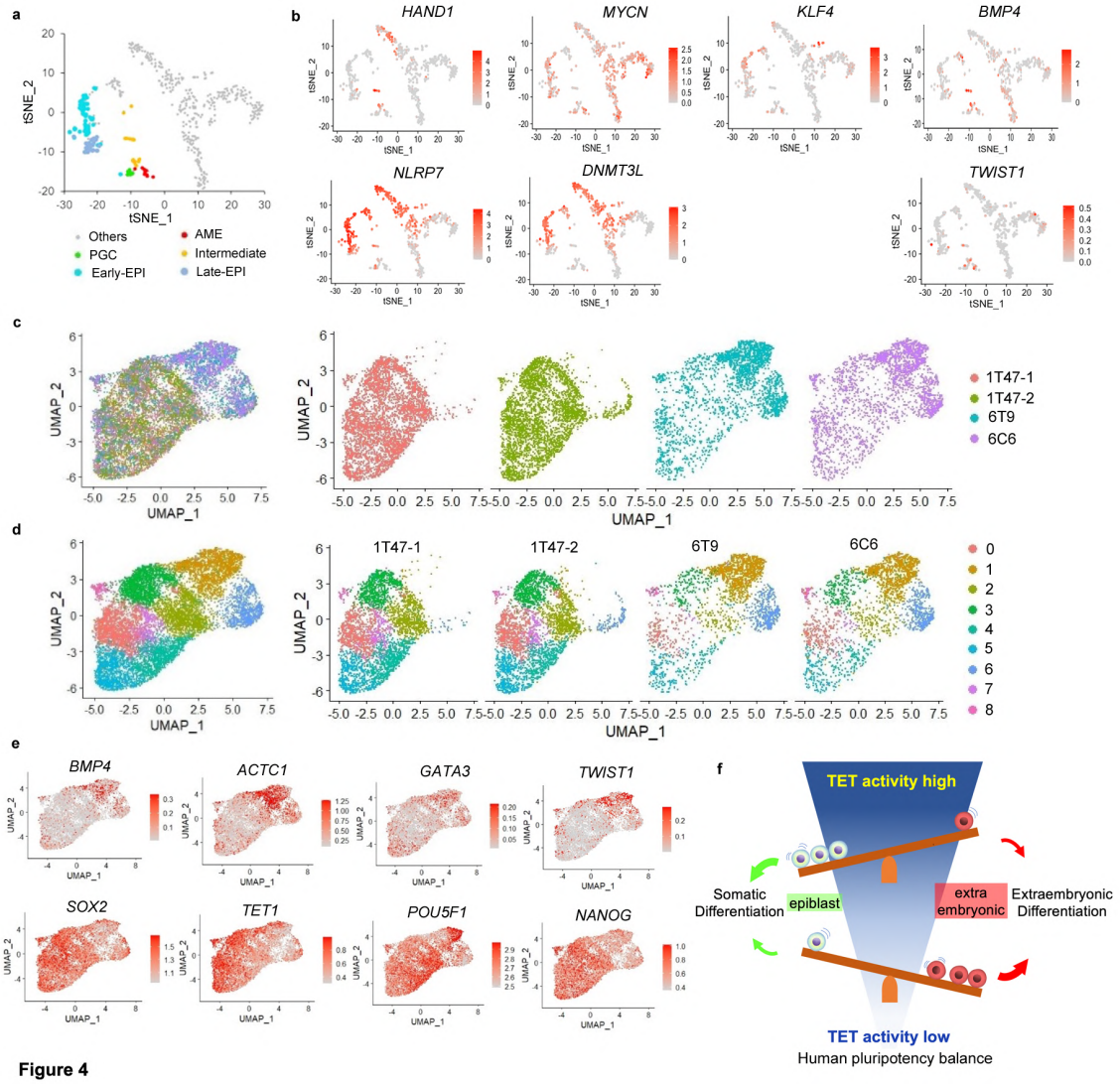


Figure 4

Figure 5

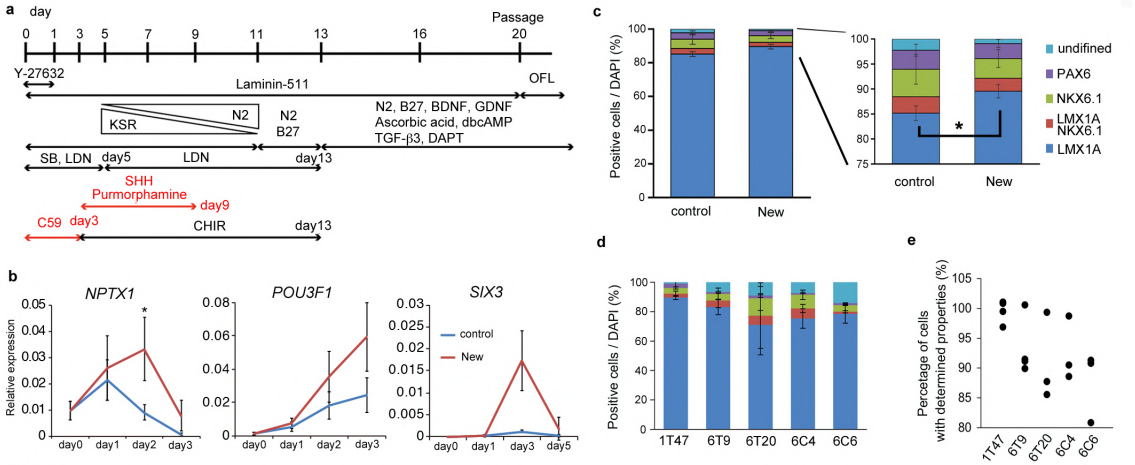


Figure 5

Figure 6

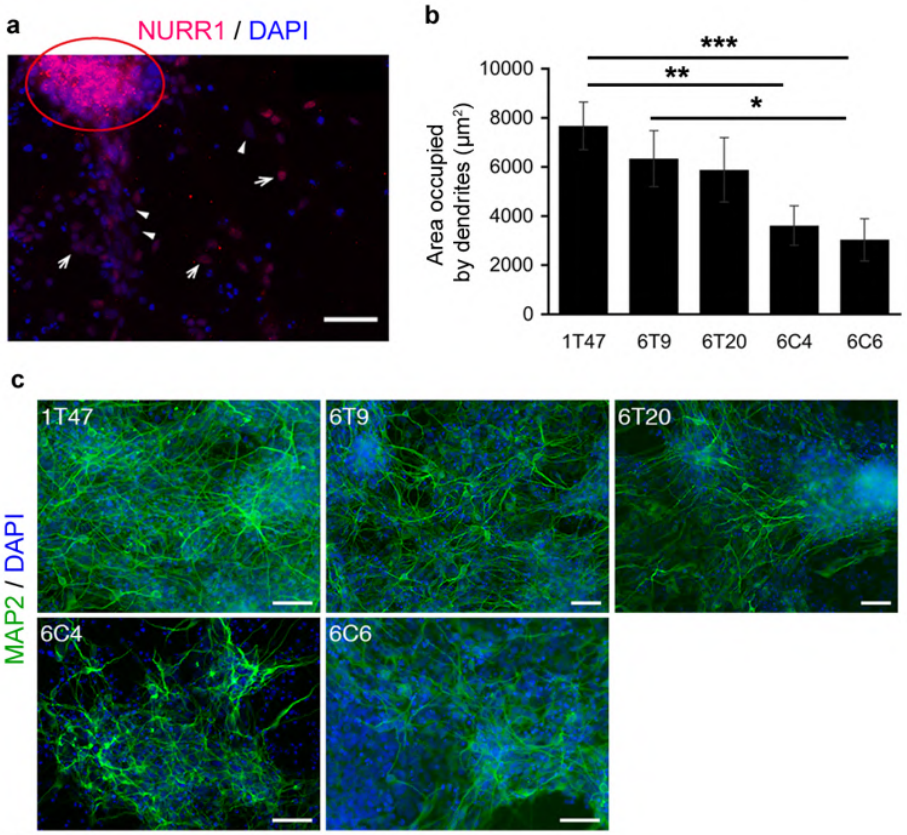
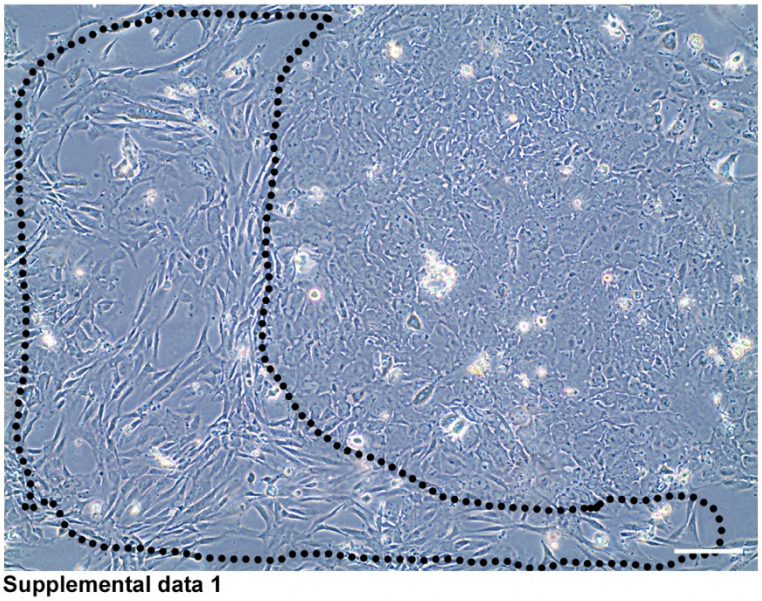


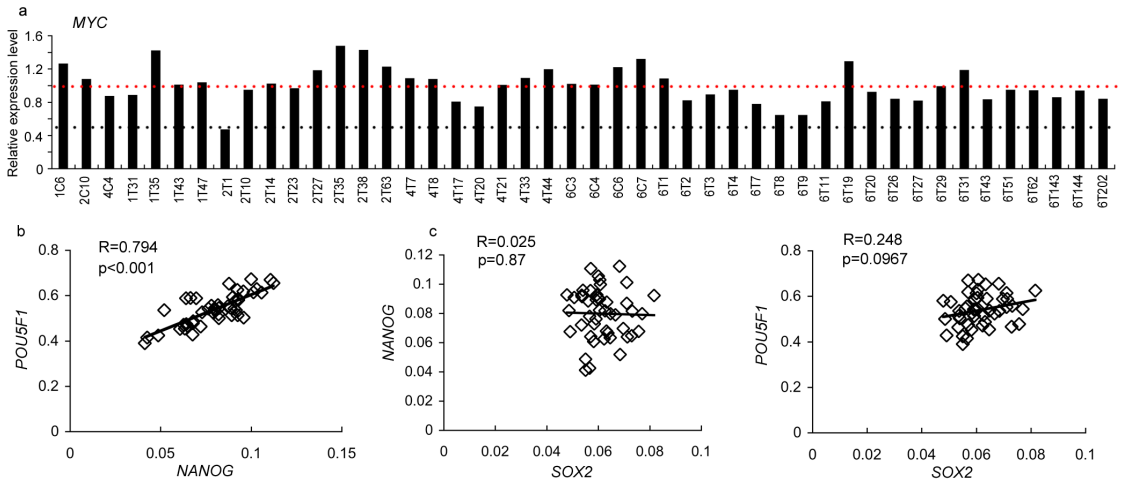
Figure 6

Supplemental data 1



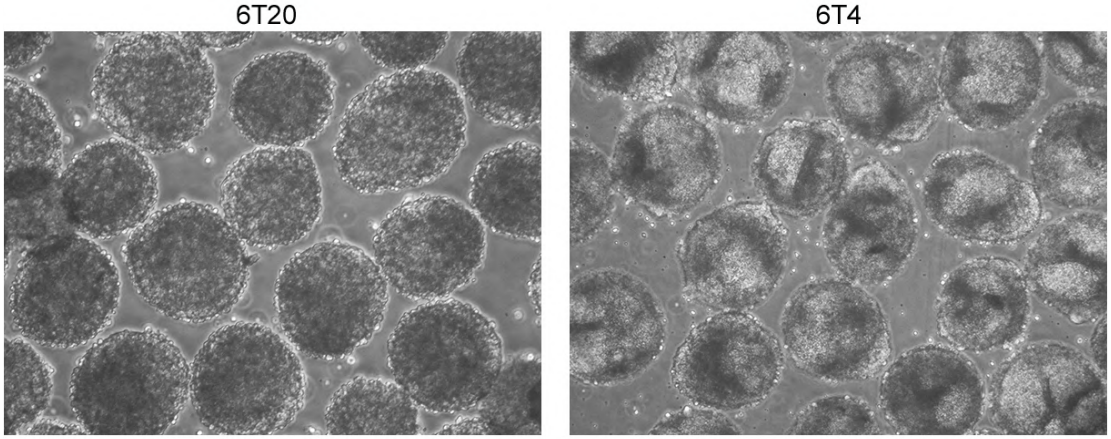
Supplemental data 1

Supplemental data 2



Sapplemental data 2

Supplemental data 3



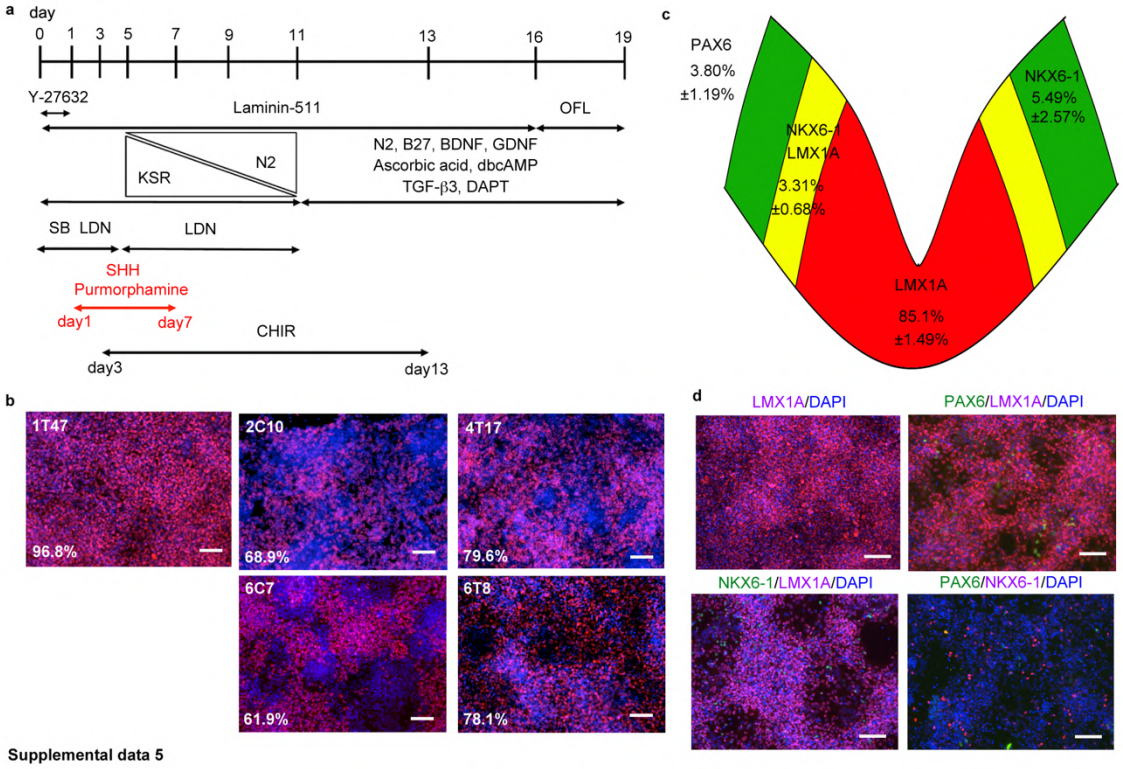
Supplemental data 3

Supplemental data 4

Cell line	Results of Karyotyping
1T35	46, XY [18] ; 47, XY, +X [2]
1T47	46, XY [20]
2T27	46, XY [5] ; 47, XY, +20 [8] ; 47, XY, +X [7]
2T63	47, XY, +X [9] ; 47, XY, +mar [7] ; 47, XY, +20 [4]
4T20	46, XY [7] ; 47, XY, +20 [13]
6T9	46, XX [20]
6T11	46, XX [6] ; 47, XX, +13 [14]
6T20	46, XX [20]
6T43	46, XX [20]
6C4	46, XX [19] ; 46, XX, -18, +mar [1]
6C6	46, XX [20]

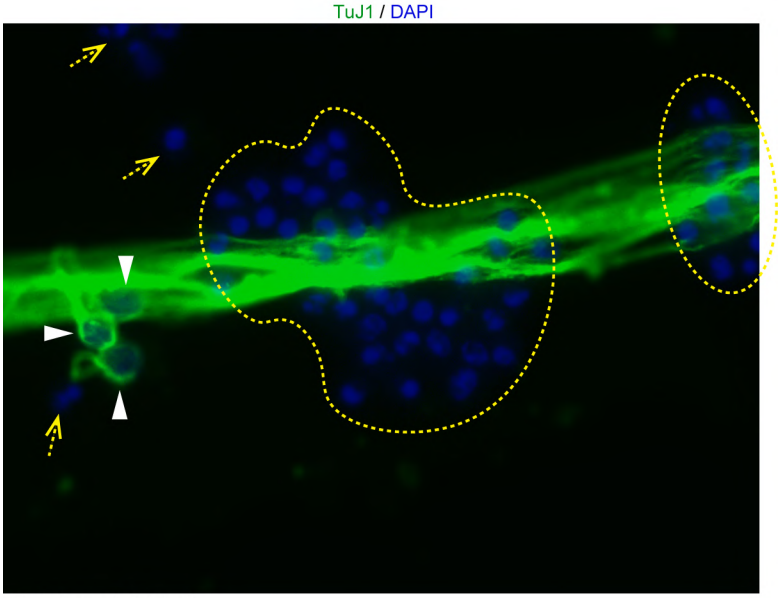
Supplemental data 4

Supplemental data 5



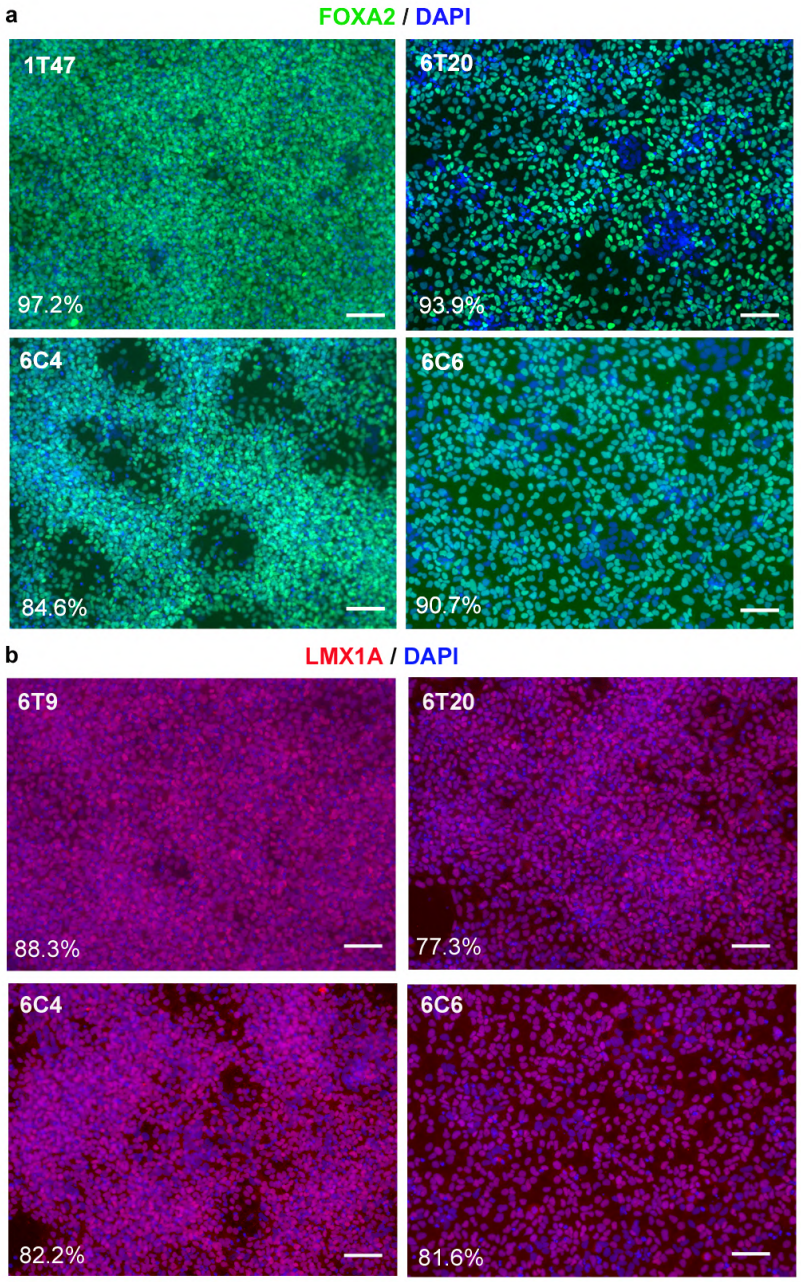
Supplemental data 5

Supplemental data 6



Supplemental data 6

Supplemental data 7



Supplemental data 7

SUPPLEMENTAL TABLE

Supplemental table 1 | Ratio of ventral midbrain markers in each mDA-differentiation experiment. The ratios of positive cells are shown as percentages for each experiment, which were performed in triplicates.

	LMX1A	NKX6.1	PAX6	NKX6.1/LMX1A	Total
1T47	92.09	6.08	3.04	0.34	100.83
	93.49	5.00	2.48	1.49	99.46
	90.83	8.50	2.46	4.92	96.87
6T9	95.11	3.99	1.72	0.24	100.58
	86.74	6.54	0.66	2.79	91.15
	82.97	10.06	1.60	4.72	89.92
6T20	93.51	3.68	2.22	0.06	99.35
	83.85	6.58	0.47	3.17	87.73
	54.45	43.69	2.93	15.49	85.58
6C4	82.83	14.88	1.42	10.54	88.60
	75.31	19.26	1.37	5.42	90.52
	88.34	14.37	0.59	4.57	98.72
6C6	85.45	5.65	1.52	1.81	90.81
	74.41	6.22	1.43	1.20	80.86
	85.06	7.13	0.90	1.79	91.30

Supplemental table 2 | Sequences of the primers used in the present study.

Target	Forward sequence	Reverse sequence
GAPDH	GCTCTCTGCTCCTCCTGTTC	GCCCAATACGACCAAATCC
NANOG	TCTCCAACATCCTGAACCTCA	TTGCTATTCTTCGGCCAGTT
POU5F1	CTTTGAGGCTCTGCAGCTTAG	TCTGCTTTGCATATCTCCTGAA
SOX2	GGGGGAATGGACCTTGTATAG	GCAAAGCTCCTACCGTACCA
KLF4	GCCGCTCCATTACCAAGA	TCTTCCCCTCTTTGGCTTG
MYC	TGCTCCATGAGGAGACACC	CTTTTCCACAGAAACAACATCG
MYCN	CCCTGAGCGATTCAGATGATGA	TTGGTGTTGGAGGAGGAACG
SOX21	GGGGCCCGGTTTGTATGTA	GAATTCACAGGCCTGCTCGC
SOX3	TCAAGAACCCCGTAGGGACA	CTGCGTTCGCACTACTCTTG
SOX1	GGCTGAGCACCACTACGACT	GCATTATAAAATTTCCCAAATCATC
PAX6	ACCCATTATCCAGATGTGTTTGCCCGAG	ATGGTGAAGCTGGGCATAGGCGGCAG
HES5	AGCACAGCAAAGCCTTCGTC	CTGCAGGCACCACGAGTAG
TBXT	GCTGTGACAGGTACCCAACC	CATGCAGGTGAGTTGTCAGAA
NLRP7	GCTGGAAGCAAGACCTGACC	CTAGCTGGGGCGATGTCATA
DPPA3	CAATTTGAGGCTCTGTCATCAG	ATCGGCGTCTTGACACAAC
DNMT3L	TCAAGGCTAACCAGCGAAAT	GCTGTGTGTGAACCTGGAGA
HHLA1	GAGTCCAACTGGTCTCCTGG	AGATCAGTAGCACCCCCTGA
ARRB1	CACGAAGACACGAACTTGGC	GGATGCAAGATCTCCCAACAG
FZD10	TGAGTTGGCTTTGCTACCCA	CCTGGGGATGGCTCATTTGT
GBX2	GGTAACTTCGACAAGGCGGAGG	GGTCGTCTTCCACCTTTGACTCG
NPTX1	CATCAATGACAAGGTGGCCAAG	TGATAGGGCGCCAAGTTCTC
OCT6	GCTCGAGAGCCACTTTCTCA	CCAGGCGCGTATACATCGT
SIX3	CCGGAAGAGTTGTCCATGTTC	CGACTCGTGTTTGTTGATGGC
HAND1	ACCAGCTACATCGCCTACCT	CCTTTTCCGCTTGCTCTCAC