Supplementary information, Data S1

Experimental Procedures

Primary Cells from Anonymous Donors

Most of human primary hematopoietic mononuclear cells (MNCs) obtained from anonymous donors were collected and frozen at AllCells, LLC (Emeryville, CA), including human MNCs expressing a high-level of the CD34 surface marker (CD34+) from newborn cord blood (CB), adult bone marrow (BM) and peripheral blood (PB) after G-CSF mobilization. CD34+ cells were isolated at AllCells or in our lab by the CD34 MACS beads (Miltenyi, Auburn, CA), although CD34-depleted MNCs also contain MNCs expressing a low-level CD34 surface expression. For selected CB and PB samples donated by the parents or patients via their doctors, we isolate MNCs using a standard protocol by Ficoll-Paque Plus (p=1.077) purchased from GE HealthCare Bio-Sciences AB (Uppsala, Sweden).

Culture Media and Conditions for Expanding Human iPSCs, and Karyotyping

The human ESCs and established iPSCs were maintained on a mitotically inactivated mouse embryonic fibroblast (MEF) feeder layer in KNOCKOUT/DMEM medium (Invitrogen) supplemented with 20% KNOCKOUT Serum Replacement (KSR) (Invitrogen), 2 mM L-glutamine (Invitrogen), 2 mM nonessential amino acids (Invitrogen), 1x antibiotic/antimycotic mix (Invitrogen), 0.1 mM β-mercaptoethanol (Sigma), and 10 ng/ml basic fibroblast growth factor (bFGF). All cytokines were purchased from Peprotech (Rocky Hill, NJ) if not otherwise indicated. Karyotyping of human ESCs and iPSCs by G-banding (300-500 bands) was examined by a certified cyto-geneticist by the method previously described (Mali et al., 2008; Ye et al., 2009; Mali et al., 2010).

Constructions of Episomal Vectors

All transgenes were cloned into the EBNA1/OriP-based pCEP4 episomal vector (Invitrogen) for reprogramming as outlined below. The cDNA for the SV40 large T antigen was PCR cloned and inserted into the pCEP4 plasmid backbone as we did previously into a lentiviral vector (Mali et al., 2008). Other vectors expressing other reprogramming transgenes individually were constructed with the cDNAs for the open reading

frames (ORFs) of human OCT4, SOX2, NANOG and LIN28 genes that were obtained by direct PCR of human ESC H9 cell cDNA. Other cDNA templates (human or mouse) were from existing plasmids. The expression cassette for expressing p53-shRNA was obtained by digesting a fragment from a pSUPER-based vector, and was inserted into the pCEP4 plasmid. The pSUPER-shRNA plasmid was purchased from Addgene (Cambridge, MA). A single EBNA1/OriP vector expressing five (mouse) cDNAs *Oct4*, *Sox2*, *Klf4*, *c-Myc* and *Lin28* that are linked by 2A sequences and under the control of the synthetic CAG promoter was constructed in the pCEP4 plasmid backbone, by cloning the fragment from the *piggyBac* transposon plasmid we previously used (Mali et al., 2010; Yusa et al., 2009). The Thomson/Yu three EBNA1/OriP plasmids used in the combination #6 were also purchased from Addgene and used in transfection at the reported ratio (Yu et al., 2009).

Human CD34+ Cell Culture and Priming for Reprogramming by Episomal Vector Transfection

Frozen human CD34+ cells from CB, adult BM and PB after G-CSF mobilization were purchased from AllCells. After thawing, CD34+ cells were cultured for 4 to 5 days under a serum-free medium (SFM): containing 50% IMDM with 50% Ham's F12 (Invitrogen), synthetic lipids (Invitrogen), ITS supplement (insulin-transferrin-selenium and 5 mg/ml BSA), 50 ug/ml of ascorbic acid and 2 mM glutamine (all from Sigma). The SFM is further supplemented with cytokines: SCF (100 ng/ml), FL (100 ng/ml), TPO (20 ng/ml) and IL-3 (10 ng/ml) for expanding CD34+ cells, similar to what we previously described (Ye et al., 2009). For reprogramming with EBNA1/OriP-containing episomal vectors, single or combinations of plasmids (up to 10 μg total) were co-transfected into $1X10^6$ human CD34+ cells via nucleofection (human CD34 cell solution [Nucleofector Kit VPA-1003] with the U-008 program, Amaxa/Lonza). Transfected CD34+ cells were cultured in one well of a 12-well plate in the same medium and cytokines for another 2 days. Subsequently, transfected cells were transferred to 3 to 6 wells of MEF-coated 12-well plates and cultured in MEF medium (DMEM + 10% FBS). Plates with seeded cells were spun at 500 rpm for 30 min at room temperature to help cells attach to MEF-coated plates. The next day (day 3), the MEF medium was replaced with the ESC medium. Small organic molecules that enhance reprogramming such as sodium butyrate (NaB, 0.25 mM) and Valproic Acid (VPA, 0.5 mM) were also added at this time when indicated. The ESC culture medium was exchanged every

other day. Starting day 9 post-transfection, MEF conditioned medium (CM) was used to sustain the development of colonies. Colonies with morphology similar to iPSC colonies were readily visible on day 6 to 10 post-transfection of CB CD34+ cells, and on day 10 to 14 post-transfection of adult BM and PB CD34+ cells.

Human MNC Culture and Priming for Reprogramming by Episomal Vector Transfection

After thawing, frozen MNCs from CB and PB were cultured in the SFM supplemented with the following cytokines and a hormone: SCF (50 ng/ml), IL-3 (10 ng/ml), EPO (2 U/ml, R&D Systems), IGF-1 (40 ng/ml) and dexamethasone (1 μM, Sigma). The media were replenished at day 3 and 6. By day 8-9 when an overt sign of cell division was observed, cells were harvested for nucleofection as well for various analyses. Approximately 3.5-fold cells were obtained for CB at day 8 and 1.5-fold for PB MNC at day 9. After optimizing nucleofection of the cultured MNCs to achieve maximal cell survival and gene transfer, we settled down with the human CD34 cell solution but with the T-016 program for one-time transfection of $2x10^6$ cultured cells with 10 μg plasmid DNA. The transfected MNCs were cultured in the MNC medium for 2 days and then further reprogrammed on MEF feeder cells as described above with cultured CD34+ cells (Ye et al., 2009).

The cultured human PB and CB MNCs for 8-9 days were also analyzed by standard hematopoietic assays, in comparisons with newly thawed stock of the same types or after one-day culture. They were analyzed by flow cytometric analysis with a panel of monoclonal antibodies and by colony-forming assays for myelo- and erythroid progenitor cells (Ye et al., 2009). In addition, the levels of fetal and adult hemoglobin genes (*HBG* and *HBB*) in adult PB MNCs (before and after the culture) were also analyzed (see below in quantitative RT-PCR assays) and by protein staining. Briefly, thawed PB MNCs or cultured cells were fixed by incubation with PBS containing 4% paraformaldehyde and 4% sucrose for 15 minutes at room temperature, and washed twice with PBS. The fixed cells were permeabilized by incubation with 0.25% Triton X-100 in PBS for 5 minutes and washed twice with PBS. Non-specific antibody binding to cells were pre-blocked by incubation with 10% bovine serum albumin in PBS for 30 minutes at room temperature and washed once. The cells were then incubated with fluorochrome-conjugated monoclonal antibodies HBB-PE (specific to HbA, Santa Cruz Biotechnology, Catalog No. sc-21757) and HBG-FITC (specific to HbF, Invitrogen, Catalog No. MHFH01) or appropriate isotype IgG controls diluted in 3% BSA/PBS for 1 hour at room temperature and

washed twice with PBS. The stained cells were analyzed using FACSCalibur flow cytometer and CellQuest software (Becton Dickinson Immunocytometry Systems, San Jose, CA) as we described previously (Ye et al., 2009).

TRA-1-60 Live Staining for Identifying Fully Reprogrammed Clones

TRA-1-60 antibody (mouse IgM, used as 1:300 of MAB4360, Millipore) and Alexa555-conjugated anti-mouse IgM secondary antibody (1:500, Invitrogen) were diluted in ESC medium and added in combination into reprogramming cell culture plates. The cell culture plates with sizable ESC-like colonies were incubated in 37°C for 1 hour before medium was changed to fresh ESC medium or CM after one wash. TRA-1-60 positive colonies were identified under an inverted fluorescence microscope, and can be picked up (within 1 day after staining) for further expansion and characterization. For quantifying reprogramming efficiency we counted numbers of TRA-1-60+, hESC-like colonies as well as TRA-1-60- colonies. The efficiency is plotted as numbers (mean +/- SEM) of TRA-1-60+ cells per unit of nucleofected cells after culture and priming (10^6 for CD34+ cells and $2x10^6$ for MNCs). When numbers of duplicates are sufficient ($n \geq 9$), we performed t-test for paired samples assuming a normal distribution and equal variances. P value of ≤ 0.05 is considered to be significant.

Immuno-staining of Undifferentiated iPSCs

To characterize clones by immuno-staining of various markers, cultured human iPSCs were fixed by 4% paraformaldehyde in PBS for 15 min, followed by permeabilization (required for detecting intracellular antigens) by 0.1% Triton X-100 for 15 min, then washed with PBS for 3 times. The fixed samples were incubated with the following primary antibodies for 2 hours at room temperature: anti-TRA-1-60 (1:300), anti-SSEA-4 (1:10, MC-813-70, mouse IgG, Developmental Studies Hybridoma Bank, Iowa City, IA), anti-NANOG (1:100, 1 µg/ml, rabbit IgG, Peprotech), anti-OCT4 (1:100, sc-5279, mouse IgG, Santa Cruz Biotechnology). After a brief wash with PBS, Alexa555 conjugated goat anti-rabbit or anti-mouse secondary antibodies (1:500, Invitrogen) were used for one-hour incubation to visualize the cells together with DAPI nuclear staining. In

order to visualize the alkaline phosphatase (AP) activity, iPSCs were fixed and stained with Sigma's FAST BCIP/NBT (B5655, Sigma). The AP assay was performed as per the manufacturer instructions.

Embryoid Body (EB) Formation, Differentiation and Immuno-staining

EBs were formed using differentiation medium similar to culture medium except withdrawing bFGF and replacing 20% KSR with 20% FBS. Human iPSCs (near confluent) from 2 wells of a 6-well plate were used for EB formation and cultured in one well of ultra-low attachment 6-well plate. After 8 days, EBs were transferred to gelatin-coated 24-well plates for additional 2-days of attachment. Immuno-staining of EBs is similar to that of human ESCs or conventional iPSCs, with the following primary antibodies used: mouse anti-ß3-tubulin (1:1000, Sigma), mouse anti-actin, smooth muscle isoform (1:500, CBL171, Millipore), rabbit anti-AFP (1:500, DAKO). Alexa555-conjugated goat anti-mouse or anti-rabbit secondary antibodies (1:500, Invitrogen) were used for visualization. The slides were last counterstained with DAPI to show cell nuclei.

Teratoma Formation Assay of Pluripotency

The use of immuno-deficient mice for the teratoma formation assay was approved by the Animal Care and Use Committee at Johns Hopkins University. Approximately five million iPSCs were harvested by Collagenase IV (Sigma) digestion (from one plates of 6-well plates), washed with PBS and resuspended in 200 μL diluted (1:1) Matrigel solution (BD). Cells were injected intra-muscularly into $Rag1^{-/-}\gamma C^{-/-}$ mice or NOD/SCID/IL2RG$(\gamma_c)^{-/-}$ immuno-deficient mice with a further reduced level of natural killer cells. Teratomas were excised 5-10 weeks after injection (Ye et al., 2009; Mali et al., 2010; Mali et al., 2008). After sectioning, slides containing various regions of teratomas were stained by H&E. Complex structures with various cell types were examined at both low and high magnitude.

Genome-wide Analysis of Promoter DNA Methylation

The promoter DNA methylation data was generated using the Infinium Human DNA Methylation27 BeadArray platform (Illumina, San Diego, California). This platform allows us to interrogate DNA methylation of 27,578 informative CpG sites at single-nucleotide resolution, with ≤1 μg of genomic DNA for each of 12

samples. The selected CpG sites are located close to the promoter region of 14,475 RefSeq genes. The methylation status at these CpG loci (near the promoter) often correlates with repression of gene expression, although under-methylation at the selected CpG loci could not guarantee active gene expression (Mali et al., 2010). Cultured human CD34+ cells from CB, adult BM and PB (2 samples each), adult BM-derived stromal cells (3 samples) and fetal fibroblasts (IMR90), neonatal cord-derived HUVEC and CB-EPCs, blood MNCs before and after culture/priming, at least 20 iPSC lines derived from these somatic cells by various vectors (Ye et al., 2009; Mali et al., 2010) (and present study), and 11 human ESC samples (8 independent lines and multiple batches or passages of H1 and H9) were used. Genomic DNA was isolated by the DNeasy® blood and tissue kit (Cat. No. 69506, Qiagen), and 1 μg DNA was in the DNA CpG methylation analysis as we previously described[20]. All data analysis and visualization was performed using MATLAB (The Math-Works, Natick, Massachusetts). The dendrogram plot (using 26,424 autosomal loci) was generated using Euclidean distances and average linkage. To validate the clustering, we computed pair-wise Euclidean distances between cultured blood cells and iPSCs/ESCs (24.08 ± 0.22, mean ± se, $n=11\times31=341$), and between cultured non-hematopoietic cells and iPSCs/ESCs (27.49 ± 0.33, $n=5\times31=155$). The DNA methylation patterns of blood cells were thus closer to iPS/ES than the non-hematopoietic cells. Based on a t-test for two samples with unequal variances and hypothesized mean difference of zero, we confirmed the different mean distance values corresponding to the above were statistically significant ($p<0.01$). Similarly, distances corresponding to the highlighted nodes a and b (in Figure 1A) were computed. Their observed differences were also significant ($p<0.01$). A similar clustering was also observed upon using Pearson correlations to compute distance metrics. K-means clustering was used to further categorize the genes into various clusters and analyze their ensemble dynamics (Mali et al., 2010), using the 26,424 autosomal loci. All the loci located in the X and Y chromosomes were excluded to avoid scoring sex-specific differences.

Genome-Wide SNP Assay to Assess Genome Integrity

The Illumina's Omni1_Quad BeadArray chip (Illumina) containing $>10^6$ informative SNPs was used. The array analysis was performed by the Johns Hopkins SNP Center as part of Center for Inherited Disease Research (CIDR, www.cidr.org). Based on 1,140,419 SNPs identified and the results with 2 control

(CIDR11993 and CIDR10860) genomic DNA samples that have been previously sequenced and included in each run, the Johns Hopkins SNP Center reported a 0.27% genotyping error rate in this run, within the normal range. Genomic DNA (~2 μg) isolated by the Qiagen DNAeasy kit from CT5 and CTN4 iPSCs (p7) that are derived from CB CD34+ cells, and the CD34- cells of the original CB donor (AllCells, #100204) were analyzed. The dis-concordance rate between CT5 or CTN4 iPSCs and the CD34- CB cells is 0.02% or 0.04%, which is much lower than the basal error rate level. In comparison, the dis-concordance rates between the CD34- cells (or iPSCs) and unrelated CIDR11993 or CIDR10860 controls are ~40%. Therefore, our data analysis suggest that CT5 or CTN4 iPSC genome DNA is essentially identical to the CD34- cells and no change is found based on the Omni1 SNP analysis.

Quanatitative RT-PCR Analysis of Pluripotency-Associated Genes

Quantitative gene expression analysis was performed using TaqMan assays from Applied Biosystems (ABI, Foster City, CA). Specifically, the probes used in this manuscript are as follows: OCT4: Hs00999632_g1; SOX2: Hs00602736_s1; MYC: Hs00153408_m1; KLF4: Hs00358836_m1; LIN28: Hs00702808_s1; BCL2: Hs00608023_m1; HBG1/2 (hemoglobin, gamma A/G): Hs00361131_g1; HBB (hemoglobin, beta): Hs00747223_g1; and GAPDH: 4352934E (a constitutive gene control). We found that the HMGA1: Hs00431242_m1 set from ABI was less sensitive than a customer-designed Taqman primers/probe previously published (Tesfaye et al., 2007). Therefore the data by using the customer-designed HMGA1 primers/probe are presented.

PCR southern blot anaysis for the presence of Vector DNA in the derived iPSCs

Total DNA (genomic and episomal) from cell extracts was isolated using the DNeasy® kit from Qiagen. Equal amounts of DNA (100 ng) isolated from naive cells (before nucleofection) were used as negative control while the transfected sample (harvested at day 2 post-transfection) were used a positive control for vector DNA detection by the 35-cycle PCR reaction. Three sets of primers used for detecting vector DNA (in either episomal or integrated form) are: EBNA1_D: 5'-TTTAATACGATTGAGGGCGTCT-3', EBNA1_U: 5'-GGTTTTGAAGGATGCGATTAAG-3', Tg_F: 5'-GCCAGGTGGGTTAAAGGAGC-3', Tg_R: 5'-GGTACTTATAGTGGCTGGGCTGT-3'; SK_F: CCATTAACGGCACACTGCCCCTGT and SK_R:

AGGACGGGAGCAGAGCGTCGCTGA (Yusa et al. 2009). The latter two primers are located at the junction the mouse *Sox2* and *Klf4* genes in the C5 reprogramming gene cassette (Figure S1), and unable to amplify human genomic DNA. Plasmid DNA, either pEB-Tg or pEB-C5, was used as a control template when appropriate, and diluted to amounts equivalent to 1, 0.2 or 0.04 copies per genome of 100 ng genomic DNA from naïve cells. The detection of the single-copy cellular gene beta-actin in the genomic DNA was employed to normalize the amount of genomic or total DNA used for PCR.

We also used Southern blot to detect vector DNA in isolated total DNA using a 445-bp EBNA1 specific probe. The probe was synthesized by PCR amplification using primers EBNA1-5': CCCAGGAGTCCCAGT AGTCAGTC and EBNA1-3': GGAATAGCAAGGGCAGTTCCTCG and the DIG-dUTP labeling kit (Roche) as previously described (Zou et al., 2009). Five μg of total DNA from controls and iPSC samples was digested by BamH I. The pEB-C5 DNA plasmid was also digested by BamH I (single cut) and used as a positive control for detection of unique vector DNA sequence. Plasmid control DNA was diluted in amounts equivalent to 0.8, 8 and 80 copies per genome in 5 μg genomic DNA used. For genomic DNA quality control of iPSC samples, we used an 848-bp *HBB* probe made by PCR using primers 5'-GACTGAGAAGAATTTGAAAGGCG and 5'-TCATCAATTCTGCCATAAATGG. The wild-type genomic DNA after Pme I/EcoR V digestion showed a single 4.3-kb band detected by the *HBB* probe, as expected. Standard Southern blotting and chemi-luminescence detection with CSPD were performed following the instruction manuals of DIG High Prime DNA Labeling and Detection Starter Kit II (Roche).

PCR analysis to Confirm the Authentic Origins of iPSC Lines (DNA Fingerprinting)

Genomic DNA isolated from various iPSC lines and their parental (or related) somatic cells types from the same donor was used for the PCR-based DNA fingerprinting assay. Two sets of primers purchased from Invitrogen) were used. Details have been described in our previous publication (Ye et al., 2009; Mali et al., 2010; Mali et al., 2008). The presence of homozygous sickle mutations in iPSCs derived from the SCD003 patient and in the parental PB MNCs was confirmed by DNA sequencing of the whole products of PCR covering the *HBB* exon 1.

PCR Analysis of the Presence of Genomic Rearrangements at TCR and IGH Loci

Validated PCR primers to analyze V(D)J genomic rearrangements in T and B cells were purchased from InVivoScribe Technologies (San Diego, CA). Genomic DNA (350-500 ng) isolated from human ESCs (a negative control) and 6 iPSC lines derived from adult PB MNCs were used as templates for PCR amplification, using 8 sets of validated primers that are based on the European BIOMED-2 study (van Dongen et al., 2003). They were designed to detect gene rearrangements at T cell receptor (TCR) or immunoglobulin heavy chain (IGH) loci in clonal cell populations. Four sets of primers will amplify the rearranged TCR beta (TCRB) and TCR gamma (TCRG) loci as shown when a clonal T cell genomic DNA is used as a positive control. Three sets of primers will amplify the rearranged IGH loci as shown when a clonal B cell genomic DNA is used as positive control. A clonal population of cells with rearranged TCR or IGH loci will yield one or two prominent amplicons within a defined size range. The last set primers (comprising 5 pairs) amplify 5 cellular DNA regions and generate DNA fragments of 100 – 550 bp, which serves as a quality control of genomic DNA and PCR amplification.

SUPPLEMENTAL REFERENCES

Mali P, Ye Z, Hommond HH, *et al*. Improved efficiency and pace of generating induced pluripotent stem cells from human adult and fetal fibroblasts. *Stem Cells* 2008; **26**:1998-2005.

Tesfaye A, Di Cello F, Hillion J, *et al.* The high-mobility group A1 gene up-regulates cyclooxygenase 2 expression in uterine tumorigenesis. *Cancer Res* 2007; **67**:3998-4004.

van Dongen JJ, Langerak AW, Brüggemann M, *et al*. Design and standardization of PCR primers and protocols for detection of clonal immunoglobulin and T-cell receptor gene recombinations in suspect lymphoproliferations: report of the BIOMED-2 Concerted Action BMH4-CT98-3936. *Leukemia* 2003; **17**:2257-2317.

Yusa K, Rad R, Takeda J, Bradley A. Generation of transgene-free induced pluripotent mouse stem cells by the piggyBac transposon. *Nat Methods* 2009; **6**:363-369.

Zou J, Maeder ML, Mali P, *et al*. Gene targeting of a disease-related gene in human induced pluripotent stem and embryonic stem cells. *Cell Stem Cell* 2009; **5**:97-110.