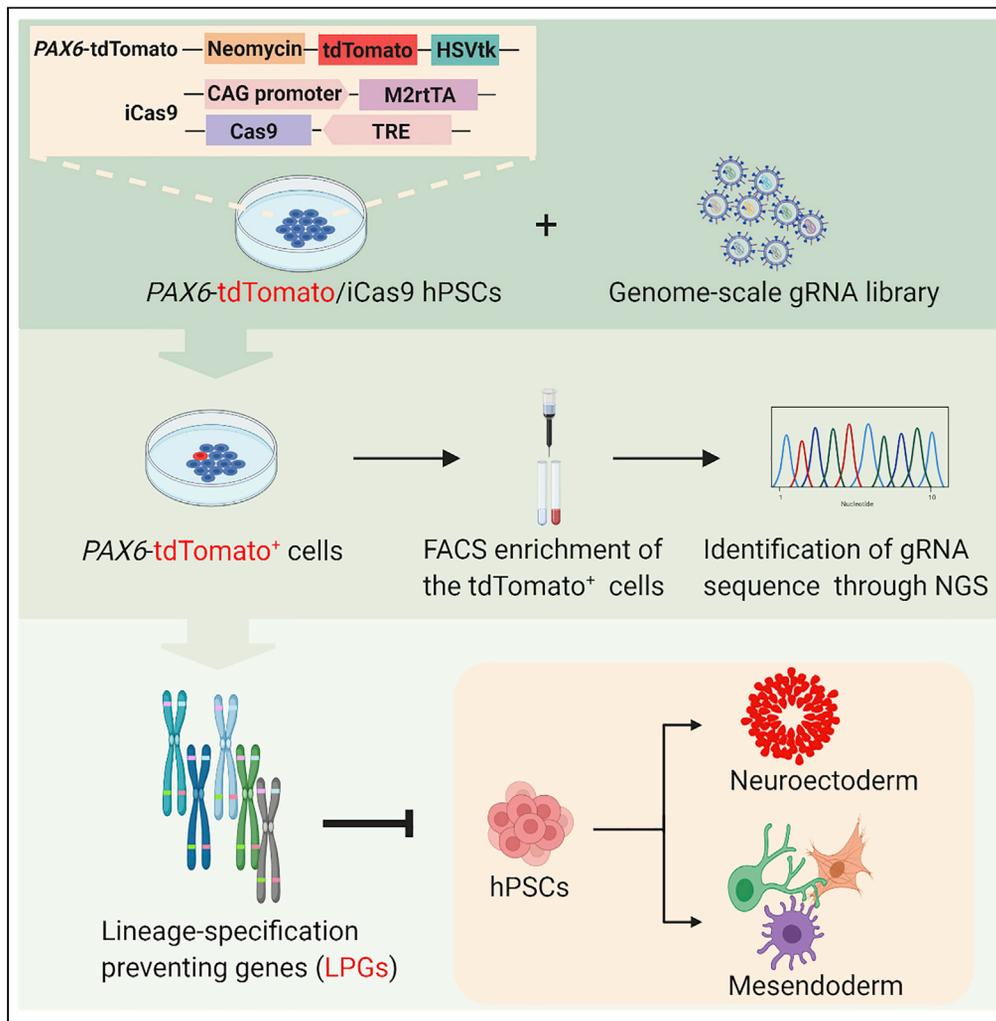


Article

Mapping germ-layer specification preventing genes in hPSCs via genome-scale CRISPR screening



Xiangjie Xu,
Yanhua Du, Lin
Ma, ..., Cizhong
Jiang, Ling Liu,
Xiaoqing Zhang

lliu@tongji.edu.cn (L.L.)
xqzhang@tongji.edu.cn (X.Z.)

Highlights

Lineage reporter and CRISPR screening are powerful tools for studying cell fates

Lineage-specification preventing genes (LPGs) are identified in hPSCs

LPGs maintain pluripotency via targeting one or multiple germ layers

LPGs are clustered into distinct functional modules

Xu et al., iScience 24, 101926
January 22, 2021 © 2020 The Author(s).
<https://doi.org/10.1016/j.isci.2020.101926>



Article

Mapping germ-layer specification preventing genes in hPSCs via genome-scale CRISPR screening

Xiangjie Xu,^{1,3,6,10} Yanhua Du,^{4,10} Lin Ma,^{1,3,6,8,10} Shuwei Zhang,^{1,3,6} Lei Shi,^{1,3,6} Zhenyu Chen,^{1,3,6} Zhongshu Zhou,^{1,3,6} Yi Hui,^{1,3,6} Yang Liu,^{1,3,6} Yujiang Fang,^{1,3,6,8} Beibei Fan,^{1,3,6} Zhongliang Liu,^{1,3,6} Nan Li,^{1,3,6} Shanshan Zhou,^{1,3,6} Cizhong Jiang,⁹ Ling Liu,^{1,3,6,7,8,*} and Xiaoqing Zhang^{1,2,3,5,6,7,11,*}

Summary

Understanding the biological processes that determine the entry of three germ layers of human pluripotent stem cells (hPSCs) is a central question in developmental and stem cell biology. Here, we genetically engineered hPSCs with the germ layer reporter and inducible CRISPR/Cas9 knockout system, and a genome-scale screening was performed to define pathways restricting germ layer specification. Genes clustered in the key biological processes, including embryonic development, mRNA processing, metabolism, and epigenetic regulation, were centered in the governance of pluripotency and lineage development. Other than typical pluripotent transcription factors and signaling molecules, loss of function of mesendodermal specifiers resulted in advanced neuroectodermal differentiation, given their inter-germ layer antagonizing effect. Regarding the epigenetic superfamily, microRNAs enriched in hPSCs showed clear germ layer-targeting specificity. The cholesterol synthesis pathway maintained hPSCs via retardation of neuroectoderm specification. Thus, in this study, we identified a full landscape of genetic wiring and biological processes that control hPSC self-renewal and trilineage specification.

Introduction

Human pluripotent stem cells (hPSCs), including human embryonic stem cells (hESCs) and human induced pluripotent stem cells (hiPSCs), can self-renew and differentiate into all three germ layers (Thomson et al., 1998; Takahashi et al., 2007). Thus, hPSCs can be considered as highly advantageous cells for studying human development, modeling inherited disorders, and replenishing degenerated tissues and cells (Soldner and Jaenisch, 2018). Sophisticated protocols have been developed to target hPSCs into the ectoderm (Ec), mesoderm (Me), or endoderm (En) lineages, primarily via cocktailing developmentally related signaling cues. Activin/Nodal and fibroblast growth factor (FGF) signaling are two dominant pathways that maintain hPSCs in an undifferentiated state, and interference with the signaling molecules of these two pathways will lead to spontaneous differentiation (Vallier et al., 2005; Xu et al., 2005). Once leaving the culture conditions favors pluripotency maintenance, hPSCs adopt a neuroectoderm (nEc) fate in the absence of extra signaling activators, and dual inhibition of bone morphogenetic protein (BMP) and transforming growth factor β (TGF- β) signaling further strengthens this nEc fate (Chambers et al., 2009; Xu et al., 2005). On the other hand, activation of the BMP/Wnt/Activin/Nodal/FGF signaling pathways is major driving forces for generating primitive streak-like cells and Me or En lineage (D'Amour et al., 2005; Tam and Loebel, 2007; Zhang et al., 2008; Martyn et al., 2018; Wang and Chen, 2016).

Extracellular signaling pathways govern hPSC fates largely through their convergence on regulating fate-determining transcription factors (TFs) (Xu et al., 2008). *OCT4* (*POU5F1*), *SOX2*, and *NANOG* are well-characterized core pluripotent genes that maintain hPSCs in pluripotency (Boyer et al., 2005). We previously showed that *PAX6* is necessary and sufficient to generate a human nEc cell fate (Zhang et al., 2010; Chen et al., 2018). Both *T* (Brachury) and *MIXL1* TFs are tightly associated with Me, whereas *GATA4* and *GATA6* are essential for the initiation of an En fate (Tam and Loebel, 2007; Faial et al., 2015).

¹Translational Medical Center for Stem Cell Therapy, Shanghai East Hospital, Tongji University School of Medicine, Shanghai 200120, China

²Key Laboratory of Reconstruction and Regeneration of Spine and Spinal Cord Injury, Ministry of Education, Shanghai 200065, China

³Key Laboratory of Neuroregeneration of Shanghai Universities, Tongji University School of Medicine, Shanghai 200092, China

⁴Department of Immunology and Microbiology, Shanghai Institute of Immunology, Shanghai Jiao Tong University School of Medicine, Shanghai 200025, China

⁵Tsingtao Advanced Research Institute, Tongji University, Qingdao 266071, China

⁶Shanghai Institute of Stem Cell Research and Clinical Translation, Shanghai 200120, China

⁷Brain and Spinal Cord Clinical Research Center, Tongji University School of Medicine, Shanghai 200092, China

⁸Department of Pathology and Pathophysiology, Tongji University School of Medicine, Shanghai 200092, China

⁹The School of Life Sciences and Technology, Tongji University, Shanghai 200092, China

¹⁰These authors contributed equally

¹¹Lead contact

*Correspondence: lliu@tongji.edu.cn (L.L.), xqzhang@tongji.edu.cn (X.Z.)
<https://doi.org/10.1016/j.isci.2020.101926>



Epigenetic machineries and metabolic states also offer proper genetic and cellular contexts to define specific cell fates (Wu et al., 2016; Xie et al., 2013; Gifford et al., 2013). Polycomb group, SWI/SNF, and MI-2/NURD proteins interact with pluripotent genes and play important roles in maintaining pluripotency (Zhang et al., 2014; Lee et al., 2006; Hu and Wade, 2012). Metabolites also have critical roles as substrates for epigenetic regulation, including acetylation and methylation (Moussaieff et al., 2015). While hPSCs are thought to depend primarily on glycolytic flux, several studies have suggested that the mitochondrial tricarboxylic acid (TCA) cycle and oxidative phosphorylation (OXPHOS) are active in hPSCs, and hPSCs consume more oxygen than the nEc cells (Birket et al., 2011; Lees et al., 2018).

Though genetic, epigenetic, and metabolic processes are implicated in defining pluripotency, a systematic understanding of the transition from pluripotency to trilineage differentiation is still lacking (Li et al., 2019; Shparberg et al., 2019; Du et al., 2017). Due to advanced genome-wide screening technologies using either RNAi or CRISPR/Cas9 gRNA libraries, we now have abundant knowledge on essential genes (EGs) for survival and proliferation and pivotal pluripotent genes for self-renewal of hPSCs by monitoring growth behavior or using OCT4-GFP as a readout reporter (Yilmaz et al., 2018; Ihry et al., 2019; Chia et al., 2010; Mair et al., 2019). To date, genome-wide screening to elucidate the genetic wiring or biological processes underlying the commitment of trilineage from hPSCs is not yet to be performed. Here, we engineered hESCs with the PAX6-tdTomato reporter and inducible CRISPR/Cas9 knockout system. Genome-wide CRISPR loss-of-function screening was then conducted, and in-depth views of the genetic wiring and biological processes governing the dynamic transition from hPSCs to trilineage were provided.

Results

Genetic engineering of lineage reporter for high-throughput screening in hPSCs

Within the three germ layers, nEc specification is independent of ectopic signaling pathways activation, referred to as a “default” cell fate induction from pluripotent stem cells (Munoz-Sanjuan and Brivanlou, 2002). We have revealed that the PAX6 TF is a human nEc cell fate determinant and thus can serve as a human nEc hallmark gene (Zhang et al., 2010; Chen et al., 2018). Moreover, nEc can be generated directly from hPSCs via overexpression of PAX6, even under strict hPSC culture conditions (Zhang et al., 2010). During early embryonic development, the expression of PAX6 is intricately regulated with cis-regulatory DNA machineries encompassing large genomic regions (Kleinjan et al., 2006; Tyas et al., 2006). We, therefore, engineered hESCs with a PAX6 reporter via CRISPR/Cas9-mediated homologous recombination (HR) to indicate PAX6 expression profiles and screen for intrinsic pathways that either drive hPSCs to three germ layer differentiations or to a nEc fate without changing culture conditions.

We first constructed a donor plasmid harboring an in-frame coding cassette of Neo-P2A-tdTomato-T2A-HSVtk flanked by 5' AND-3' homology arms of PAX6 (Figure 1A). A single guide RNA (gRNA) targeting exon 4 of PAX6 DNA near the ATG start codon was designed, and the cleavage efficacy was verified in HEK293 cells after co-transfection with Cas9 expression plasmids (Figure S1A). Sanger sequencing results showed that all 5 predicted loci which showed high similarity with the target sequences were identified intact with no mutations, suggesting no off-target effects of the designed gRNA (Figures S1B and S1C). After correctly HR, the putative integrated hESCs would express Neo-P2A-tdTomato-T2A-HSVtk protein under the tight control of endogenous PAX6 gene expression regulatory machineries. After protease cleavage, tdTomato mirrors an endogenous PAX6 expression and nEc fate, whereas neomycin (Neo) and HSVtk could serve as positive selection and negative depletion tools, respectively, when neomycin or ganciclovir are supplied to the cells. Genomic polymerase chain reaction (PCR) revealed that clones 2 and 11 were two monoallelic HR lines (Figures 1A and 1B). Southern blot analysis confirmed precise HR with no off-target recombination (Figure 1C). Both clone 2 and clone 11 retained typical hESC morphology and growth behavior compared with wild-type hESCs. After guiding the lines to a nEc fate with our standard differentiation protocol (Zhang and Zhang, 2010; Chen et al., 2018; Zhang et al., 2001; Li et al., 2009; Liu et al., 2019; Chi et al., 2016), both lines were normally specified to columnar neuroepithelial cells arranged in neural rosette-like structures, and cells started to express tdTomato from day 6 (Figure S1D). Immunostaining further revealed that tdTomato was co-labeled with endogenous PAX6 in the differentiated nEc cells (Figure 1D). Therefore, we successfully generated a nEc reporter line in hPSCs (referred to herein as PAX6-tdTomato reporter line), which can be used for lineage specification studies.

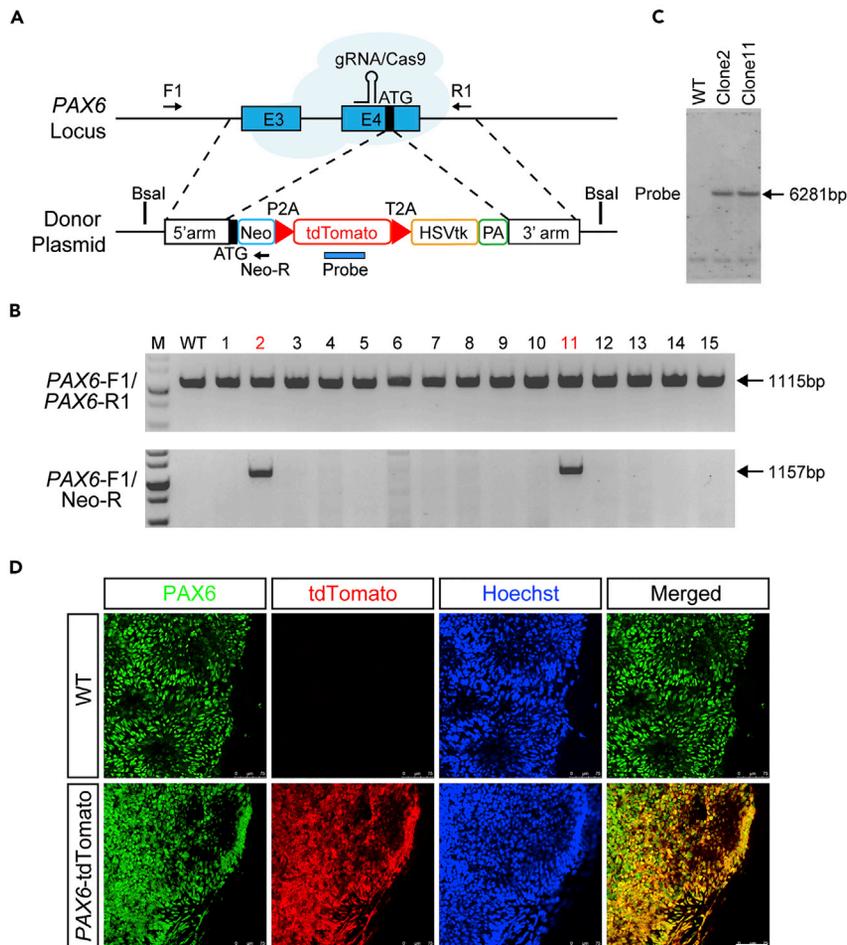


Figure 1. Establishment and characterization of the PAX6-tdTomato reporter line

(A) Schematic view of the targeting strategy through gRNA-guided CRISPR/Cas9 system. The black box indicates the start codon (ATG) of the *PAX6* gene. Arrows indicate primer sets used in genomic DNA PCR. Probe for Southern blot analysis is marked with blue bar at indicated genetic regions after homologous recombination (HR). E3, exon3; E4, exon4; PA, poly(A) signal; Neo, neomycin resistance gene.

(B) Genomic DNA PCR results of individual colonies after gene targeting. Primer sets of *PAX6*-F1 and *PAX6*-R1 were used for detecting the wild-type allele, and primer sets of *PAX6*-F1 and Neo-R were used for detecting the HR allele. Colonies #2 and #11 were monoallelically targeted.

(C) Southern blot analysis showed precise but no off-target recombination.

(D) *PAX6*-tdTomato hESCs were differentiated toward a neuroectoderm (nEc) fate. Immunostaining studies at day 10 showed that tdTomato was co-labeled with endogenous *PAX6*. Scale bar, 75 μ m.

Construction of doxycycline-inducible Cas9 expression cassettes in *PAX6*-Tdtomato line

We modified the two-component inducible Cas9 (iCas9) system (Gonzalez et al., 2014) in the *PAX6*-tdTomato line. One donor plasmid containing a doxycycline-inducible Cas9 expression cassette (3 \times TRE-FLAG-Cas9) and another carrying a constitutive reverse tetracycline transactivator (M2rtTA) expression cassette (CAG-M2rtTA) were constructed (Figure 2A). Both donor plasmids were electroporated into the clone 11 *PAX6*-tdTomato reporter line together with left and right TALEN plasmids targeting *AAVS1* (also known as PPP1R12C) loci. After genomic DNA PCR, we identified that clone 4 had both transgenes integrated in the *AAVS1* loci biallelically (Figure 2B). Western blotting and immunostaining confirmed the robust induction of Cas9 expression upon doxycycline treatment (Figures 2C and 2D). In addition, the constructed *PAX6*-tdTomato/iCas9 line showed typical hPSCs morphology and OCT4 and NANOG expression, as expected (Figures 2D and S2A). After doxycycline treatment, over 90% cells showed both NANOG and Cas9 staining, suggesting high induction efficiency (Figures S2B and S2C).

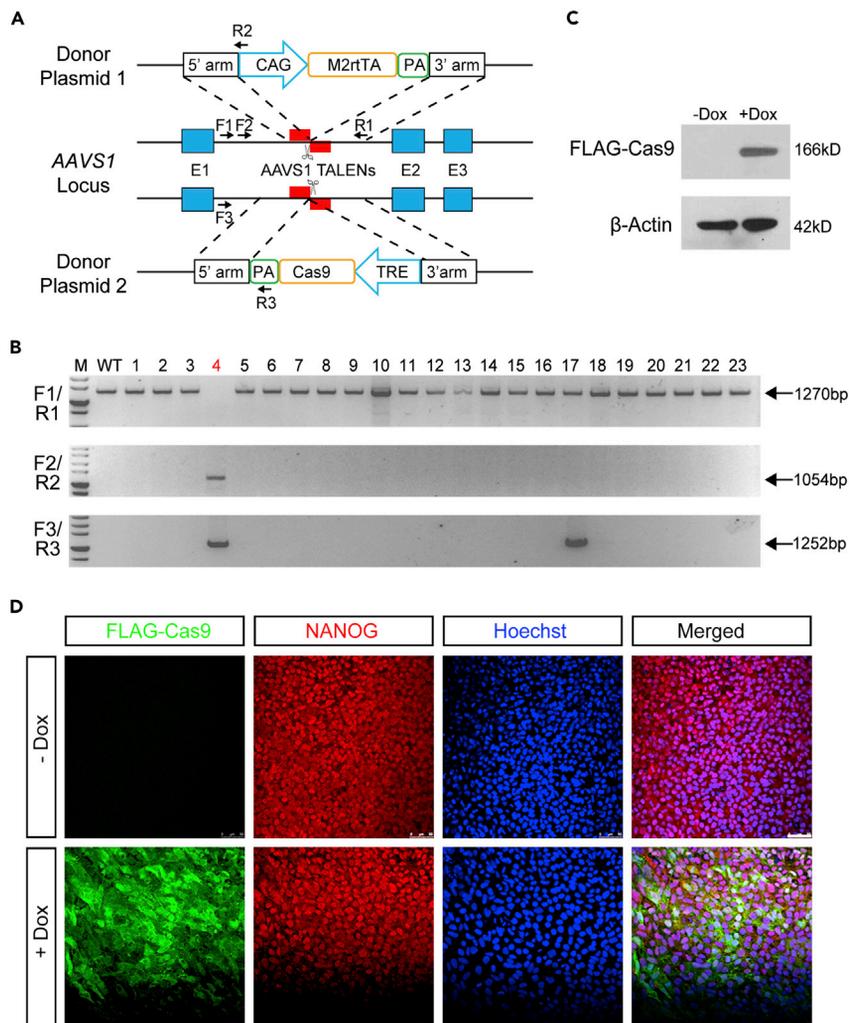


Figure 2. Construction of inducible Cas9 expression systems in PAX6-tdTomato reporter hESCs

(A) Schematic diagram for constructing inducible Cas9 (iCas9) expression cassette in PAX6-tdTomato line through TALEN (marked in red bars)-mediated gene targeting at the AAVS1 loci. Genotyping PCR primer sets are labeled with arrows. E1, exon1; E2, exon2; E3, exon3; CAG, CMV enhancer/chicken β -actin promoter; M2rtTA, reverse tetracycline transactivator; TRE, tetracycline response element; PA, poly(A) signal.

(B) Genomic DNA PCR results showed colony #4 was biallelically targeted and had both CAG-M2rtTA and 3xTRE-Flag-Cas9 expression cassettes integrated at the AAVS1 loci. Primer sets of F1 and R1 were used for detecting the wild-type allele, primer sets of F2 and R2 were used for detecting the recombinated CAG-M2rtTA allele, and primer sets of F3 and R3 were used for detecting the recombinated 3xTRE-FLAG-Cas9 allele, respectively.

(C) Western blot analysis of Cas9 expression in PAX6-tdTomato/iCas9 hESCs treated with or without doxycycline.

(D) Immunostaining results of FLAG-Cas9 and NANOG expression in PAX6-tdTomato/iCas9 hESCs treated with or without doxycycline. Scale bar, 75 μ m.

To determine the capacity of the iCas9 hPSCs for DNA cleavage, we designed gRNA pairs targeting the human *NF1* gene (Liu et al., 2016). The PAX6-tdTomato/iCas9 hESCs were then electroporated with the double gRNAs of *NF1*. After treatment with doxycycline, we observed \sim 70% (5/7) knockout efficiency based on genomic DNA PCR (Figure S2D). We also designed a gRNA targeting *P53* gene (Liu et al., 2016). The gRNA was packaged into lentivirus, which was then used to infect the PAX6-tdTomato/iCas9 cells. We observed prominent occurrence of indels surrounding the gRNA-guided cleavage site when doxycycline was supplied (Figure S2E). These results indicate that the iCas9 system constructed through HR at the genetically open AAVS1 loci is tight and functional and thus is suitable for large-scale CRISPR screening.

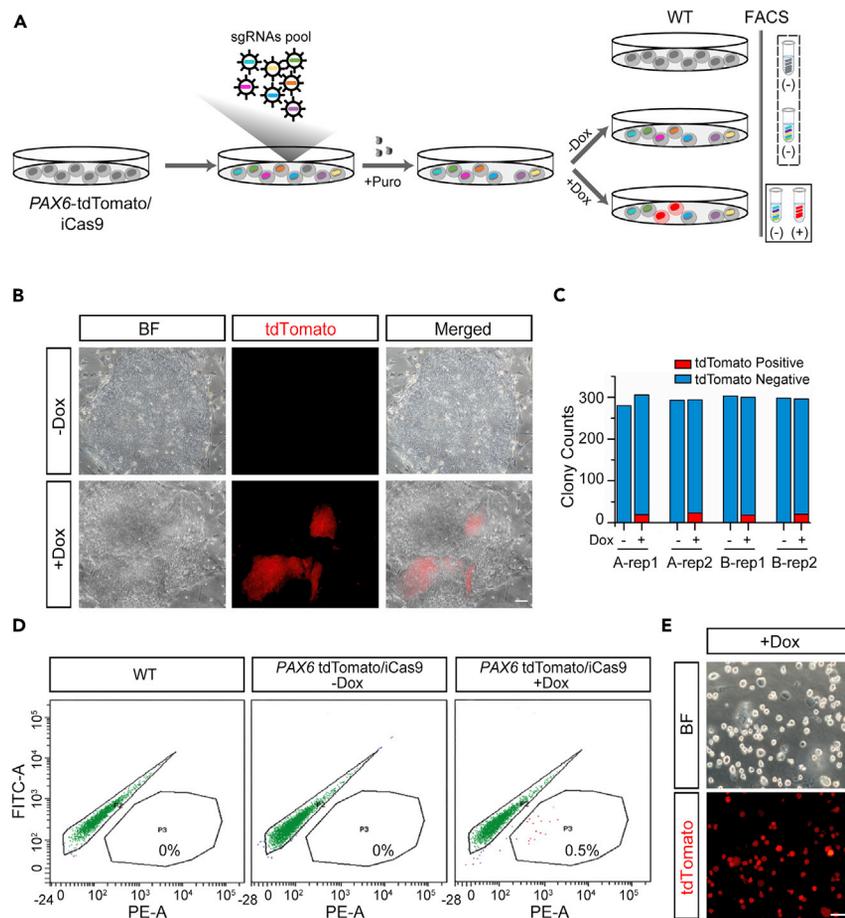


Figure 3. Genome-scale CRISPR screening in PAX6-tdTomato/iCas9 hESCs

(A) Schematic illustration of the genome-scale CRISPR screening strategy. PAX6-tdTomato/iCas9 hESCs were transduced with lentiviral particles of either A pool or B pool gRNA libraries at an MOI of 0.3. Puromycin was applied for 4 days to eliminate non-infected cells. Doxycycline treatment for another 5 days led to target gene loss of functions, and loss of function of germ layer specification preventing genes therefore caused lineage differentiation and aberrant expression of tdTomato under hPSC culture conditions. Wild-type (WT) cells and PAX6-tdTomato cells with no doxycycline treatment did not show tdTomato expression, and they served as negative controls. tdTomato-positive cells induced by doxycycline were purified by FACS. DNA isolation, two-step PCR amplification, and next-generation sequencing were subsequently used for elucidating gRNA enrichment in tdTomato-positive cells.

(B) Epithelial cell clusters expressing tdTomato within the hESCs colony appeared after doxycycline treatment. Scale bar, 100 μ m.

(C) Counts of colonies comprising at least one tdTomato-positive cluster of a 6-well plate with or without doxycycline treatment.

(D) FACS of tdTomato-positive cells after doxycycline treatment.

(E) Fluorescent images showed that FACS-sorted cells had uniform tdTomato expression. Scale bar, 100 μ m.

FACS-based forward genome-wide CRISPR screening

We hypothesized that genetic loss of function in the PAX6-tdTomato line which caused tdTomato expression would indicate the role of this gene in preventing either tril lineage or nEc specification. Therefore, we performed genome-wide loss-of-function screening using the high-throughput CRISPR/Cas9 platform in the PAX6-tdTomato/iCas9 line (Figure 3A). Lentiviral libraries expressing gRNAs (GeCKO) targeting all coding genes and microRNAs were prepared (Shalem et al., 2014). To avoid spontaneous differentiation, the PAX6-tdTomato/iCas9 hESCs were digested into small clusters (not single cells) and infected with lentiviral libraries at a multiplicity of infection (MOI) of 0.3 to ensure high enrichment in the proportion of cells infected with only one viral particle and thus target a unique gene. As the gRNA lentiviral libraries had a puromycin drug-resistant gene, we removed non-infected cells by 4 days puromycin treatment. The

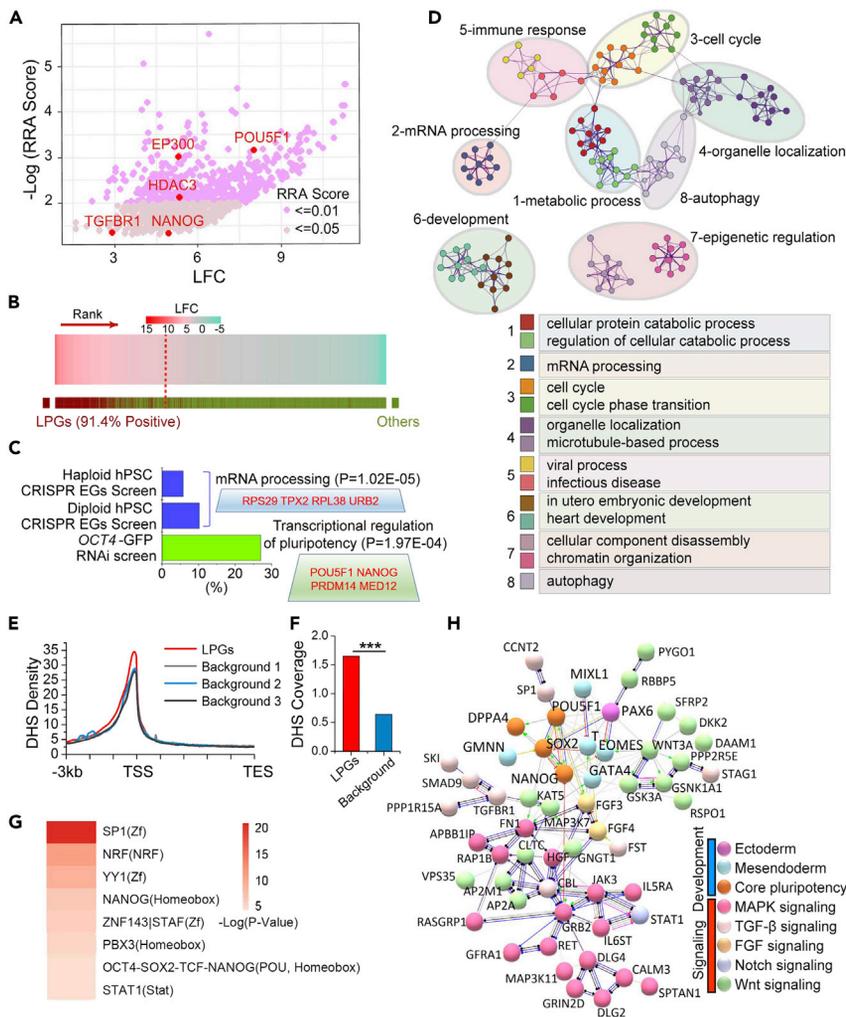


Figure 4. Identification of lineage-specification prevention genes in hPSCs

(A) MAGeCK algorithm was used to estimate the statistical significance of enriched gRNA in tdTomato-positive cells after doxycycline induction. gRNA-associated genes were ranked through robust rank aggregation (RRA) and were termed lineage preventing genes (LPGs). Log₂ fold change (LFC) was used to calculate the enrichment index as compared with the library. Each dot of the scatterplot represented one single LPG and genes known for regulating self-renewal of hPSCs were labeled in red.

(B) Over 90% of LPGs were efficiently enriched in the second sub-pool library screening.

(C) Percentage of overlapped genes of LPGs and reported essential genes (EGs) related to hPSC proliferation and survival or hPSC self-renewal-required genes. Gene Ontology (GO) analyses showed these overlapped genes were mostly related to mRNA processing and transcriptional regulation of pluripotency, with representative genes of each GO term been boxed and marked in red.

(D) LPGs clustered into functional modules as analyzed by the Metascape tool. Enriched terms retrieved from GO, KEGG pathway, and Reactome Gene Sets of all LPGs were assigned to modules based on Kappa-statistical similarities among their gene memberships. Eight modules were defined and each module represented a group of similar functional categories, and edges were connected where terms with similarity above 0.3. Terms with the best p values were depicted as network nodes, and the size of nodes represented number of enrichment of genes.

(E) LPGs and randomly selected background genes at equal size were subjected to DNase hypersensitive sites (DHSs) analysis. As compared with all 3 background controls, LPGs showed significantly higher DHS density around transcriptional start sites (TSS), indicating their active or poised transcriptional state.

(F) Statistics of DHS coverage of LPGs and background controls. Chi-square (χ^2) test was used to examine the significant difference ($\chi^2 = 892.2$; $df = 1$; $p = 0$).

(G) DHSs of LPGs were enriched in binding sites of typical pluripotent transcription factors as revealed by HOMER.

(H) Protein-protein interaction (PPI) network analysis generated by STRING revealed that LPGs within the embryonic development module comprised core pluripotent transcription factors and mesoderm specifiers, both heavily

Figure 4. Continued

interacted with the nEc determinant *PAX6*. Signaling molecules related to Wnt, TGF- β , FGF/MAPK, and Notch pathways are early embryonic developmental cues, which showed a tight link with pluripotent genes and lineage specifiers. This PPI analysis pointed to an antagonistic effect within early lineage specifiers in hPSCs, given loss of function of mesendoderm specifiers led to spontaneous nEc differentiation.

PAX6-tdTomato/iCas9 hESCs infected with the gRNA lentiviral libraries were subsequently treated with doxycycline for another 5 days. Without doxycycline treatment, we did not see tdTomato-positive cells containing colonies, suggesting high quality hESC maintenance under strict culture conditions. Furthermore, doxycycline treatment repetitively yielded tdTomato-positive clusters within the hESC colonies at a frequency of 4%–5% (Figures 3B and 3C). Notably, the tdTomato-expressing cells showed columnar neuroepithelial morphology, forming clusters of similar size and containing distinct borders with surrounding tdTomato-negative hPSCs (Figure 3B). These tdTomato-positive cells completely lost pluripotent marker OCT4 but were positive for SOX2, a known TF also for facilitating nEc fate decision (Wang et al., 2012) (Figure S3A). These results indicate that the tdTomato-expressed clusters were guided to a nEc fate and their uniform clonal appearance implied a single-cell origin likely caused by a loss-of-function mutation of a gene. Fluorescence-activated cell sorting (FACS) experiment showed that approximately 0.5% of entire hESC population was tdTomato positive after doxycycline treatment (Figure 3D). We then enriched the tdTomato-positive cells of interest using FACS and validated their tdTomato expression under the microscope immediately and their nEc characteristics by analyzing the expression of nEc marker genes including *PAX6*, *SOX1*, *DLK1*, *MEIS2*, and *ZIC1* at the mRNA level (Figures 3E and S3B). After validation, the total genomic DNA of these tdTomato-positive cells was extracted, the fragments containing the integrated gRNA were amplified via PCR, and target gRNAs were profiled through sequencing.

Recovered gRNAs are enriched in biological processes essential for hPSC maintenance and lineage specification

To evaluate the CRISPR/Cas9 screening performance, we first verified the integrity of the gRNA pools (A pool and B pool as designed) transduced in the *PAX6*-tdTomato/iCas9 hESCs without doxycycline treatment, with results showing more than 90% coverage (Figure S4A). The Gini index was then calculated to reflect disparities in the distribution of gRNAs. Both the A and B pool in the doxycycline-untreated cells showed a low Gini index value, thus representing highly uniform transduction with no bias (Figure S4B). In contrast, gRNAs retrieved from doxycycline-induced tdTomato-positive cells showed a high Gini index value, indicating target gene selectivity and enrichment (Figure S4B). Furthermore, high repeatability between independent screens reinforced the reliability of the forward screen system based on lineage reporter and inducible gene knockout (Figure S4C).

The MAGeCK algorithm (Li et al., 2014) was used to score the statistical significance (using a negative binomial test) of enrichment for individual gRNAs in tdTomato-positive cells by comparison with the overall gRNA pool. Significantly, enriched gRNAs were defined with an over 2-fold change and false discovery rate (FDR) below 0.05 (Figure S5A). Genes considered relevant were identified by their targeting gRNAs shown in the screen ranked consistently higher (by significance) using robust rank aggregation (RRA). Based on these criteria, we identified ~4% (829/19,062) of overall genes in the coding genome as essential and their loss of function in hPSCs triggered lineage specification (Figure 4A). We therefore termed these genes as lineage-specification preventing genes (LPGs). To further validate these LPGs, we designed a sub-pool library, which included 182 non-targeting control gRNAs, and 7,324 gRNAs for the 829 LPGs and for those genes which were identified for at least one screen but showed no statistical significance. The sub-pool gRNAs were then verified by next-generation sequencing and showed a Gini index value at 0.029, suggesting uniform distribution (Figure S5B). Two sub-pool validation screenings were then performed, and more than 90% of previously identified LPGs were preferentially enriched with over 2-fold changes and FDRs below 0.05 (Figure 4B). The top 100 LPGs were listed in Table S1.

Both *OCT4* and *NANOG* were among the LPGs, coinciding with their pivotal role in pluripotency (Boyer et al., 2005). *TGFBR1*, receptor of the Activin/Nodal signaling pathway, which is essential for hPSC maintenance, was also among the identified LPGs (Vallier et al., 2005; Xu et al., 2008). In addition, *HDAC3* and *EP300*, which are known to facilitate self-renewal while inhibiting lineage specification of hPSCs, were also detected within the LPGs (Qiao et al., 2015). Importantly, the identified LPGs were distributed across all chromosomes without enrichment in specific chromosomal regions (Figure S5C). These results suggest

that the genome-wide CRISPR screening tool revealed a plethora of candidate LPGs and the mutation of which caused hPSC pluripotency exit and germ layer entry.

EGs critical for pluripotent stem cell survival and proliferation have been defined in both diploid and haploid hPSCs through CRISPR loss-of-function screening (Yilmaz et al., 2018; Mair et al., 2019). We made an intersection between LPGs and EGs and found that less than 10% of LPGs overlapped with the EGs, and these overlapped genes were enriched in the biological function of RNA processing (Figure 4C). With the OCT4-GFP reporter, genes required for hPSC self-renewal were also identified through RNAi library screening (Chia et al., 2010). LPGs shared far more genes with self-renewal-required genes than with the common genes of LPGs and EGs (Figures 4C and S5D). In addition, the shared LPGs and self-renewal-required genes were mostly clustered in the transcriptional regulation of pluripotency (Figures 4C and S5D). These results suggested that our CRISPR screening successfully retrieved a group of genes essential for hPSC self-renewal. However, given the advanced lineage reporter design and forward screening system, we also discovered those previously undefined genes essential for other biological processes, which shape trilineage specification.

We further annotated the LPGs to determine their biological function using the Panther Classification System (Mi et al., 2005). Biological process analysis showed strong enrichment in metabolic processes, biological regulation, and cellular component organization pathways (Figure S5E). Protein classification analysis showed enrichment in nucleic acid binding, transcription factors, enzyme modulators, and signaling molecules (Figure S5F). Additionally, we assigned various functional categories to hierarchical clusters based on Kappa-statistical similarities among their gene memberships using Metascape (Tripathi et al., 2015). This revealed that the LPGs were mostly categorized into eight functional modules, among which embryonic development, epigenetic regulation, metabolic regulation, and mRNA processing were the most prominent (Figures 4D and S5G).

To further delineate the regulatory networks controlling LPG expression, we profiled the DNase hypersensitive sites (DHSs) for these genes. DHSs are regions of open chromatin that make DNA accessible, and these accessible chromatin zones are functionally related to transcriptional activity via recruitment of dominant TFs. Here, compared with random or background genes, LPGs harbored significantly higher DHS coverage through chi-square tests (Figures 4E and 4F), suggesting their open chromatin states and active or poised transcription. The DNA sequences around the DHSs of all LPGs were subsequently subjected to TF binding motif analysis. We found that the top motifs were well-characterized TF binding sites associated with pluripotency, such as SP1 of the Z-finger domain, OCT4/NANOG/SOX2 core pluripotency factors, PBX3 of homeobox domain, NF- κ B repressing factor (NRF), and Yin Yang-1 (YY1) (Figure 4G). These results suggest that LPGs are largely transcriptionally active genes in hPSCs and are tightly regulated by core pluripotency factors, reinforcing the fidelity and biological relevance of the identified LPGs. Although we found specific motif enrichment in modules such as metabolism, autophagy, and immune response, the SP1, OCT4/NANOG/SOX2, and YY1 binding motifs appeared in multiple modules (Figure S5H), reflecting their broad effect in regulating pluripotency and lineage specification.

Mesendodermal lineage specifiers antagonize nEc cell fate

Those LPGs within the embryonic development module could be clearly separated into a large-sized and a small-sized group. The large-sized group contained core pluripotency TFs, such as OCT4 and NANOG, and key molecules related to TGF- β , FGF/MAPK, and Notch signaling, which largely overlapped with pluripotency genes retrieved from previous RNAi screening using the OCT4-GFP reporter (Chia et al., 2010). Strikingly, the small-sized group was centered by typical lineage specifiers, such as T and MIXL1 for Me and GATA4 and EOMES for En. Protein-protein interaction (PPI) network analysis showed the interplay of these pluripotency genes and Me or En specifiers (Figure 4H). Deng and colleagues revealed that a combination of nEc, Me, and En lineage specifiers can efficiently reprogram fibroblasts into iPSCs (Shu et al., 2013). It has been reported that germ layer genes and pluripotent genes are co-expressed in a small population of hPSCs, and these cells still preserved their clonogenicity and potency to differentiate into all three germ layers (Allison et al., 2018; Han et al., 2010; Hough et al., 2014). Thus, expression of trilineage specifiers in hPSCs could act against each other to maintain a state of equilibrium in hPSCs, and upsetting such a balance will cause the differentiation of hPSCs toward antagonistic lineage.

Mutations of epigenetic machineries lead to hPSC pluripotency exit or lineage differentiation

Epigenetic regulation, especially chromatin modification and remodeling, plays an important role in hPSC maintenance and differentiation (Gifford et al., 2013; Xie et al., 2013). We identified a functional module for

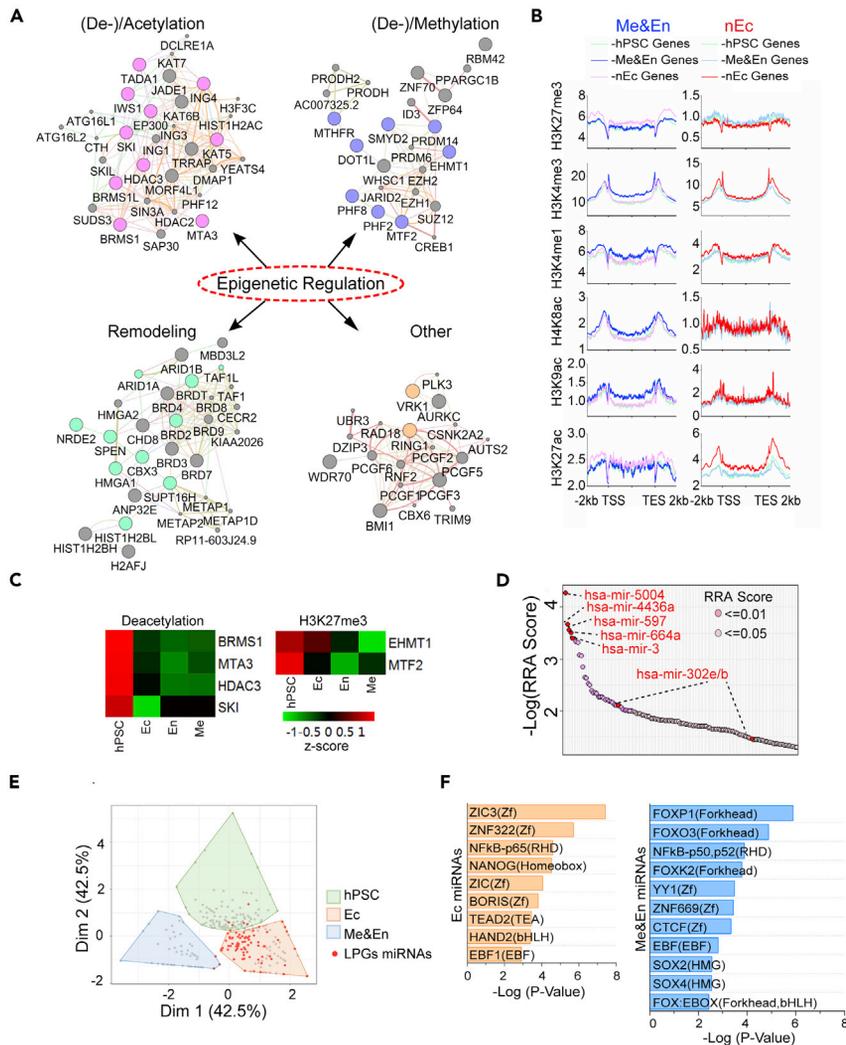


Figure 5. Comprehensive epigenome contributes to hPSC maintenance

(A) Four epigenetic regulation clusters as labeled were highlighted in LPGs. Networks were constructed by using GeneMANIA integrated with Cytoscape visualization. Within each tightly inter-connected cluster, colored nodes represented the genes identified in LPGs.

(B) Profiling histone modifications (HMs) of hESC-, mesendoderm (Me&En)-, and nEc-specific genes in hPSCs differentiated Me&En or nEc lineages. Histone acetylation, H3K4me1, and H3K4me3 of lineage-specific genes were significantly increased, and H3K27me3 were decreased during corresponding lineage specification.

(C) Heatmap showing a decreased expression of core regulators for histone deacetylation and H3K27me3 along with trilineage specification of hPSCs. The expression level was normalized to RPKM and z-scored.

(D) miRNAs within the scope of LPGs were identified by MAGECK algorithm. The significance of each miRNA was ranked through RRA. Each dot of the scatterplot represented one single miRNA and top 5 miRNA, as well as hsa-mir-302e/b known for regulating self-renewal of hPSCs were labeled in red.

(E) Among all miRNAs in miRBae, hPSCs-, Ec-, or Me&En-related miRNAs were defined by using PCA classifier according to the enrichment score of all 3 categories based on GO annotation of its target genes. Red dots represented miRNAs shown in defined LPGs, and they were highly enriched in the Ec-related group.

(F) Motif analyses with HOMER for the promoter regions of Ec- or Me&En-related miRNAs.

epigenetic regulation in LPGs, mostly categorized into families of histone acetylation or de-acetylation, methylation or de-methylation, chromatin remodelers, and others related to ubiquitination or phosphorylation (Figure 5A). Many of these genes were downregulated during lineage specification (Figure S6A). To understand the roles of these epigenetic regulations in trilineage specification, we profiled the histone modifications (HMs) in gene categories of hPSCs-, mesendoderm-, and neural lineage-specific genes.

Results showed that during mesendoderm and nEc lineage specification, the levels of histone acetylation/H3K4me3 in promoter regions and H3K4me1/H3K27ac in enhancer regions of lineage-specific genes were significantly increased, whereas those of H3K27me3 were decreased (Figures 5B and S6B). Heatmap analysis confirmed an obvious decrease in histone deacetylases, such as *BRMS1*, *MTA3*, *HDAC3*, and *SKI*, and methyl transferases for H3K27, including *EHMT1* and *MTF2* along with trilineage differentiation of hPSCs (Figure 5C). These results suggest that active epigenetic machineries are repressed, while repressive epigenetic machineries are highly active in lineage genes in hPSCs. Thus, loss of function of genes centered in these epigenetic machineries caused aberrant activation of lineage genes in hPSCs and therefore led to spontaneous lineage differentiation (Figure S6C).

hPSCs-expressed miRNAs are clustered and maintain pluripotency by targeting trilineage-specific genes

We identified a cluster of miRNAs (85/1864) within the screened LPGs (Figure 5D). Among these miRNAs, the *hsa-mir-302* family is reported to play a role in hPSC maintenance by targeting *COUP-TFII*, a TF expressed in the nEc (Rosa and Brivanlou, 2011). To profile the function of miRNAs in LPGs, we annotated the top five miRNAs through miRBase (Griffiths-Jones et al., 2006) and performed Gene Ontology (GO) analysis for their target genes. We found that most of the target genes were highly enriched in the nEc lineage and were associated with nEc development (Figures S6D and S6E). Indeed, when we sub-grouped the miRNA superfamily into hPSC-, mesendoderm-, and ectoderm-related groups according to their target genes, we found that most miRNAs in our defined LPGs were in the Ec-related group (Figure 5E). These data suggested that hPSC-expressed miRNAs act as differentiation barriers, which specifically inhibit differentiation of an individual lineage. In line with this hypothesis, the promoter regions of Ec-related miRNAs primarily exhibited binding sites for Zinc finger family TFs and *NANOG*, whereas the mesendoderm-related miRNAs were enriched with distinct binding motifs of TFs, such as *SOX2/4* and forkhead family TFs (Figure 5F).

Metabolic pathways profoundly regulate hPSC maintenance and lineage specification

An accumulation of evidence has demonstrated that metabolic pathways play important roles in hPSC maintenance (Wu et al., 2016). Here, genes belonging to the metabolic pathways were the largest superfamily in the identified LPGs. GO analysis showed that these metabolism-related genes were mostly clustered in functional modules of the TCA cycle, energy production, and lipid or amino acid metabolism (Figure 6A).

While hPSCs generally use glycolysis for energy production, a switch from glycolysis to OXPHOS is observed during their differentiation (Zhang et al., 2016). Within the LPGs, we were surprised to find a cluster of mitochondrial electron transport complex genes, considered as the endpoint of the TCA cycle. More importantly, these mitochondrial TCA-related genes were highly expressed in undifferentiated hPSCs but exhibited decreased expression upon differentiation (Figure 6B), implying their important role in hPSCs. They may act like *c-Myc*, an important modulator for mitochondrial biogenesis and is key for hPSC pluripotency maintenance and hiPSC reprogramming (Takahashi et al., 2007; Cliff et al., 2017). Although systematic research is still required to fully elucidate the impact of bioenergetic pathways on the fate of hPSCs, our data suggest that the mitochondrial TCA cycle and OXPHOS are absolutely required for the self-renewal of hPSCs, as loss of function of genes in these pathways caused spontaneous lineage differentiation.

We also found that most LPGs related to amino acid metabolism were downregulated during the process of hPSC differentiation (Figure 6C), implying an important role of amino acid metabolism in maintaining hPSCs (Shiraki et al., 2014). Surprisingly, genes related to cholesterol synthesis were highly enriched in LPGs (8 genes). These cholesterol synthesis genes were highly expressed in undifferentiated hPSCs and Me or En lineages but were downregulated when a nEc cell fate was adopted (Figure 6D). This implies that, different from other metabolic pathways, the cholesterol synthesis pathway is a preset barrier and maintains hPSCs via retardation of nEc specification.

Discussion

With a nEc-specific reporter and inducible CRISPR/Cas9 expression system, we performed genome-scale loss-of-function screening in hPSCs to identify dominant genes that determine pluripotency and lineage

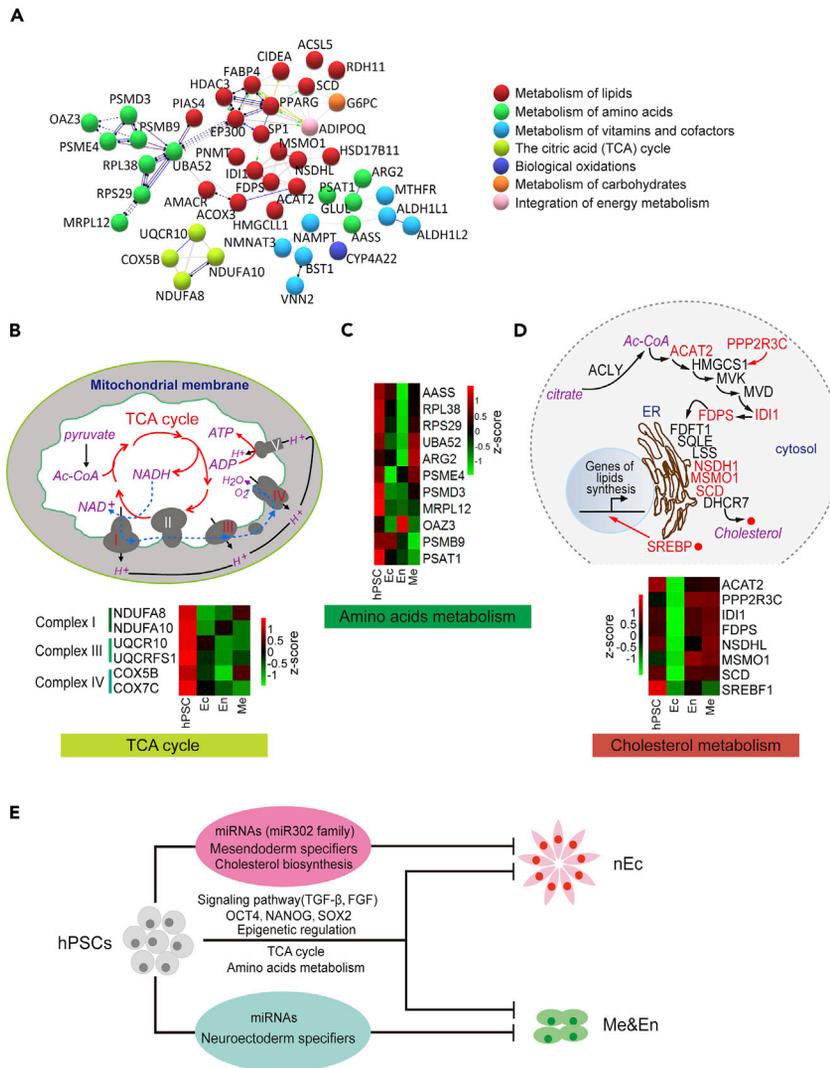


Figure 6. Cholesterol synthesis pathway restricts nEc specification from hPSCs

(A) PPI network analysis generated by STRING revealed that LPGs within the metabolism module comprised key molecules function in multiple metabolic pathways.

(B) Genes related to TCA cycle and oxidative phosphorylation were enriched in LPGs, and heatmap showed decreased expression of these LPGs along trilineage differentiation of hPSCs. The expression level was normalized to RPKM and z-scored.

(C) Heatmap showed decreased expression of LPGs related to amino acid metabolism along trilineage differentiation of hPSCs. The expression level was normalized to RPKM and z-scored.

(D) Genes of cholesterol synthesis pathway that were enriched in LPGs were labeled in red. Heatmap showed that the expression of these cholesterol synthesis pathway-related LPGs was specifically downregulated when the hPSCs were specified toward a nEc fate. The expression level was normalized to RPKM and z-scored.

(E) A schematic model for hPSC self-renewal and trilineage differentiation. Extracellular signaling, epigenetic regulation, transcriptional regulation, and metabolic pathways play a synergistic role in regulating hPSC self-renewal and trilineage differentiation. Histone modifications share common mechanisms in leveraging the transcriptional activity of genes belonging to all three lineages. The TCA cycle, the respiratory chain, and the amino acid metabolism are all important for maintaining an undifferentiated state of hPSCs, though glycolytic flux is highly active. A mutual antagonistic effect occurs in Me&En and nEc lineage specifiers, and a balance of these lineage specifiers is required for the self-renewal of hPSCs. hPSC-enriched miRNAs showed clear lineage gene targeting preference. The cholesterol synthesis pathway plays a pivotal role in maintaining hPSCs via restraining the nEc differentiation.

specification. We discovered various LPGs for hPSCs and loss of function of which led to either pluripotency exit or specific lineage commitment. Our study demonstrated that lineage-based reporters combined with CRISPR/Cas9 whole-genome screening offer a powerful strategy for revealing the biological processes and gene networks underlying specific cell fate determination, which is a key question in developmental biology and the basis for regenerative medicine. The identified LPGs could also serve as an invaluable resource for future studies on cellular and molecular events regarding hPSC self-renewal and trilineage development.

Our study re-emphasized the pivotal role of Notch/TGF- β /FGF signaling and *OCT4/NANOG* core transcriptional networks in regulating pluripotency of hPSCs (Vallier et al., 2005; Xu et al., 2008; Boyer et al., 2005). Our screening also revealed that common cellular biological processes, such as cellular organization, cell cycle regulation, and RNA processing, also play universal and important roles in hPSCs. Some of these functional modules overlap with EGs crucial for hPSC survival and proliferation. As demonstrated by our screening, there was a specific module for immunological responses within the LPGs. Eggenberger et al. showed that ectopic activation of the canonical type I interferon antiviral pathway drives hPSCs away from pluripotency and induces aberrant trilineage differentiation, supporting the conclusion that immune responses are also tightly associated with pluripotency and lineage development (Eggenberger et al., 2019).

Epigenetic modifiers establish a fundamental genomic environment for precise gene expression control and cellular context homeostasis. We identified that HMs and chromatin remodelers had profound effects on hPSC self-renewal and lineage differentiation, based on the following observations: (1) HM-related machineries and chromatin remodelers in LPGs were largely downregulated during lineage specification; (2) active HMs in the promoter and enhancer regions of lineage specific genes were enriched, whereas repressive HMs were de-enriched during corresponding lineage commitment; thus we concluded that HMs and chromatin remodelers maintained hPSCs in an undifferentiated state by establishing transcriptional barriers to restrict lineage gene expression. Notably, those miRNAs highly expressed in hPSCs retained pluripotency by targeting lineage genes. However, these miRNAs could be clearly sub-grouped into specific lineage-related clusters. Moreover, the promoter regions of the ectoderm-related miRNAs were enriched in binding sites of *NANOG*, the core pluripotency factor that antagonizes Ec but favors definitive En development. With our analysis, we also found that the mesendoderm-related miRNAs were enriched with binding motifs for *SOX2*, a known TF for facilitating nEc but inhibiting Me or En specification (Wang et al., 2012).

A small but intriguing cluster within the identified LPGs was the key mesendoderm specifiers, i.e., *GATA4*, *EOMES*, *MIXL1*, and *T*. Deng and his colleagues revealed that *GATA3* and *DLX3* can substitute *OCT4* and *SOX2* to reprogram mouse fibroblasts into iPSCs. Furthermore, they established a “seesaw” model, which demonstrated the balance between pluripotency factors and/or counteracting lineage specifiers (Shu et al., 2013). The same model also applies in humans, with lineage-specific genes able to efficiently reprogram human fibroblasts into hiPSCs (Mai et al., 2018; Shu et al., 2015). With the advent of single-cell RNA-seq technology, pluripotent stem cells are considered relatively heterogeneous, and lineage-specific genes are also expressed in some subpopulations of pluripotent stem cells responsible for clonal expansion (Allison et al., 2018; Hough et al., 2014). Thus, we hypothesized that, at least, in some populations of pluripotent stem cells, lineage specifiers counteract with each other and therefore maintain their pluripotency. In the future, the same strategy by applying mesendoderm-specific reporters could be used to identify Ec including nEc specifiers, LPGs antagonizing mesendoderm differentiation.

Here, genes within the metabolic pathway accounted for the largest family among identified LPGs. Although glycolytic flux has been shown to be crucial for the maintenance of hPSCs (Folmes et al., 2011; Panopoulos et al., 2012), we identified LPGs that were clustered in the functional module of mitochondrial TCA cycle and OXPHOS. Multiple studies have shown that proteins function in the TCA cycle and OXPHOS are higher in hPSCs than in their differentiated derivatives (Varum et al., 2011). Thus, the mitochondrial TCA cycle and OXPHOS and are also indispensable for hPSC pluripotency. Surprisingly, loss of function of genes within the cholesterol synthesis pathway led to nEc differentiation of hPSCs. It has been reported that the expression of pluripotent gene *OCT4* was decreased when the cholesterol biosynthesis pathway was inhibited (Matsuzaki et al., 2018). Moreover, the cholesterol synthesis pathway is highly active in hPSCs and Me and En lineages but low in nEc. Thus, both glycolysis and OXPHOS bioenergetic pathways are crucial for the maintenance of hPSCs, and cholesterol synthesis process is required for human pluripotency via retardation of nEc specification.

In summary, this study provides a full landscape of biological processes underlying hPSC self-renewal and trilineage differentiation (Figure 6E). Pathways involved in bioenergetic production, HMs and chromatin remodelers, immune responses, RNA processing, as well as core pluripotency-related signaling and transcriptional networks, function synergistically to maintain a pluripotent state in hPSCs via inhibition of trilineage differentiation. However, hPSC-enriched miRNAs, lineage specifiers, and the cholesterol synthesis pathway exhibit germ layer selectivity, and these biological processes promote or antagonize specific germ layers along the maintenance of hPSCs or early lineage differentiation.

Limitations of the study

In this paper, we performed CRISPR/Cas9 screen in hPSCs to identify those genes that prevent lineage specification. However, when taking the trilineage cell fate decision in count, we only used the *PAX6* reporter and lacked parallel screen with the mesendodermal reporters. The uncovered library of lineage specification preventing genes is a valuable source for studying mechanisms underlying hPSC self-renewal and trilineage specification. However, specific validation is apparently required when comes to individual genes.

Resource availability

Lead contact

Xiaoqing Zhang (xqzhang@tongji.edu.cn)

Materials availability

Requests for materials and reagents should be directed to the Lead contact.

Data availability

All relevant data are available from the correspondence authors upon reasonable request. The accession number for the CRISPR/Cas9 screening data reported in this paper is GEO: GSE132309.

Methods

All methods can be found in the accompanying [Transparent Methods supplemental file](#).

Supplemental information

Supplemental Information can be found online at <https://doi.org/10.1016/j.isci.2020.101926>.

Acknowledgments

This work was supported by grants from the National Key Research and Development Program of China (Grant no. 2018YFA0108000 and 2019YFA0110300), the National Natural Science Foundation of China (Grant No. 82025020, 31400934, 31771132, 31872760, 31801204, and 31800858), the Science and Technology Commission of Shanghai Municipality (19JC1415100), the Shanghai Municipal Education Commission (C120114), China Postdoctoral Science Foundation (Grant No. 2017M621526), the Fundamental Research Funds for the Central Universities, and the Major Program of Development Fund for Shanghai Zhangjiang National Innovation Demonstration Zone (Stem Cell Strategic Biobank and Clinical Translation Platform of Stem Cell Technology, ZJ2018-ZD-004). Graphical abstract of this study was created with [BioRender.com](#).

Author contributions

X.Z. and L.L. conceived and designed the project. X.X., Y.D., and L.M. performed most of the experiments and analyses. SW.Z., L.S., Z.C., Z.Z., Y.H., Y.L., Y.F., B.F., Z.L., N.L., SS.Z., and L.L., helped to set up the genetic engineering system and performed Western blot, immunostaining, and genomic DNA PCR experiments. C.J. helped with bioinformatics analyses. X.X., Y.D., and X.Z. wrote the manuscript. All authors discussed the results and commented on the manuscript.

Declaration of interests

The authors declare no competing interests.

Received: February 14, 2020

Revised: August 17, 2020

Accepted: December 7, 2020

Published: January 22, 2021

References

- Allison, T.F., Smith, A.J.H., Anastassiadis, K., Sloane-Stanley, J., Biga, V., Stavish, D., Hackland, J., Sabri, S., Langerman, J., Jones, M., et al. (2018). Identification and single-cell functional characterization of an Endodermally biased pluripotent substate in human embryonic stem cells. *Stem Cell Reports* 10, 1895–1907.
- Birket, M.J., Orr, A.L., Gerencser, A.A., Madden, D.T., Vitelli, C., Swistowski, A., Brand, M.D., and Zeng, X. (2011). A reduction in ATP demand and mitochondrial activity with neural differentiation of human embryonic stem cells. *J. Cell Sci.* 124, 348–358.
- Boyer, L.A., Lee, T.I., Cole, M.F., Johnstone, S.E., Levine, S.S., Zucker, J.R., Guenther, M.G., Kumar, R.M., Murray, H.L., Jenner, R.G., et al. (2005). Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell* 122, 947–956.
- Chambers, S.M., Fasano, C.A., Papapetrou, E.P., Tomishima, M., Sadelain, M., and Studer, L. (2009). Highly efficient neural conversion of human ES and iPSC cells by dual inhibition of SMAD signaling. *Nat. Biotechnol.* 27, 275–280.
- Chen, Z., Ren, X., Xu, X., Zhang, X., Hui, Y., Liu, Z., Shi, L., Fang, Y., Ma, L., Liu, Y., et al. (2018). Genetic engineering of human embryonic stem cells for precise cell fate tracing during human lineage development. *Stem Cell Reports* 11, 1257–1271.
- Chi, L., Fan, B., Zhang, K., Du, Y., Liu, Z., Fang, Y., Chen, Z., Ren, X., Xu, X., Jiang, C., et al. (2016). Targeted differentiation of regional ventral neuroprogenitors and related neuronal subtypes from human pluripotent stem cells. *Stem Cell Reports* 7, 941–954.
- Chia, N.Y., Chan, Y.S., Feng, B., Lu, X., Orlov, Y.L., Moreau, D., Kumar, P., Yang, L., Jiang, J., Lau, M.S., et al. (2010). A genome-wide RNAi screen reveals determinants of human embryonic stem cell identity. *Nature* 468, 316–320.
- Cliff, T.S., Wu, T., Boward, B.R., Yin, A., Yin, H., Glushka, J.N., Prestegard, J.H., and Dalton, S. (2017). MYC controls human pluripotent stem cell fate decisions through regulation of metabolic flux. *Cell Stem Cell* 21, 502–516 e9.
- D'Amour, K.A., Agulnick, A.D., Eliazer, S., Kelly, O.G., Kroon, E., and Baetge, E.E. (2005). Efficient differentiation of human embryonic stem cells to definitive endoderm. *Nat. Biotechnol.* 23, 1534–1541.
- Du, Y., Liu, Z., Cao, X., Chen, X., Chen, Z., Zhang, X., Zhang, X., and Jiang, C. (2017). Nucleosome eviction along with H3K9ac deposition enhances Sox2 binding during human neuroectodermal commitment. *Cell Death Differ.* 24, 1121–1131.
- Eggenberger, J., Blanco-Melo, D., Panis, M., Brennand, K.J., and Tenover, B.R. (2019). Type I interferon response impairs differentiation potential of pluripotent stem cells. *Proc. Natl. Acad. Sci. U S A* 116, 1384–1393.
- Faial, T., Bernardo, A.S., Mendjan, S., Diamanti, E., Ortmann, D., Gentsch, G.E., Mascetti, V.L., Trotter, M.W., Smith, J.C., and Pedersen, R.A. (2015). Brachyury and SMAD signalling collaboratively orchestrate distinct mesoderm and endoderm gene regulatory networks in differentiating human embryonic stem cells. *Development* 142, 2121–2135.
- Folmes, C.D.L., Nelson, T.J., Martinez-Fernandez, A., Arrell, D.K., Lindor, J.Z., Dzeja, P.P., Ikeda, Y., Perez-Terzic, C., and Terzic, A. (2011). Somatic oxidative bioenergetics transitions into pluripotency-dependent glycolysis to facilitate nuclear reprogramming. *Cell Metab.* 14, 264–271.
- Gifford, C.A., Ziller, M.J., Gu, H., Trapnell, C., Donaghey, J., Tsankov, A., Shalek, A.K., Kelley, D.R., Shishkin, A.A., Issner, R., et al. (2013). Transcriptional and epigenetic dynamics during specification of human embryonic stem cells. *Cell* 153, 1149–1163.
- Gonzalez, F., Zhu, Z., Shi, Z.D., Lelli, K., Verma, N., Li, Q.V., and Huangfu, D. (2014). An iCRISPR platform for rapid, multiplexable, and inducible genome editing in human pluripotent stem cells. *Cell Stem Cell* 15, 215–226.
- Griffiths-Jones, S., Grocock, R.J., Van Dongen, S., Bateman, A., and Enright, A.J. (2006). miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Res.* 34, D140–D144.
- Han, D.W., Tapia, N., Joo, J.Y., Greber, B., Arauzo-Bravo, M.J., Bernemann, C., Ko, K., Wu, G., Stehling, M., Do, J.T., and Scholer, H.R. (2010). Epiblast stem cell subpopulations represent mouse embryos of distinct pregastrulation stages. *Cell* 143, 617–627.
- Hough, S.R., Thornton, M., Mason, E., Mar, J.C., Wells, C.A., and Pera, M.F. (2014). Single-cell gene expression profiles define self-renewing, pluripotent, and lineage primed states of human pluripotent stem cells. *Stem Cell Reports* 2, 881–895.
- Hu, G., and Wade, P.A. (2012). NuRD and pluripotency: a complex balancing act. *Cell Stem Cell* 10, 497–503.
- Ihry, R.J., Salick, M.R., Ho, D.J., Sondey, M., Kommineni, S., Paula, S., Raymond, J., Henry, B., Frias, E., Wang, Q., et al. (2019). Genome-scale CRISPR screens identify human pluripotency-specific genes. *Cell Rep.* 27, 616–630 e6.
- Kleinjan, D.A., Seawright, A., Mella, S., Carr, C.B., Tyas, D.A., Simpson, T.I., Mason, J.O., Price, D.J., and Van Heyningen, V. (2006). Long-range downstream enhancers are essential for Pax6 expression. *Dev. Biol.* 299, 563–581.
- Lee, T.I., Jenner, R.G., Boyer, L.A., Guenther, M.G., Levine, S.S., Kumar, R.M., Chevalier, B., Johnstone, S.E., Cole, M.F., Isono, K., et al. (2006). Control of developmental regulators by Polycomb in human embryonic stem cells. *Cell* 125, 301–313.
- Lees, J.G., Gardner, D.K., and Harvey, A.J. (2018). Mitochondrial and Glycolytic Remodeling during Nascent Neural Differentiation of Human Pluripotent Stem Cells. *Development* 145, dev168997.
- Li, Q.V., Dixon, G., Verma, N., Rosen, B.P., Gordillo, M., Luo, R., Xu, C., Wang, Q., Soh, C.L., Yang, D., et al. (2019). Genome-scale screens identify JNK-JUN signaling as a barrier for pluripotency exit and endoderm differentiation. *Nat. Genet.* 51, 999–1010.
- Li, W., Xu, H., Xiao, T., Cong, L., Love, M.I., Zhang, F., Irizarry, R.A., Liu, J.S., Brown, M., and Liu, X.S. (2014). MAGECK enables robust identification of essential genes from genome-scale CRISPR/Cas9 knockout screens. *Genome Biol.* 15, 554.
- Li, X.J., Zhang, X., Johnson, M.A., Wang, Z.B., Lavaute, T., and Zhang, S.C. (2009). Coordination of sonic hedgehog and Wnt signaling determines ventral and dorsal telencephalic neuron types from human embryonic stem cells. *Development* 136, 4055–4063.
- Liu, L., Chen, Z., Zhang, X., Li, S., Hui, Y., Feng, H., Du, Y., Jin, G., Zhou, X., and Zhang, X. (2019). Protection of ZIKV infection-induced neuropathy by abrogation of acute antiviral response in human neural progenitors. *Cell Death Differ.* 26, 2607–2621.
- Liu, Z., Hui, Y., Shi, L., Chen, Z., Xu, X., Chi, L., Fan, B., Fang, Y., Liu, Y., Ma, L., et al. (2016). Efficient CRISPR/Cas9-Mediated versatile, predictable, and donor-free gene knockout in human pluripotent stem cells. *Stem Cell Reports* 7, 496–507.
- Mai, T., Markov, G.J., Brady, J.J., Palla, A., Zeng, H., Sebastiano, V., and Blau, H.M. (2018). NKX3-1 is required for induced pluripotent stem cell reprogramming and can replace OCT4 in mouse and human iPSC induction. *Nat. Cell Biol.* 20, 900.
- Mair, B., Tomic, J., Masud, S.N., Tonge, P., Weiss, A., Usaj, M., Tong, A.H.Y., Kwan, J.J., Brown, K.R., Titus, E., et al. (2019). Essential gene profiles for human pluripotent stem cells identify uncharacterized genes and substrate dependencies. *Cell Rep.* 27, 599–615 e12.
- Martyn, I., Kanno, T.Y., Ruzo, A., Siggia, E.D., and Brivanlou, A.H. (2018). Self-organization of a human organizer by combined Wnt and Nodal signalling. *Nature* 558, 132–135.
- Matsuzaki, T., Matsumoto, S., Kasai, T., Yoshizawa, E., Okamoto, S., Yoshikawa, H.Y., Taniguchi, H., and Takebe, T. (2018). Defining lineage-specific membrane fluidity signatures

that regulate adhesion kinetics. *Stem Cell Reports* 11, 852–860.

Mi, H., Lazareva-Ulitsky, B., Loo, R., Kejarawal, A., Vandergriff, J., Rabkin, S., Guo, N., Muruganujan, A., Doremieux, O., Campbell, M.J., et al. (2005). The PANTHER database of protein families, subfamilies, functions and pathways. *Nucleic Acids Res.* 33, D284–D288.

Moussaieff, A., Kogan, N.M., and Aberdam, D. (2015). Concise review: energy metabolites: key mediators of the epigenetic state of pluripotency. *Stem Cells* 33, 2374–2380.

Munoz-Sanjuan, I., and Brivanlou, A.H. (2002). Neural induction, the default model and embryonic stem cells. *Nat. Rev. Neurosci.* 3, 271–280.

Panopoulos, A.D., Yanes, O., Ruiz, S., Kida, Y.S., Diep, D., Tautenhahn, R., Herrerias, A., Batchelder, E.M., Plongthongkum, N., Lutz, M., et al. (2012). The metabolome of induced pluripotent stem cells reveals metabolic changes occurring in somatic cell reprogramming. *Cell Res.* 22, 168–177.

Qiao, Y., Wang, R., Yang, X., Tang, K., and Jing, N. (2015). Dual roles of histone H3 lysine 9 acetylation in human embryonic stem cell pluripotency and neural differentiation. *J. Biol. Chem.* 290, 2508–2520.

Rosa, A., and Brivanlou, A.H. (2011). A regulatory circuitry comprised of miR-302 and the transcription factors OCT4 and NR2F2 regulates human embryonic stem cell differentiation. *EMBO J.* 30, 237–248.

Shalem, O., Sanjana, N.E., Hartenian, E., Shi, X., Scott, D.A., Mikkelsen, T., Heckl, D., Ebert, B.L., Root, D.E., Doench, J.G., and Zhang, F. (2014). Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science* 343, 84–87.

Shiraki, N., Shiraki, Y., Tsuyama, T., Obata, F., Miura, M., Nagae, G., Aburatani, H., Kume, K., Endo, F., and Kume, S. (2014). Methionine metabolism regulates maintenance and differentiation of human pluripotent stem cells. *Cell Metab.* 19, 780–794.

Shparberg, R.A., Glover, H.J., and Morris, M.B. (2019). Modeling mammalian commitment to the neural lineage using embryos and embryonic stem cells. *Front. Physiol.* 10, 705.

Shu, J., Wu, C., Wu, Y.T., Li, Z.Y., Shao, S.D., Zhao, W.H., Tang, X., Yang, H., Shen, L.J., Zuo, X.H., et al. (2013). Induction of pluripotency in mouse somatic cells with lineage specifiers. *Cell* 153, 963–975.

Shu, J., Zhang, K., Zhang, M.J., Yao, A.Z., Shao, S.D., Du, F.X., Yang, C.Y., Chen, W.H., Wu, C., Yang, W.F., et al. (2015). GATA family members as inducers for cellular reprogramming to pluripotency. *Cell Res.* 25, 169–180.

Soldner, F., and Jaenisch, R. (2018). Stem cells, genome editing, and the path to translational medicine. *Cell* 175, 615–632.

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

Tam, P.P., and Loebel, D.A. (2007). Gene function in mouse embryogenesis: get set for gastrulation. *Nat. Rev. Genet.* 8, 368–381.

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

Tripathi, S., Pohl, M.O., Zhou, Y., Rodriguez-Frandsen, A., Wang, G., Stein, D.A., Moulton, H.M., Dejesus, P., Che, J., Mulder, L.C., et al. (2015). Meta- and orthogonal integration of influenza "OMICs" data defines a role for UBR4 in virus budding. *Cell Host Microbe* 18, 723–735.

Tyas, D.A., Simpson, T.I., Carr, C.B., Kleinjan, D.A., Van Heyningen, V., Mason, J.O., and Price, D.J. (2006). Functional conservation of Pax6 regulatory elements in humans and mice demonstrated with a novel transgenic reporter mouse. *BMC Dev. Biol.* 6, 21.

Vallier, L., Alexander, M., and Pedersen, R.A. (2005). Activin/Nodal and FGF pathways cooperate to maintain pluripotency of human embryonic stem cells. *J. Cell Sci.* 118, 4495–4509.

Varum, S., Rodrigues, A.S., Moura, M.B., Momcilovic, O., Easley, C.A., Ramalho-Santos, J., Van Houten, B., and Schatten, G. (2011). Energy metabolism in human pluripotent stem cells and their differentiated counterparts. *PLoS One* 6, e20914.

Wang, L., and Chen, Y.G. (2016). Signaling control of differentiation of embryonic stem cells toward mesendoderm. *J. Mol. Biol.* 428, 1409–1422.

Wang, Z., Oron, E., Nelson, B., Razis, S., and Ivanova, N. (2012). Distinct lineage specification roles for NANOG, OCT4, and SOX2 in human embryonic stem cells. *Cell Stem Cell* 10, 440–454.

Wu, J., Ocampo, A., and Belmonte, J.C.I. (2016). Cellular metabolism and induced pluripotency. *Cell* 166, 1371–1385.

Xie, W., Schultz, M.D., Lister, R., Hou, Z., Rajagopal, N., Ray, P., Whitaker, J.W., Tian, S., Hawkins, R.D., Leung, D., et al. (2013). Epigenomic analysis of multilineage differentiation of human embryonic stem cells. *Cell* 153, 1134–1148.

Xu, R.H., Peck, R.M., Li, D.S., Feng, X., Ludwig, T., and Thomson, J.A. (2005). Basic FGF and suppression of BMP signaling sustain undifferentiated proliferation of human ES cells. *Nat. Methods* 2, 185–190.

Xu, R.H., Sampsel-Barron, T.L., Gu, F., Root, S., Peck, R.M., Pan, G., Yu, J., Antosiewicz-Bourget, J., Tian, S., Stewart, R., and Thomson, J.A. (2008). NANOG is a direct target of TGFbeta/activin-mediated SMAD signaling in human ESCs. *Cell Stem Cell* 3, 196–206.

Yilmaz, A., Peretz, M., Aharoni, A., Sagi, I., and Benvenisty, N. (2018). Defining essential genes for human pluripotent stem cells by CRISPR-Cas9 screening in haploid cells. *Nat. Cell Biol.* 20, 610–619.

Zhang, J., Khvorostov, I., Hong, J.S., Oktay, Y., Vergnes, L., Nuebel, E., Wahjudi, P.N., Setoguchi, K., Wang, G., Do, A., et al. (2016). UCP2 regulates energy metabolism and differentiation potential of human pluripotent stem cells (vol 30, pg 4860, 2011). *Embo J.* 35, 899.

Zhang, P., Li, J., Tan, Z., Wang, C., Liu, T., Chen, L., Yong, J., Jiang, W., Sun, X., Du, L., et al. (2008). Short-term BMP-4 treatment initiates mesoderm induction in human embryonic stem cells. *Blood* 111, 1933–1941.

Zhang, S.C., Wernig, M., Duncan, I.D., Brustle, O., and Thomson, J.A. (2001). In vitro differentiation of transplantable neural precursors from human embryonic stem cells. *Nat. Biotechnol.* 19, 1129–1133.

Zhang, X., Huang, C.T., Chen, J., Pankratz, M.T., Xi, J., Li, J., Yang, Y., Lavaute, T.M., Li, X.J., Ayala, M., et al. (2010). Pax6 is a human neuroectoderm cell fate determinant. *Cell Stem Cell* 7, 90–100.

Zhang, X., Li, B., Li, W., Ma, L., Zheng, D., Li, L., Yang, W., Chu, M., Chen, W., Mailman, R.B., et al. (2014). Transcriptional repression by the BRG1-SWI/SNF complex affects the pluripotency of human embryonic stem cells. *Stem Cell Reports* 3, 460–474.

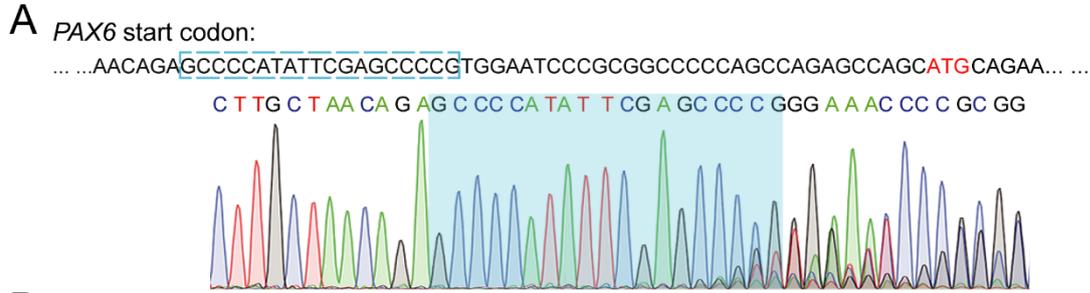
Zhang, X.Q., and Zhang, S.C. (2010). Differentiation of neural precursors and dopaminergic neurons from human embryonic stem cells. *Methods Mol. Biol.* 584, 355–366.

Supplemental Information

Mapping germ-layer specification preventing genes in hPSCs via genome-scale CRISPR screening

Xiangjie Xu, Yanhua Du, Lin Ma, Shuwei Zhang, Lei Shi, Zhenyu Chen, Zhongshu Zhou, Yi Hui, Yang Liu, Yujiang Fang, Beibei Fan, Zhongliang Liu, Nan Li, Shanshan Zhou, Cizhong Jiang, Ling Liu, and Xiaoqing Zhang

Supplemental information



B

Match name	Coordinate	Guide sequence + PAM	Mismatches
PAX6 gRNA	chr11:31827964 to 31827986	AGCCCCATATTCGAGCCCCG <u>TGG</u>	4
PAX6 gRNA off-1	chr6:56740611 to 56740633	A <u>CCCCAA</u> ATTCAGCCCCA <u>TGG</u>	4
PAX6 gRNA off-2	chr4:12373222 to 12373244	AGCCCCAAGTTCG <u>IGCCCCA</u> AGG	4
PAX6 gRNA off-3	chr19:17393106 to 17393128	A <u>ACCCAC</u> CCCTCGAGCCCCG <u>AGG</u>	4
PAX6 gRNA off-4	chr8:141312272 to 141312294	AG <u>ACCCAC</u> AT <u>CC</u> AGCCCCG <u>GGG</u>	4
PAX6 gRNA off-5	chr19:40235907 to 40235929	<u>IGCCCC</u> AT <u>ITTC</u> AAGCCCC <u>AGG</u>	4

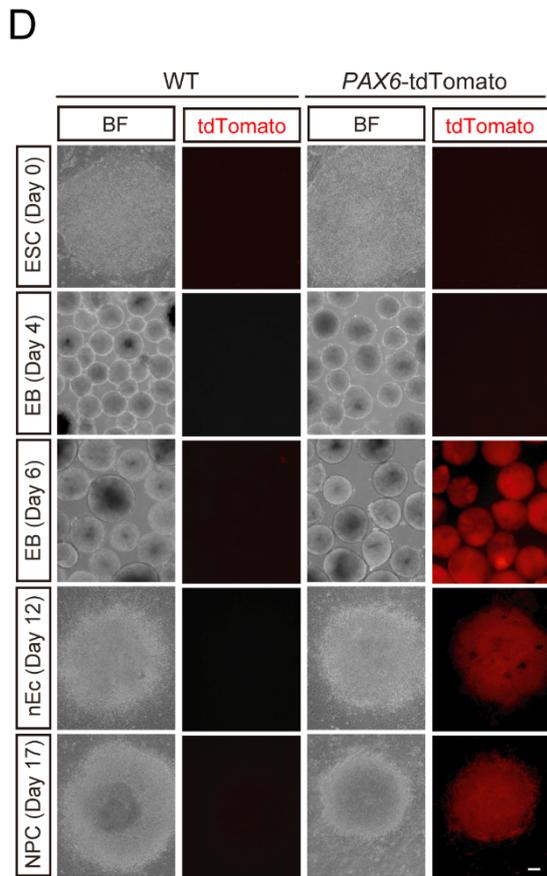
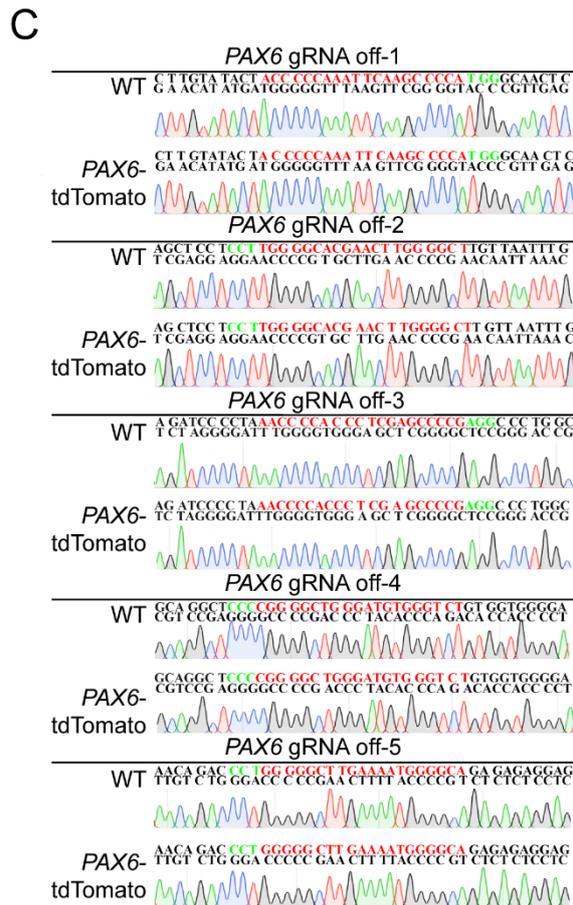


Figure S1. Targeting gRNA design, cleavage efficacy verification, and characterization of the *PAX6*-tdTomato reporter line. Related to Figure 1. (A) A gRNA targeting the start codon (ATG, labeled in red) region of human *PAX6* gene were designed. The gRNA sequences were boxed in blue. To evaluate the cleavage efficacy of the designed gRNA, it was transiently transfected into HEK 293FT cells together with the Cas9 expressing plasmid. After transfection for 3 days, the genomic DNA was extracted, PCR amplified and sent to Sanger sequencing. Overlapped peaks surrounding the targeting site represented non-homologous end joining repair after correct DNA cleavage. **(B)** Top 5 possible off target sites were predicted through the website (<http://crispor.tefor.net/>). **(C)** Sanger sequencing showed all 5 predicted loci were identified with Sanger sequencing showed all 5 predicted loci were identified with no off-target mutations. **(D)** Bright field (BF) and tdTomato fluorescent images of WT or *PAX6*-tdTomato hESCs along with neural differentiation. tdTomato started to express in the reporter line from day 6-10, when a neuroectoderm fate was specified. EB, embryoid body; nEc, neuroectoderm; NPC, neural progenitor. Scale bar, 100 μ m.

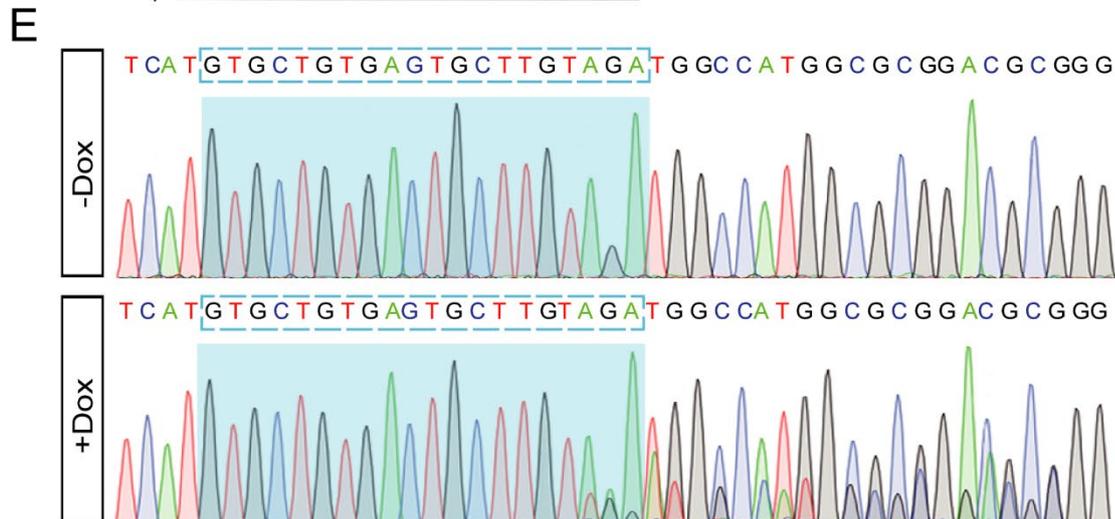
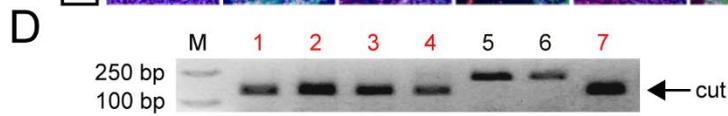
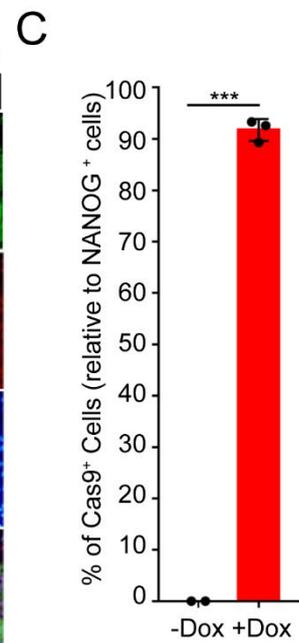
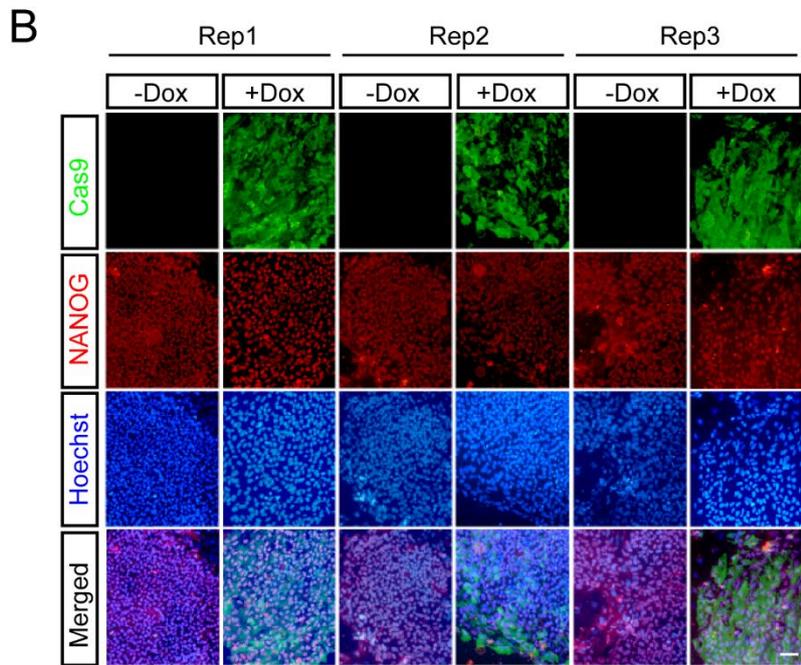
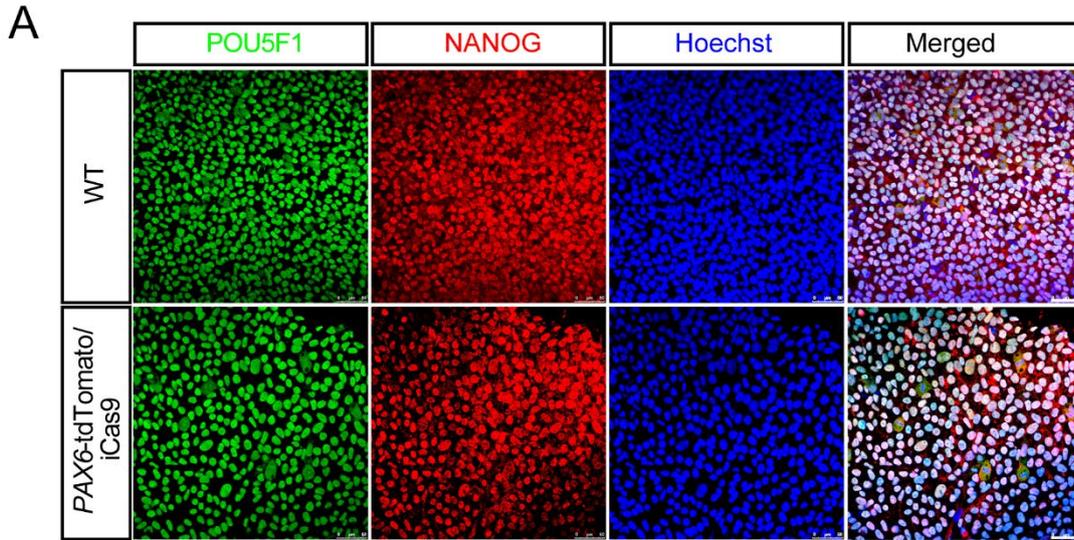


Figure S2. Characterization of the *PAX6*-tdTomato/iCas9 hESCs. Related to Figure 2. (A) Immunostaining images showed that *PAX6*-tdTomato/iCas9 hESCs remained pluripotency as they had typical POU5F1 (OCT4) and NANOG expression. Scale bars, 50 μ m. **(B)** Immunostaining images showed that most of the cells were double positive for NANOG and Cas9 after doxycycline treatment, Scale bar, 50 μ m. **(C)** Quantification of the percentage of NANOG and Cas9 double positive cells. Data are presented as mean \pm SEM of three independent experiments, **** p <0.0001, unpaired two-tailed Student's *t* test. **(D)** Two gRNAs targeting *NFI* were designed and electroporated into *PAX6*-tdTomato/iCas9 hESCs. Genomic DNA PCR showed that after treatment with doxycycline, there were ~70% (5/7) knockout efficiency, suggesting the robust cleavage efficacy of the inducible Cas9 system. **(E)** A gRNA targeting *P53* gene was designed and packaged into lentivirus. After lentiviral infection of the *PAX6*-tdTomato/iCas9 hESCs, doxycycline was used to induce Cas9 expression. Genomic DNA PCR followed by Sanger sequencing revealed prominent occurrence of indels surrounding the gRNA-guided cleavage site.

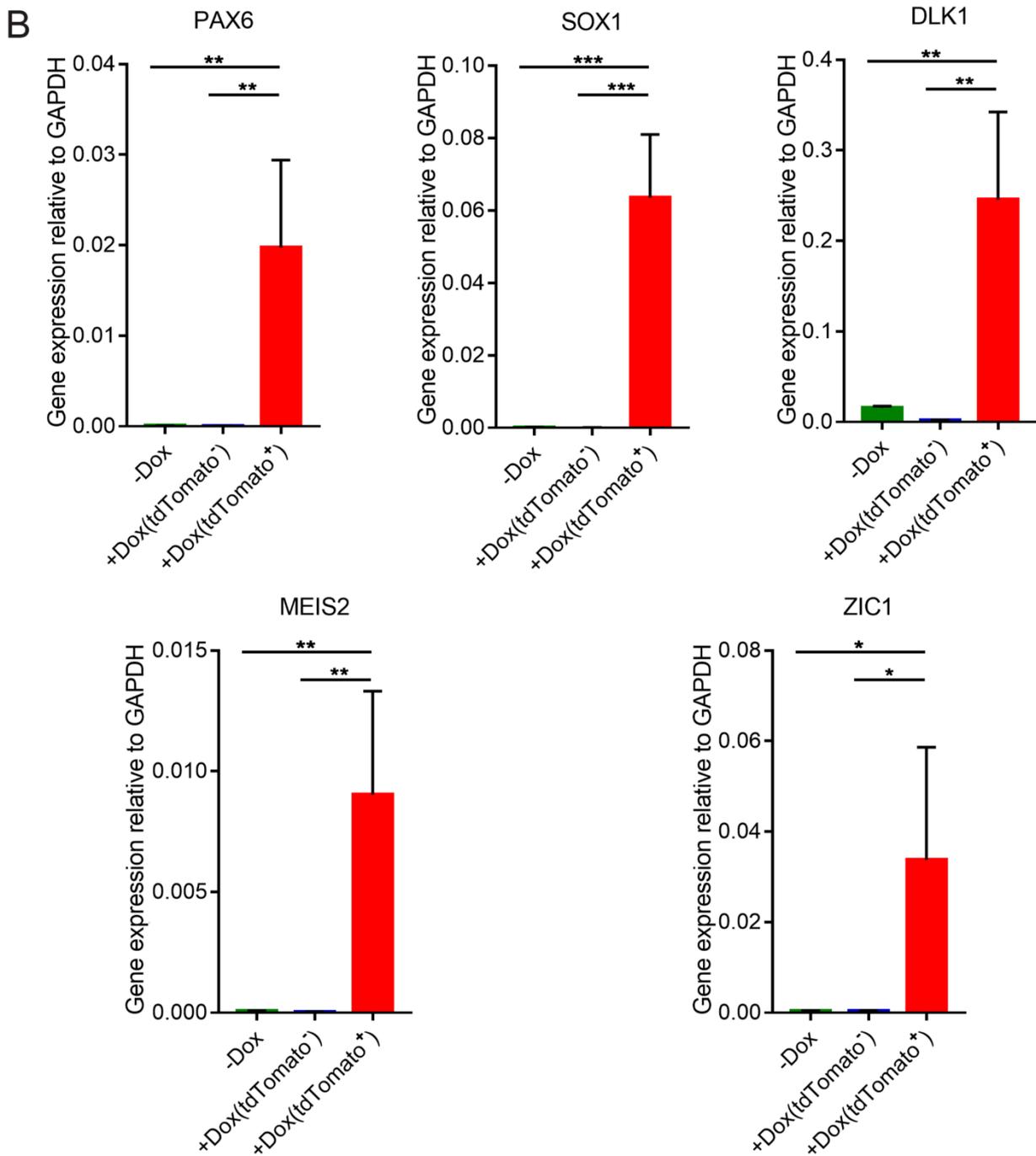
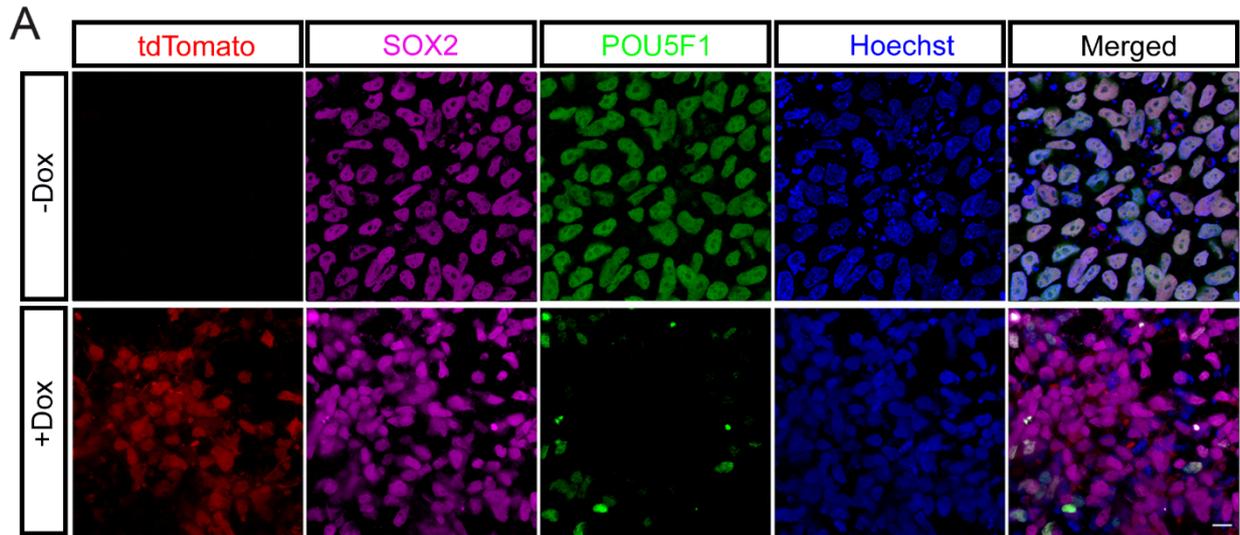


Figure S3. Characterization of the transformed tdTomato-positive cells. Related to Figure 3.

(A) After infection of the *PAX6*-tdTomato/iCas9 hESCs by gRNA libraries, these cells expressed pluripotent markers SOX2 and OCT4, and tdTomato was not expressed in the absence of doxycycline treatment. Doxycycline treatment induced clustered tdTomato expression in some hPSC colonies, and the tdTomato-positive cells showed complete absence of OCT4, but strong expression of nEc marker SOX2. Scale bar, 10 μ m. **(B)** FACS-enriched tdTomato-positive cells exhibited high levels of nEc genes including *PAX6*, *SOX1*, *DLK1*, *MEIS2* and *ZIC1*. Data are presented as Mean \pm SEM. Student's *t* test: $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

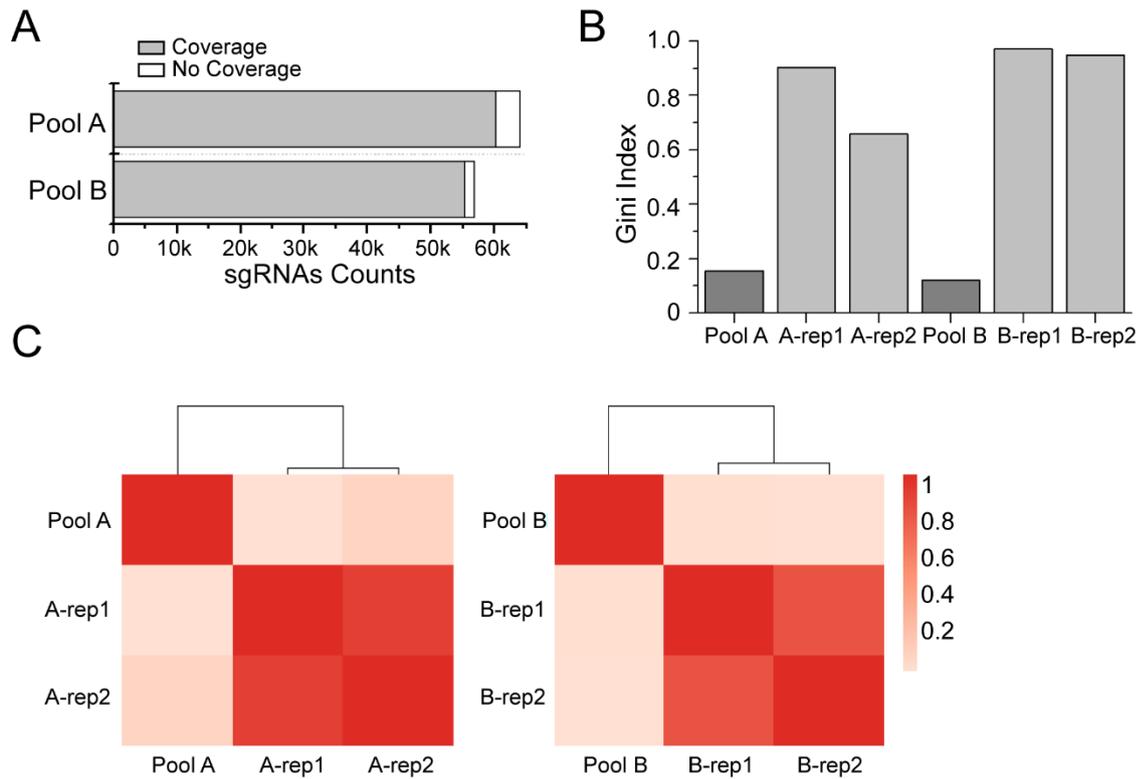


Figure S4. Performance evaluation of the CRISPR/Cas9 screening system. Related to Figure 4. (A) The coverage of gRNA pools revealed by PCR amplification and deep sequencing. (B) Gini index was calculated by the count module of MAGeCK tool to evaluate the level of disparities of the distribution of gRNAs retrieved from either pools or FACS enriched tdTomato cells. (C) Pearson correlation demonstrated the high repeatability between independent screenings of both A and B pool.

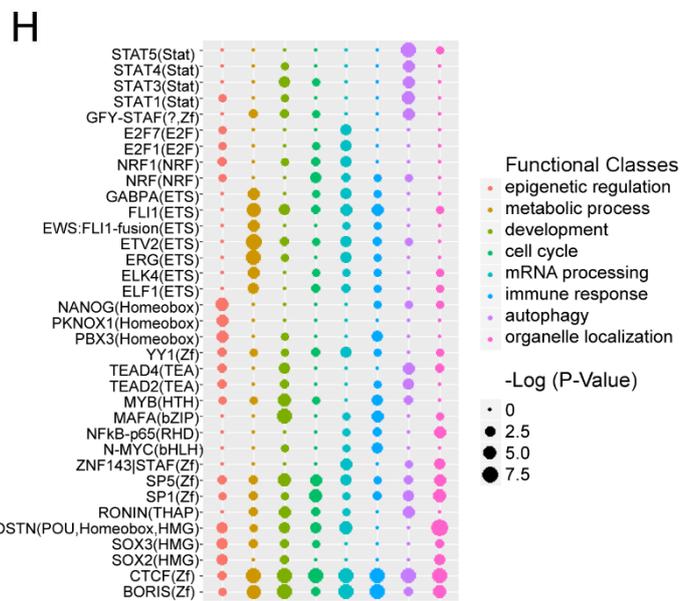
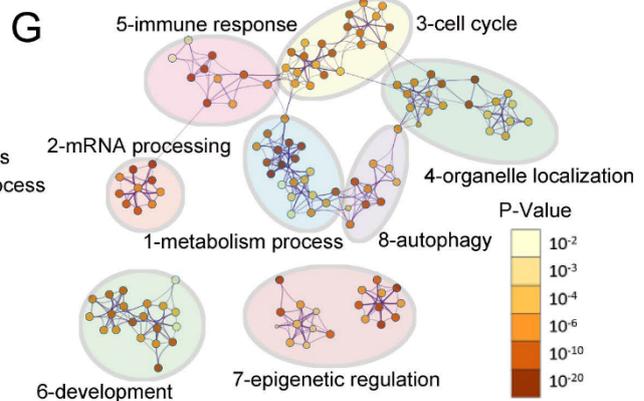
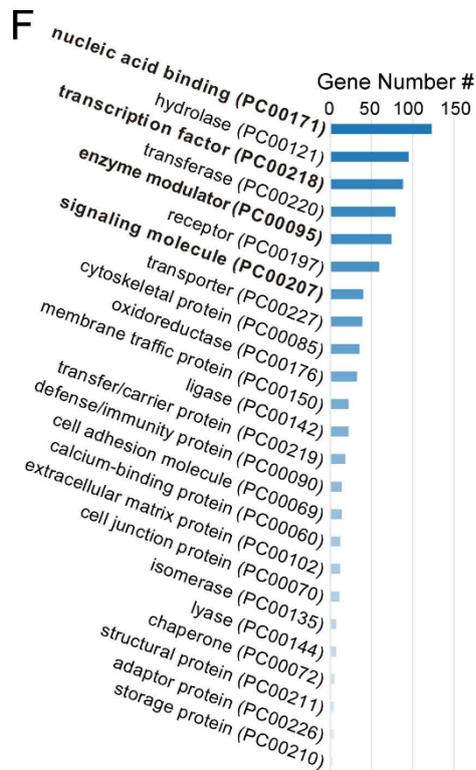
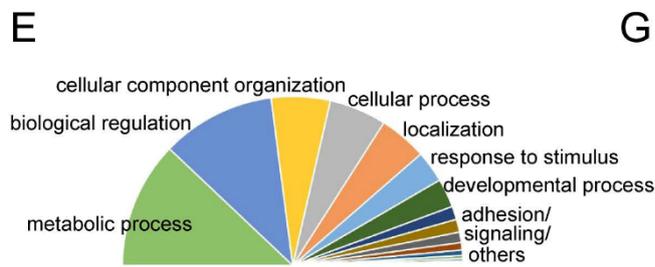
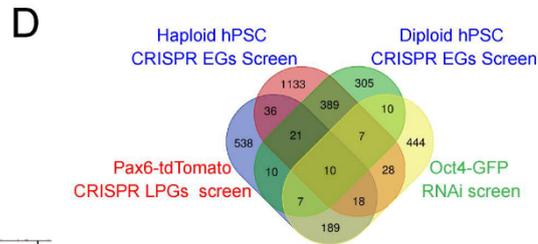
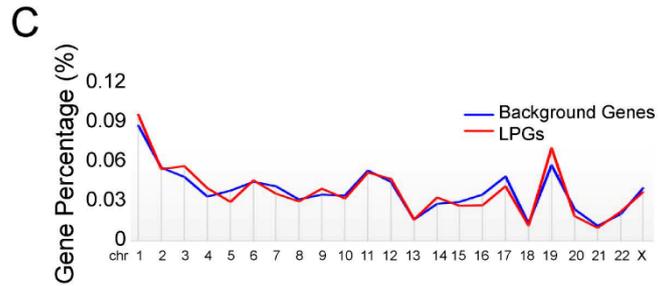
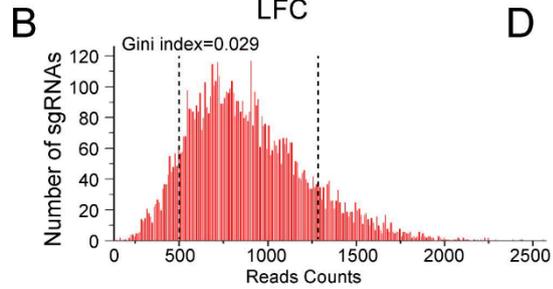
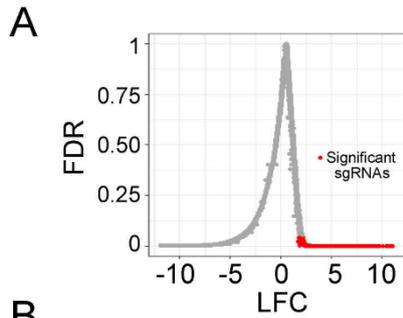


Figure S5. Characterization and functional annotation of lineage-specification preventing genes uncovered by genome-scale CRISPR/Cas9 screening. Related to Figure 4. (A) MAGeCK algorithm was used to estimate the statistical significance of enriched gRNA in tdTomato-positive cells after doxycycline induction. gRNAs had an over 2-fold change (LFC over 1) and false discovery rate (FDR) below 0.05 were considered statistically significant, and were labeled in red. (B) The designed sub-pool library showed a low Gini index value at 0.029. (C) Identified LPGs showed no chromosomal distribution preference as compared with genome background. (D) Number of overlapped genes of LPGs and reported essential genes (EGs) related to hPSC proliferation and survival or hPSC self renewal-required genes. (E) Biological processes analysis by using PATHER showed enrichment of biological functions of LPGs in metabolic process, biological regulation and cellular component organization. (F) Gene annotation by using PATHER showed LPGs were largely related to nucleic acid binding, transcription factors, and enzyme modulators. (G) LPGs clustered into functional modules as analyzed by the Metascape tool. Enriched terms retrieved from GO, KEGG pathway and Reactome Gene Sets of all LPGs were assigned to modules based on Kappa-statistical similarities among their gene memberships. 8 modules were defined and each module represented a group of similar functional categories, and edges were connected where terms with similarity above 0.3. Terms showed p-values lower than 0.01 were displayed, and the size of nodes represented number of enrichment of genes. (H) DHSs of LPGs categorized in each functional module were subjected to motif analysis through HOMER. LPGs of all 8 modules shared common binding motifs for transcription factors, including *CTCF*, *SPI* and *BORIS* (Z-finger proteins), and *OCT4/NANOG/SOX2* core pluripotency factors. LPGs of autophagy module showed more preferential transcriptional regulation by JAK-STAT pathway. ETS family transcription factor binding motifs, however, were largely enriched in DHSs of LPGs categorized in the metabolism process module.

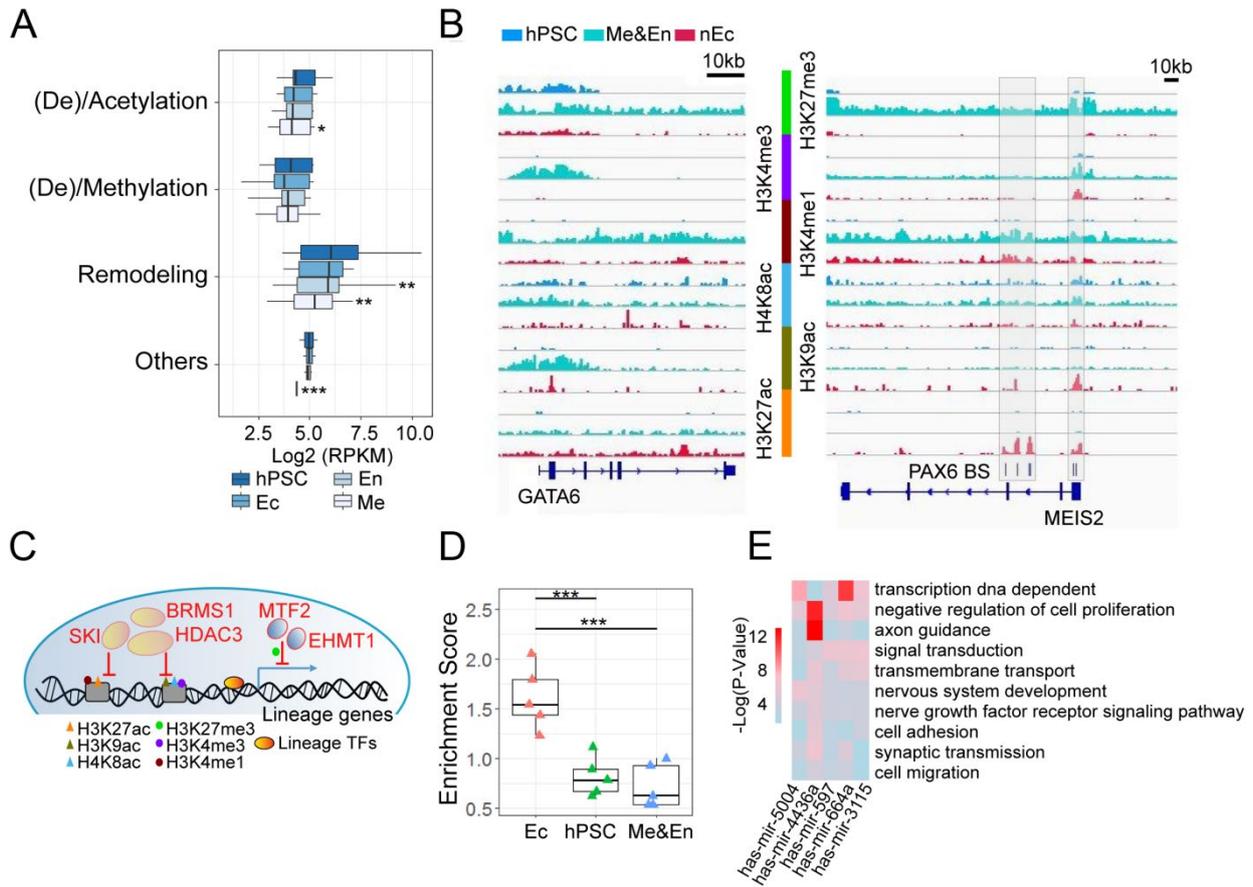


Figure S6. Epigenome characterization of hPSCs and differentiated trilineage. Related to Figure 5. (A) LPGs in the functional module of epigenetic regulation were mostly categorized into families of histone acetylation or de-acetylation, methylation or de-methylation, chromatin remodelers, and others related to ubiquitination or phosphorylation. Boxplot showed that LPGs belonging to each category were down-regulated during trilineage specification. Unpaired *t* test, **p* < 0.5, ***p* < 0.01, ****p* < 0.001. (B) Genome browser view of ChIP-seq tracks of various HMs as well as PAX6 binding for *GATA6* (an En representative gene) or *MEIS2* (a nEc representative gene). Grey box indicated that nEc determinant transcription factor PAX6 bound to the *MEIS2* gene loci, where active HMs gradually enriched during nEc specification from hPSCs. En specifier, *GATA6*, showed similar pattern of enrichment of active HMs in the promoter region during Me&En differentiation. (C) Modelling the functional pattern of epigenetic regulation during lineage specification of hPSCs. In hPSCs, lineage-specific genes harbor less active HMs but rich repressive HMs to restrict their transcription and maintain the pluripotent state. During trilineage differentiation, the epigenetic states of these lineage specification genes are either erased or rewritten, which offers an accessible epigenetic environment for lineage transcription factor

binding and active gene transcription. Epigenetic erasers or writers in LPGs were highlighted and labeled in red. **(D)** The top 5 miRNAs in LPGs were subjected to GO annotation based on their target genes. The GO annotation enrichment score in hPSC, Ec or Me&En categories of each miRNA was then calculated, respectively. Data are presented as mean value \pm standard deviation (SD). Unpaired *t* test, ****p* < 0.001. **(E)** GO analysis for the targeted genes of the top 5 miRNAs shown in LPGs demonstrated that they were preferentially nEc development related.

Supplemental table

Table S1: List of the top 100 LPGs. Related to Figure 4.

Gene	LFC	p-value	FDR	
<i>ICMT</i>	15.417	0	0.000124	
<i>TMC8</i>	15.289	0	0.000323	
<i>FBXW10</i>	14.425	0	0.000223	
<i>ACBD6</i>	14.0765	0	0.000422	
<i>NDUFA10</i>	13.9278	0	3.09E-05	
<i>LZTFL1</i>	13.9054	0	0.000968	
<i>MPDU1</i>	13.7926	0	0.001762	
<i>RWDD2B</i>	13.5787	0	0.001861	
<i>RNASE7</i>	13.2728	0	0.000472	
<i>S100A6</i>	13.0246	0	0.002505	
<i>RAB3A</i>	12.9256	0	0.003694	
<i>TPMT</i>	12.8477	0	0.000869	
<i>SUB1</i>	12.8221	0	0.002307	
<i>MAG</i>	12.8103	0	0.005969	
<i>GGCT</i>	12.6698	0	0.004931	
<i>DNAJB5</i>	12.6436	0	0.003793	
<i>PHF17</i>	12.6314	0	0.00305	
<i>GAS7</i>	12.5627	0	0.002059	
<i>ACTG2</i>	12.5345	0	0.00409	
<i>RCBTB1</i>	12.5048	0	0.009522	
<i>MCPH1</i>	12.4841	0	0.001036	
<i>TGM4</i>	12.483	0	0.008832	
<i>PALB2</i>	12.4546	0	0.003892	
<i>SF3A1</i>	12.2949	0	0.004189	
<i>NFASC</i>	12.276	0	0.00962	
<i>BTN2A2</i>	12.2449	0	0.008339	
<i>TCF7</i>	12.1475	0	0.006809	
<i>SAE1</i>	12.1059	0	0.003595	
<i>C2orf16</i>	12.0101	0	0.003298	
<i>ARL8A</i>	11.9762	0	0.006118	

<i>GDF11</i>	11.9591	0	0.000166	
<i>DHRS12</i>	11.9252	0	0.005129	
<i>SEMA3D</i>	11.861	0	0.000273	
<i>TMEM45B</i>	11.7938	0	0.009817	
<i>PRPF40A</i>	11.7455	0	0.004634	
<i>RGS16</i>	11.6805	0	0.004535	
<i>C5orf46</i>	11.64	0	0.006167	
<i>KRTAP6-3</i>	11.6385	0	0.006068	
<i>MTHFR</i>	11.4662	0	0.000172	
<i>POU5F1</i>	11.4097	0	0.000868	
<i>DLG4</i>	11.3839	0	0.002332	
<i>UFSP1</i>	11.263	0	0.007747	
<i>TBC1D12</i>	11.2552	0	0.008231	
<i>TBC1D22A</i>	11.2534	0	0.008345	
<i>C9orf3</i>	11.1813	0	0.003086	
<i>NME5</i>	11.1617	0	0.010113	
<i>UPRT</i>	11.1449	0	6.26E-05	
<i>FBXL2</i>	11.1032	0	0.001298	
<i>ZNF467</i>	11.0776	0	0.009128	
<i>SON</i>	11.0016	0	0.009769	
<i>FRS2</i>	10.9609	0	0.006448	
<i>TFPI</i>	10.954	0	0.000495	
<i>TMEM82</i>	10.9498	0	0.008437	
<i>ITIH1</i>	10.9304	0	0.000231	
<i>CRYGC</i>	10.8664	0	0.010064	
<i>PEX14</i>	10.7643	0	0.009916	
<i>TPPI</i>	10.7258	0	0.00587	
<i>CSF1R</i>	10.6781	0	0.004376	
<i>ITGAI1</i>	10.5561	0	0.008036	
<i>HYAL3</i>	10.4267	0	0.000198	
<i>UBAC1</i>	10.2691	0	0.005524	
<i>SMTNL2</i>	10.269	0	0.001216	
<i>FSD2</i>	10.1531	0	0.00173	
<i>FAM49A</i>	10.0049	0	0.000283	
<i>TUBA8</i>	9.9478	0	0.004647	
<i>KRTAP12-2</i>	9.9444	0	0.000638	
<i>PROZ</i>	9.9439	0	0.007401	
<i>C13orf35</i>	9.9324	2.07E-250	0.000693	
<i>CBLL1</i>	9.9275	1.85E-176	0.001326	
<i>GJD3</i>	9.8974	1.21E-277	0.000601	
<i>GPR78</i>	9.7785	1.02E-227	0.001093	
<i>NR1D2</i>	9.7254	4.71E-144	0.004577	
<i>C14orf79</i>	9.6542	8.49E-218	0.00116	
<i>GPR137C</i>	9.6525	1.45E-168	0.001653	

<i>PPFIA1</i>	9.5933	7.39E-242	0.002751	
<i>DGAT2L6</i>	9.5858	0	0.005116	
<i>ZNF711</i>	9.4445	5.35E-84	0.009615	
<i>IL21</i>	9.3884	0	0.005016	
<i>KIAA1958</i>	9.3593	1.41E-72	0.004107	
<i>ZBTB5</i>	9.29	0	0.003546	
<i>ZNF513</i>	9.1867	3.21E-160	0.008495	
<i>ORIQ1</i>	9.0188	1.20E-103	0.004655	
<i>PVRL1</i>	9.012	2.71E-166	0.006648	
<i>MIB1</i>	9.0115	2.10E-136	0.005234	
<i>PRDM14</i>	8.9661	1.99E-89	0.052331	
<i>OLFM1</i>	8.964	4.19E-95	0.004662	
<i>SCD</i>	8.9237	2.52E-42	0.003688	
<i>DSG3</i>	8.8235	6.39E-50	0.005757	
<i>ZBTB46</i>	8.7564	1.58E-146	0.004515	
<i>ACPL2</i>	8.5596	1.39E-124	0.000859	
<i>HDX</i>	8.4177	1.58E-50	0.00432	
<i>HAT1</i>	8.3975	7.97E-46	0.006027	
<i>EP300</i>	8.3869	1.24E-97	0.004939	
<i>ESAM</i>	8.2473	9.79E-23	0.00144	
<i>ZNF138</i>	7.9516	1.62E-51	0.005761	
<i>ELTD1</i>	7.8844	2.58E-29	4.12E-05	
<i>MMP15</i>	7.8773	1.43E-89	0.005228	
<i>ADA</i>	7.8278	1.43E-30	0.001537	
<i>T</i>	7.8049	0	0.001539	
<i>PPAPDC2</i>	7.3005	1.00E-11	0.007956	

TRANSPARENT METHODS

Human ESC (hESC) Culture

hESCs (WA09, Passages 25-45, WiCell Agreement No.14-W0377) were cultured on a feeder layer of irradiated mouse embryonic fibroblasts (MEF) as described previously (Zhang and Zhang, 2010; Chi et al., 2016; Zhang et al., 2001). Culture medium for hESCs (hESCM) was: DMEM/F12, 20% knockout serum replacer, 1×Non-Essential Amino Acids (NEAA), 1×GlutaMAX, 0.1 mM β-mercaptoethanol, supplied with 4 ng/ml FGF2. Cells were passaged every 5 days with the ratio of 1:6 with dispase (Gibco, 17105) digestion.

Neural Differentiation

Detailed procedure for embryoid body (EB)-based differentiation of hPSCs toward a PAX6 positive nEc and cortical neural progenitors was described previously (Zhang and Zhang, 2010; Chi et al., 2016; Zhang et al., 2001). Briefly, hESCs were digested with dispase, lifted from the MEF layers, and triturated into 100-200 μ m pieces. EBs formed after 4 days of suspension culture in hESCM. EBs were then transferred into the neural induction medium (NIM) (DMEM/F12 with 1 \times NEAA, 1 \times N2 supplement and 2 mg/ml heparin) for another two days and then plated on laminin-coated culture surfaces in NIM. nEc formed at day 6-10 post differentiation.

Lentivirus Packaging

HEK 293FT cells (Invitrogen) of 70% to 80% confluence in a 10 cm dish were transiently transfected with 10 μ g gRNA lentiviral vectors, 5 μ g VSVG and 7.5 μ g PAX2 via the calcium phosphate precipitation method. Fresh DMEM with 10% FBS was then supplied 14 hr after transfection. 48 hr later, lentivirus particles-containing supernatant was harvested and filtered through a 0.45 μ m cellulose acetate filter, followed by ultracentrifugation (Beckman) with a speed of 55000 g at 16 $^{\circ}$ C for 3 hours. Lentivirus particles-containing pellet was reconstituted in hESCM.

Construction of the *PAX6*-tdTomato Reporter Cell Line

A donor plasmid containing the 5' homology arm of *PAX6* (849bp), the neomycin (Neo)-P2A-tdTomato-T2A-HSVtk expressing cassette, the 3' homology arm of *PAX6* (820bp) in sequential order was constructed. Primer sets for amplifying the 5' and 3' homology arms were as follows: 5'-CTCAGCTCTTGGCCTCTACTCCTTA-3' (5'arm-F), 5'-GCTGGCTCTGGCTGGGGGCC-3' (5'arm-R); 5'-CAGAACAGTAAGTGCCTCTGGTCT-3' (3'arm-F), 5'-AGGCTCCCAGGTCGGAGCTCTAGA-3' (3'arm-R).

The gRNA GCCCCATATTCGAGCCCCG sequence, targeting a region close to the ATG start

codon of human *PAX6* gene was designed through the Zhanglab website (<https://zlab.bio/guide-design-resources>) and cloned into the gRNA expressing vector as previously described (Chen et al., 2018; Chi et al., 2016). Cleavage efficiency of the designed gRNA was tested in HEK 293FT cells through co-transfection with Cas9 expressing plasmids (Addgene #44719). After 60~65 hr of transient transfection, genomic DNA was extracted and the DNA fragments surrounding the cleavage site of *PAX6* gene was PCR amplified for Sanger sequencing. Overlapped peaks indicate efficient double strand breaks made by the cleavage followed by indels generated after DNA repair.

The method to generate *PAX6*-tdTomato HR line in hPSCs was similar to that of *PAX6*-Cre or *FOXA2*-GFP lines as we previously described (Chi et al., 2016; Chen et al., 2018). hESCs were pretreated with 1 mM Y27632, a Rho kinase (ROCK)-inhibitor, for at least 3 hr, and digested into single cells with pre-warmed trypsin. Then the cells were electroporated with 5 µg CAG-Cas9, 5 µg gRNA, 40 µg Neo-P2A-tdTomato-T2A-HSVtk donor, and 5 µg EF1α-puromycin plasmid in 200 µl of electroporation buffer (5 mM KCl, 25 mM MgCl₂, 15 mM HEPES, 102.94 mM Na₂HPO₄, 47.06 mM NaH₂PO₄, pH 7.2) by using a Gene Pulser Xcell system (Bio-Rad) at 250 V, 500 mF in a 0.4-cm cuvette (Phenix Research Products). Electroporated cells were plated on MEF feeder layer supplied with 1 mM Y27632 for 1 day in hESCM. Puromycin (0.5 µg/ml) was treated from day 2 to day 4 post electroporation for positive selection. Positive HR lines were validated by genomic DNA PCR. Primer sets for wild-type alleles are as follows: 5'-CTTCCCCTGGTCTCCAAACTTCAG-3' (*PAX6*-F1), 5'-TCTCCAGTATCGAGAAGAGCCA-3' (*PAX6*-R1). Primer sets for HR alleles are as follows: *PAX6*-F1, 5'-AGTGACAACGTCGAGCACAGCTG-3' (Neo-R).

Construction of *PAX6*-tdTomato/iCas9 Line

AAVS1-CAG-M2rtTA donor was directly modified from AAVS1-Neo-M2rtTA plasmid (Addgene #60843) by removing the SA-Neo expression cassette. AAVS1-3×TRE-FLAG-Cas9 donor plasmid was modified from Puro-Cas9 plasmid (Addgene #58409) by removing the SA-puromycin cassette. *PAX6*-tdTomato reporter hESCs were electroporated with 5 µg each left and right TALEN plasmids (Addgene #59025, Addgene #59026), 20 µg each donor plasmids (AAVS1-CAG-M2rtTA, AAVS1-3×TRE-FLAG-Cas9), and 5 µg EF1α-puromycin plasmids at 250 V, 500 mF. Puromycin was supplied from day 2 to day 4 post electroporation for positive selection. HR lines were validated by genomic DNA PCR. Primer sets for the wild-type alleles are as follows: 5'-CTCCGCATTGGAGTCGCTTTA-3' (F1), 5'-ACAGGAGGTGGGGGTTAGAC-3' (R1). Primer sets for HR alleles of CAG-M2rtTA are as follows: 5'-ACCGTAAGTTATGTAACGCGGA-3' (F2), 5'-CTGGCCATTGTCACTTTGCG-3' (R2). Primer sets for HR alleles of 3×TRE-FLAG-Cas9 are as follows: 5'-CTCTTCCGATGTTGAGCCCC-3' (F3), 5'-AGCAATAGCATCACAAATTTACAA-3' (R3). Validated *PAX6*-tdTomato/iCas9 line was treated with 1 µg/ml doxycycline (Sigma) for 48 hr, and immunostaining and Western blot analyses were further performed to confirm the inducible expression of FLAG-tagged Cas9.

To test the genome editing efficacy of the CRISPR/iCas9 system, we designed paired gRNAs for *NF1*, 5'-TTGTGCTCAGTACTGACTT-3' (*NF1*-gRNA-1), 5'-ATTCTTTAAAATAGTAGTG-3' (*NF1*-gRNA-2). 5 µg each of *NF1*-gRNA-1 as well as *NF1*-gRNA-2 plasmids were electroporated into *PAX6*-tdTomato/iCas9 line together with 5 µg EF1α-puromycin vector. Cells were then shortly selected with puromycin and Cas9 expression was induced by 1 µg/ml doxycycline. Remained colonies after two weeks continuous culture were subjected to genomic DNA PCR with the primer sets: 5'-TCGTTTTTAAGGATAAGCTGTTAACG-3' (*NF1*-F), 5'-AGCAAATTCCCCAAAACACAGTAAC-3' (*NF1*-R). Wild type *NF1* allele leads to a PCR

product of 232bp, while the mutated allele (direct ligation of the blunt ends caused by the two double strand breaks) leads to a PCR product of 156 bp. The gRNA targeting *P53* (*P53*-gRNA, 5'-GTGCTGTGACTGCTTGTAGA-3') were constructed in the lentiCRISPR v2 backbone (Addgene #52961) with the EF1 α -Cas9 cassette been replaced with EF1 α -blastcidin. Packaged *P53* gRNA lentiviral particles were then used to infect *PAX6*-tdTomato/iCas9 hPSCs, and 10 μ g/ml blastcidin (InvivoGen) was used for positive selection. Doxycycline was supplied for 3 days to induce the expression of Cas9, and the genomic DNA PCR followed by Sanger sequencing was performed to identify indels surrounds the *P53* cleavage site. Primer sets used for genomic DNA PCR are as follows: 5'-GCTCGCTAGTGGGTTGCAG-3' (*P53*-F), 5'-GTCATCCAAATACTCCACACGC-3' (*P53*-R).

Genome-wide CRISPR Screening

The lentiviral gRNA library for genome-wide CRISPR/Cas9 screening was purchased from Addgene (#1000000049), which contains 123,411 different gRNAs (A pool, 65,383 gRNAs; B pool, 58,028 gRNAs). The quality and coverage of the gRNA library was verified with next generation sequencing. Both A and B libraries were then packaged into lentiviral particles in HEK 293FT cells as described above, and the titer of the virus was evaluated through the survival rate of the infected cells with puromycin selection after serial dilution.

PAX6-tdTomato/iCas9 hESC were treated with dispase and triturated into small clumps. A total of 10^7 cells were subsequently infected with either A or B libraries of lentiviral particles at a multiplicity of infection (MOI) of 0.3. For either A or B pool, the transduced cells were then plated on 6 6-well plates on MEF layer. 24 hr after virus infection, cells were treated with puromycin (0.5 μ g/ml) for 4 days to eliminate non-infected cells. 5 plates were applied with doxycycline for 5 days

to induce Cas9 expression and gene loss-of-function, with 1 plate left untreated to serve as control. Both the control and doxycycline treated cells were passaged once at a ratio of 1:5. And when the cells became confluent, they were digested with pre-warmed trypsin to create a single cell suspension and tdTomato-positive cells were collected by FASC sorting (BD FACSVerse™ System, BD Biosciences).

Genomic DNA of the sorted tdTomato positive cells were extracted with QuickExtract DNA Extraction Solution (Epicentre, QE09050) and the gRNAs were amplified by two-step PCR. The primer sets to amplify lentiCRISPR gRNAs for the first round PCR are as follows: 5'-AATGGACTATCATATGCTTACCGTAACTTGAAAGTATTTTCG-3', 5'-CTTTAGTTTGTATGTCTGTTGCTATTATGTCTACTATTCTTTCC-3'. PCR reactions were setup with the PrimeSTAR GXL (Takara) system with the PCR program as 98°C 10s, 55°C 10s, 68°C 20s, 35 cycles. The PCR products were then gel purified and subjected to the second round PCR to add illumine adaptor and barcode. Primer sets for the second round PCR are as follows: 5'-AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTT AAGTAGAGGCTTTATATATCTTGTGGAAAGGACGAAACACC-3', 5'-CAAGCAGAAGACGGCATAACGAGATTCGCCTTGGTGAAGTTCAGACGTGTGCTCTTCCGATCTCCGACTCGGTGCCACTTTTTCAA-3'. Amplicons from the second round PCR were then gel purified, quantified, and sequenced using the pair-ended 150 bp sequencing protocol (Illumina Hiseq-x-ten system).

Sub-pool library validation screen

In order to construct a sub-pool gRNA library for validation, we modified the lentiCRISPR v2 (Addgene, #52961) by replacing Cas9 with blastocidin. 7,506 sgRNAs, including 7,324 selected hit

gRNAs from the primary screen targeting 3,941 genes, and 182 non-targeting control sgRNAs, were individually cloned into the modified lentiCRISPR v2 vector and verified by next generation sequencing.

Lentivirus was produced with the sub-pool library. *PAX6*-tdTomato/iCas9 hESCs were treated with dispase and triturated into small clumps. A total of 1.5×10^6 cells were subsequently infected with sub-pool lentiviral particles at a multiplicity of infection (MOI) of 0.3. 24 hr after viral infection, cells were treated with blastcidin (4 μ g/ml) for 4 days to eliminate non-infected cells. Cells were then supplied with doxycycline for 5 days to induce Cas9 expression and gene loss-of-function mutation. Both the untreated control and doxycycline treated cells were passaged once at a ratio of 1:5. And when the cells became confluent, they were digested with pre-warmed trypsin to create a single cell suspension and tdTomato-positive cells were collected by FACS sorting (BD FACSVerserTM System, BD Biosciences).

Genomic DNA of the sorted tdTomato positive cells were extracted and the gRNAs were amplified through PCR. The primer sets to amplify lentiCRISPR gRNAs for the first round PCR were: 5'-AATGGACTATCATATGCTTACCGTAACTTGAAAGTATTTTCG-3', 5'-CGGATCAATTGCCGACCCCTCCCCCAACTTCTCGGGGACTGTG-3'. The purified PCR products were performed second PCR to add illumina adaptor and barcode with the following primer sets: 5'-AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTTAAGTAGAGGCTTTATATATCTTGTGGAAAGGACGAAACACC-3', 5'-CAAGCAGAAGACGGCATAACGAGATTCGCCTTGGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCCGACTCGGTGCCACTTTTTCAA-3'. Amplicons from the second round PCR were then gel purified,

quantified, and sequenced using the pair-ended 150 bp sequencing protocol (Illumina Hiseq-x-ten system).

Immunocytochemistry

For immunocytochemistry analysis, coverslip cultured cells were fixed with 4% paraformaldehyde for 10 min at room temperature. After adequate washing with PBS, cells were incubated in a blocking buffer containing 10% donkey serum and 0.1% TritonX-100 for 1 hr at room temperature followed by primary antibody incubation at 4°C overnight. On the next day coverslips were washed with PBS and incubated with corresponding fluorescently conjugated secondary antibodies (1:1000, Jackson ImmunoResearch) for 1 hr at room temperature. Nuclei were counterstained with Hoechst 33258. Primary antibodies used in this study were: FLAG (1:500, mouse IgG, Sigma), PAX6 (1:1,000, mouse IgG, DSHB), tdTomato (1:200, rabbit IgG, ROCKLAND, DsRed), NANOG (1:1,000, goat IgG, R&D), SOX2 (1:1,000, goat IgG, R&D), OCT4 (1:1,000, mouse IgG, Santa Cruz).

mRNA extraction and Real-time PCR

Total RNA was isolated and reverse transcribed into cDNA using the EZ-press Cell to cDNA Kit PLUS II (EZBioscience, B0003C). The cDNA was subjected to real-time PCR using the TB GreenTM Premix Ex TaqTM II (Takara, PR820A). Primer oligonucleotides used for real-time PCR were as follows:

Gene	Forward Primer	Reverse Primer
<i>PAX6</i>	TCTTTGCTTGGGAAATCCG	CTGCCCGTTCAACATCCTTAG
<i>DLK1</i>	TCCTGAAGGTGTCCATGAAAG	GTGGTTGTAGCGCAGGTTG
<i>MEIS2</i>	CCAGGGGACTACGTTTCTCA	TAACATTGTGGGGCTCTGTG
<i>SOX1</i>	GTTTTTTGTAGTTGTTACCGC	GCATTTACAAGAAATAATAC

<i>ZIC1</i>	AGCCACGATGCTCCTGGACGC	TGGCCCAGGGCCGCAGCAG
<i>GAPDH</i>	GAAGGTGAAGGTCTGGAGTC	ATGGTGATGGGATTTC

Western Blot

Cells were lysed in RIPA buffer supplied with protein protease inhibitors overnight. Protein concentrations were measured with BCA kit (Thermo Scientific). 30 µg of total proteins were separated by SDS-PAGE, transferred to nitrocellulose membranes and blotted for FLAG (1:1,000, mouse IgG, Sigma, F3165) and β-Actin (1:5,000, mouse IgG, Sigma, A5316).

Southern Blot Analysis

The DNA probe targeting tdTomato sequence was PCR amplified from the *PAX6*-tdTomato donor vector with the PCR DIG Probe Synthesis Kit (Roche). Primer sets were as follows: 5'-TGGTGAGCAAGGGCGAGGAG-3' and 5'-TGCCGGTGCTGCCGGTGCCAT-3'. The efficacy of the DIG-labeled probe was evaluated with dot blot. 10 µg of genomic DNA extracted from *PAX6*-tdTomato hESCs was digested over-night with BsaI and subjected to agarose gel electrophoresis in TBE buffer. Gels were then denatured, neutralized, and transferred over night by capillarity on Hybond-N membranes (GE Healthcare) using 20×SSC transfer buffer. Hybridization was carried out overnight at 50°C and then was blocked with Roche blocking buffer. DIG signals were detected with an AP-conjugated DIG-Antibody (Roche) using CDP-Star (Roche) as a substrate for chemiluminescence.

Data Analysis for CRISPR/iCas9 Screening

We sequenced in pair-ended mode with 150-bp read length, first in read paired (R1) matched TAAGTAGAGGCTTTATATATCTTGTGGAAAGGACGAAACACCG (43 nt), followed by 20 nt gRNA (a), and GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAAC

TTGAAAAAGTGGCACCGAGTCGGAGATCGGAAGAGCA (87 nt); second in read paired (R2) matched CCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAAC (73 nt); followed by 20nt gRNA (a'), and CGGTGTTTCGTCCTTCCACAAGATATATAAAGCCTCTACTTAAGATCGGAAGAGCG (57 nt). For data analysis, the sequencing reads of gRNAs from different samples were first identified by barcode using cutadapt (version 1.11) with parameters $-u$ (43, -87 for R1; 73, -57 for R2). Bowtie2-build function of Bowtie2 (version 2.3.4.3) (Langmead and Salzberg, 2012) was applied on the gRNA sequences of GeCKO library to generate Burrows-Wheeler index. The gRNA sequences were then retrieved and counted by aligning processed reads of each sample to the gRNA library using Bowtie2 with default parameters and only the reads with unique alignment were reported. The count table for each gRNA was then normalized relative to the total number of reads in each of conditions and two libraries were combined. As a result, we got the normalized counts of each gRNA in control sample (unscreened cell library) and KO-screened sample respectively. The MAGeCK algorithm (Li et al., 2014) was used to estimate the statistical significance (using a negative binomial test) of enrichment for each gRNA in the KO-screened group compared to control group. LPGs were then identified by looking for genes whose gRNAs were ranked consistently higher (by significance) using robust rank aggregation (RRA). The negative controls were incorporated in the MAGeCK analysis to generate null distributions and calculate the p-value and FDR for each gene.

Sub-pool validation screen analysis

We sequenced in pair-ended mode with 150-bp read length, first in read paired (R1) matched TAAGTAGAGGCTTTATATATCTTGTGGAAAGGACGAAACACCG (43 nt), followed by 20 nt gRNA, and

TGCTCTTCCGATCTCCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTAT
TTTAACTTGCTATTTCTAGCTCTAAAAC (87 nt); second in read paired (R2) matched
CCAAGCAGAAGACGGCATAACGAGATtcGccttGGTGACTGGAGTTCAGACGTGTGCTCTT
CCGATCTCCGACTCGGTGCCACTTTTTCAA (73 nt); followed by 20nt gRNA, and
CGCTCTTCCGATCTTAAGTAGAGGCTTTATATATCTTGTGGAAAGGACGAAACACCG
(57 nt). For data analysis, the sequencing reads of gRNAs from different samples were first
identified by barcode using cutadapt (version 1.11) with parameters -u (43, -87 for R1; 73, -57 for
R2). With Bowtie2-build function, we constructed the gRNA sub-pool Burrows-Wheeler index.
The gRNA sequences were aligned to the gRNA sub-pool Burrows-Wheeler index with Bowtie2
(version 2.3.4.3) and the reads with unique alignment were used for the gRNAs reads counting.
Each gRNA were calculated as LFCs of gRNA counts between control sample (unscreened cell
library) and KO-screened sample. The gRNAs counts were normalized by comparing to median of
non-targeting control gRNAs. We averaged the LFCs of two replicates and genes with an over 2-
fold change and FDR below 0.05 were positive.

High-throughput Sequencing (NGS) Data Analysis

Data of DNase-seq (DNase I hypersensitive sites sequencing) used to identify the location of
regulatory regions was from Gene Expression Omnibus (GEO) database (GSM878612,
GSM878613). Regulatory motifs of the DHs of LPGs were predicted by using HOMER software
(Heinz et al., 2010). To depict the epigenetic regulation of the committed nEc fate from hESCs, we
performed the profiles of various histone modifications by using computeMatrix and plotProfile
commands of deepTools (Ramirez et al., 2014), and the ChIP-seq data was from Roadmap
Epigenomics project. To reveal the changes of mRNA expression levels of LPGs during different

lineage specification from hESCs, we download RNA-seq data from GEO database (Gifford et al., 2013). Expression levels for genes were normalized to Reads Per Kilobase per Million mapped reads (RPKM) and row z-scored when performing heatmap profiles.

Functional Annotation of Defined LPGs

We performed functional classification of LPGs by using the Panther Classification System (Mi et al., 2005). And then we assigned various functional categories to clusters of genes using Metascape (Tripathi et al., 2015) with default analysis parameters. In details, all the statistically enriched terms identified from Gene Ontology (GO), KEGG pathway and Reactome Gene Sets were hierarchically clustered into a tree based on Kappa-statistical similarities among their gene memberships. Clusters which represent a group of similar functional categories were defined by a kappa score of 0.3 as threshold. A subset of representative terms from each cluster was automatically selected by Metascape and converted into a network, where terms with similarity above 0.3 are connected by edges. More specifically, terms with the best p-values from each of the clusters were depicted as network nodes, with the constraint of having a maximum of 15 terms per cluster and 250 terms in total.

Gene Network Construction

The online database resource search tool (STRING) for retrieving interactive genes provides protein-protein interaction information, including the prediction and comparison of inter-genomic interactions (Franceschini et al., 2013). Only protein associations with combined confidence score above 0.5 were retained and used to assign weights to each network link. Different clusters of gene-encoded proteins were represented by nodes of different colors in the network, and line shape indicated the predicted mode of action between proteins referring to <https://string-db.org/> (version

10.5). We used GeneMANIA integrated with the Cytoscape network visualization (Shannon et al., 2003) to construct interactional network for epigenetic regulators. GeneMANIA identifies the most related genes to a query gene set using a guilt-by-association approach (Montoyo et al., 2014). In the network, colored nodes but not grey represented the genes we screened out and the lines indicated the interactions between genes.

Statistical Analysis

Chi-square (χ^2) test was used in [Figure 4E](#) to test the statistical significance of treat-group (LPGs) when compared to random or background group. Unpaired two-tailed Student's *t* test was used in [Figures S2C, S3B, 5D and S6A](#) for analyzing the statistical significance between every two-groups ($n \geq 3$ for each group), and data were represented as mean value \pm standard deviation (SD). Single, double and triple asterisks represented $P < 0.05$, 0.01 , and 0.001 , respectively, wherein. * $P < 0.05$ was considered statistically significant and ** $P < 0.01$ was extremely significant.

SUPPLEMENTAL REFERENCES

- CHEN, Z., REN, X., XU, X., ZHANG, X., HUI, Y., LIU, Z., SHI, L., FANG, Y., MA, L., LIU, Y., TERHEYDEN-KEIGHLEY, D., LIU, L. & ZHANG, X. 2018. Genetic Engineering of Human Embryonic Stem Cells for Precise Cell Fate Tracing during Human Lineage Development. *Stem Cell Reports*, 11, 1257-1271.
- CHI, L., FAN, B., ZHANG, K., DU, Y., LIU, Z., FANG, Y., CHEN, Z., REN, X., XU, X., JIANG, C., LI, S., MA, L., GAO, L., LIU, L. & ZHANG, X. 2016. Targeted Differentiation of Regional Ventral Neuroprogenitors and Related Neuronal Subtypes from Human Pluripotent Stem Cells. *Stem Cell Reports*, 7, 941-954.
- FRANCESCHINI, A., SZKLARCZYK, D., FRANKILD, S., KUHN, M., SIMONOVIC, M., ROTH, A., LIN, J., MINGUEZ, P., BORK, P., VON MERING, C. & JENSEN, L. J. 2013. STRING v9.1: protein-protein interaction networks, with increased coverage and integration. *Nucleic Acids Res*, 41, D808-15.
- GIFFORD, C. A., ZILLER, M. J., GU, H., TRAPNELL, C., DONAGHEY, J., TSANKOV, A., SHALEK, A. K., KELLEY, D. R., SHISHKIN, A. A., ISSNER, R., ZHANG, X., COYNE, M., FOSTEL, J. L., HOLMES, L., MELDRIM, J., GUTTMAN, M., EPSTEIN, C., PARK, H., KOHLBACHER, O., RINN, J., GNIRKE, A., LANDER, E. S., BERNSTEIN, B. E. & MEISSNER, A. 2013. Transcriptional and epigenetic dynamics during specification of human embryonic stem cells. *Cell*, 153, 1149-63.
- HEINZ, S., BENNER, C., SPANN, N., BERTOLINO, E., LIN, Y. C., LASLO, P., CHENG, J. X., MURRE, C., SINGH, H. & GLASS, C. K. 2010. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell*, 38, 576-89.
- LANGMEAD, B. & SALZBERG, S. L. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods*, 9, 357-9.

- LI, W., XU, H., XIAO, T., CONG, L., LOVE, M. I., ZHANG, F., IRIZARRY, R. A., LIU, J. S., BROWN, M. & LIU, X. S. 2014. MAGeCK enables robust identification of essential genes from genome-scale CRISPR/Cas9 knockout screens. *Genome Biol*, 15, 554.
- MI, H., LAZAREVA-ULITSKY, B., LOO, R., KEJARIWAL, A., VANDERGRIF, J., RABKIN, S., GUO, N., MURUGANUJAN, A., DOREMIEUX, O., CAMPBELL, M. J., KITANO, H. & THOMAS, P. D. 2005. The PANTHER database of protein families, subfamilies, functions and pathways. *Nucleic Acids Res*, 33, D284-8.
- MONTOJO, J., ZUBERI, K., RODRIGUEZ, H., BADER, G. D. & MORRIS, Q. 2014. GeneMANIA: Fast gene network construction and function prediction for Cytoscape. *F1000Res*, 3, 153.
- RAMIREZ, F., DUNDAR, F., DIEHL, S., GRUNING, B. A. & MANKE, T. 2014. deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res*, 42, W187-91.
- SHANNON, P., MARKIEL, A., OZIER, O., BALIGA, N. S., WANG, J. T., RAMAGE, D., AMIN, N., SCHWIKOWSKI, B. & IDEKER, T. 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*, 13, 2498-504.
- TRIPATHI, S., POHL, M. O., ZHOU, Y., RODRIGUEZ-FRANSEN, A., WANG, G., STEIN, D. A., MOULTON, H. M., DEJESUS, P., CHE, J., MULDER, L. C., YANGUEZ, E., ANDENMATTEN, D., PACHE, L., MANICASSAMY, B., ALBRECHT, R. A., GONZALEZ, M. G., NGUYEN, Q., BRASS, A., ELLEDGE, S., WHITE, M., SHAPIRA, S., HACHOEN, N., KARLAS, A., MEYER, T. F., SHALES, M., GATORANO, A., JOHNSON, J. R., JANG, G., JOHNSON, T., VERSCHUEREN, E., SANDERS, D., KROGAN, N., SHAW, M., KONIG, R., STERTZ, S., GARCIA-SASTRE, A. & CHANDA, S. K. 2015. Meta- and Orthogonal Integration of Influenza "OMICS" Data Defines a Role for UBR4 in Virus Budding. *Cell Host Microbe*, 18, 723-35.
- ZHANG, S. C., WERNIG, M., DUNCAN, I. D., BRUSTLE, O. & THOMSON, J. A. 2001. In vitro differentiation of transplantable neural precursors from human embryonic stem cells. *Nat Biotechnol*, 19, 1129-33.
- ZHANG, X. Q. & ZHANG, S. C. 2010. Differentiation of neural precursors and dopaminergic neurons from human embryonic stem cells. *Methods Mol Biol*, 584, 355-66.