iScience, Volume *24*

Supplemental Information

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

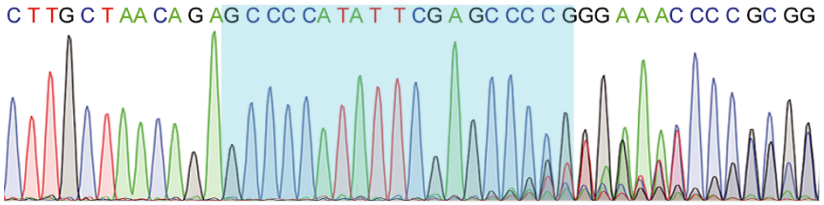
Xiangjie Xu, Yanhua Du, Lin Ma, Shuwei Zhang, Lei Shi, Zhenyu Chen, Zhongshu Zhou, Yi Hui, Yang Liu, Yujiang Fang, Beibei Fan, Zhongliang Liu, Nan Li, Shanshan Zhou, Cizhong Jiang, Ling Liu, and Xiaoqing Zhang

Supplemental information

A

PAX6 start codon:

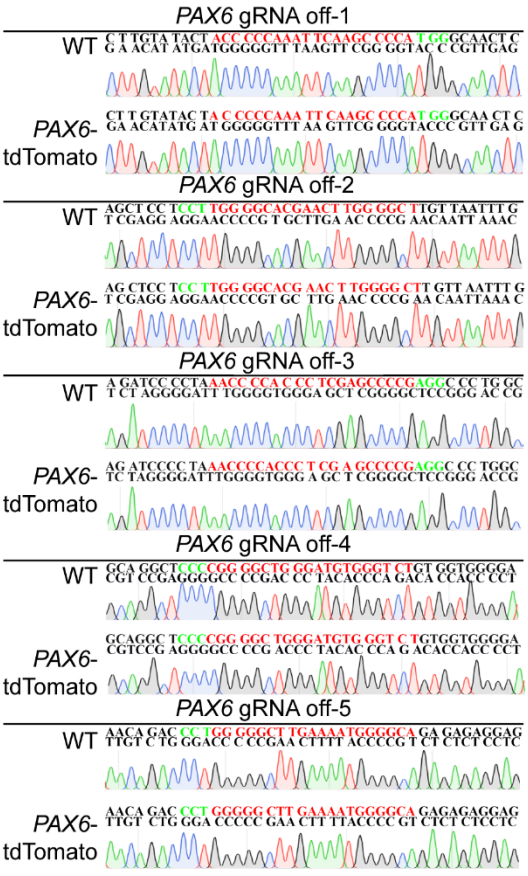
... ...AACAGAGCCCCATATTCGAGCCCCGTGGAATCCCGCGGCCCCCAGCCAGAGCCAGCATGCAGAA... ...



B

Match name	Coordinate	Guide sequence + PAM	Mismatches
PAX6 gRNA	chr11:31827964 to 31827986	AGCCCCATATTCGAGCCCCGTGG	4
PAX6 gRNA off-1	chr6:56740611 to 56740633	ACCCCCAAATTCAAGCCCCATGG	4
PAX6 gRNA off-2	chr4:12373222 to 12373244	AGCCCCAAGTTCGTGCCCCAAGG	4
PAX6 gRNA off-3	chr19:17393106 to 17393128	AACCCCACCCTCGAGCCCCGAGG	4
PAX6 gRNA off-4	chr8:141312272 to 141312294	AGACCCACATCCCAGCCCCGGGG	4
PAX6 gRNA off-5	chr19:40235907 to 40235929	TGCCCCATTTTCAAGCCCCCAGG	4

C



D

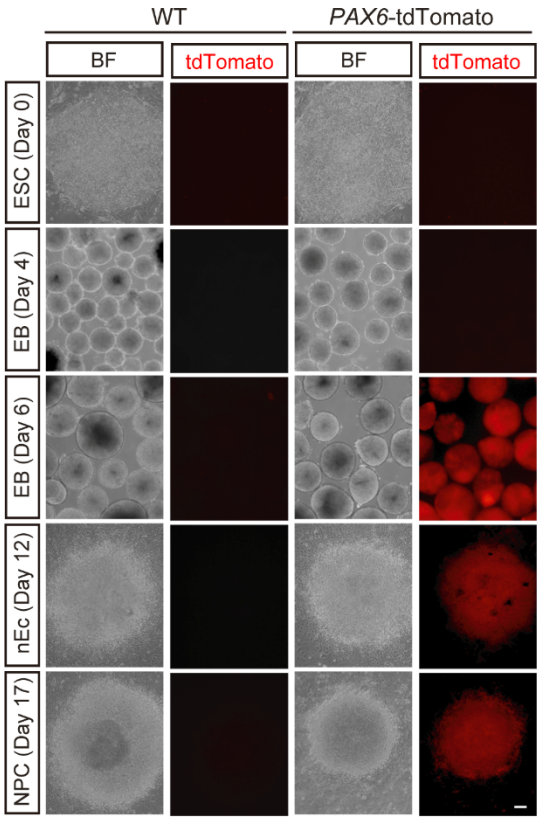


Figure S1. Targeting gRNA design, cleavage efficacy verification, and characterization of the *PAX6*-tdTomato reporter line. Related to Figure 1. (A) A gRNA targeting the start codon (ATG, labeled in red) region of human *PAX6* gene were designed. The gRNA sequences were boxed in blue. To evaluate the 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 μm.

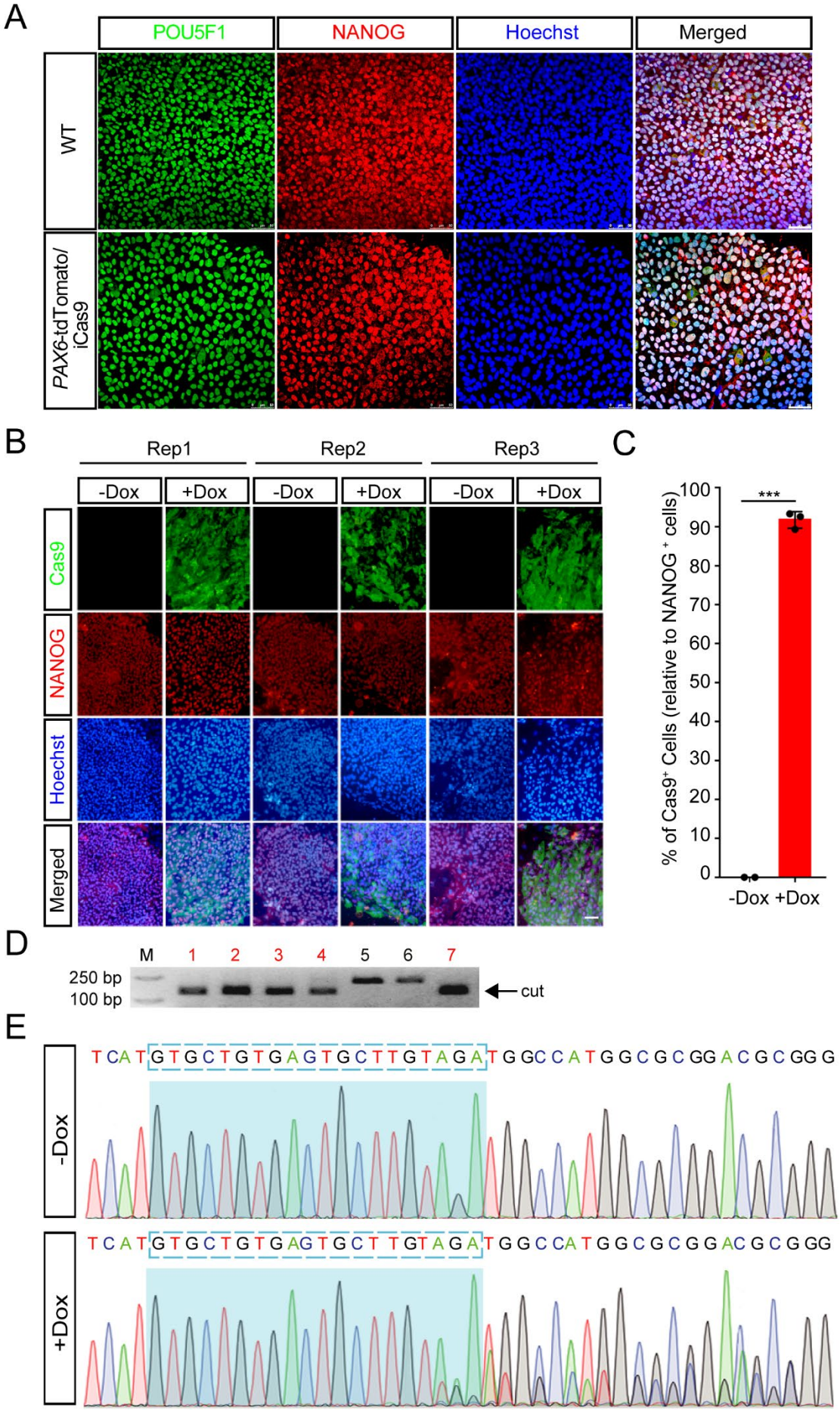
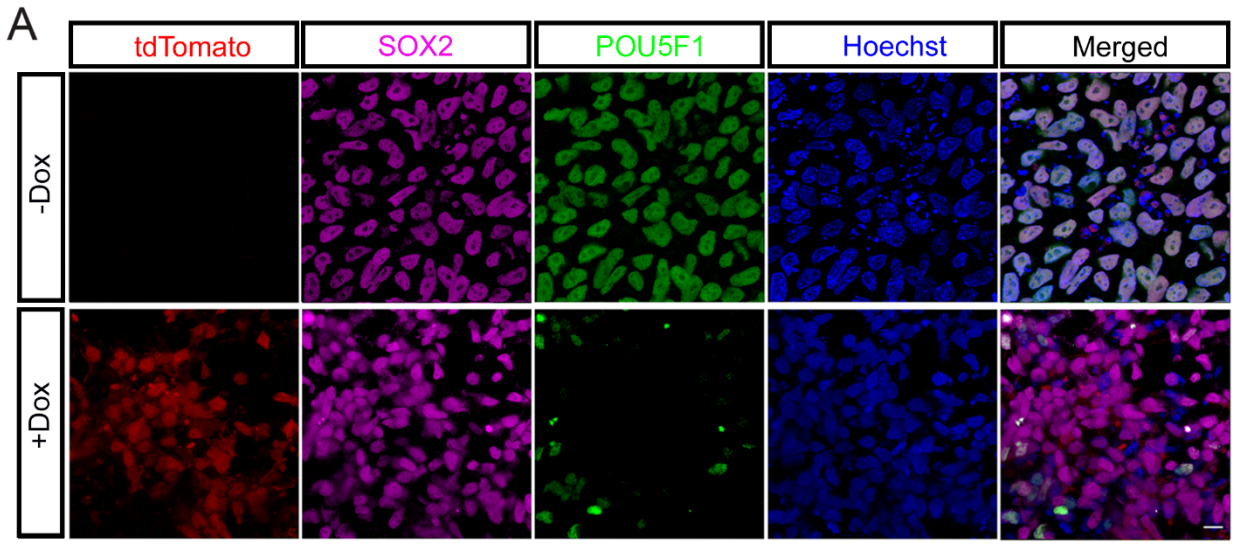

A
POU5F1
NANOG
Hoechst
Merged
WT
PAX6-tdTomato/ iCas9
B
Rep1
Rep2
Rep3
-Dox
+Dox
Cas9
NANOG
Hoechst
Merged
C
% of Cas9+ Cells (relative to NANOG + cells)
***
-Dox +Dox
D
M 1 2 3 4 5 6 7
250 bp
100 bp
cut
E
-Dox
+Dox
TCATGTGCTGTGAGTGCTTGTAGATGGCCATGGCGCGGACGCGGG
TCATGTGCTGTGAGTGCTTGTAGATGGCCATGGCGCGGACGCGGG

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 μm. **(B)** Immunostaining images showed that most of the cells were double positive for NANOG and Cas9 after doxycycline treatment, Scale bar, 50 μm. **(C)** Quantification of the percentage of NANOG and Cas9 double positive cells. Data are presented as mean±SEM of three independent eperiments, $^{****}p<0.0001$, unpaired two-tailed Student's *t* test. **(D)** Two gRNAs targeting *NF1* 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.


A
tdTomato
SOX2
POU5F1
Hoechst
Merged
-Dox
+Dox

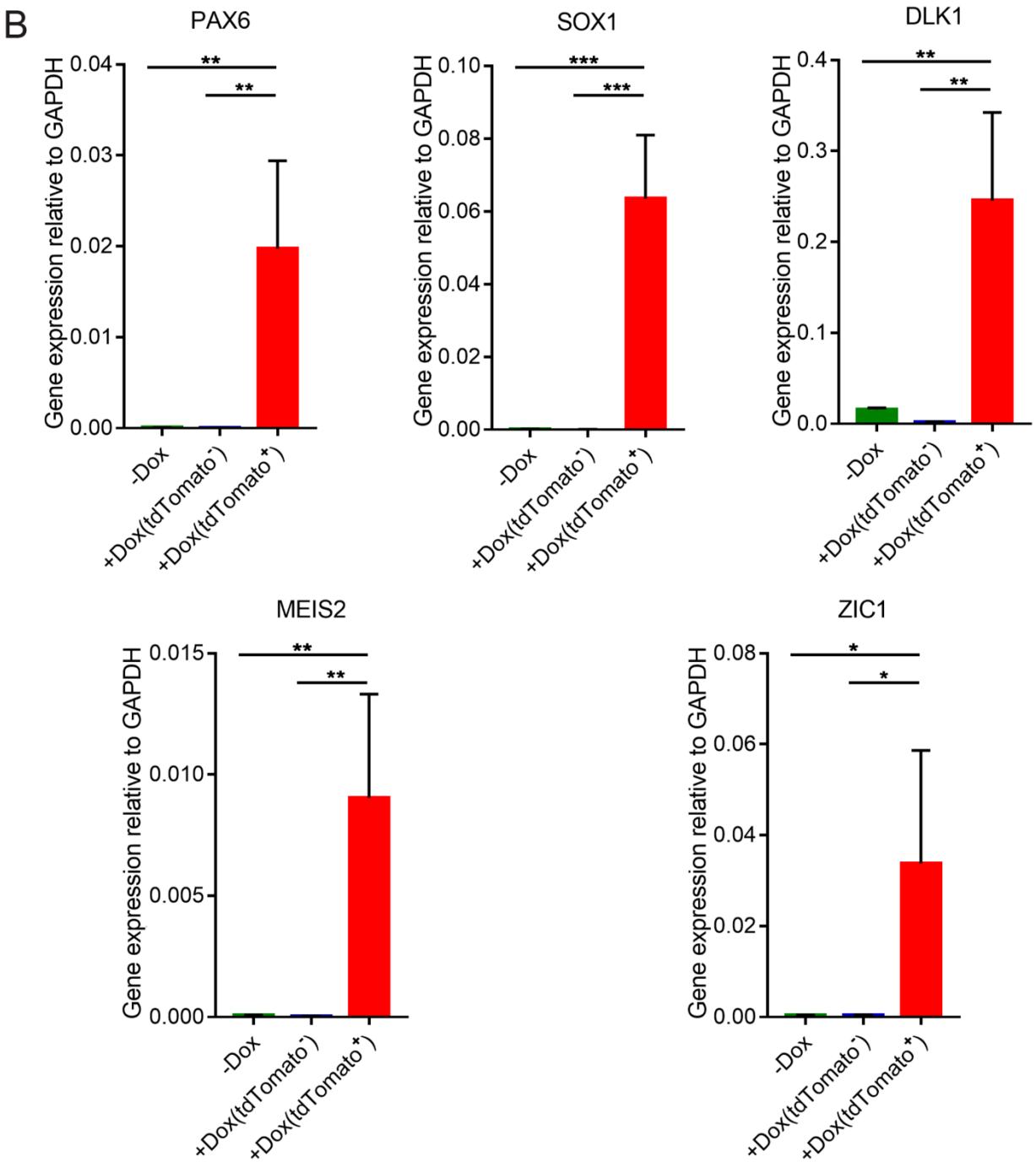

B
PAX6
SOX1
DLK1
MEIS2
ZIC1
Gene expression relative to GAPDH
-Dox
+Dox(tdTomato-)
+Dox(tdTomato+)

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 μm. **(B)** FACS-enriched tdTomato-positive cells exhibited high levels of nEc genes including *PAX6*, *SOX1*, *DLK1*, *MEIS2* and *ZIC1*. Data are presented as Mean±SEM. Student's *t* test: $p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

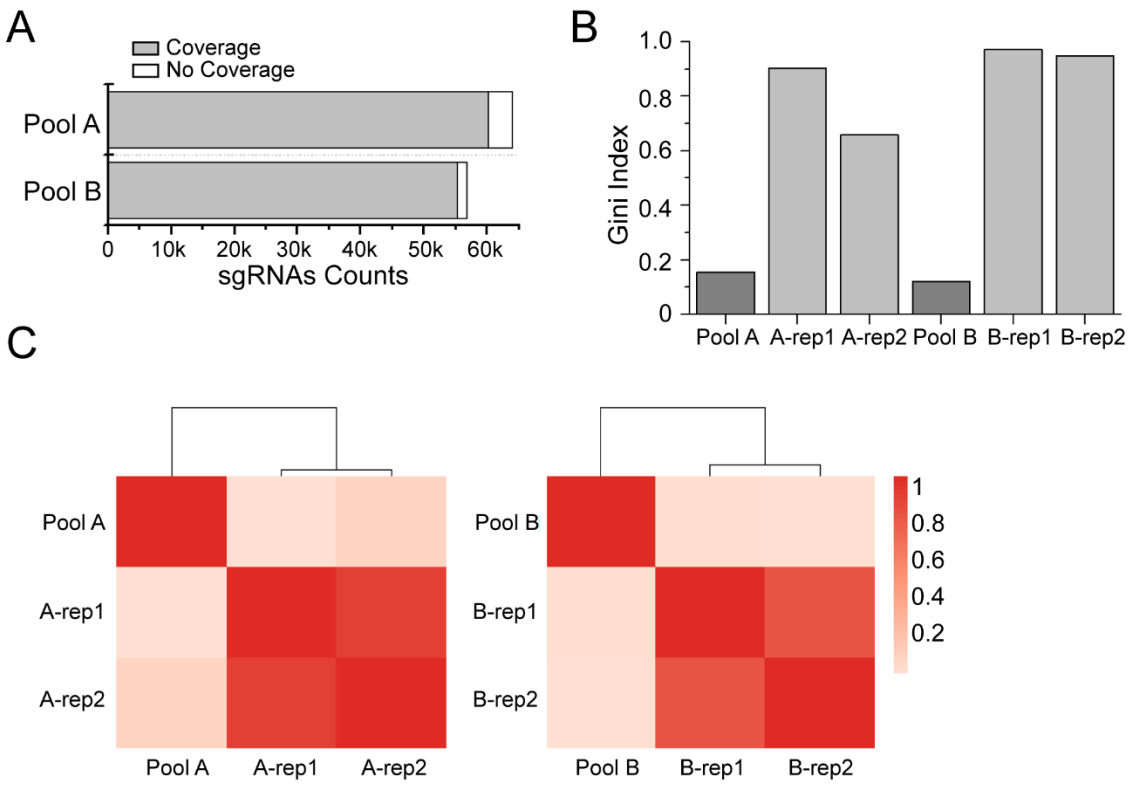


Figure S4. Performance evaluation of the CRISPR/Cas9 screening system. Related to Figure 4. (**A)** The coverage of gRNA pools revealed by PCR amplification and deep sequencing. **(B)** Gini index was calculated by the count module of MAGeCK tool to evaluate the level of disparities of the distribution 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.

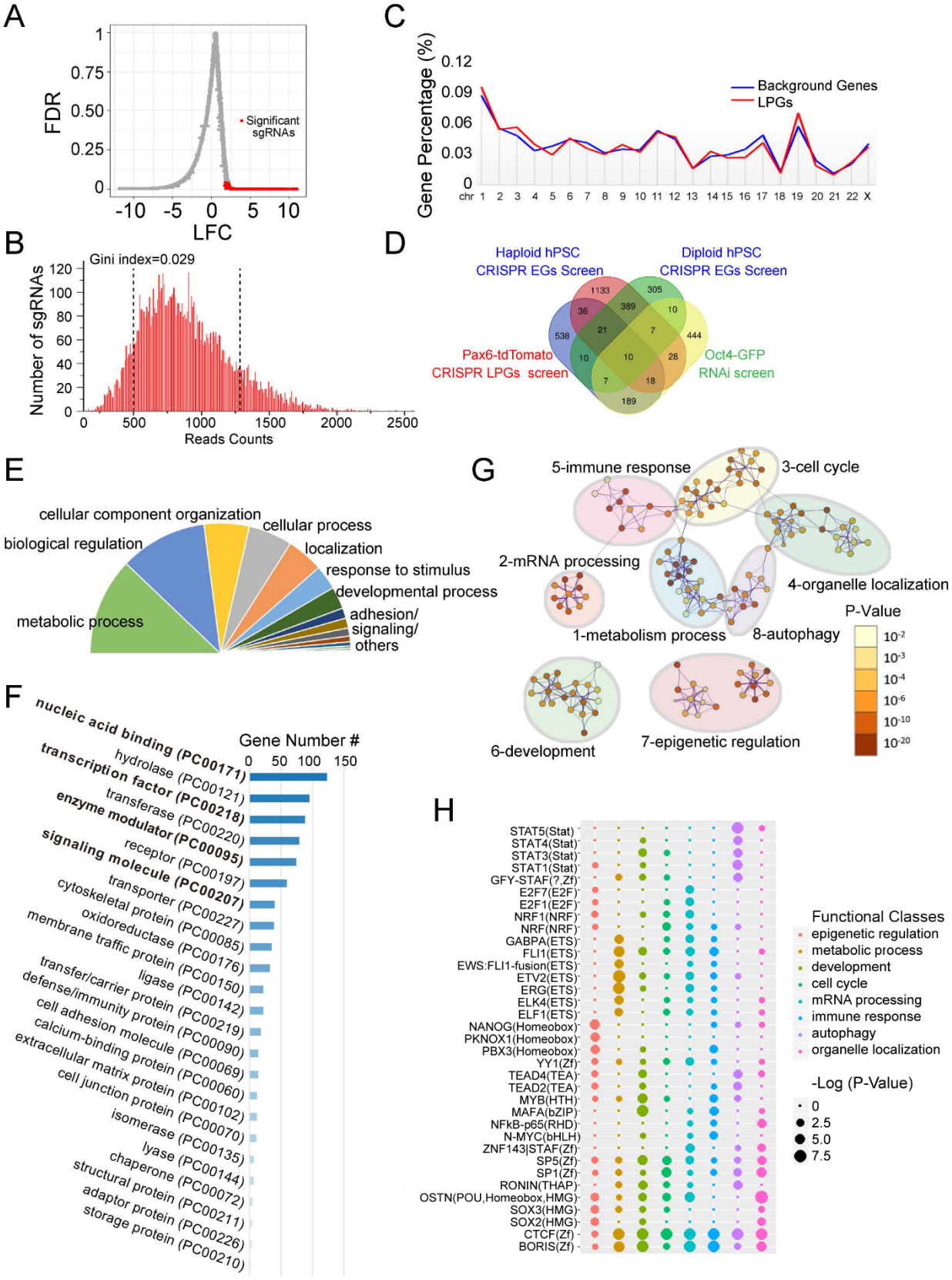

A
FDR
LFC
Significant sgRNAs
B
Gini index=0.029
Number of sgRNAs
Reads Counts
C
Gene Percentage (%)
Background Genes
LPGs
chr
D
Haploid hPSC CRISPR EGs Screen
Diploid hPSC CRISPR EGs Screen
Pax6-tdTomato CRISPR LPGs screen
Oct4-GFP RNAi screen
E
cellular component organization
cellular process
biological regulation
localization
response to stimulus
developmental process
metabolic process
adhesion/ signaling/ others
F
Gene Number #
nucleic acid binding (PC00171)
hydrolase (PC00121)
transcription factor (PC00218)
transferase (PC00220)
enzyme modulator (PC00095)
receptor (PC00197)
signaling molecule (PC00207)
transporter (PC00227)
cytoskeletal protein (PC00085)
oxidoreductase (PC00176)
membrane traffic protein (PC00150)
ligase (PC00142)
transfer/carrier protein (PC00219)
defense/immunity protein (PC00090)
cell adhesion molecule (PC00069)
calcium-binding protein (PC00060)
extracellular matrix protein (PC00102)
cell junction protein (PC00070)
isomerase (PC00135)
lyase (PC00144)
chaperone (PC00072)
structural protein (PC00211)
adaptor protein (PC00226)
storage protein (PC00210)
G
5-immune response
3-cell cycle
2-mRNA processing
4-organelle localization
1-metabolism process
8-autophagy
P-Value
6-development
7-epigenetic regulation
H
STAT5(Stat)
STAT4(Stat)
STAT3(Stat)
STAT1(Stat)
GFY-STAF(?,Zf)
E2F7(E2F)
E2F1(E2F)
NRF1(NRF)
NRF(NRF)
GABPA(ETS)
FLI1(ETS)
EWS:FLI1-fusion(ETS)
ETV2(ETS)
ERG(ETS)
ELK4(ETS)
ELF1(ETS)
NANOG(Homeobox)
PKNOX1(Homeobox)
PBX3(Homeobox)
YY1(Zf)
TEAD4(TEA)
TEAD2(TEA)
MYB(HTH)
MAFA(bZIP)
NFkB-p65(RHD)
N-MYC(bHLH)
ZNF143|STAF(Zf)
SP5(Zf)
SP1(Zf)
RONIN(THAP)
OSTN(POU,Homeobox,HMG)
SOX3(HMG)
SOX2(HMG)
CTCF(Zf)
BORIS(Zf)
Functional Classes
epigenetic regulation
metabolic process
development
cell cycle
mRNA processing
immune response
autophagy
organelle localization
-Log (P-Value)

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, SP1* 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.

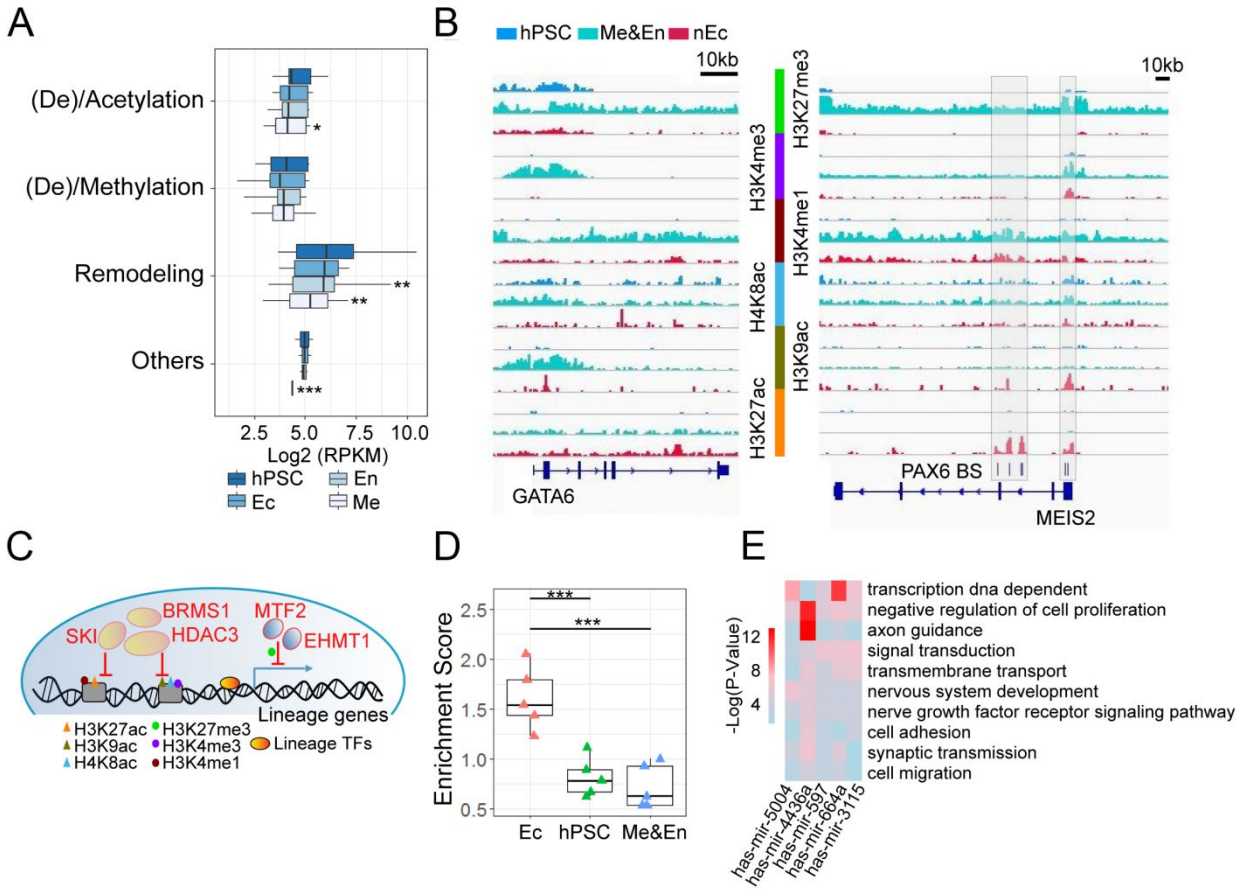


Figure S6. Epigenome characterization of hPSCs and differentiated trilineage. Related to Figure 5. (A) LPGs in the functional module of epigenetic regulation were mostly categorized into families of histone acetylation or de-acetylation, methylation or de-methylation, chromatin remodelers, and others 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 ± 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	
ICMT	15.417	0	0.000124	
TMC8	15.289	0	0.000323	
FBXW10	14.425	0	0.000223	
ACBD6	14.0765	0	0.000422	
NDUFA10	13.9278	0	3.09E-05	
LZTFL1	13.9054	0	0.000968	
MPDU1	13.7926	0	0.001762	
RWDD2B	13.5787	0	0.001861	
RNASE7	13.2728	0	0.000472	
S100A6	13.0246	0	0.002505	
RAB3A	12.9256	0	0.003694	
TPMT	12.8477	0	0.000869	
SUB1	12.8221	0	0.002307	
MAG	12.8103	0	0.005969	
GGCT	12.6698	0	0.004931	
DNAJB5	12.6436	0	0.003793	
PHF17	12.6314	0	0.00305	
GAS7	12.5627	0	0.002059	
ACTG2	12.5345	0	0.00409	
RCBTB1	12.5048	0	0.009522	
MCPH1	12.4841	0	0.001036	
TGM4	12.483	0	0.008832	
PALB2	12.4546	0	0.003892	
SF3A1	12.2949	0	0.004189	
NFASC	12.276	0	0.00962	
BTN2A2	12.2449	0	0.008339	
TCF7	12.1475	0	0.006809	
SAE1	12.1059	0	0.003595	
C2orf16	12.0101	0	0.003298	
ARL8A	11.9762	0	0.006118	

GDF11	11.9591	0	0.000166	
DHRS12	11.9252	0	0.005129	
SEMA3D	11.861	0	0.000273	
TMEM45B	11.7938	0	0.009817	
PRPF40A	11.7455	0	0.004634	
RGS16	11.6805	0	0.004535	
C5orf46	11.64	0	0.006167	
KRTAP6-3	11.6385	0	0.006068	
MTHFR	11.4662	0	0.000172	
POU5F1	11.4097	0	0.000868	
DLG4	11.3839	0	0.002332	
UFSP1	11.263	0	0.007747	
TBC1D12	11.2552	0	0.008231	
TBC1D22A	11.2534	0	0.008345	
C9orf3	11.1813	0	0.003086	
NME5	11.1617	0	0.010113	
UPRT	11.1449	0	6.26E-05	
FBXL2	11.1032	0	0.001298	
ZNF467	11.0776	0	0.009128	
SON	11.0016	0	0.009769	
FRS2	10.9609	0	0.006448	
TFPI	10.954	0	0.000495	
TMEM82	10.9498	0	0.008437	
ITIH1	10.9304	0	0.000231	
CRYGC	10.8664	0	0.010064	
PEX14	10.7643	0	0.009916	
TPP1	10.7258	0	0.00587	
CSF1R	10.6781	0	0.004376	
ITGA11	10.5561	0	0.008036	
HYAL3	10.4267	0	0.000198	
UBAC1	10.2691	0	0.005524	
SMTNL2	10.269	0	0.001216	
FSD2	10.1531	0	0.00173	
FAM49A	10.0049	0	0.000283	
TUBA8	9.9478	0	0.004647	
KRTAP12-2	9.9444	0	0.000638	
PROZ	9.9439	0	0.007401	
C13orf35	9.9324	2.07E-250	0.000693	
CBLL1	9.9275	1.85E-176	0.001326	
GJD3	9.8974	1.21E-277	0.000601	
GPR78	9.7785	1.02E-227	0.001093	
NR1D2	9.7254	4.71E-144	0.004577	
C14orf79	9.6542	8.49E-218	0.00116	
GPR137C	9.6525	1.45E-168	0.001653	

PPFIA1	9.5933	7.39E-242	0.002751	
DGAT2L6	9.5858	0	0.005116	
ZNF711	9.4445	5.35E-84	0.009615	
IL21	9.3884	0	0.005016	
KIAA1958	9.3593	1.41E-72	0.004107	
ZBTB5	9.29	0	0.003546	
ZNF513	9.1867	3.21E-160	0.008495	
OR1Q1	9.0188	1.20E-103	0.004655	
PVRL1	9.012	2.71E-166	0.006648	
MIB1	9.0115	2.10E-136	0.005234	
PRDM14	8.9661	1.99E-89	0.052331	
OLFM1	8.964	4.19E-95	0.004662	
SCD	8.9237	2.52E-42	0.003688	
DSG3	8.8235	6.39E-50	0.005757	
ZBTB46	8.7564	1.58E-146	0.004515	
ACPL2	8.5596	1.39E-124	0.000859	
HDX	8.4177	1.58E-50	0.00432	
HAT1	8.3975	7.97E-46	0.006027	
EP300	8.3869	1.24E-97	0.004939	
ESAM	8.2473	9.79E-23	0.00144	
ZNF138	7.9516	1.62E-51	0.005761	
ELTD1	7.8844	2.58E-29	4.12E-05	
MMP15	7.8773	1.43E-89	0.005228	
ADA	7.8278	1.43E-30	0.001537	
T	7.8049	0	0.001539	
PPAPDC2	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 μm 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×NEAA, 1×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 μg gRNA lentiviral vectors, 5 μg VSVG and 7.5 μg 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 μm cellulose acetate filter, followed by ultracentrifugation (Beckman) with a speed of 55000 g at 16 °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 $MgCl_2$, 15 mM HEPES, 102.94 mM Na_2HPO_4, 47.06 mM NaH_2PO_4, 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'-CTTCCGCATTGGAGTCGCTTTA-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'-AGCAATAGCATCACAAATTTCACAA-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 EF1α-Cas9 cassette been replaced with EF1α-blastcidin. Packaged *P53* gRNA lentiviral particles were then used to infect *PAX6*-tdTomato/iCas9 hPSCs, and 10 μg/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 hESCs 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 μg/ml) for 4 days to eliminate non-infected cells. 5 plates were upplied 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'-AATGGACTATCATATGCTTACCGTAACTTGAAAGTATTTCG-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'-CAAGCAGAAGACGGCATACGAGATTCGCCTTGGTGACTGGAGTTCAGACGTGTGCTC TTCCGATCTCCGACTCGGTGCCACTTTTTCAA-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 blastcidin. 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 μg/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 FASC sorting (BD FACSVerse™ 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'-AATGGACTATCATATGCTTACCGTAACTTGAAAGTATTTCG-3', 5'-CGGATCAATT

GCCGACCCCTCCCCCCAACTTCTCGGGGACTGTG-3'. The purified PCR products were performed second PCR to add illumina adaptor and barcode with the following primer sets: 5'-AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTT

AAGTAGAGGCTTTATATATCTTGTGGAAAGGACGAAACACC-3', 5'- CAAGCAGAAGAC

GGCATACGAGATTCGCCTTGGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCCGA

CTCGGTGCCACTTTTTCAA-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 ImmunoReearch) 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 trancribed 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 Green™ Premix Ex Taq™ II (Takara, PR820A). Primer oligonucleotides used for real-time PCR were as follows:

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

ZIC1	AGCCACGATGCTCCTGGACGC	TGGCCCAGGGCCGCAGCAG
GAPDH	GAAGGTGAAGGTCGGAGTC	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 CCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTG CTATTTCTAGCTCTAAAAC (73 nt); followed by 20nt gRNA (a'), and CGGTGTTTCGTCCTT TCCACAAGATATATAAAGCCTCTACTTAAGATCGGAAGAGCG (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 CCAAGCAGAAGACGGCATACGAGATtcGccttGGTGACTGGAGTTCAGACGTGTGCTCTT 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 form 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 (Montojo 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 (x^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 werepresented as mean value ± standard deviation (SD). Single, double and triple asterisks represented $P < 0.05$, 0.01, and 0.001, respectively, wherein. $*P < 0.05$ was is 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., VANDERGRIFF, 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-FRANDSEN, 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., HACOHEN, 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.