TET1 can attenuate human pluripotency-related heterogeneities and enhance differentiation capabilities

Makoto Motono[1,2,4,*], Keiko Hiraki-Kamon[1,4], Masayoshi Kamon[1], Hidenori Kiyosawa[3], Yoichi Kondo[2], and Hidemasa Kato[1,5,*]

[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.

[5]Lead Contact

[*]Correspondence: makoto.motono@ompu.ac.jp (M.M.); hidemasa.kato.zz@ehime-u.ac.jp (H. Kato)

Abstract

Induced pluripotent stem cells (iPSCs) exhibit inconsistent differentiation potential, negatively impacting their downstream application. Herein, we improved their reprogramming by adding TET1, a DNA demethylase, to produce TET1-iPSCs (T-iPSCs) with enhanced epithelialized morphology and differentiation capabilities. By systematically comparing the transcriptomes of high-differentiation T-iPSCs and conventional iPSCs (C-iPSCs), we discovered a subpopulation of C-iPSCs in G_0/G_1 phase exhibiting low *TET1* expression and an extraembryonic gene signature. An extraembryonic gene signature is common to various iPSCs. Interestingly, this signature was attributable to Polycomb derepression and TET1 insufficiency. Deviation of the C-iPSCs from the cell-cycling embryonic gene profile was accompanied by DNA hypermethylation of poised and Polycomb-repressed enhancers, leading to the derepression of extraembryonic genes. TET1-facilitated reprogramming can ameliorate this deviation. This study unveils a novel aspect of TET1 for establishing human pluripotency by remedying various inherent heterogeneities resulting from the conventional iPSC reprogramming method.

Introduction

Human pluripotent stem cells (hPSCs) exhibit variable capabilities for differentiation, and this is usually more pronounced in human induced pluripotent stem cells (hiPSCs)[1,2] than in human embryonic stem cells (hESCs)[3]. Heterogeneity can be acquired during the reprogramming process[4] of hiPSCs. Attempts to improve human reprogramming have focused on increasing the efficiency of iPSC colony derivation[5,6], and methods to eliminate heterogeneity remain scarce. Although a recent reprogramming study using a transient-naive-treatment strategy corrected aberrant epigenetic memories inherent to hiPSCs and improved their differentiation capabilities[7], the underlying molecular mechanisms remain unclear. In mice, combinations of other factors can be substituted for the standard Yamanaka's reprogramming cocktail[8-10]. Tet1, a DNA demethylating dioxygenase, can replace Pou5f1[11], boosting the reprogramming efficiencies[12], and increasing the differentiation potential of hiPSCs by demethylating enhancers[13]. The mesenchymal-to-epithelial transition (MET) is an essential step[14] of reprogramming, in which Tet1 plays a crucial role[15]. Compared with mice, the onset of *TET1* expression during human reprogramming is more delayed and less intense, which hinders MET and leads to epigenetic aberrations[13,16]. Therefore, the Yamanaka factors alone may be insufficient to demethylate the genome of human mesenchymal cells to attain genuine pluripotency.

To treat Parkinson's disease, iPSC-derived ventral midbrain (VM) progenitors could

provide a transplantable source for the missing ventral midbrain dopaminergic (mDA) neurons[17,18]. However, the global expression profiles of iPSC-derived mDA cells differ substantially from their *in vivo* counterparts[19]. Thus, further improvements are needed to the pipeline for obtaining human ventral mDA neurons.

Herein, we propose a novel hiPSC production pipeline that adds TET1 to the Yamanaka factors, yielding TET1-iPSCs (T-iPSCs). This pipeline integrates differentiation tests of individual clones, selecting hiPSCs with enhanced differentiation capabilities. Assessment of the differentiation capabilities and corresponding systematic transcriptomics revealed inter- and intraclonal cell heterogeneities, linked to differentiation defects of conventional hiPSCs. Furthermore, DNA methylome and histone modification analyses support the hypothesis that this gene dysregulation is inherent to the conventional reprogramming process.

Results

TET1-aided human cell reprogramming: the T-iPSCs

During human iPSC reprogramming, the establishment of pluripotent DNA methylation patterns is delayed compared with mice[16]. By reanalyzing previously published single-cell transcriptome data from human reprogramming[20], we found that *TET1* expression is pronounced during a successful primed reprogramming, followed by *DNMT3B* upregulation

(Extended Data Fig. 1a). At the early stage of reprogramming, the *TET1* expression level is higher on the reprogramming trajectory than on the non-reprogramming trajectory (Extended Data Fig. 1b; day 4). Given the known pleiotropic roles of TET1 in reprogramming and pluripotency[13,21,22], these results indicate that *TET1* expression may be a bottleneck in human reprogramming. We generated hiPSCs from fibroblasts using episomal vectors carrying conventional reprogramming factors[23,24] alone (C-iPSCs) and with the addition of TET1 (T-iPSCs) (Fig. 1a). We used a culture medium that supports the mouse primed epiblast stem cells for propagation[25]. This method facilitated the single-cell passage of cells after trypsin digestion and plating onto a non-feeder surface, allowing for better discrimination of the morphological aspects of the clones (Fig. 1b and Methods). During colony formation, most emerging T-iPSCs exhibited an epithelialized morphology characteristic of simple columnar epithelium. These epithelialized colonies were rare for C-iPSCs, which typically bore a mesenchymal morphology (Fig. 1c,d). Upon expansion, we observed that some clones acquired mesenchymal-like morphologies (Fig. 1e). As the epithelialization of mesenchymal cells during reprogramming results from TET activities[15,16], we aborted any mesenchymal clones for both C-iPSCs and T-iPSCs. In this manner, we established 25 C-iPSC and 145 T-iPSC clones (from 140 and 420 colonies, respectively) (Fig. 1f,g). Episomal vector introduction by electroporation is versatile but may trigger genome integration[23]. Using 0.03 as the vector-free threshold (Methods), we obtained 46 vector-free clones (7 C-iPSCs and 39 T-iPSCs) from four

independent parallel inductions.

To depict the transcriptional signatures of the vector-free hiPSC clones at the earliest point after establishment, we quantitatively analyzed the endogenous expression of the "pluripotency-associated" factors in all the clones (Fig. 1h and Extended Data Figs. 1c and 2a). We observed no significant difference between C-iPSCs and T-iPSCs for *NANOG*, *POU5F1,* and *SOX2* (Fig. 1h). Despite the addition of TET1 during reprogramming, the vector-free T-iPSC clones did not show significantly higher endogenous *TET1* levels compared with C-iPSCs (Fig. 1h). *KLF4* and *MYCN* showed considerable fluctuations (Extended Data Fig. 2a). Notably, this analysis was performed in low-passage nascent iPSC lines prone to transcriptional noise[26]. Therefore, the averaged expression of the pluripotency-associated factors indicated the overall uniformity of the hiPSC clones obtained and that all clones conformed to the qualitative criteria for pluripotency.

The default neural induction model showed interclonal heterogeneity of the differentiation capability

HiPSCs have inherently unpredictable differentiation defects[4]. To set a tentative standard for the differentiation capability, we chose the default model of neural induction ("default differentiation"), an evolutionarily conserved developmental trajectory among vertebrates (Fig.

2a). This model suggests that dissociated early embryonic cells default to neural induction when incubated in a developmentally neutral culture medium devoid of any signaling molecules[27-30]. We selected well-expandable C-iPSCs (1C6) and T-iPSCs (1T35, 1T47) from the initial reprogramming lot and pilot-differentiated them to determine a differentiation timeline. This showed that, after day 18 of default differentiation, 1T35 and 1T47 cells turned positive for the neural markers PAX6 and SOX1[31,32], while 1C6 cells failed to express these markers (Fig. 2b).

We then analyzed the gene expression changes during default differentiation using these clones (1C6, 1T35, and 1T47) and evaluated the neural gene expression levels on day 8, when the differentiation efficiency diverges (Extended Data Fig. 2b). We analyzed *HES5* expression, a neuroectodermal marker, in addition to *PAX6* and *SOX1* (Extended Data Fig. 2a)[33], and included indicators of improper default differentiation: *NANOG* and *POU5F1* are direct inhibitors of neural induction[34], while *TBXT* is expressed in the mesendoderm and non-neural ectoderm[35] in humans. 1T35 and 1T47 outperformed 1C6 for all neural markers (Fig. 2c). Differentiated 1C6 cells expressed higher levels of *TBXT*, *NANOG,* and *POU5F1*, providing further evidence of their deviation from default neurogenesis (Fig. 2d and Extended Data Fig. 2b).

Based on these results, we rejected clones that could not outperform 1C6 (Fig. 2b–d) and named the remaining hiPSC clones “default-elite” clones (their karyotypes are shown in

Extended Data Fig.3). Figure 2e summarizes the number of clones that met the criteria at each step of our iPSC production pipeline. Based on these criteria, only 0.00005% of fibroblasts used for C-iPSC induction acquired the default differentiation capability. A previous report using the same reprogramming factors produced 0.006% of iPSC clones[23], and thus our pipeline sets a standard that is more than one hundred times more stringent (Figs. 1g and 2e). TET1-assisted reprogramming elevated the efficiency to 0.0002%, a four-fold increase in the production of hiPSC clones with enhanced differentiation potential.

Transcriptome features affecting differentiation efficiency

Next, we explored gene expression markers of the undifferentiated state, which would predict the default differentiation capability of established clones, and investigated the correlations between the gene expression levels before and after default differentiation to expand the gene panel. The classical pluripotency-associated genes showed no significant correlations with neural differentiation markers (Extended Data Fig. 2c). No gene that positively predicted the differentiation capability was identified. However, the fluctuating expression of endogenous *KLF4* and *MYCN* in the undifferentiated state correlated negatively with the neural markers upon differentiation (Extended Data Fig. 2a,c). Because KLF4 is a factor in differentiation resistance[2,36,37], we investigated KLF4-related factors (NLRP7 and DNMT3L)[16] and a mesoderm and extraembryonic marker (HAND1)[38]. We found that these factors negatively

correlated with the neural markers (Extended Data Fig. 2c). Although all our hiPSC clones qualitatively met the pluripotency criteria (Fig. 1h and Extended Data Figs. 1c and 2a), the default differentiation survey revealed that the expression of pluripotency-associated factors did not indicate the differentiation capabilities of the hiPSC clones.

Neuronal differentiation beyond the default neurogenesis

To explore the differentiation potential beyond the early neural stages assessed by default differentiation, we selected precursors of the ventral midbrain (VM)[17]. We modified a published protocol[17] and achieved high efficiency in inducing FOXA2, a VM precursor marker, using the 1T47 clone (over 95%; Extended Data Fig. 4a,b). The differentiation efficiency for LMX1A, another VM precursor marker, was also the highest for 1T47 (Fig. 3a,b, and Extended Data Fig. 4c)[39]. However, some cells lacked LMX1A, NKX6-1, or PAX6, collectively all the markers of VM precursors[39,40]. The proportion of these "undefined" cells, presumably induced to cells outside the VM, was significantly higher for 6C6 than for 1T47 (Fig. 3c, and Supplementary Table 1). Furthermore, the dendrites of the neurons derived from 1T47 were longer than those from 6C6 (Fig. 3d,e).

Interclonal heterogeneity of hiPSC clones

A recently proposed reference iPSC clone shows high differentiation potential, especially for neural cells[41]. However, the specific properties responsible for this potential remain unclear, hindering the replication of similar iPSC clones. Our preliminary results designated 1T47, which consistently demonstrated elevated differentiation potentials, as the reference clone. We aimed to identify the distinguishing features unique to 1T47 (T-iPSC^{1T47}) compared with the other default-elite clones. We selected 6C6 (C-iPSC^{6C6}) as a control representing an average C-iPSC clone (6C4 was aneuploid; Extended Data Fig. 3). Bulk RNA-sequencing (RNA-seq) analysis revealed differentially expressed genes (DEGs) between these clones in the undifferentiated state (Fig. 4a). Gene ontology (GO) analysis revealed few GO terms associated with the DEGs expressed at higher levels in T-iPSC^{1T47} (T-iPSC^{1T47}-high-DEGs) (Fig. 4b). However, the C-iPSC^{6C6}-high-DEGs were systematically related to epithelial and mesenchymal traits (extracellular matrix organization, epithelial cell differentiation, and burn wound healing; Fig. 4b). Examples of mesenchyme-related genes include epithelial-mesenchymal transcription factor (*TWIST1*), mesenchymal cytoskeletal components (*MYL7*, *ACTC1*, *ANXA1,* and *VIM*), and mesoderm/mesenchyme-inducing signaling molecule (*BMP4*) (Fig. 4c)[42,43]. The C-iPSC^{6C6}-high-DEGs were also related to aging (p53 transcriptional gene network; Fig. 4b). Examples of aging-associated genes include cell cycle inhibitors (*CDKN1A/2A/2B*) and the senescence-associated secretory phenotype (SASP) genes (*GDF15, IL6, IL1A,* and *IGFBP7*) (Fig. 4c)[44-46].

Next, we explored the versatility of the mesenchymal-related and aging-associated gene expression among hiPSC clones. Reanalysis of array data from 49 hiPSC clones[1] revealed minimal variation in the pluripotency-associated genes between the clones, but significant variation in the mesenchymal-related and aging-associated gene expression (Extended Data Fig. 5a,b). Previously assigned differentiation-defective markers[2] also showed substantial interclonal expression heterogeneities (Extended Data Fig. 5c,d). These results suggest that this heterogeneity is a characteristic observed in a considerable number of hiPSC clones, and thus is not unique to our selected reference C-iPSC^{6C6} clone.

Mesenchymal gene noise in C-iPSCs is related to extraembryonic lineages

This observed mesenchymal propensity was surprising because hiPSCs are supposedly equivalent to post-implantation epiblasts, an epithelialized tissue[47]. Because of the difficulties in obtaining human embryonic gene profiles from the implantation to gastrulation stages, we used published single-cell RNA-seq (scRNA-seq) data of the peri-gastruloids[48] to determine which lineages express the C-iPSC^{6C6}-high-DEGs during human development. We integrated the peri-gastruloid scRNA-seq data with data obtained from T-iPSC^{1T47}, C-iPSC^{6C6}, and a standard hESC H9[49]. Peri-gastruloids exhibit developmental traits from epiblast to definitive mesoderm and endoderm (Extended Data Fig. 6a–c). Cluster annotations using putative lineage markers have led to some controversial issues. A significant population of the peri-gastruloids

expressing *ACTC1*, *BMP4*, *HAND1*, *MYL7*, *TWIST1*, and *VIM*, which are mesodermal markers, did not co-express the definitive mesodermal markers *MESP1/2* and *MIXL1* (Extended Data Fig. 6c,d)[48]. Notably, this population exclusively expressed the C-iPSC^{6C6}-high-DEG *ANXA1*, a marker associated with human extraembryonic mesoderm (Extended Data Fig. 6d)[43]. The annotations for T-iPSC^{1T47} and hESC H9 co-segregated close to the epiblast clusters, while C-iPSC^{6C6} was uniquely annotated close to the extraembryonic mesoderm and primitive endoderm cluster (Extended Data Fig. 6a,e). These data indicate that the mesenchyme-related C-iPSC^{6C6}-high-DEGs are likely linked to the extraembryonic mesoderm rather than the definitive mesoderm (Fig. 4c and Extended Data Fig. 6c,d). To consolidate the developmental contexts of hPSCs, we referred to cultured monkey blastocysts up to the gastrulation stage by focusing on the C-iPSC^{6C6}-high-DEGs (Extended Data Fig. 7a)[38]. C-iPSC^{6C6}-high-DEGs (*ACTA1*, *ACTC1*, *ANXA1*, *BMP4*, *MYL7*, *TWIST1*, and *VIM*) and the differentiation-defective marker (*HAND1*) were highly expressed in the amnion or the extraembryonic mesoderm (Fig. 4a,c and Extended Data Figs. 2c, and 7b)[38]. Importantly, the expression levels of *ACTA1*, *ANXA1*, *MYL7*, and *TWIST1* were low in the definitive mesodermal lineages (Extended Data Fig. 7b). These results indicate that C-iPSC^{6C6} expresses extraembryonic rather than embryonic mesodermal genes.

Unveiling intraclonal heterogeneities in C-iPSCs

Integrated scRNA-seq analyses of non-sorted hiPSC clones revealed a C-iPSC^{6C6}-specific cell population, predominantly residing in the G_0/G_1 cell cycle phase (Fig. 4d). In this study, we defined cells with markedly higher *CDKN2A* expression and lower *MKI67* expression (a cell proliferation marker) than the other cell cycle phases (G_1, S, and G_2/M) as being in the G_0-phase (Extended Data Fig.8a). Based on this criterion, 7% of C-iPSC^{6C6} were in the G_0 phase, while none of the T-iPSC^{1T47} were (Fig. 4e). The *TET1*, *DNMT3B*, and *SOX2* expression levels in G_0 phase were lower than in the other cell cycle phases (Fig. 4d,f,g and Extended Data Fig. 8b–d). Of note, some of the C-iPSC^{6C6}-specific population highly expressed extraembryonic genes, particularly in the G_0 phase (Fig. 4d,f,g and Extended Data Fig. 8b–d). Given the lower expression of *TET1* in cells expressing the cell cycle inhibitors and the extraembryonic markers, we reanalyzed the RNA-seq data for *TET1/2/3*-triple-knockout hESCs (TET-TKO), where all the TETs were knocked out in HUES8 hESCs[50]. TET-TKO exhibited augmented expression of the cell cycle inhibitors and the extraembryonic markers (Fig. 4h). Of note, *CDKN2A* and *TWIST1* expression were observed in late OSKM-reprogrammed cells, even after sorting with pluripotent surface markers (Extended Data Fig. 8e)[7,20]. Moreover, the differentiation-defective markers (*ABHD12B*, *LINC-ROR*, and *HHLA1*)[2] were overexpressed in TET-TKO (Fig. 4h) and C-iPSC^{6C6} compared with the wild-type (WT) hESC and T-iPSC^{1T47}, respectively (Extended Data Fig. 8f). Collectively, this indicates that we have revealed a previously unrecognized trait of conventional hiPSCs, in which a subpopulation emerges, mainly in the G_0 phase where *TET1*

expression is lower, exhibiting intraclonal heterogeneity of mesenchymal/extraembryonic gene noise (Fig. 4i).

We then determined whether the observed differentiation superiority of T-iPSC^{1T47} (Fig. 3) is a trade-off for their inability to differentiate into extraembryonic lineages[51]. We proactively differentiated T-iPSC^{1T47} by adding BMP4, which induces extraembryonic cells[52-54]. *BMP4*, itself a marker of extraembryonic cells, was self-induced in T-iPSC^{1T47} at a much higher level than in C-iPSC^{6C6} (Extended Data Fig. 9a). BMP4 also induced other extraembryonic markers such as *TWIST1* and *ANXA1* (Extended Data Figs. 7b and 8b), but only in T-iPSC^{1T47}, suggesting that T-iPSC^{1T47} have not lost the potential to differentiate into extraembryonic lineages, despite possessing a superior neural differentiation capability (Extended Data Fig. 9a). Furthermore, we investigated whether T-iPSC^{1T47} have a biased propensity to differentiate toward the ectodermal lineage and found that T-iPSC^{1T47} also showed higher efficiency than C-iPSC^{6C6} for differentiating into mesendodermal lineages (Extended Data Fig. 9b,c). These results indicate that T-iPSCs are not biased towards the neural lineage and can differentiate into the three germ layers and extraembryonic tissues.

TET1 insufficiency leads to hypermethylation of poised extraembryonic gene enhancers

To determine the possible involvement of TET1 insufficiency (Fig. 4f,g,i), we compared the DNA methylomes of WT hESCs, its TET-TKO[50], fibroblasts[36], T-iPSC^{1T47}, and C-iPSC^{6C6} using whole genome bisulfite sequencing (WGBS). We examined the differentially methylated regions (DMRs) and identified 4769 C-iPSC^{6C6}- and 1244 T-iPSC^{1T47}-hypermethylated DMRs (hyper-DMRs) (≥10% difference in DNA methylation). DMRs were observed around the C-iPSC^{6C6}-high-DEG loci of extraembryonic markers, aging-related markers, and *KLF4* (Fig. 5a and Extended Data Fig. 10a). Assigning the DMRs to the neighboring gene structure showed a pronounced enrichment within the gene body, exceeding the promoter region (Fig. 5b). Functional annotation of the C-iPSC^{6C6}-hyper-DMR-associated genes revealed an association with developmental morphogenesis, particularly related to cell patterning and epithelial–mesenchymal plasticity (Fig. 5c)[55]. In contrast, the T-iPSC^{1T47}-hyper-DMRs showed no such trend (Extended Data Fig. 10b). To investigate the impact of DNA methylation on gene expression, we examined the proportion of DEGs associated with the DMRs of the two hiPSC clones. The genes linked to hyper-DMRs tended to be downregulated in T-iPSC^{1T47} but upregulated in C-iPSC^{6C6} (Fig. 5d and Supplementary Table 2). For the C-iPSC^{6C6}-hyper-DMRs, paradoxically, DNA hypermethylation facilitated gene expression, while DNA methylation is generally considered to repress gene expression[56]. This can be partly resolved by analyzing the epigenetic statuses of nearby histones, particularly the Polycomb repression mark, H3K27me3. H3K27me3 (and a lack of H3K27ac) was observed in the C-iPSC^{6C6}-hyper-

DMRs in standard hPSCs (Extended Data Fig. 10c). These regions were also enriched for a putative enhancer mark (H3K4me1), categorizing them as "poised enhancers"[57]. H3K27me3 was enriched in the C-iPSC^{6C6}-hyper-DMRs associated with the C-iPSC^{6C6}-high-DEGs, but T-iPSC^{1T47} showed no such trend (Fig. 5e and Extended Data Fig. 10d). Notably, these DMRs associated with the DEGs avoided gene promoters and predominantly resided in enhancer regions (Fig. 5f and Extended Data Fig. 10e). We found gene body-embedded poised enhancers for *ACTA1*, *BMP4*, *GDF15*, *TWIST1*, and *KLF4* (Fig. 5a and Extended Data Fig. 10a). Hypermethylation of poised enhancers could cause target gene derepression because DNA methylation is mutually exclusive to Polycomb repressive complex 2 (PRC2) binding[58]. These results suggest that TET1 insufficiency could lead to hypermethylation of poised enhancers and the paradoxical derepression of the related genes. Many C-iPSC^{6C6}-high-DEGs were upregulated when Polycomb repression was compromised by knockout of EZH2, a PRC component, in hESCs (EZH2-KO) (Fig. 5g,h)[59]. We identified 127 genes common to the C-iPSC^{6C6}-high-DEGs, the genes expressed at higher levels in EZH2-KO (EZH2-KO-high-DEGs), and those at higher levels in TET-TKO (TET-TKO-high-DEGs) (Fig. 5i and Supplementary Table 3). GO analysis of these common genes indicated an enrichment of epithelial–mesenchymal transition-related terms (Fig. 5j). These findings suggest that the hypermethylation of poised enhancers in C-iPSC^{6C6} resulting from TET1 insufficiency cause

the derepression of cell cycle inhibitors and extraembryonic genes by evading Polycomb-mediated repression.

Aberrant methylation is acquired during iPSC reprogramming

Aberrant DNA methylation may be acquired during the reprogramming process itself[7,60,61]. The C-iPSC^{6C6}-hyper-DMRs within the gene bodies of *ACTA1, GDF15*, and *TWIST1* were hypomethylated in fibroblasts (Fig. 5a), suggesting that these regions have likely acquired this methylation during reprogramming. We reanalyzed published Reduced Representation of Bisulfite Sequencing (RRBS) data from conventionally reprogrammed iPSCs[16] and found 1732 autosomal enhancers (of 40,838) that were poised in hPSCs and acquired *de novo* DNA methylation during reprogramming (Fig. 5k). We observed that 1615 of the hPSC-poised enhancers were hypermethylated in C-iPSC^{6C6} compared with fibroblasts, including those around C-iPSC^{6C6}-high-DEGs (Fig. 5a). Approximately 70% of the enhancers were located within gene bodies, and *de novo* methylated poised enhancers (hiPSC-hyper-poised-enhancers) were similarly distributed (Fig. 5l). These results indicated that a significant number of poised enhancers in hiPSCs acquired *de novo* methylation in the gene body as the sole consequence of reprogramming. Because gene-body methylation primarily occurs in correlation with gene expression[62], we assessed their expression during reprogramming. As expected, these extraembryonic genes were transiently expressed during reprogramming (Extended Data Fig.

10f)[20]. The DNA methylation pattern in hPSCs is constantly regulated by counteracting methylating (*DNMT3A/B*) and demethylating (*TET1*) activities[50,63,64], and this balance may be disrupted in a subset of cells with insufficient *TET1* and *DNMT3B* activity, both during reprogramming (Extended Data Fig. 1a) and in the G_0 phase in hiPSCs (Fig. 4f).

Discussion

We found that mouse definitive mesodermal genes were also expressed in the primate extraembryonic lineages. The expression of extraembryonic markers in the hPSCs suggests a different scenario to the primed pluripotency proposed for mouse-PSCs, where cells are halted before differentiation of the germ lineage[65]. Considering the variation in *CDKN2A* and extraembryonic gene expression observed in large-scale hiPSC datasets, a subset of cells expressing these genes is expected to reside in the G_0 phase. We found that cells in the G_0/G_1 phase had lower levels of *TET1* expression (Fig. 4d,f,g), which may lead to hypermethylation of the poised enhancers (Fig. 5a,e). This hypermethylation of nearby poised enhancers may cause the of extraembryonic genes because DNA methylation is mutually exclusive to H3K27me3[58]. This observation is consistent with previous studies showing that hyper-DMRs in extraembryonic tissue frequently overlap with PRC2-regulated genes in mouse pluripotent cells[66], and that EZH2-KO exhibit extraembryonic gene dysregulation[59]. The impact of the transient-naive-treatment method, reducing *BMP4* expression, can be reconciled with the

induction of TET1 during the process[7]. *BMP4* expression is commonly found among hiPSCs[7] and may assist in an autocrine/paracrine fashion to aggravate the aberrant extraembryonic gene expression[35,38,67]. The bystander effect of secreted proteins such as BMP4 from cell cycle-arrested cells is strongly reminiscent of cellular senescence. Indeed, these G_0 cells exhibited SASP (Fig. 4c), a key feature of senescence. The appearance of G_0/senescent cells was unexpected in stem cells. However, considering the recent revelation that aging-related hypermethylated gene loci are Polycomb domains[68,69], a wide survey of the "ages" of various hPSCs would be informative. Data from single-cell transcriptomics and DNA methylomics suggest that prolonged low expression of TET1 leads to hypermethylation of poised enhancers located in the gene body, resulting in the derepression of extraembryonic genes, presumably contributing to the cell cycle arrest and further aggravating the TET1 insufficiency.

We also revisited a DNA methylation paradox, that gene-body methylation is associated with gene transcription[56]. Active genes acquire gene-body methylation as RNA polymerase II recruits the H3K36me3-methyltransferase SETD2, further attracting DNMT3B[62,70]. Gene-body methylation then functions to prune spurious intragenic transcription[71]. Notably, these H3K4me3-attracting non-methylated CpG islands within the gene body would require silencing through a Polycomb-mediated mechanism, resulting in a bivalent chromatin signature[62]. An intriguing hypothesis is that TET1 is involved in resetting such "faulty" gene derepression and the ensuing DNA methylation as a premise of the bivalent chromatin for

stringent gene silencing[72], a cardinal feature of pluripotency[73].

Previous studies have attributed residual *KLF4* expression to differentiation defects or epigenetic abnormalities of hiPSCs[2,36,37]. Our default differentiation screening showed a negative correlation between *KLF4* expression and differentiation outputs, providing the first quantitative and statistical evidence consolidating these previous findings (Extended Data Fig. 2c)[2,36,37]. Another study demonstrated that TETs demethylated bivalent promoters and poised enhancers, but the effects of TETs on gene expression were not well documented[22]. Our reanalysis revealed that *KLF4* is highly expressed in TET-TKO and EZH2-KO, which is attributable to the hypermethylation of poised enhancers within the *KLF4* gene body (Extended Data Fig. 10a). Previously assigned differentiation-defective markers were also highly expressed in TET-TKO (Fig. 4h), suggesting that these genes are targets of TETs. These results suggest that the observed interclonal heterogeneities among hiPSC clones are rooted in TET1 deficiency.

This study highlights the importance of standardization when obtaining target cell types, as the outcome is affected by both the selection of appropriate hiPSC clones (Figs. 1 and 2a) and the differentiation protocol (Extended Data Fig. 4a). Our findings suggest that the eight-day default differentiation assay may not fully discern significant differences in the capacity of the hiPSC lines for more advanced-stage differentiation (Fig. 3b–e), despite all the lines selected exhibiting high differentiation efficiency at the early differentiation stages (Fig. 3a,b and

Extended Data Fig. 4b,c). This indicates the importance of developing appropriate differentiation protocols for standardized clones with high differentiation capability, highlighting the current scarcity of universally applicable differentiation protocols[74].

PSCs are known for a shortened G_1 phase, leading to an S–M-pivoting cell cycle[75,76], a feature partly inherited from the elite fast-dividing somatic cells from which iPSCs originate[76-79]. We found that hiPSC clones exhibit intraclonal heterogeneity with cells in G_0/senescence expressing *CDKN2A* and low *TET1*. *CDKN2A* is also upregulated in TET-TKO and EZH2-KO. This raises the possibility that these epigenetic controls are reciprocally linked to cell cycle regulation. TET1 and PRC2 function to establish bivalent chromatin, a cardinal feature of pluripotency[50]. Any cell not attaining adequate pluripotency during development may be "outcompeted" to not contribute to development through premature senescence. Therefore, it is possible that sub-optimized hPSCs would suffer from this developmental "failsafe" machinery and manifest heterogeneities.

In summary, we have developed a new hiPSC induction system coupled with default differentiation screening, showing the quasi-total induction of VM progenitors from T-iPSCs. Our results strongly suggest that the current method of creating hiPSCs induces heterogeneities due to TET1 insufficiency during and after the reprogramming process. The development of strategies to avoid these heterogeneities will be critical for advancing the practical application of hiPSCs, and T-iPSCs may be a promising avenue for accomplishing this.

Acknowledgments

We thank Y. Moriyama, Y. Tokuzawa, M. Bagheri, S. Funayama, R. Hamada, 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. We thank Catherine Perfect, MA (Cantab), from Edanz (https://jp.edanz.com/ac), for editing a draft of this manuscript.

Conflict of interest

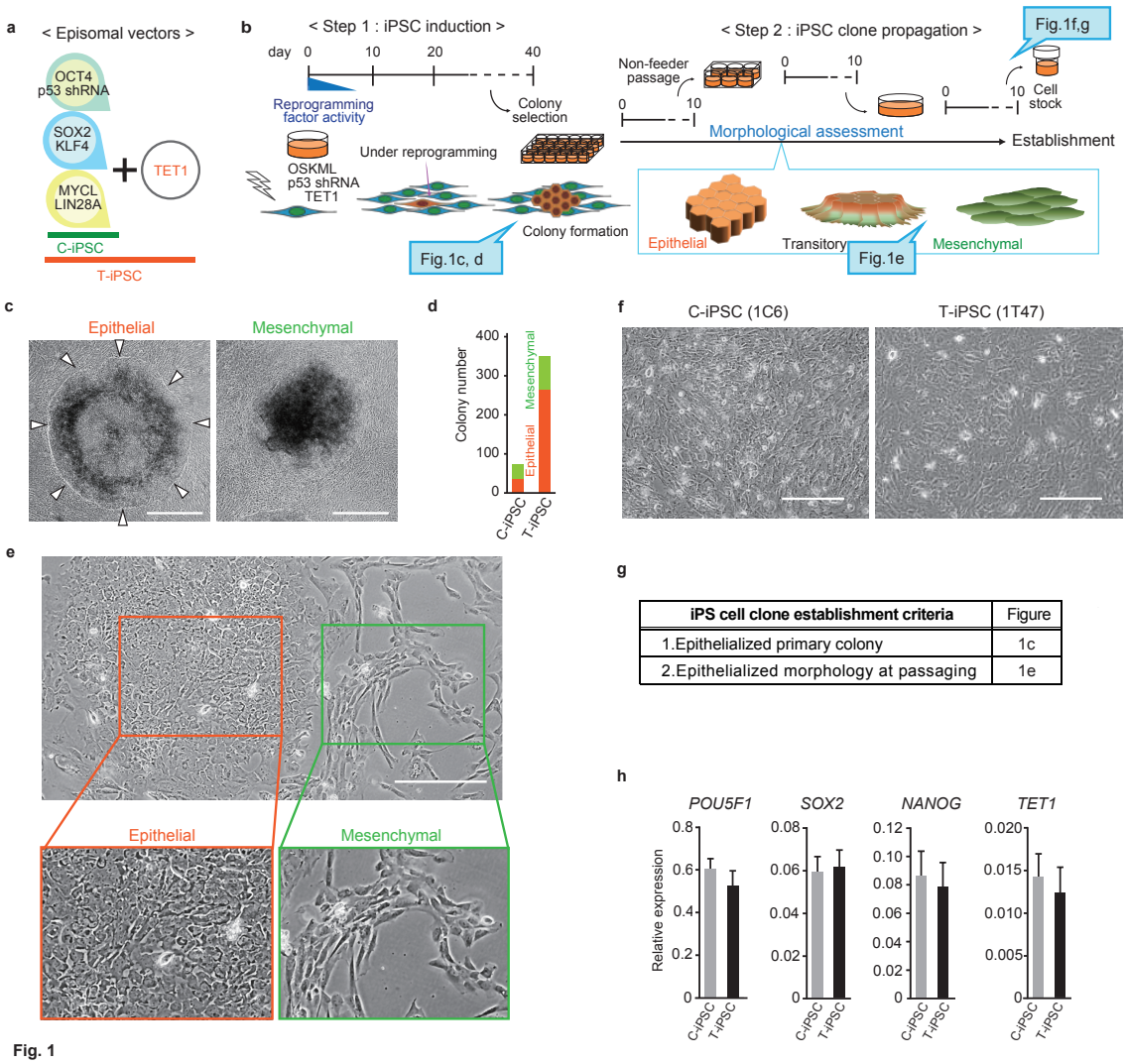
K.H-K. and H. Kato are inventors on patent no. WO2014-069479, which covers the procedure for T-iPSC production. The other authors declare that they have no competing interests.

Author contributions

H. Kato conceived the project. M. M., K. H-K., and H. Kato wrote the original draft. M. M., K. H-K., M. K., and H. Kato performed culture and differentiation experiments and analyzed data. K. H-K. and M. K. analyzed sequencing data and performed bioinformatic analyses

under the supervision of H. Kato. M. M., K. H-K., M. K., H. Kiyosawa, Y. K., and H. Kato have reviewed, edited, and approved the final version of the manuscript.

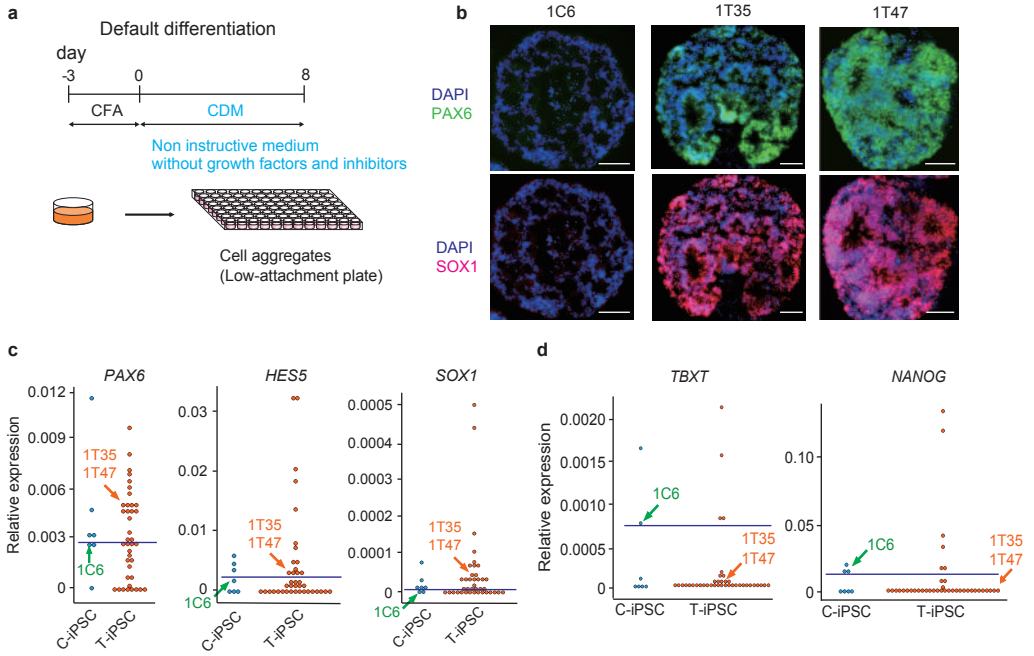
Figure Legends



iPS cell clone establishment criteria	Figure
1.Epithelialized primary colony	1c
2.Epithelialized morphology at passaging	1e

Fig. 1: HiPSC-production pipelines. a, Episomal vector combinations for producing C-iPSCs and T-iPSCs. The TET1 vector was only used in the production of T-iPSCs. OCT4 is also known as POU5F1. C-iPSC, conventional iPSC. **b**, Timeline and procedure for hiPSC production (Methods). HiPSCs were induced by electroporation of episomal vectors into fibroblasts in Step 1. HiPSC colonies were expanded on feeder cells after colony-picking. Next, hiPSCs were replated on a vitronectin-coated dish without feeder cells to facilitate

morphological assessment in Step 2. HiPSCs were cryopreserved when they reached 90% confluency in 100-mm dishes. OSKML: OCT4, SOX2, KLF4, MYCL, and LIN28A. **c**, Characteristic morphologies of the emerging hiPSC colonies in Step 1. "Epithelial" denotes hiPSC colonies with sharp boundaries, and "Mesenchymal" those with fuzzy contours. Non-reprogrammed background fibroblasts served as feeder cells for the hiPSCs. Arrowheads indicate epithelial-like colony edges. Scale bar = 50 μm. **d**, Colony numbers with epithelial or mesenchymal morphologies. There were 33 epithelial and 35 mesenchymal colonies for C-iPSCs, and 279 epithelial and 76 mesenchymal colonies for T-iPSCs. These data were collected from three independent iPSC inductions. **e**, Characteristic morphologies of a subconfluent culture during the expansion of a picked colony in Step 2. Clones, where mesenchymal-like morphologies reappeared during expansion, were excluded. Scale bar = 100 μm. **f**, Characteristic morphologies of established C-iPSC and T-iPSC clones. Scale bar = 50 μm. **g**, Criteria during the iPSC establishment processes. **h**, Relative expression of the pluripotency markers and *TET1* in all vector-free clones. The y-axis indicates relative expression levels normalized to those of *GAPDH*. Data are presented as the mean ± standard deviation (C-iPSC, n = 7; T-iPSC, n = 39).



e

The number of	C-iPSC	T-iPSC	Criteria
Induced fibroblasts	4,000,000	4,000,000	
Established clones	25 (0.00063%)	145 (0.0036%)	Cell morphology (Fig.1f)
Vector-free clones	7 (0.00018%)	39 (0.00098%)	< 0.03 copy/genome (Supplementary Fig. 1a)
Elite clones	2 (0.000050%)	9 (0.00023%)	Fig. 2c,d and Extended Data Figs. 2a and 3
Elites (normal karyotype)	1 (0.000025%)	4 (0.00010%)	Extended Data Fig. 3

Fig. 2

Fig. 2: Marked differences in the default differentiation capability between hiPSC clones. **a**, Schematic representation of the default differentiation timeline and procedure used to evaluate the differentiation capabilities of C-iPSC and T-iPSC clones (Methods). HiPSCs were adaptation-cultured in CFA, a KSR-reduced medium, before default differentiation. Cell aggregates were generated on v-bottom 96-well plates in the non-instructive medium, CDM. The CFA and CDM are described in the Methods. **b**, Representative immunostaining images with DAPI (blue), PAX6 (green), and SOX1 (red) on day 18 of default differentiation. The clone names are indicated at the top of each image set. 1C6 is a C-iPSC; 1T35 and 1T47 are T-

iPSCs. Scale bars = 100 μm. **c**, Relative expression levels of neural markers on day 8 of default differentiation measured using RT-qPCR (C-iPSC, n = 7; T-iPSC, n = 39). Each dot represents the expression level of the corresponding clone. The horizontal lines in each panel indicate tentative threshold values for the default differentiation capability. Arrows mark the values for 1C6, 1T35, and 1T47. The y-axis indicates relative expression levels normalized to those of *GAPDH*. The default differentiation surveys were performed from iPSC clones between passage numbers 5 and 8. **d**, Relative expression levels of *TBXT* and *NANOG* on day 8 of default differentiation (C-iPSC, n = 7; T-iPSC, n = 39). The lines indicate tentative thresholds for default differentiation defectiveness. Arrows mark the values of 1C6, 1T35, and 1T47. The y-axis indicates relative expression levels normalized to those of *GAPDH*. **e**, Summary of cells' number (ratio) meeting our tentative criteria in each iPSC production process.

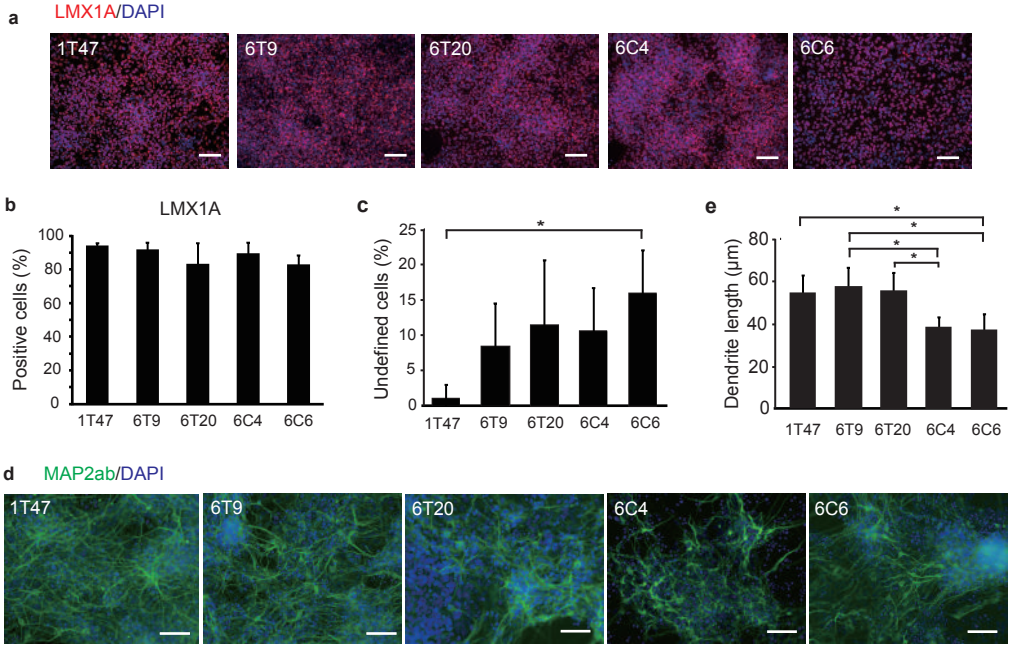


Fig. 3

Fig. 3: Rationale-based induction of ventral midbrain (VM) progenitors revealed differences in the differentiation capability between default-elite hiPSCs. a, Immunostainings of the hiPSC-derived VM progenitors with DAPI (blue) and the VM progenitor marker, LMX1A (red), on day 19 for default-elite clones. Scale bars = 50 μm. **b**, Ratio of the hiPSC-derived VM progenitors based on the midbrain markers (LMX1A) on day 19. The clone names are indicated under each column. The data are presented as the mean ± standard deviation (n = 3–4). **c**, Ratio of the undefined cells on day 19 (related immunostainings in Extended Data Fig. 4c). The data are presented as the mean ± standard deviation (n = 3-4). One-way analysis of variance and Tukey's multiple-comparison tests were performed to determine the significance of the observed differences. A significant difference was observed only between 1T47 and 6C6. *$P < 0.05$. **d**, Representative immunostaining images with DAPI

(blue) and MAP2ab (green) for morphological evaluation on day 29. Scale bars = 50 μm. **e**, Lengths of MAP2ab-positive dendrites on day 29. The values are presented as mean ± standard deviation (n = 3). One-way analysis of variance and Tukey's multiple-comparison tests were performed to determine the significance of the observed differences. *$P < 0.05$.

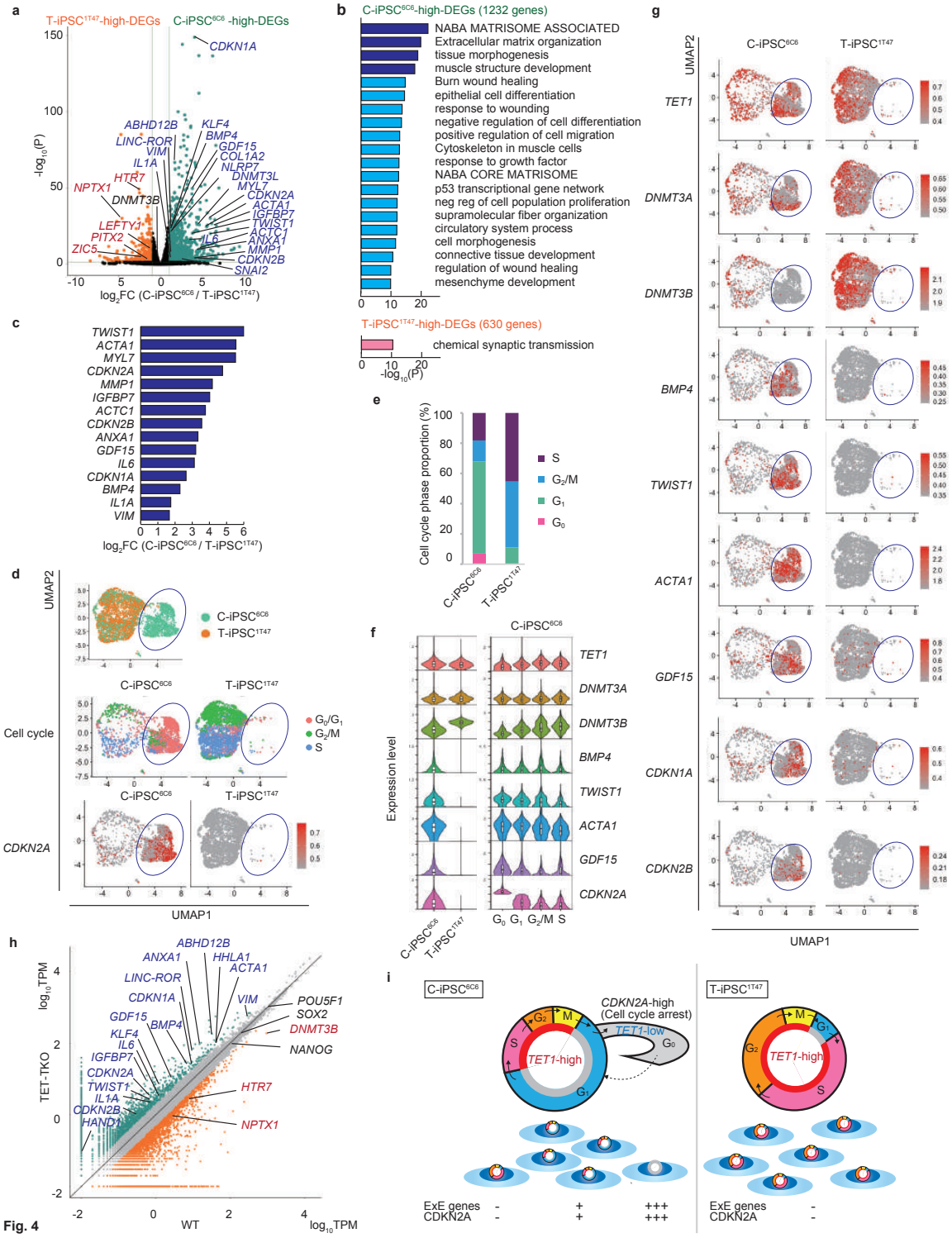


Fig. 4: Characterization of intraclonal heterogeneities in C-iPSCs. a, Volcano plot showing the differentially expressed genes (DEGs) between C-iPSC^{6C6} and T-iPSC^{1T47} in the

undifferentiated state. Green and orange dots indicate genes with a $\log_2$(fold change (FC)) > 1.0 in C-iPSC^{6C6} and T-iPSC^{1T47} ($P < 0.05$), respectively. Genes marked in blue and red indicate upregulated genes in C-iPSC^{6C6} and T-iPSC^{1T47}, respectively. **b**, GO terms of the DEG sets between C-iPSC^{6C6} and T-iPSC^{1T47}. The upper and lower panels show the GO term(s) for C-iPSC^{6C6}-high-DEGs and T-iPSC^{1T47}-high-DEGs, respectively. GO terms with $-\log_{10}$(P) > 10 are listed. **c**, Bar plots of the relative gene expression of the mesenchyme-associated and aging-associated C-iPSC^{6C6}-high-DEGs. **d**, UMAP plot showing C-iPSC^{6C6} and T-iPSC^{1T47} scRNA-seq data. The upper panel shows the distribution of C-iPSC^{6C6} and T-iPSC^{1T47} cells. The middle panel shows the cell cycle phase distribution in C-iPSC^{6C6} and T-iPSC^{1T47}, determined using Seurat's default settings. The lower panel shows *CDKN2A* gene expression. The blue circles indicate the C-iPSC^{6C6}-specific population. **e**, Bar plots showing the proportion of the cell cycle phases in C-iPSC^{6C6} and T-iPSC^{1T47}. **f**, Violin plots showing the distribution of the gene expression levels of the regulatory genes for DNA methylation, extraembryonic-associated genes, and aging-associated genes. The left panel shows the distribution of the gene expression levels in C-iPSC^{6C6} and T-iPSC^{1T47}. The right panel shows the gene expression levels at each cell cycle phase in C-iPSC^{6C6}. **g**, UMAP plot showing the expression of the regulatory genes for DNA methylation, extraembryonic-associated genes, and aging-associated genes in C-iPSC^{6C6} (left) and T-iPSC^{1T47} (right). The blue circles indicate the C-iPSC^{6C6}-specific population. **h**, Correlation of gene expression between the wild-type hESC HUES8

(WT) and TET1/2/3 triple-knockout cells (TET-TKO)[50]. Genes with more than two-fold differences between WT and TET-TKO are highlighted in orange (upregulated in WT) and in green (upregulated in TET-TKO). Genes marked in blue or red indicate upregulated genes in C-iPSC^{6C6} or T-iPSC^{1T47}, respectively. Genes were plotted by $\log_{10}$(TPM) in WT and TET-TKO. *HHLA1*, *ABHD12B*, and *LINC-ROR* are previously assigned differentiation-defective markers[2]. TPM: Transcripts Per Million. **i,** The cell cycle affects *TET1* and extraembryonic (ExE) gene expression. C-iPSC^{6C6} exhibits a pronounced cell cycle arrest in the G_0/G_1 phase, whereas T-iPSC^{1T47} shuttles between S and G_2/M. The G_0/G_1 cells express lower levels of *TET1* (*TET1*-low/*CDKN2A*-high) and are the sources of ExE gene expression.

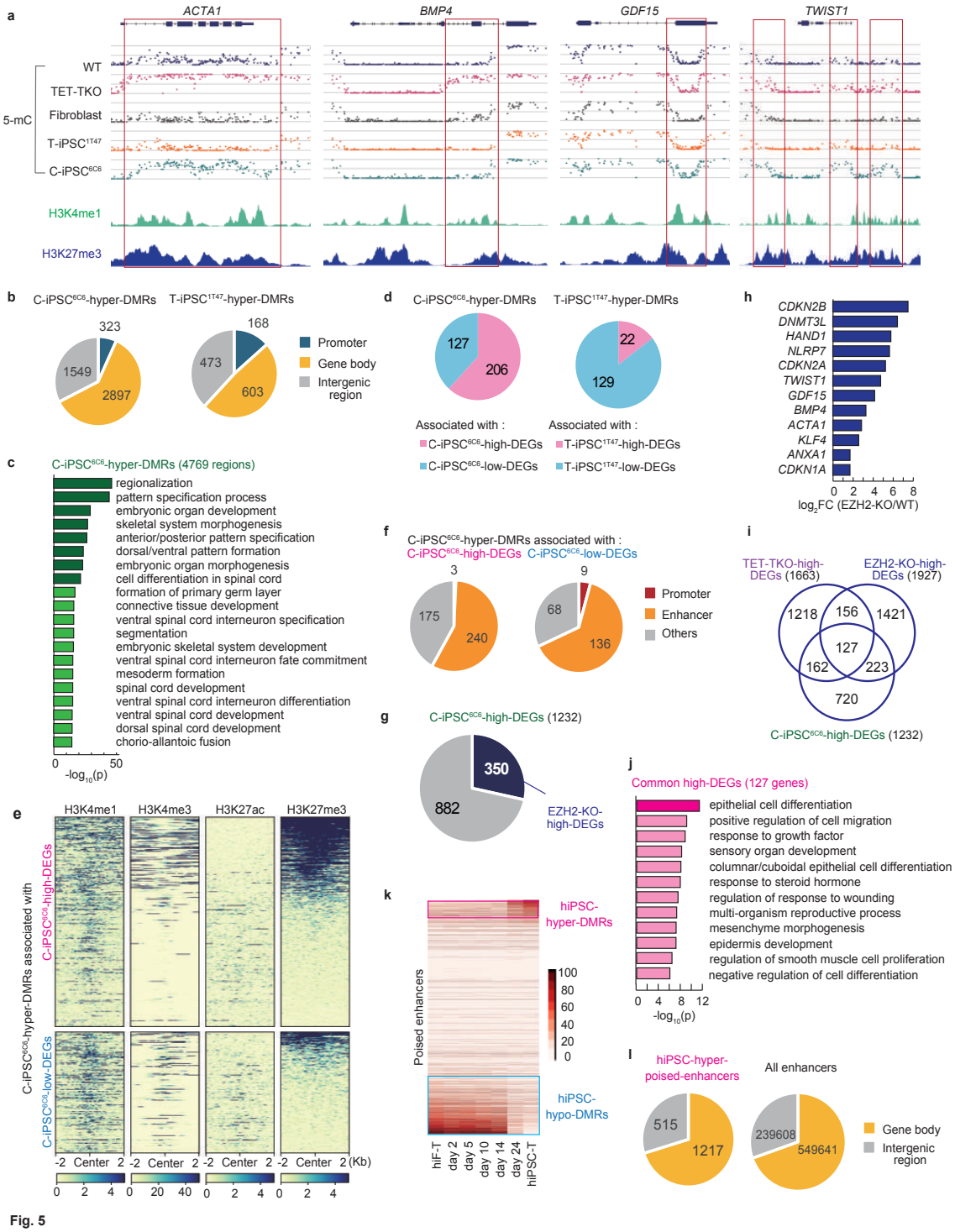


Fig. 5

Fig. 5: Hypermethylation of poised enhancers on extraembryonic gene loci caused by TET insufficiencies. a, Genome browser tracks for the WGBS data (upper panel, 5-mC) and ChIP-seq data (lower panel) of extraembryonic genes in hESCs, TET-TKO, fibroblasts, T-

iPSC^{1T47}, and C-iPSC^{6C6}. The ChIP-seq reads for H3K4me1 and H3K27me3 were obtained from hESC H1. DMRs between C-iPSC^{6C6} and T-iPSC^{1T47} are indicated by red rectangles. WT, hESC HUES8; TET-TKO, TET1/2/3-triple-knockout HUES8. **b**, Distribution of the hyper-DMRs in the genomes of C-iPSC^{6C6} and T-iPSC^{1T47}. **c**, Functional enrichment of the genes related to the C-iPSC^{6C6}-hyper-DMRs using GREAT. **d**, DMRs between C-iPSC^{6C6} and T-iPSC^{1T47} associated with DEGs between C-iPSC^{6C6} and T-iPSC^{1T47}. The pink and blue sections represent the DMRs associated with upregulated (high-DEGs) and downregulated (low-DEGs) genes, respectively. **e**, Heatmap of ChIP-seq signals of histone modifications (hESC H1) in C-iPSC^{6C6}-hyper-DMRs associated with the DEGs between C-iPSC^{6C6} and T-iPSC^{1T47}. The rows are sorted in the order of H3K27me3-densities. **f**, C-iPSC^{6C6}-hyper-DMRs associated with DEGs between C-iPSC^{6C6} and T-iPSC^{1T47} on the cis-regulatory elements. **g**, EZH2-KO-high-DEGs that are also C-iPSC^{6C6}-high-DEGs. The gene expression profile of EZH2-KO was obtained from a previous study[59]. EZH2-KO-high-DEGs were identified by the same threshold as C-iPSC^{6C6}-high-DEGs ($\log_2$(FC) > 1.0 and $P < 0.05$). A significant proportion of genes upregulated in C-iPSC^{6C6} were induced upon the loss of Polycomb repression in EZH2-KO. **h**, Bar plots of selected common gene expression levels to the C-iPSC^{6C6}-high-DEGs and EZH2-KO-high-DEGs[59]. WT, hESC H9; EZH2-KO, EZH2-knockout H9. **i**, Venn diagram showing the overlap of genes between the C-iPSC^{6C6}-high-DEGs, TET-TKO-high-DEGs, and EZH2-KO-high-DEGs. **j**, GO terms of the high-DEGs common to the C-iPSC^{6C6}-high-DEGs, TET-

TKO-high-DEGs, and EZH2-KO-high-DEGs. **k**, Heatmap of the DNA methylation profile at poised enhancers during human reprogramming[16]. Although more poised enhancers lose methylation than those that acquire it, over 1000 enhancers gained methylation during reprogramming. In the comparison between hiF-T and hiPSC-T, hiPSC-hyper-DMRs and hiPSC-hypo-DMRs are indicated in pink and blue, respectively. hiF-T, the TERT-immortalized fibroblast used for reprogramming; hiPSC-T, reprogrammed hiPSC; days 2, 5, 10, 14, and 24, the number of days since the start of hiPSC induction[16]. **l**, HiPSC-hyper-poised-enhancers located in gene bodies and intergenic regions. All enhancers described are only those on autosomes. According to our criteria, this RRBS data covers about 40% of the poised enhancers on autosomes detectable by WGBS, suggesting the potential for a greater number of poised enhancers to be highly methylated.

Methods

HiPSC induction and culture

Human fibroblasts (from the 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 using a previously reported method with episomal vectors[23]. Episomal vectors (Supplementary Table 4) 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 alongside the three episomal vectors upon reprogramming. The cells (between passages 3 and 9) were transfected using a 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. The next day, the culture medium was replaced with Essential 8™ Medium (E8M; Gibco) supplemented with 100 nM hydrocortisone (Sigma-Aldrich) and 1× P/S. E8M, which contains TGF-β, was selected because of its potentiating effect of activating SMAD2/3 during reprogramming[80]. The medium was changed every other day. E8M

without hydrocortisone was used from day ten of iPSC induction. The iPSC colonies were picked up, triturated into E8M supplemented with 10 μM Y-27632 (FUJIFILM Wako), and plated onto STO feeder cells (RIKEN BRC, RCB0536; RRID: CVCL_3420) treated with mitomycin C (FUJIFILM Wako). The culture medium was replaced the following day with KnockOut™ DMEM/F12 (Thermo Fisher Scientific) supplemented with 20% KnockOut™ Serum Replacement (KSR, Gibco), 2 mM GlutaMAX™-I (Glu; Gibco), 1× MEM non-essential amino acids (NEAA, Gibco), 0.1 mM β-mercaptoethanol (2-ME; FUJIFILM Wako), 5 ng/mL fibroblast growth factor 2 (FGF2; Peprotech), 5 mM NaOH (FUJIFILM Wako), and 1× P/S. After the colonies had been picked, the medium was changed daily. Grown iPSC colonies were dissociated with 0.01% trypsin/0.1 mM EDTA buffer (trypsin/EDTA; Nacalai Tesque) and plated onto 5 μg/mL vitronectin (Invitrogen)-coated six-well plates in KnockOut™ DMEM/F12 supplemented with 20% KSR, 2 mM Glu, 1× NEAA, 0.1 mM 2-ME, 15 ng/mL FGF2, 10 ng/mL Activin A (R&D Systems), 5 mM NaOH, and 1× P/S (KFA medium) supplemented with 10 μM Y-27632. The culture medium was then replaced daily with the KFA medium without Y-27632. KFA medium was routinely used throughout this study to expand and maintain hiPSCs. We considered clones that reached >90% confluency in a 100-mm dish (Becton Dickinson) without a mesenchymal morphology to be “established.” Half the cells were cryopreserved in cell stock solution (10% dimethyl sulfoxide (Nacalai Tesque) and 90% FBS). Genomic DNA was obtained from the remaining cells to analyze the genomic integration

of the plasmid vectors. On average, colonies required 28 days to form and 17 days for the iPSC clones to be expanded sufficiently to prepare frozen stocks.

Setting a threshold for genomic integration of the episomal vectors

As episomal vectors are electroporated for introduction, genomic integration is inevitable. Although residual reprogramming vectors can negatively influence cell differentiation[1] or potentiate carcinogenesis upon transplantation[81], no explicit cutoff has been proposed for the residual vector when using episomal vectors. Reanalysis with quantitative polymerase chain reaction (qPCR) showed that conventional PCR only detected clones with >0.16 plasmid copies/genome as positive. The mean copy number of conventional PCR-deduced "vector-free" clones was 0.0062 ± 0.0067 per genome (mean ± standard deviation; Supplementary Fig. 1a). Using a default differentiation paradigm (Fig. 2), we found that a residual episomal copy number of <0.03 per genome did not influence the differentiation output (Supplementary Fig. 1b). We thus set the vector-free threshold to 0.03.

Differentiation of hiPSCs by default differentiation

Pilot experiments using hiPSCs and attempts to derive neural cells from monkey ESCs[29] revealed that KSR-containing culture media are incompatible with default differentiation.

Therefore, we gradually changed the KSR medium to a chemically defined medium (CDM) before differentiation to facilitate the efficient derivation of neural cells[29]. 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 FGF2, 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 (Thermo Fisher Scientific), 2 mM Glu, 7 μg/mL Insulin (FUJIFILM Wako), 15 μg/mL transferrin (Sigma-Aldrich), 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, 1 × NEAA, 80 nM sodium selenite (Sigma-Aldrich), and 1× P/S. Half the media was carefully replaced with CFA without Y-27632 daily. On the third day of replication, cells were dissociated with trypsin/EDTA and plated at 500 cells per well in a 96-well low-binding Nunclon™ Sphera™ plate (Thermo Fisher Scientific) in 150 μL of CDM supplemented with 10 μM Y-27632. The day on which cells were seeded was designated 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; the medium was changed every three days using the same procedure as on day 5. The induced aggregates were collected and analyzed on differentiation days 8 and 18.

Differentiation of ventral midbrain neural progenitor cells

Three different media were used to induce VM neural progenitors. The first medium tested was a differentiation medium (DM) composed of DMEM (Nacalai Tesque), 15% KSR, 2 mM Glu, 1 mM sodium pyruvate (Sigma-Aldrich), 10 μM 2-ME, and 0.1× P/S. The second medium tested was N2 medium (N2M), composed of DMEM/F12 (FUJIFILM Wako), 1× N-2 Supplement (Gibco), 2 mM Glu, 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× B-27™ Supplement (Gibco), 2 mM Glu, and 0.1× P/S. Default-elite cell lines were cultured and passaged two or three times in the KFA medium. 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 iMatrix-511 (Laminin; Matrixome) 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 on which cells were seeded was designated as differentiation day 0. 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, and 25% DM and 75% N2M, respectively. On differentiation day 11, the medium was changed to N2M supplemented with 0.5× B27. LDN was added daily to the medium from days 0 to 11, SB from days 0 to 4, and C59 from days 0 to 2. Ventralization was initiated by adding 100 ng/mL recombinant Sonic

Hedgehog/Shh (C24II) (R&D Systems) and 2 μM purmorphamine (FUJIFILM Wako) from days 3 to 8. Furthermore, 3 μM CHIR99021 (CHIR; Sigma-Aldrich) was added to the medium from days 3 to 12. Beginning on differentiation day 13, the medium was changed to NIM supplemented with 10 ng/mL brain-derived neurotrophic factor (BDNF; FUJIFILM Wako), 10 ng/mL glial-derived neurotrophic factor (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 (dbcAMP; Sigma-Aldrich), and 5 μM DAPT (Tocris Bioscience). The cells were dissociated with Accumax (Innovative Cell Technologies, Inc.) on day 16. 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 in NIM supplemented with the above growth factors and 10 μM Y-27632. Two or three days later, half of the medium without Y-27632 was changed. The medium was changed every day from days 0 to 15, and half was changed every two or three days from day 16 onward.

Extraembryonic differentiation of hiPSCs

We based our extraembryonic differentiation protocol on one aiming at inducing the human amnion from hESCs[67]. Our pilot experiments suggested that the induced tissues also included extraembryonic mesenchymal cells, recently confirmed in an assembloid human peri-

implantation model[54]. This differentiation directly followed our feeder-free KFA maintenance culture. At the differentiation start point, cells were trypsin-passaged and replated onto vitronectin-coated dishes in a culture where FGF2 was omitted from KFA and replaced by 50 ng/mL BMP4 (R&D Systems) (with 10 μM Y-27632 for the first 24 h). Differentiation culture was carried out for four days with daily medium changes.

Mesendoderm differentiation of hiPSCs

Early BMP or WNT signaling induces extraembryonic and germline lineages in non-murine mammals[82]. We introduced an adaptation culture step in which both BMP and WNT signaling were inhibited (BMWi). KFA-cultures were trypsin-passaged onto vitronectin-coated dishes, followed by two rounds of BMWi-passaging in CFA containing 200 nM LDN and 1 μM XAV939 (FUJIFILM Wako). From this point, mesoderm and endoderm differentiation were processed independently.

Mesoderm differentiation

The downstream mesoderm differentiation protocol was conducted as per recent protocols inducing skeletal muscle[83,84]. BMWi/CFA-adapted hiPSCs were trypsin-triturated and plated at 2×10^4 cells/cm^2 onto vitronectin-coated dishes in CDM containing 3 ng/mL Activin A, 5 ng/mL BMP4, and 10 μM Y-27632 (MesDiff medium1; induction start point). Twelve hours

later, the medium was changed to MesDiff medium2 (MesDiff medium1 supplemented with 3 μM CHIR). The differentiation medium was changed to MesDiff medium3 (CDM medium containing 3 μM CHIR, 10 μM SB, and 200 nM LDN) 24 h post-induction, and the cells were further cultured for an additional three days with daily medium-changes of MesDiff medium3 until sampling.

Endoderm differentiation

The BMWi/CFA-adapted hiPSCs were trypsin-triturated and plated at 2×10^4 cells/cm^2 onto vitronectin-coated dishes in CDM containing 100 ng/mL Activin A, 5 ng/mL BMP4 and 10 μM Y-27632 (EndDiff medium1; induction start point). Twelve hours later, the medium was changed to EndDiff medium2 (EndDiff medium1 supplemented with 3 μM CHIR). The differentiation medium was changed to EndDiff medium3 (CDM medium containing 100 ng/mL Activin A, 3 μM CHIR, 200 nM LDN, and 10 μM Y-27632) 24 h post-induction. An additional day with a ROCK inhibitor helped cell attachment in this two-dimensional culture. Two days post-induction, CHIR and Y-27632 were omitted from EndDiff medium3 (EndDiff medium4), and the cells were cultured for an additional two days with daily medium-changes of EndDiff medium4 until sampling.

Immunofluorescence of the differentiated cells

On day 18 of default differentiation, the aggregates were washed with phosphate-buffered saline (PBS; TaKaRa) and fixed in 4% paraformaldehyde (PFA; FUJIFILM Wako). The aggregates were embedded in OCT compound (Sakura Finetek), cryosectioned at 10 μm using a Cryostat (Leica Biosystems) and attached to a MAS-coated glass slide (MATSUNAMI). Individual induced cells on days 19, 29, and 42 of VM differentiation and undifferentiated cells were fixed in 4% PFA, washed with PBS, and preserved in PBS at 4°C until analysis.

The samples were blocked in 5% (v/v) normal donkey serum (NDS; Millipore)/0.3% (v/v) Triton X-100 (Nacalai Tesque)/PBS (NDS-TX) for 1 h and then incubated overnight with primary antibodies diluted in 1% NDS-TX at 4°C (Supplementary Table 5). The samples were washed with 0.05% (v/v) Tween-20 (Sigma-Aldrich)/PBS, incubated with secondary antibodies diluted in 1% NDS-TX for 2 h, washed with 0.05% Tween-20/PBS, and incubated with DAPI (Invitrogen) prepared in 1% NDS-TX for 10 min. Finally, the samples were washed with 0.05% Tween-20/PBS and mounted with an anti-photobleaching mounting medium comprising 2.4 g Mowiol (Millipore), 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 (Sigma-Aldrich). The stained cells were imaged using a Biorevo BZ-9000 fluorescence microscope (Keyence). Given the fluorescence filter constraints, we were unable to perform simultaneous three-color immunostaining and nuclear staining simultaneously. 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 using the ImageJ software (ver 1.53v). Statistical significance was determined by a Student's unpaired *t*-test using R software (ver 3.4.3). $P < 0.05$ indicated statistically significant differences.

Evaluating MAP2-positive induced cells

The lengths of the MAP2-positive neurites from the cell body were measured using ImageJ. Error bars represent the standard deviation of three biological replicates. Statistical significance was determined using one-way analysis of variance (ANOVA) and Tukey's multiple-comparison tests using R software. $P < 0.05$ indicated statistically significant differences.

PCR and qPCR analyses of integrated episomal vectors

Genomic DNA (gDNA) was extracted using a DNeasy Blood & Tissue Kit (QIAGEN). Conventional (semi-quantitative) PCR was performed using Ex Taq DNA polymerase (TaKaRa). Quantitative PCR (qPCR) was performed using Thunderbird SYBR qPCR Mix (Toyobo) and CFX Connect Real-Time System (Bio-Rad). The cycle threshold (Ct) values of the episomal vector were normalized against gDNA (measured using the CFTR primer set). The primer sequences for gDNA detection are listed in Table S3. The pEP4-SF1 and pEP4-

SR1 primers (Supplementary Table 4) were used to detect the episomal vector backbones for semi-quantitative PCR[85]. For qPCR, the pEP4-SR1 primer was replaced with the pEP4-SR3 primer, designed in-house (Supplementary Table 4). The template for conventional PCR and qPCR was 50 ng and 20 ng gDNA, respectively. gDNA extracted from male human dermal fibroblasts was used as a negative control.

hiPSC karyotyping

Nihon Gene Research Laboratories, Inc. (Japan) karyotyped the C-iPSC and T-iPSC lines by performing Giemsa testing.

Reverse transcription quantitative PCR (RT-qPCR) analysis of differentiated cells

Total RNA was purified from the samples using an RNeasy Plus Mini Kit (QIAGEN). Over 50 ng of total RNA was reverse transcribed using ReverTra Ace (Toyobo) and oligo-dT primers (Toyobo). qPCR analysis was performed using Thunderbird SYBR qPCR Mix and a CFX Connect Real-Time System. The following thermocycling conditions were used for qPCR: 95°C for 30 s, 45 cycles of 95°C for 5 s, and 60°C for 30 s. The data were assessed using the delta-Ct method and normalized to glyceraldehyde-3-phosphate dehydrogenase (*GAPDH*) expression. Error bars represent the standard deviation of the biological replicates. The

sequences of the primers used for qPCR are listed in Supplementary Table 4. Statistical significance was determined using Student's unpaired *t*-test or one-way ANOVA and Tukey's multiple-comparison tests using R software. $P < 0.05$ indicated statistically significant differences.

Correlation analysis

Statistical analyses were performed using R software. 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 the two groups. $P < 0.05$ indicated statistically significant differences.

Evaluating standards for default-elite clones

Considering the conducted correlation analyses of differentiation markers and differentiation efficiencies between 1C6, 1T35 and 1T47, we set our default-elite thresholds to *HES5*/*GAPDH* > 0.002 and *TBXT*/*GAPDH* < 0.0001 detected by qPCR.

RNA-seq processing

RNA-seq data analyses were performed for T-iPSC^{1T47} and C-iPSC^{6C6} cultured in an undifferentiated culture condition (KFA medium). Total RNA was purified using a miRNeasy Mini Kit according to the manufacturer's instructions. Then, the library preparation and sequencing were performed by Macrogen Japan Corp. Briefly, RNA-seq libraries were prepared using a TruSeq Stranded Total RNA LT Sample Prep Kit with Ribo-Zero Human/Mouse/Rat (Illumina). By harvesting cells at different passage numbers, we generated biological triplicate data for T-iPSC^{1T47} and C-iPSC^{6C6}. The RNA-seq libraries were sequenced on an Illumina NovaSeq 6000, generating 100-base paired-end reads.

RNA-seq data analysis

RNA-seq reads were aligned to the human reference genome (GRCh38/hg38) using STAR software (ver 2.7.0a) with default parameters[86]. We calculated the transcripts per million mapped reads (TPM) using RSEM (ver 1.3.1) with default parameters[87] and considered their expression levels. The DEGs between T-iPSC^{1T47} and C-iPSC^{6C6}, as well as between EZH2-KO and WT hESCs, were calculated using the R package DESeq2 (ver 1.38.3)[88] with the cutoffs of absolute log2(FC) > 1 and adjusted *P* value < 0.05. The DEGs between TET-TKO and WT hESCs were calculated using a cutoff of absolute log2(FC) > 1. The WT HUES8

(SRX5496558), TET-TKO (SRX5496559)[50], WT H9 (SRX1521198, SRX1521199 and SRX1521200), and EZH2-KO (SRX1521204, SRX1521205 and SRX1521206)[59] data were re-mapped to the human reference genome (GRCh38/hg38) and analyzed as mentioned above. We only considered autosomal chromosomes for RNA-seq data analysis.

Isolation and RNA-sequencing of single cells

RNA-sequencing of single cells was performed for T-iPSC^{1T47} and C-iPSC^{6C6} cultured in an undifferentiated culture condition (KFA medium). Azenta Life Sciences performed library preparation and sequencing. Briefly, single-cell RNA-sequencing (scRNA-seq) was performed using the 10× Chromium platform and Chromium 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.

Single-cell RNA-seq data analysis

Alignment of reads to the human reference genome (GRCh38/hg38) and the preparation of the raw gene expression matrix were conducted using the Cell Ranger (ver 3.0.1) pipeline[89]. scRNA-seq integration analysis was performed using the Seurat package (ver 4.0.0)[90]. C-iPSC^{6C6} cells and T-iPSC^{1T47} cells were excluded if they contained a portion of mitochondrial

reads greater than 20% or had fewer than 200 genes detected. The resulting matrix contained 2119 C-iPSC^{6C6} and 3249 T-iPSC^{1T47}. Unique molecular identifier (UMI) counts were normalized, followed by variable gene identification, anchor identification, data integration, and data scaling with the default settings of the Seurat package. The cell cycle phases (G_0/G_1, S, and G_2/M) were determined using Seurat's default settings. Cells in the G_0 phase were selected from the G_0/G_1 phase subset based on *CDKN2A* expression (>1). For Uniform Manifold Approximation and Projection (UMAP), we used the first 15 principal component analysis dimensions with the default settings using the RunUMAP function. Gene expression profiles were visualized on a UMAP plot using the FeaturePlot function (min.cutoff = q50, max.cutoff = q80). For clustering, we used a resolution of 0.5 with the default settings.

To reanalyze scRNA-seq data from peri-gastruloids (GSE232861), we conducted integration with scRNA-seq data from T-iPSC^{1T47}, C-iPSC^{6C6}, and hESC H9 (E-MTAB-6819). Cells were excluded if they had fewer than 200 genes detected. We randomly selected 100 cells from each of the T-iPSC^{1T47} and C-iPSC^{6C6} datasets, and then merged them. Additionally, we selected 200 cells cultured in primed culture conditions from the hESC H9 data. The resulting matrix contained 200 cells (C-iPSC^{6C6} and T-iPSC^{1T47}), 200 cells (hESC H9), and 12,008 cells (peri-gastruloids). UMI counts were normalized, followed by variable gene identification with the default settings of the Seurat package. For integration with peri-gastruloids, merged hiPSCs, and hESCs in primed culture condition, we identified anchors using the default settings, except

for anchor.features = 1500 and k.filter = 10. Integration and data scaling were performed with the default settings. For t-distributed stochastic neighbor embedding (tSNE), we used the first 14 principal component analysis dimensions with the default settings using the RunTSNE function. For clustering, we used a resolution of 0.7 with the default settings using the FindClusters function and identified tissue-specific markers expressed in the clusters using the FindAllMarkers function with the default settings.

The scRNA-seq data of the human reprogramming process[20] and the cultured monkey blastocysts[38] were reanalyzed using the interactive online tools.

Microarray data reanalysis

The normalized signal intensity data for 49 hiPSCs (GSE42455) were downloaded from NCBI by the R package GEOquery (ver 2.62.2)[91]. The normalized values were converted to the original numbers to calculate the coefficient of variation.

Chromatin immunoprecipitation sequencing (ChIP-seq) data analysis

The ENCODE[87] hESC H1 signal *P*-values for H3K4me1 (ENCFF335ZGP), H3K4me3 (ENCFF149EKJ), H3K27ac (ENCFF771GNB), and H3K27me3 (ENCFF578BZC) were used for characterization of the DMRs. We also used the ENCODE[92] hiPSC19.11 signal P-values

for H3K4me1 (ENCFF394MDE), H3K4me3 (ENCFF678JTM), H3K27ac (ENCFF976IQG), and H3K27me3 (ENCFF776DCF). Genome tracks were created using the pygenometracks (ver 3.8)[93] to visualize CpG-specific methylation levels and histone ChIP-seq peaks throughout the genome. Heatmaps of the histone marks were created using deeptools (ver 2.0)[94].

Characterization of promoters, enhancers, and gene bodies.

BED files for promoter-like signature (PLS) and enhancer-like signature (ELS) were downloaded from SCREEN (ENCODE3)[95]. Promoter-associated DMRs were identified by intersecting with PLS using bedtools (ver 2.31.0)[96]. DMRs within gene bodies were identified by intersecting non-promoter-associated DMRs with GENCODE V45 genes[97].

WGBS processing

gDNA was extracted using the DNeasy Blood & Tissue Kit according to the manufacturer's instructions. Library preparation and sequencing were then performed by Macrogen Japan Corp. Briefly, bisulfite conversion was carried out using the EZ DNA Methylation-Gold Kit (ZYMO RESEARCH) according to the manufacturer's instructions. The eluted DNA was processed using the Accel-NGS Methyl-seq DNA library kit (Swift Biosciences) following the manufacturer's recommendations. The libraries were sequenced on an Illumina NovaSeq 6000,

generating 150-base paired-end reads. WGBS samples have at least a 99.5% sodium bisulfite conversion rate.

WGBS data analysis

Raw sequencing reads were trimmed using Trim Galore! with the recommended conditions for the Accel-NGS Methyl-seq DNA library kit and aligned to the human reference genome (GRCh38/hg38) using Bismark (ver 0.22.3)[98] with default parameters. Bismark eliminated duplicated reads, and methylation calls were made using Bismark methylation extractor with minimum sequencing depth = 4 at the CpG site. The CpG coverage against the human reference genome (GRCh38/hg38) was 88% for T-iPSC^{1T47} and 87% for C-iPSC^{6C6}. The differentially methylated regions (DMRs) between T-iPSC^{1T47} and C-iPSC^{6C6} were identified using the R package DSS (ver 2.48.0)[99] (≥10% difference of CpG methylation, p.threshold = 0.02, minCG = 15, minlen = 300). Averaged methylation levels of CpGs in each enhancer were calculated using bedtools (ver 2.31.0) [96]. The hypermethylated poised enhancers in C-iPSC^{6C6} compared with fibroblasts were identified using the following criteria: at least 10% higher CpG methylation in C-iPSC^{6C6} and less than 5% higher CpG methylation in T-iPSC^{1T47}, both compared with fibroblasts. We only considered autosomal chromosomes for the DNA methylation analysis.

RRBS data reanalysis

We used RRBS data from a previous study (GSE63419) for a time-series of hiPSCs established using the conventional iPSC reprogramming method to characterize the methylation profiles during reprogramming[16]. Data from hiF-T (the fibroblasts used for reprogramming), d2 (dox+), d5 (dox+), d10 (dox+ and SSEA3+), d14 (dox+ and SSEA3+), d24 (dox− and TRA-1-60+), and hiPSC-T were used for the analysis. Raw sequencing reads were trimmed using Trim Galore! with the recommended conditions for RRBS and aligned to the human reference genome (GRCh38/hg38) using Bismark (ver 0.22.3)[98] with default parameters. Methylation calls were made using Bismark methylation extractor with minimum sequencing depth = 4 at the CpG site and the "--ignore_r2 2" option. The poised enhancers in hPSCs were defined as ELS that overlapped with H3K27me3 ChIP-seq peaks (hESC H1, ENCFF084QDP). Averaged methylation levels of CpGs in each enhancer were calculated using bedtools (ver 2.31.0)[96]. The hiPSC-hyper-poised-enhancers were identified using the criterion of ≥10% hyper-CpG methylation in hiPSC-T compared with hiF-T. Heatmaps of the DNA methylation profiles were created using R software.

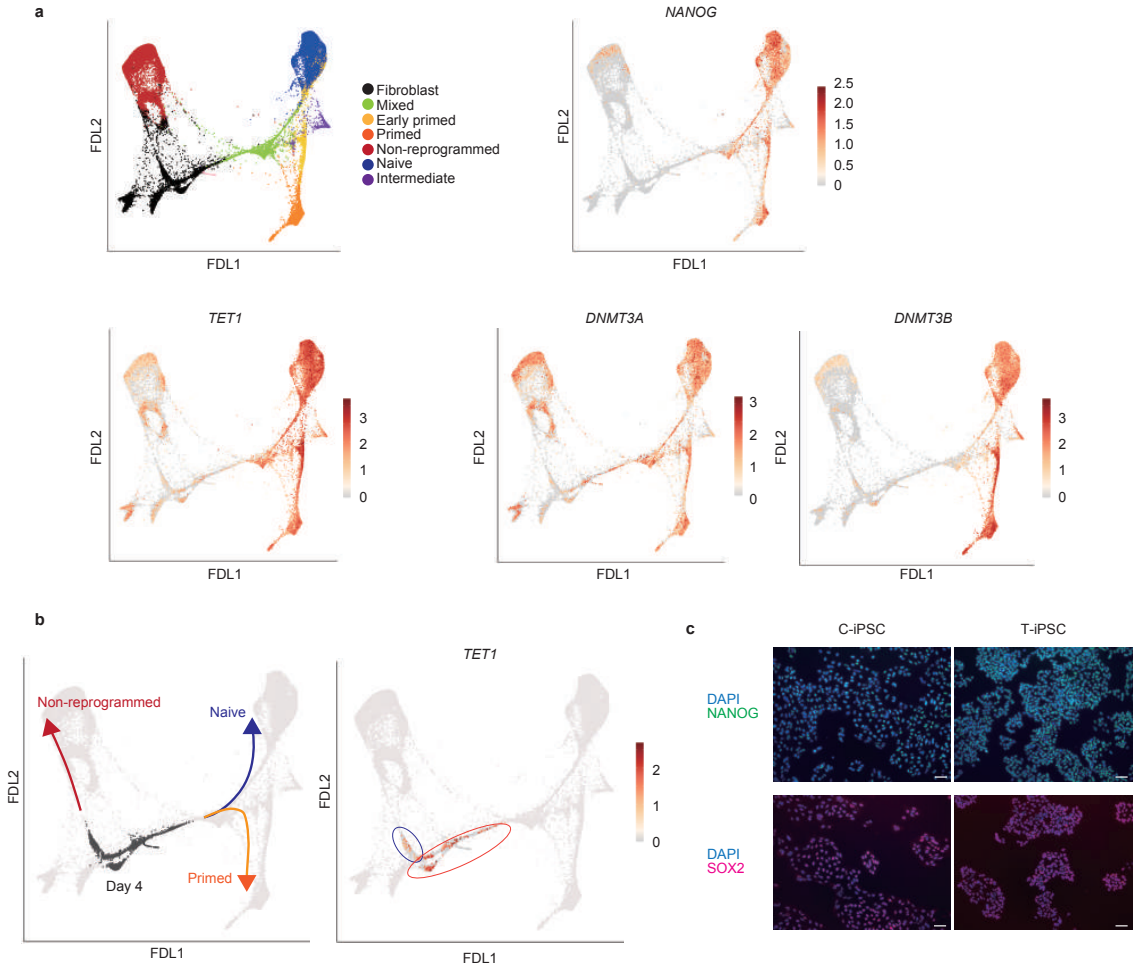
Gene ontology analysis of the DEGs and the DMR neighboring genes

The DEGs were analyzed using Metascape (ver 3.5)[100] for GO analysis. GREAT (ver 4.0.4)[101] was used to find the genes near DMRs and to perform GO analysis. GREAT was run under the "single nearest gene" association strategy. The DMRs associated with the DEGs were selected based on the nearest genes to the DMRs and the DEGs were identified between C-iPSC^{6C6} and T-iPSC^{1T47}.

Data availability

All data supporting the conclusions in the paper are present in the paper or the Supplementary Materials. The sequencing data generated in the present study are available through the DNA Data Bank of Japan (DDBJ), DRA013769 (scRNA-seq), DRA014828 (bulk RNA-seq) and DRA017148 (WGBS).

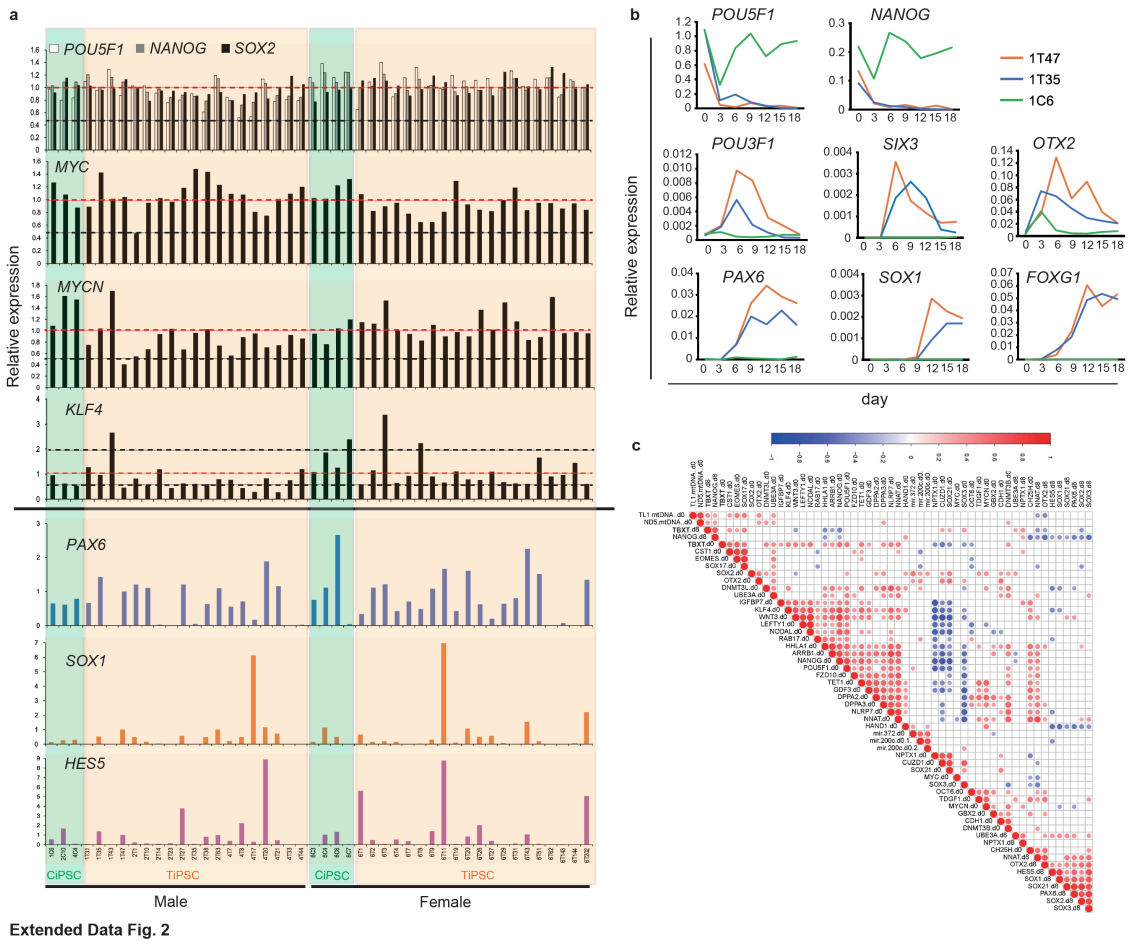
Extended data



Extended Data Fig. 1

Extended Data Fig. 1: Gene expression profiles of *NANOG*, *TET1*, and *DNMT3*s during naive and primed reprogramming from human fibroblasts to iPSCs. a, Force-directed layout (FDL) visualization of the fibroblast, naive, and primed reprogramming scRNA-seq libraries[20]. Naive and primed reprogramming were conducted in bulk until day 8, after which the two trajectories diversified (Extended Data Fig. 1b, left panel). Predicted cell states using defined gene signatures on FDL (upper-left panel). *NANOG*, *TET1,* and *DNMT3A/B* expression

during reprogramming superimposed on FDL plots. *DNMT3B*, a CpG *de novo* methylase, was specifically induced under pluripotency acquisition (compared with the *DNMT3A* expression profile). **b**, FDL visualization of scRNA-seq libraries from the early stage of reprogramming. Black dots showing FDL of the day 4 library. Arrows show the cell trajectories to non-reprogrammed, naive, and primed pluripotent cell states (left panel). Gene expression of *TET1* on day 4 on FDL. Blue and red circles indicate cells projecting toward non-reprogrammed and reprogramming, respectively. *TET1* expression levels within the blue circle are lower than those in the red circle and were soon attenuated in non-reprogramming cells (right panel). **c**, Representative immunostaining data with DAPI (blue), NANOG (green), and SOX2 (red) in established C-iPSCs and T-iPSCs. Scale bars = 50 μm.



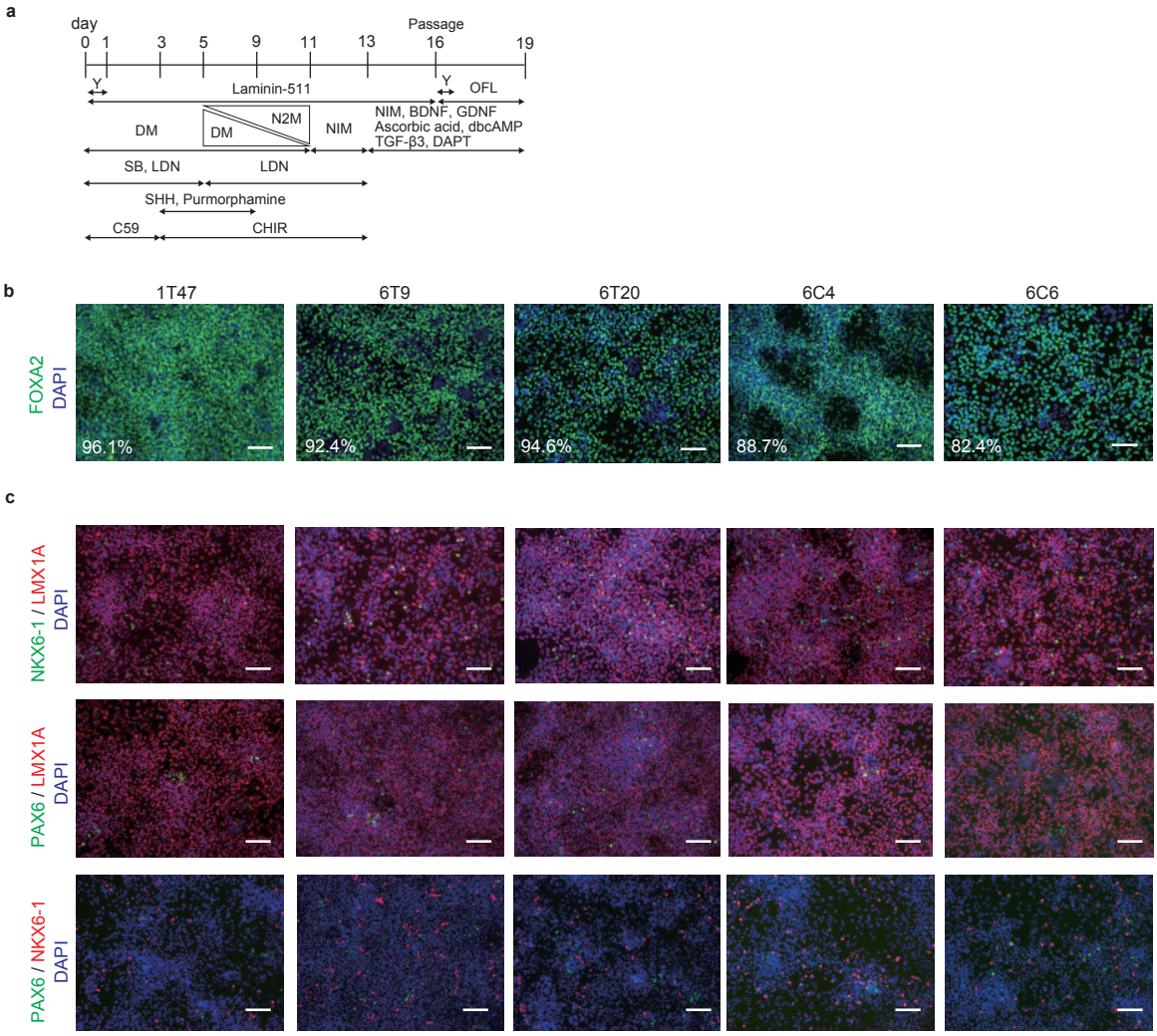
Extended Data Fig. 2: Marker profiling of each C-iPSC and T-iPSC before and after default differentiation. a, Relative expression levels of pluripotency-associated markers on day 0 of default differentiation (upper panel), and neural markers on day 8 of default differentiation (lower panel). Expression levels on day 0 were normalized to the mean value of expression levels in all clones. Expression levels on day 8 were normalized to those for the 1T47 clone. The x-axis shows the clone names. The red and black dotted lines show the average expression levels and half and twice the average, respectively (n = 1). **b**, Changes in gene expression during default differentiation. The x-axis indicates the number of days since the start of default differentiation. Expression levels were normalized to those of *GAPDH*. The

1C6 clone exhibited marked differentiation resistance by maintaining the pluripotent marker expression. Day 8 of default differentiation was the earliest reliable time to assess the differentiation capabilities of various clones (n = 1). **c**, Correlation matrix for marker expression levels and the default differentiation behaviors of all vector-free hiPSC clones. The colored circles indicate all significant correlations ($P < 0.05$) between the two markers. Red and blue ones indicate a positive and negative correlation, respectively. Color intensity and circle size show the correlation strength (R).

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]

Extended Data Fig. 3

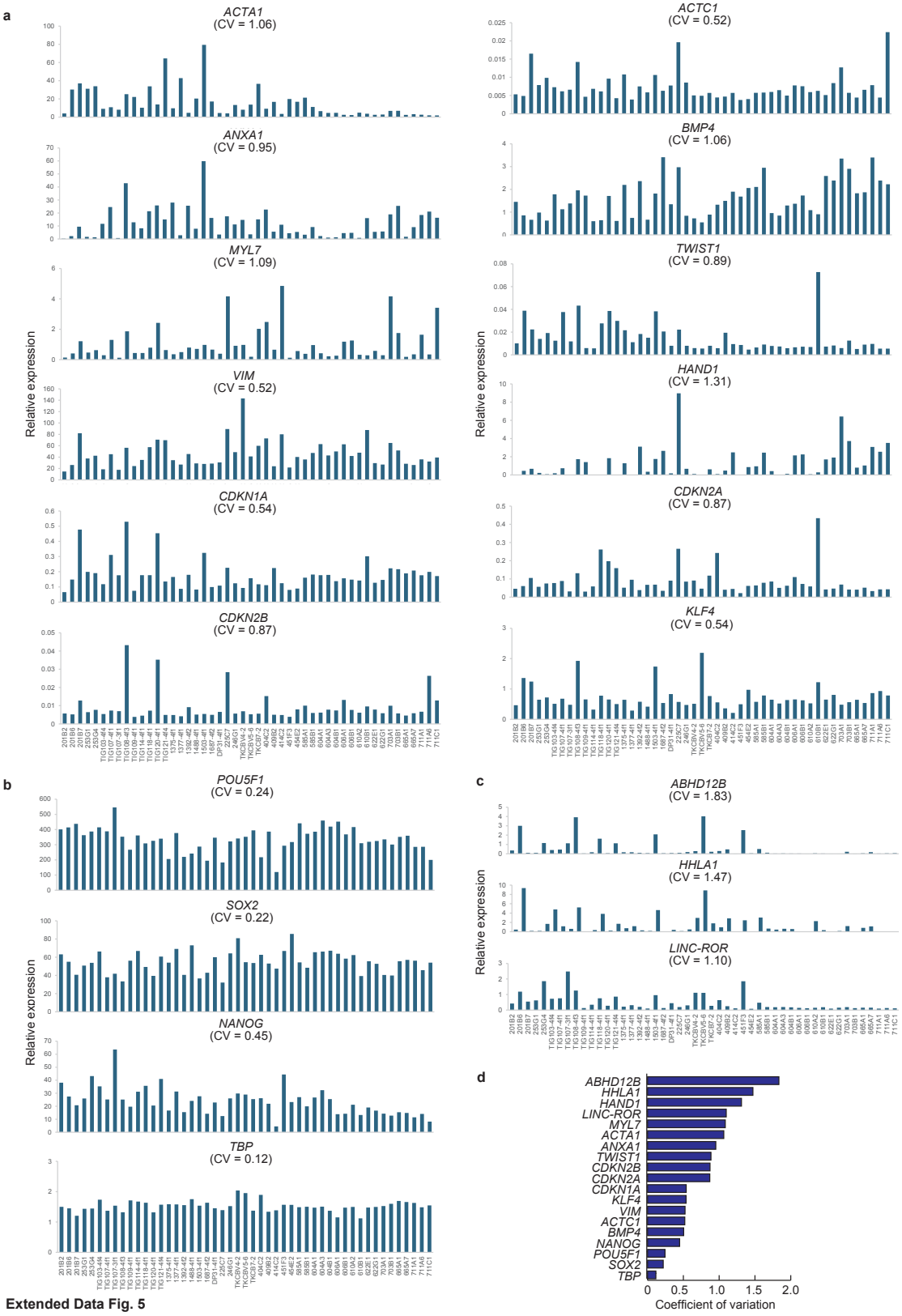
Extended Data Fig. 3: Karyotyping of default-elite clones. The numbers in square brackets represent the number of observed karyotypes among the 20 nuclei counted. +mar, marker chromosome of unknown origin.



Extended Data Fig. 4

Extended Data Fig. 4: Differentiation capabilities of default-elite clones for the VM progenitors. a, Differentiation protocol for the induction of VM progenitors. Triple inhibition of the TGFβ-, BMP-, and WNT-signaling pathways was executed by SB, LDN, and C59, respectively. Y, Y-27632; SB, SB431542; LDN, LDN193189; CHIR, CHIR99021; C59, WNT-C59; dbcAMP, N6,2-O-dibutyryladenosine 3, 5-cyclic monophosphate; OFL, ornithine, fibronectin, and Laminin-511. KSR-medium (DM) was gradually changed to N2-medium (N2M) from days 5 to 11, and NIM (neural induction medium) was changed from day 11

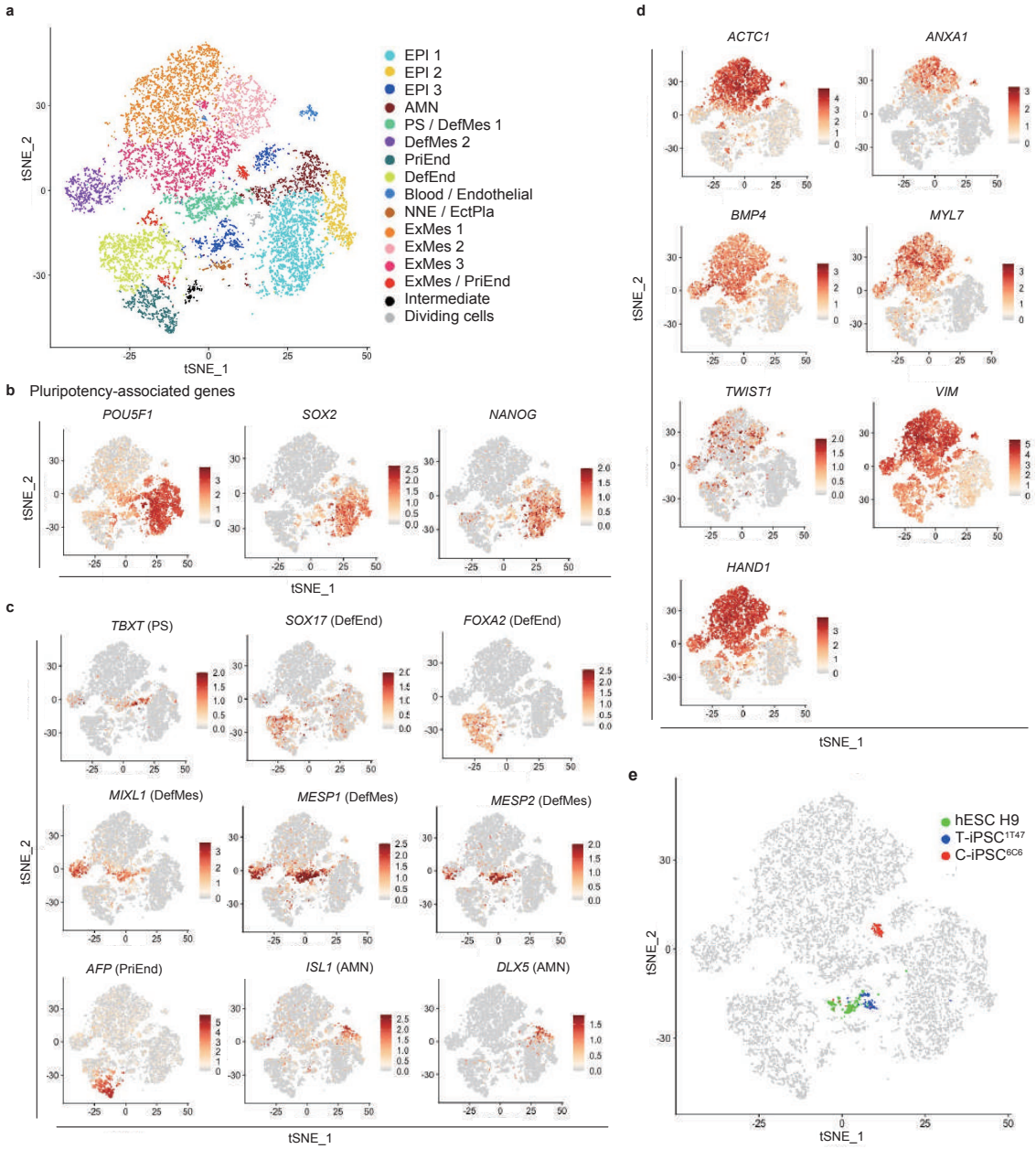
(Methods). **b**, Immunostaining of the VM-region marker, FOXA2, in default-elite clones. The numbers indicate the percentage of positive cells. Blue, DAPI; green, FOXA2. Scale bars, 50 μm. **c**, Immunostaining of other VM-region markers in default-elite clones. Blue, DAPI; green, NKX6-1 or PAX6; red, LMX1A or NKX6-1. Scale bars = 50 μm.



Extended Data Fig. 5

Extended Data Fig. 5: Interclonal heterogeneity of extraembryonic genes and cell cycle

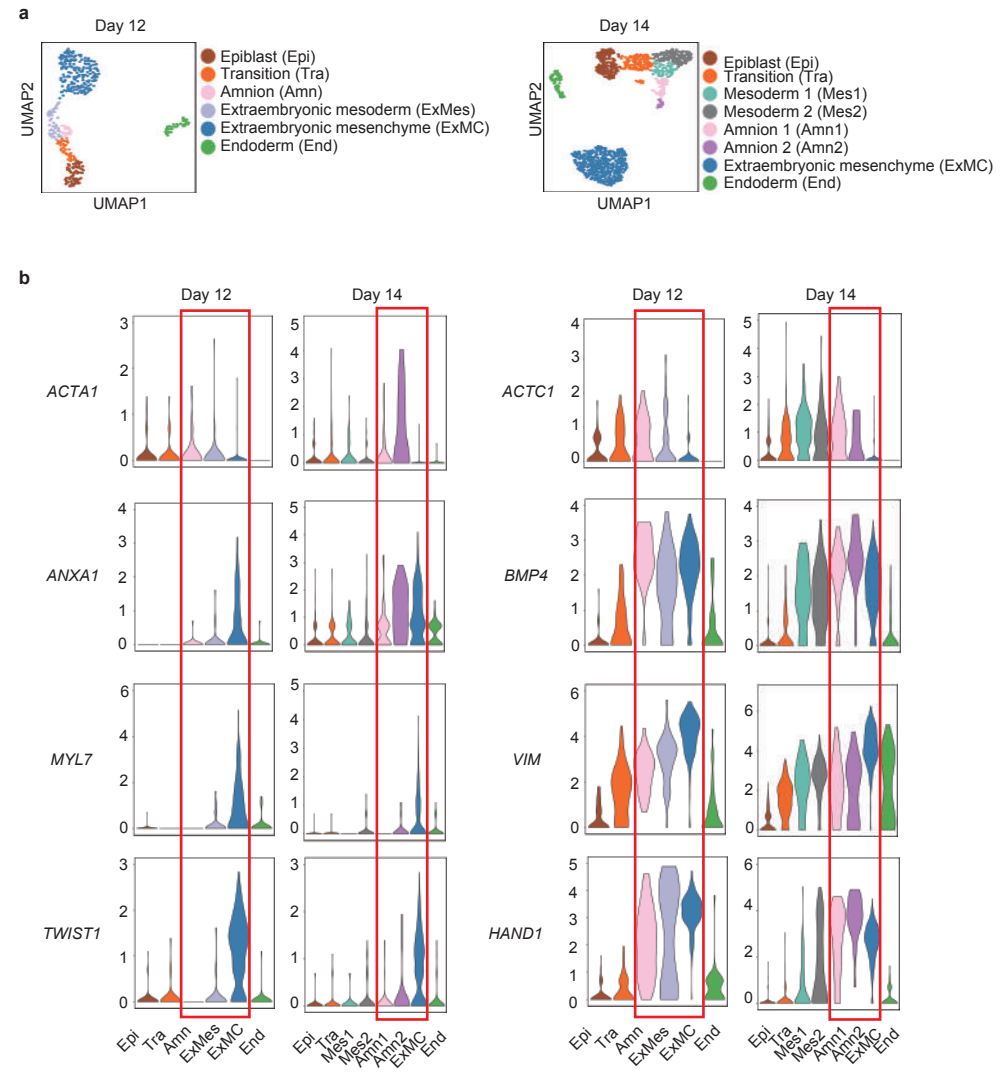
regulators in 49 hiPSCs. a, Bar plots of the relative expression levels of extraembryonic genes and cell cycle regulators. Because the normalized values of microarray data were in logarithmic form, they were converted to the original number to calculate the coefficient of variation (CV). **b**, Bar plots of the relative expression levels of pluripotency-associated genes and the housekeeping gene *TBP*. **c**, Bar plots of the relative expression levels of previously assigned differentiation-defective markers. **d**, Bar plots of the coefficients of variation of gene expression for extraembryonic genes and cell cycle regulators in 49 hiPSCs.



Extended Data Fig. 6

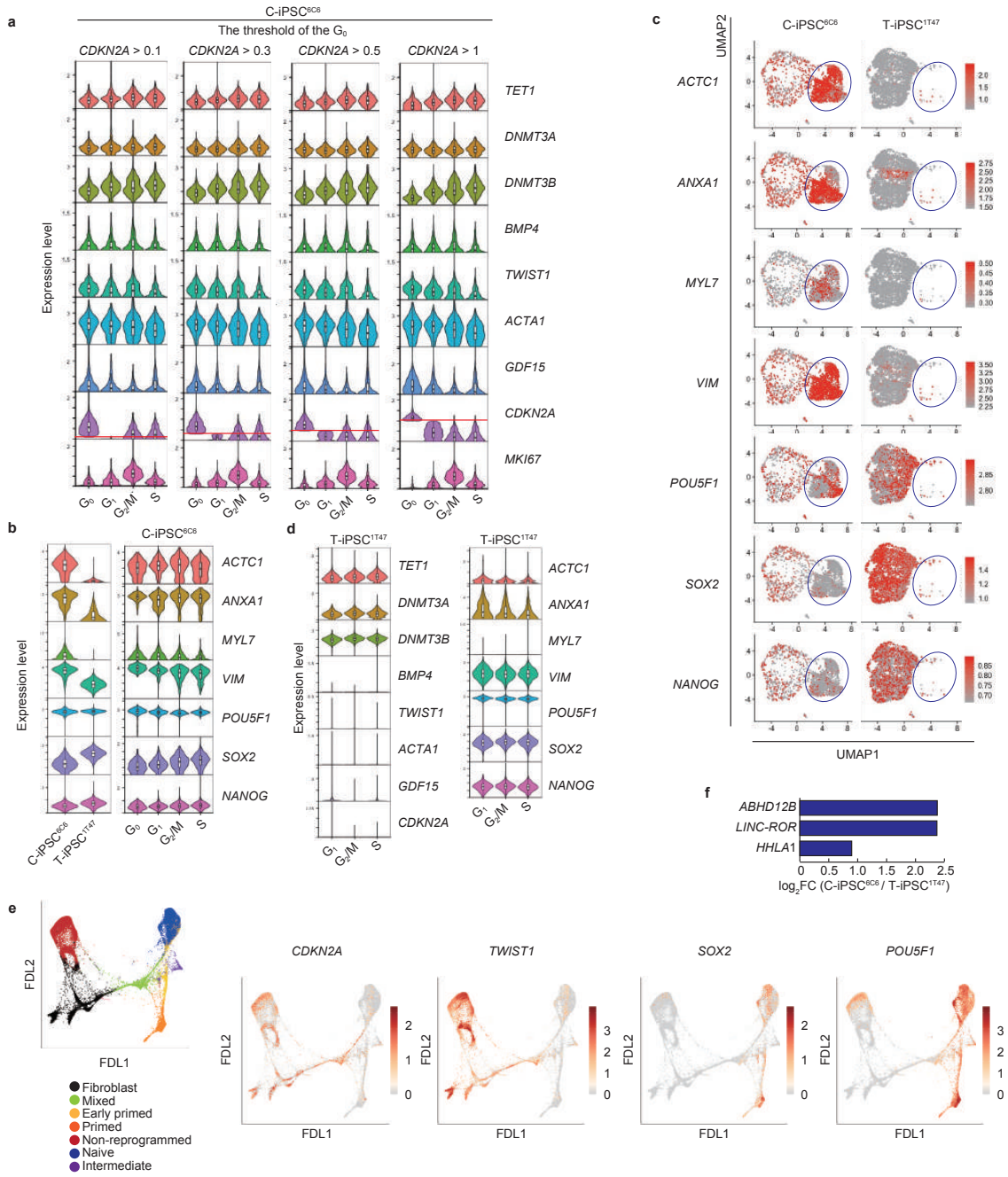
Extended Data Fig. 6: Characterization of mesenchymal genes associated with extraembryonic lineages by reanalysis of human peri-gastruloids. a, Single-cell transcriptomic integration of human peri-gastruloids[48], T-iPSC[1T47], C-iPSC[6C6], and hESC H9[49]. tSNE plot of human peri-gastruloids integrated with T-iPSC[1T47], C-iPSC[6C6], and hESC H9. EPI,

epiblast; AMN, amnion; PS, primitive streak; DefMes, definitive mesoderm; PriEnd, primitive endoderm; DefEnd, definitive endoderm; NNE, non-neural ectoderm; EctPla, ectoderm placode; ExMes, extraembryonic mesoderm. **b**, tSNE plot showing pluripotency-associated gene expression. **c**, tSNE plot showing the gene expression of *TBXT* (PS), *MESP1/2* and *MIXL1* (DefMes), *AFP* (PriEnd), *SOX17* and *FOXA2* (DefEnd), and *ISL1* and *DLX5* (AMN). **d**, tSNE plot showing mesodermal and extraembryonic gene expression. **e**, tSNE plot showing the location of T-iPSC^{1T47}, C-iPSC^{6C6}, and hESC H9.



Extended Data Fig. 7

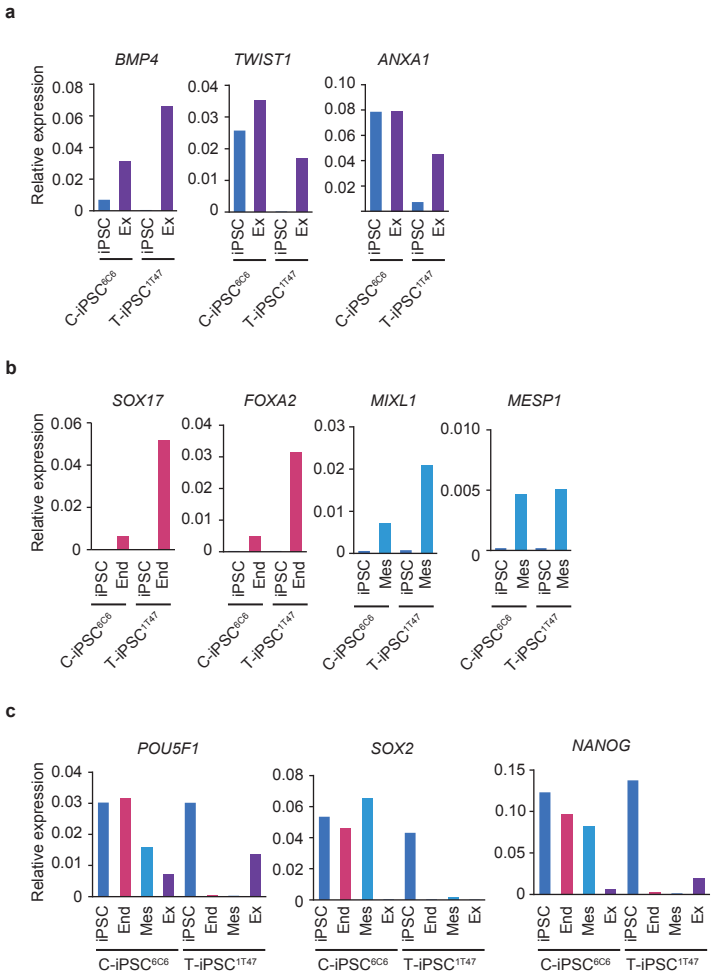
Extended Data Fig. 7: C-iPSC^{6C6}-high-DEGs are expressed in the extraembryonic lineages during early monkey development. a, UMAP plot of cells from the *in vitro* cultured monkey embryos on days 12 and 14. Each cell type is indicated by its coloring[38]. Day-12 monkey embryos are phenotypically comparable to hPSCs. **b**, Violin plot of the relative expression levels of extraembryonic-associated genes. Red rectangles surround the extraembryonic lineage plots.



Extended Data Fig. 8

Extended Data Fig. 8: G_0 phase annotation and characteristics of C-iPSC^{6C6} and T-iPSC^{1T47} in each cell cycle. a, Violin plots showing the distribution of gene expression levels of the regulatory genes for DNA methylation, extraembryonic, aging-, and cell cycle-associated

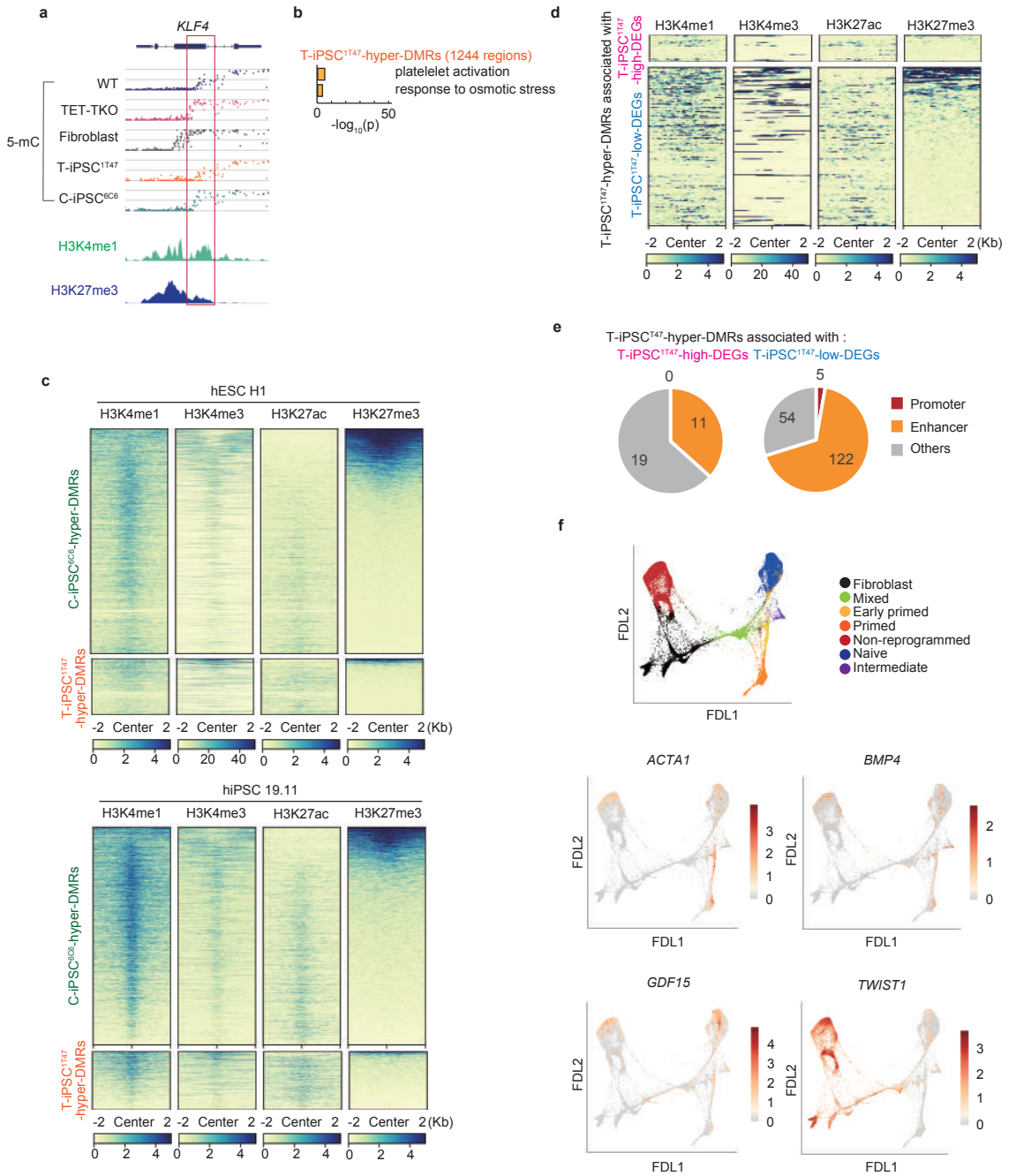
genes in each cell cycle phase in C-iPSC^{6C6}. Each panel represents the differences in the distribution of gene expression levels for each factor when different expression levels of *CDKN2A* define the G_0 phase. We defined *CDKN2A* > 1 as the G_0 phase, as cells expressing *CDKN2A* > 1 are rarely observed in the G_2/M or S phases. **b**, Violin plots showing the distribution of gene expression levels of extraembryonic and pluripotency-associated genes. The left panel shows the distribution of the gene expression levels in C-iPSC^{6C6} and T-iPSC^{1T47}. The right panel shows the gene expression levels in each cell cycle phase in C-iPSC^{6C6}. **c**, UMAP plot showing the gene expression levels of extraembryonic- and pluripotency-associated genes in C-iPSC^{6C6} and T-iPSC^{1T47}. Blue circles indicate the C-iPSC^{6C6}-specific population. **d**, Violin plots showing the distribution of the regulatory genes for DNA methylation, extraembryonic, aging-, and pluripotency-associated genes in each cell cycle phase in T-iPSC^{1T47}. **e**, The left panel represents a duplication of the FDL visualization of the fibroblasts, naive, and primed reprogramming scRNA-seq libraries shown in Extended Data Fig. 1a[20]. The remaining panels represent *CDKN2A*, *TWIST1*, *SOX2*, and *POU5F1* expression during reprogramming superimposed on the FDL plots. **f**, Bar plots of the gene expression levels of previously assigned differentiation-defective markers in C-iPSC^{6C6} compared with T-iPSC^{1T47}.



Extended Data Fig. 9

Extended Data Fig. 9: Directed differentiation into the extraembryonic and mesendodermal lineages in C-iPSC^{6C6} and T-iPSC^{1T47}. a, Relative expression levels of extraembryonic-associated genes in C-iPSC^{6C6} and T-iPSC^{1T47} before and after differentiation into the extraembryonic lineage on day 4 (n = 1). iPSC, before differentiation; Ex, after differentiation into the extraembryonic lineage. **b**, Relative expression levels of marker genes after day 4 of directed differentiation into endoderm (End) and mesoderm (Mes) (n = 1). **c**, Relative expression levels of pluripotency-associated genes before and after differentiation into

various lineages (n = 1). Pluripotency-associated markers were retained after differentiation in C-iPSC^{6C6}, indicating differentiation defects.



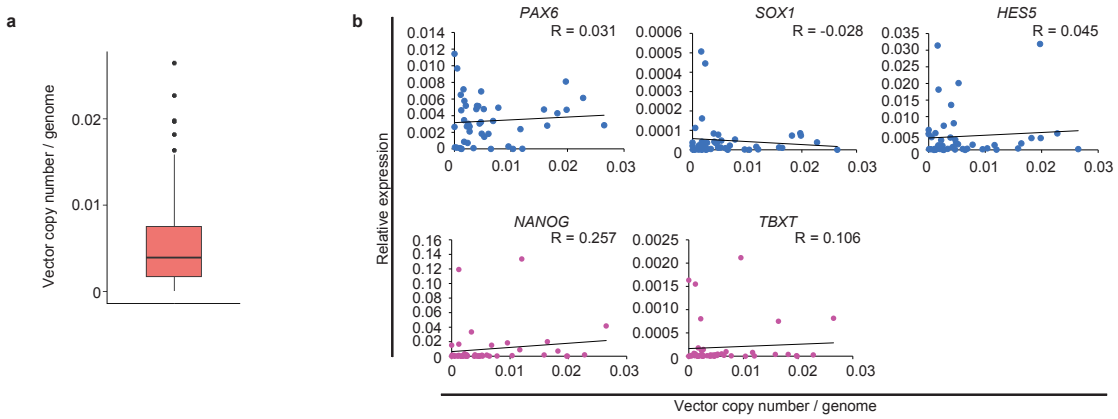
Extended Data Fig. 10

Extended Data Fig. 10: Characterization of DMRs between C-iPSC^{6C6} and T-iPSC^{1T47} and the extraembryonic-associated gene expression dynamics during reprogramming. a, Genome browser tracks for the WGBS data (upper panel, 5-mC) and ChIP-seq data (lower panel) of *KLF4*. The ChIP-seq reads for H3K4me1 and H3K27me3 were obtained from hESC

H1. The DMRs between C-iPSC^{6C6} and T-iPSC^{1T47} are shown by a red rectangle. WT, hESC HUES8; TET-TKO, TET1/2/3-triple-knockout HUES8. **b**, Functional enrichment of the genes associated with the T-iPSC^{1T47}-hyper-DMRs using GREAT. **c**, Heatmap of ChIP-seq signals of histone modifications in C-iPSC^{6C6}-hyper-DMRs and T-iPSC^{1T47}-hyper-DMRs. The upper panel shows histone modifications in hESC H1. The lower panel shows histone modifications in hiPSC 19.11. The rows are sorted in order of the H3K27me3-densities. **d**, Heatmap of the ChIP-seq signals of histone modifications (hESC H1) in T-iPSC^{1T47}-hyper-DMRs associated with the DEGs between C-iPSC^{6C6} and T-iPSC^{1T47}. The rows are sorted in order of the H3K27me3-densities. **e**, Proportion of the T-iPSC^{1T47}-hyper-DMRs associated with the DEGs between C-iPSC^{6C6} and T-iPSC^{1T47} on the cis-regulatory elements. **f**, The upper panel represents a duplication of the FDL visualization of fibroblasts, naive, and primed reprogramming scRNA-seq libraries, as shown in Extended Data Fig. 1a[20]. The remaining panels represent the gene expression dynamics on FDL of C-iPSC^{6C6}-high-DEGs during reprogramming.

Supplementary Information

Supplementary Figure



Supplementary Fig.1

Supplementary Fig. 1: Correlation between the episomal vector copy number and the expression levels of neural markers, *NANOG*, or *TBXT*. a, Box plot showing copy numbers of integrated reprogramming episomal vectors in vector-free clones as called by conventional PCR (n = 46). The genomic DNA qPCR-deduced vector copy numbers are shown as ratios to the halves of the *CTFR* locus counts. **b**, Correlation between the episomal vector copy number and neural markers, *NANOG*, or *TBXT* (n = 46). The correlation coefficients are shown in the figures. The y-axis indicates the relative expression levels normalized to those of *GAPDH*. The x-axis indicates the copy number of integrated reprogramming episomal vectors per genome.

Supplementary Tables

Supplementary Table 1. Ratios of markers in and around the VM region in each differentiation experiment.

Supplementary Table 2. Gene lists of the DMRs associated with DEGs.

Supplementary Table 3. Gene lists of the genes common to the C-iPSC^{6C6}-high-DEGs, the TET-TKO-high-DEGs, and the EZH2-KO-high-DEGs.

Supplementary Table 4. Primer sequences used in this study.

Supplementary Table 5. Primary antibodies used in this study.

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**, 20569-74 (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**, 12426-31 (2014).
3. Bock, C. *et al.* Reference Maps of human ES and iPS cell variation enable high-throughput characterization of pluripotent cell lines. *Cell* **144**, 439-52 (2011).
4. Ohnuki, M. & Takahashi, K. Present and future challenges of induced pluripotent stem cells. *Philos Trans R Soc Lond B Biol Sci* **370**, 20140367 (2015).
5. Esteban, M.A. *et al.* Vitamin C enhances the generation of mouse and human induced pluripotent stem cells. *Cell Stem Cell* **6**, 71-9 (2010).
6. Wang, L. *et al.* Enhanced human somatic cell reprogramming efficiency by fusion of the MYC transactivation domain and OCT4. *Stem Cell Res* **25**, 88-97 (2017).
7. Buckberry, S. *et al.* Transient naive reprogramming corrects hiPS cells functionally and epigenetically. *Nature* **620**, 863-872 (2023).
8. Hou, P. *et al.* Pluripotent stem cells induced from mouse somatic cells by small-molecule compounds. *Science* **341**, 651-4 (2013).
9. Buganim, Y. *et al.* The developmental potential of iPSCs is greatly influenced by reprogramming factor selection. *Cell Stem Cell* **15**, 295-309 (2014).
10. Shu, J. *et al.* Induction of pluripotency in mouse somatic cells with lineage specifiers. *Cell* **153**, 963-75 (2013).

11. Gao, Y. *et al.* Replacement of Oct4 by Tet1 during iPSC induction reveals an important role of DNA methylation and hydroxymethylation in reprogramming. *Cell Stem Cell* **12**, 453-69 (2013).
12. Kim, K.P. *et al.* Permissive epigenomes endow reprogramming competence to transcriptional regulators. *Nat Chem Biol* **17**, 47-56 (2021).
13. 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 e9 (2020).
14. Schiebinger, G. *et al.* Optimal-Transport Analysis of Single-Cell Gene Expression Identifies Developmental Trajectories in Reprogramming. *Cell* **176**, 928-943 e22 (2019).
15. Hu, X. *et al.* Tet and TDG mediate DNA demethylation essential for mesenchymal-to-epithelial transition in somatic cell reprogramming. *Cell Stem Cell* **14**, 512-22 (2014).
16. Cacchiarelli, D. *et al.* Integrative Analyses of Human Reprogramming Reveal Dynamic Nature of Induced Pluripotency. *Cell* **162**, 412-424 (2015).
17. 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**, 204-9 (2015).
18. Kirkeby, A. *et al.* Preclinical quality, safety, and efficacy of a human embryonic stem cell-derived product for the treatment of Parkinson's disease, STEM-PD. *Cell Stem Cell* **30**, 1299-1314 e9 (2023).
19. La Manno, G. *et al.* Molecular Diversity of Midbrain Development in Mouse, Human, and Stem Cells. *Cell* **167**, 566-580 e19 (2016).
20. Liu, X. *et al.* Reprogramming roadmap reveals route to human induced trophoblast stem cells. *Nature* **586**, 101-107 (2020).
21. Cheng, S. *et al.* The intrinsic and extrinsic effects of TET proteins during gastrulation. *Cell* **185**, 3169-3185 e20 (2022).
22. Verma, N. *et al.* TET proteins safeguard bivalent promoters from de novo methylation in human embryonic stem cells. *Nat Genet* **50**, 83-95 (2018).
23. Okita, K. *et al.* A more efficient method to generate integration-free human iPS cells. *Nat Methods* **8**, 409-12 (2011).
24. Chen, G. *et al.* Chemically defined conditions for human iPSC derivation and culture. *Nat Methods* **8**, 424-9 (2011).
25. Brons, I.G. *et al.* Derivation of pluripotent epiblast stem cells from mammalian embryos. *Nature* **448**, 191-5 (2007).
26. Chin, M.H. *et al.* Induced pluripotent stem cells and embryonic stem cells are distinguished by gene expression signatures. *Cell Stem Cell* **5**, 111-23 (2009).
27. Ozair, M.Z., Kintner, C. & Brivanlou, A.H. Neural induction and early patterning in vertebrates. *Wiley Interdiscip Rev Dev Biol* **2**, 479-98 (2013).
28. 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).
29. Ono, T. *et al.* A single-cell and feeder-free culture system for monkey embryonic stem cells. *PLoS One* **9**, e88346 (2014).

30. 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).
31. Chambers, S.M. *et al.* Highly efficient neural conversion of human ES and iPS cells by dual inhibition of SMAD signaling. *Nat Biotechnol* **27**, 275-80 (2009).
32. Zhang, X. *et al.* Pax6 is a human neuroectoderm cell fate determinant. *Cell Stem Cell* **7**, 90-100 (2010).
33. Tchieu, J. *et al.* A Modular Platform for Differentiation of Human PSCs into All Major Ectodermal Lineages. *Cell Stem Cell* **21**, 399-410.e7 (2017).
34. Loh, K.M. & Lim, B. A precarious balance: pluripotency factors as lineage specifiers. *Cell Stem Cell* **8**, 363-9 (2011).
35. Sasaki, K. *et al.* The Germ Cell Fate of Cynomolgus Monkeys Is Specified in the Nascent Amnion. *Dev Cell* **39**, 169-185 (2016).
36. Lister, R. *et al.* Hotspots of aberrant epigenomic reprogramming in human induced pluripotent stem cells. *Nature* **471**, 68-73 (2011).
37. Kim, H. *et al.* miR-371-3 expression predicts neural differentiation propensity in human pluripotent stem cells. *Cell Stem Cell* **8**, 695-706 (2011).
38. Yang, R. *et al.* Amnion signals are essential for mesoderm formation in primates. *Nat Commun* **12**, 5126 (2021).
39. Andersson, E. *et al.* Identification of intrinsic determinants of midbrain dopamine neurons. *Cell* **124**, 393-405 (2006).
40. 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**, 353-72 (2013).
41. Pantazis, C.B. *et al.* A reference human induced pluripotent stem cell line for large-scale collaborative studies. *Cell Stem Cell* **29**, 1685-1702 e22 (2022).
42. Winnier, G., Blessing, M., Labosky, P.A. & Hogan, B.L. Bone morphogenetic protein-4 is required for mesoderm formation and patterning in the mouse. *Genes Dev* **9**, 2105-16 (1995).
43. Pham, T.X.A. *et al.* Modeling human extraembryonic mesoderm cells using naive pluripotent stem cells. *Cell Stem Cell* **29**, 1346-1365 e10 (2022).
44. Conte, M. *et al.* GDF15, an emerging key player in human aging. *Ageing Research Reviews* **75**(2022).
45. Kumari, R. & Jat, P. Mechanisms of Cellular Senescence: Cell Cycle Arrest and Senescence Associated Secretory Phenotype. *Front Cell Dev Biol* **9**, 645593 (2021).
46. Domen, A. *et al.* Cellular senescence in cancer: clinical detection and prognostic implications. *J Exp Clin Cancer Res* **41**, 360 (2022).
47. Nakamura, T. *et al.* A developmental coordinate of pluripotency among mice, monkeys and humans. *Nature* **537**, 57-62 (2016).
48. Liu, L. *et al.* Modeling post-implantation stages of human development into early organogenesis with stem-cell-derived peri-gastruloids. *Cell* **186**, 3776-3792 e16 (2023).
49. Messmer, T. *et al.* Transcriptional Heterogeneity in Naive and Primed Human Pluripotent Stem Cells at Single-Cell Resolution. *Cell Rep* **26**, 815-824 e4 (2019).

50. Charlton, J. *et al.* TETs compete with DNMT3 activity in pluripotent cells at thousands of methylated somatic enhancers. *Nat Genet* **52**, 819-827 (2020).
51. Nakanoh, S. *et al.* Human surface ectoderm and amniotic ectoderm are sequentially specified according to cellular density. *Sci Adv* **10**, eadh7748 (2024).
52. Chhabra, S. & Warmflash, A. BMP-treated human embryonic stem cells transcriptionally resemble amnion cells in the monkey embryo. *Biol Open* **10**(2021).
53. Chhabra, S., Liu, L., Goh, R., Kong, X. & Warmflash, A. Dissecting the dynamics of signaling events in the BMP, WNT, and NODAL cascade during self-organized fate patterning in human gastruloids. *PLoS Biol* **17**, e3000498 (2019).
54. Weatherbee, B.A.T. *et al.* Pluripotent stem cell-derived model of the post-implantation human embryo. *Nature* (2023).
55. Haerinck, J., Goossens, S. & Berx, G. The epithelial-mesenchymal plasticity landscape: principles of design and mechanisms of regulation. *Nat Rev Genet* **24**, 590-609 (2023).
56. Jones, P.A. The DNA methylation paradox. *Trends in Genetics* **15**, 34-37 (1999).
57. Rada-Iglesias, A. *et al.* A unique chromatin signature uncovers early developmental enhancers in humans. *Nature* **470**, 279-83 (2011).
58. Blackledge, N.P., Rose, N.R. & Klose, R.J. Targeting Polycomb systems to regulate gene expression: modifications to a complex story. *Nat Rev Mol Cell Biol* **16**, 643-649 (2015).
59. Collinson, A. *et al.* Deletion of the Polycomb-Group Protein EZH2 Leads to Compromised Self-Renewal and Differentiation Defects in Human Embryonic Stem Cells. *Cell Rep* **17**, 2700-2714 (2016).
60. Ma, H. *et al.* Abnormalities in human pluripotent cells due to reprogramming mechanisms. *Nature* **511**, 177-83 (2014).
61. Roost, M.S. *et al.* DNA methylation and transcriptional trajectories during human development and reprogramming of isogenic pluripotent stem cells. *Nat Commun* **8**, 908 (2017).
62. Jeziorska, D.M. *et al.* DNA methylation of intragenic CpG islands depends on their transcriptional activity during differentiation and disease. *Proc Natl Acad Sci U S A* **114**, E7526-E7535 (2017).
63. Shipony, Z. *et al.* Dynamic and static maintenance of epigenetic memory in pluripotent and somatic cells. *Nature* **513**, 115-9 (2014).
64. Lando, D. *et al.* Enhancer-promoter interactions are reconfigured through the formation of long-range multiway hubs as mouse ES cells exit pluripotency. *Mol Cell* **84**, 1406-1421 e8 (2024).
65. Smith, A. Formative pluripotency: the executive phase in a developmental continuum. *Development* **144**, 365-373 (2017).
66. Smith, Z.D. *et al.* Epigenetic restriction of extraembryonic lineages mirrors the somatic transition to cancer. *Nature* **549**, 543-547 (2017).
67. Zheng, Y. *et al.* Controlled modelling of human epiblast and amnion development using stem cells. *Nature* **573**, 421-425 (2019).
68. Lu, A.T. *et al.* Universal DNA methylation age across mammalian tissues. *Nat Aging* **3**, 1144-1166 (2023).
69. Moqri, M. *et al.* PRC2-AgeIndex as a universal biomarker of aging and rejuvenation. *Nat Commun* **15**,

5956 (2024).

70. Baubec, T. *et al.* Genomic profiling of DNA methyltransferases reveals a role for DNMT3B in genic methylation. *Nature* **520**, 243-7 (2015).

71. Neri, F. *et al.* Intragenic DNA methylation prevents spurious transcription initiation. *Nature* **543**, 72-77 (2017).

72. Chrysanthou, S. *et al.* The DNA dioxygenase Tet1 regulates H3K27 modification and embryonic stem cell biology independent of its catalytic activity. *Nucleic Acids Res* **50**, 3169-3189 (2022).

73. Bernstein, B.E. *et al.* A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* **125**, 315-26 (2006).

74. Lippmann, E.S. *et al.* Deterministic HOX patterning in human pluripotent stem cell-derived neuroectoderm. *Stem Cell Reports* **4**, 632-44 (2015).

75. Burdon, T., Smith, A. & Savatier, P. Signalling, cell cycle and pluripotency in embryonic stem cells. *Trends in Cell Biology* **12**, 432-438 (2002).

76. Eastman, A.E. & Guo, S. The palette of techniques for cell cycle analysis. *FEBS Lett* (2020).

77. Guo, S. *et al.* Nonstochastic reprogramming from a privileged somatic cell state. *Cell* **156**, 649-62 (2014).

78. Shakiba, N. *et al.* Cell competition during reprogramming gives rise to dominant clones. *Science* **364**(2019).

79. Jain, N. *et al.* Retrospective identification of cell-intrinsic factors that mark pluripotency potential in rare somatic cells. *Cell Systems* **15**, 109-133.e10 (2024).

80. 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).

81. Ohnishi, K. *et al.* Premature termination of reprogramming in vivo leads to cancer development through altered epigenetic regulation. *Cell* **156**, 663-77 (2014).

82. Kobayashi, T. *et al.* Principles of early human development and germ cell program from conserved model systems. *Nature* **546**, 416-420 (2017).

83. Xi, H. *et al.* A Human Skeletal Muscle Atlas Identifies the Trajectories of Stem and Progenitor Cells across Development and from Human Pluripotent Stem Cells. *Cell Stem Cell* **27**, 158-176 e10 (2020).

84. Zhao, M. *et al.* Induced Fetal Human Muscle Stem Cells with High Therapeutic Potential in a Mouse Muscular Dystrophy Model. *Stem Cell Reports* **15**, 80-94 (2020).

85. International Stem Cell, I. *et al.* Characterization of human embryonic stem cell lines by the International Stem Cell Initiative. *Nat Biotechnol* **25**, 803-16 (2007).

86. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21 (2013).

87. Li, B. & Dewey, C.N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**, 323 (2011).

88. Love, M.I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550 (2014).

89. Zheng, G.X. *et al.* Massively parallel digital transcriptional profiling of single cells. *Nat Commun* **8**, 14049 (2017).

90. Hao, Y. *et al.* Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573-3587 e29 (2021).
91. Davis, S. & Meltzer, P.S. GEOquery: a bridge between the Gene Expression Omnibus (GEO) and BioConductor. *Bioinformatics* **23**, 1846-7 (2007).
92. Bernstein, B.E. *et al.* The NIH Roadmap Epigenomics Mapping Consortium. *Nat Biotechnol* **28**, 1045-8 (2010).
93. Lopez-Delisle, L. *et al.* pyGenomeTracks: reproducible plots for multivariate genomic datasets. *Bioinformatics* **37**, 422-423 (2021).
94. Ramirez, F. *et al.* deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res* **44**, W160-5 (2016).
95. Consortium, E.P. *et al.* Expanded encyclopaedias of DNA elements in the human and mouse genomes. *Nature* **583**, 699-710 (2020).
96. Quinlan, A.R. & Hall, I.M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-2 (2010).
97. Harrow, J. *et al.* GENCODE: the reference human genome annotation for The ENCODE Project. *Genome Res* **22**, 1760-74 (2012).
98. Krueger, F. & Andrews, S.R. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* **27**, 1571-2 (2011).
99. Wu, H. *et al.* Detection of differentially methylated regions from whole-genome bisulfite sequencing data without replicates. *Nucleic Acids Res* **43**, e141 (2015).
100. Zhou, Y. *et al.* Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat Commun* **10**, 1523 (2019).
101. Tanigawa, Y., Dyer, E.S. & Bejerano, G. Which TF is functionally important in your open chromatin data? *PLoS Comput Biol* **18**, e1010378 (2022).