Supplemental Information

A Roadmap for Human Liver Differentiation from Pluripotent Stem Cells

Lay Teng Ang, Antson Kiat Yee Tan, Matias I. Autio, Su Hua Goh, Siew Hua Choo, Kian Leong Lee, Jianmin Tan, Bangfen Pan, Jane Jia Hui Lee, Jen Jen Lum, Christina Ying Yan Lim, Isabelle Kai Xin Yeo, Chloe Jin Yee Wong, Min Liu, Jueween Ling Li Oh, Cheryl Pei Lynn Chia, Chet Hong Loh, Angela Chen, Qingfeng Chen, Irving L. Weissman, Kyle M. Loh, and Bing Lim
SUPPLEMENTAL INFORMATION

Cell culture
Human ESC lines H1, H7, H9, HES2, HES3, ESi035, H1 FAH-Clover knock-in reporter line, H9 EFLA-BCL2-2A-GFP, and H9 EFLA-BCL2-2A-GFP: UBC-idTomato-2A-Luciferase (Loh et al., 2016) and human iPSC line BJC1 were maintained in feeder-free conditions using mTeSR1 medium (StemCell Technologies, 05851/2). The cell lines were confirmed to be mycoplasma-free before differentiation. The human ESCs or iPSCs were dissociated as single cells using Accutase (Millipore) or TrypLE express (Gibco) and seeded in mTeSR1 medium containing 1µM Thiazovin at a density of 10k to 21k per cm² of growth area 1 day before differentiation. Cell numbers of other human PSC should be optimized for differentiation. It is crucial to ensure that cell density is not too low or too high and that cells are uniformly distributed in the well, because these factors affect differentiation efficiency. Cryopreserved primary adult human hepatocytes (Thermofisher Scientific, HU8055) were thawed with CHRM (Gibco, CM7000), plated, and grown in matrigel sandwich cultures (Gibco, CM3000, CM4000), as per manufacturer’s instructions.

Differentiation of hPSCs into hepatocyte-like cells
In this study, generation of hepatocyte-like cells from hPSCs involves six consecutive lineage choices. First, anterior primitive streak was specified using Defined Endoderm Induction Medium A for 24 hours, followed by Defined Endoderm Induction medium B for 24 hours to induce definitive endoderm (Gibco, A27654SA) (Loh et al., 2014). The definitive endoderm cells were then differentiated into posterior foregut (PGF) using 1µM A8301, 10ng/mL FGF2, 30ng/mL BMP4, and 2µM ATRA or 75nM TTNPB in CDM3 medium for another 24 hours. To further differentiate PGF to liver bud progenitors, one of two related types of differentiation conditions were used on days 4-6 of differentiation: (1) “ABBr/Fs” denote 10ng/mL ACTIVIN, 30ng/mL BMP4, 1mM 8-bromo-cAMP or 1µM Forskolin in CDM3 medium for 3 days (day 4 to 6) or (2) 1µM C59 on top of ABBr/Fs for 2 days (day 4 to 5) followed by 1µM CHIR99201 on top of ABBr/Fs for 1 day (day 6). Thereafter, for the next 6 days (day 7 to 12), 10ng/mL BMP4, 10ng/mL Oncostatin M (OSM), 10µM Dexamethasone, 2µM RO4929097 or 10µM DAPT, 10µM Forskolin, 10µg/mL human recombinant Insulin, Ascorbic acid-2 phosphate (AAP; 200µg/mL), and amino acid concentrate in CDM4 base medium. Optionally, 1µM SB505124 was added for 48 hours on days 7 and 8. Subsequently, hPSC-derived hepatocyte-like cells were further treated with 10µM Dexamethasone, 10µM Forskolin, 10µg/mL human recombinant Insulin, 200µg/mL AAP, and 2µM RO4929097 or 10µM DAPT in CDM5 base medium for 6 days (day 13 to 18).

Generation of midgut/hindgut and pancreatic endoderm cells
Midgut/hindgut cells were specified from the definitive endoderm cells using 50 to 100 ng/mL FGF2, 10 ng/mL BMP4 and 3µM CHIR99201 on top of base medium CDM2 for 4 days (Loh et al., 2014). Early pancreatic endoderm cells were specified from DE using 1µM C59, 250 nM DM3189, PD0325901, 2µM ATRA in CDM3 for 1 day followed by 10ng/mL ACTIVIN, 1µM C59, 250nM DM3189, PD0325901, 2µM ATRA, and 150nM SANT1 in CDM3 for 3 days.

Preparation of base medium
The composition of CDM2 base medium has been previously reported (Loh et al., 2014; Loh et al., 2016; Touboul et al., 2010). CDM3 base medium comprises 10% KOSR, 0.1% PVA, IMDM/F12 (1:1), 1% concentrated lipids and 1% Pen Strep. CDM4 base medium comprises of amino acid supplements, 15 µg/mL Transferrin, 1% Glutamax, IMDM/F12 (1:1), 1% Pen Strep and 1% concentrated lipids. Amino acids were purchased from Sigma Aldrich, dissolved as per manufacturer’s recommendation, and combined to form amino acid-rich concentrated 10x medium. These amino acid supplements include glycine, L-alanine, L-arginine hydrochloride, L-asparagine, L-aspartic acid, L-cysteine hydrochloride, L-histidine hydrochloride, L-lysine hydrochloride, L-methionine, L-phenylalanine, L-proline, L-serine, L-threonine and L-valine. On days 13-18 of differentiation, one of two related types of basal media were used: (1) CDM4 basal medium, supplemented with 10% KOSR and without the additional amino acid-rich mixture or (2) CMRL medium + 10% KOSR. For the sake of brevity "CDM5" is used to interchangeably refer to either of these two types of basal media throughout the manuscript.

Comparison of liver progenitor differentiation protocols
Human ESC lines (H1 and H9) were used to differentiate into liver genes for comparison of SR2 and other liver differentiation strategies (Figures 3A-B). Day-6 human liver progenitors generated by SR2 were used for comparison of gene and protein expression with liver progenitors generated from other protocols. The same initial endoderm-inducing conditions (‘ACP’, ‘ADP’) were used to generate endoderm cells (Gibco, A27654SA) (Loh et al., 2014) prior to further differentiation into liver cells using 3 liver-inducing protocols (Figures 3A-B) (SR2, and other methods by Zhao et al., 2013 and Si Tayeb et al., 2010).
Based on Zhao et al., 2013, hPSC-derived endoderm cells were treated with 20ng/mL FGF7, 10µM SB421542 in RPMI containing 0.5mg/mL Albumin Fraction V and 1% Insulin/Transferrin/Selenium (ITS) for 2 days. Subsequently, the cells were treated with 20ng/mL FGF7, 20ng/mL FGF2, 50ng/mL BMP4 in DMEM/F12 containing 1mM Glutamax, 0.1mM Beta-mercaptoethanol, 1% NEAA and 1% B27 for 2 or 5 more days to generate day-6 liver progenitors (Figures 3A-B) or day 9 liver progenitors (Figures 5A-B). To further differentiate the day-9 liver progenitors into hepatocyte-like cells, day-9 liver progenitors were treated with 20ng/mL HGF, 50ng/mL BMP4 in DMEM/F12 with 1mM Glutamax, 0.1mM Beta-mercaptoethanol, 1% NEAA and 1% B27 for 5 days. Finally, these cells were further treated with 10ng/mL OSM and 0.1µM Dexamethasone in Hepatocyte Culture Medium (Lonza) for 4 days to generate day-18 hepatocyte-like cells (Figures 5A-B).

Following Si Tayeb et al., 2010, hPSC-derived endoderm cells were treated with 10ng/mL FGF2 and 20ng/mL BMP4 in RPMI containing 2% B27 for 4 to 5 days to generate day-6 liver progenitors (Figures 3A-B) or day-7 liver progenitors (Figures 5A-B) for 6 days to generate day-18 hepatocyte-like cells (Figures 5A-B).

Based on Avior et al., 2015 and Chen et al., 2012, human ESCs were treated with 100ng/mL ACTIVIN, 50ng/mL WNT3A and 10ng/mL HGF for 3 days in 2% B27 RPMI supplemented with 1mM Glutamax. Later, cells were treated with 20% KOSR, 1% NEAA, 1mM Glutamax, 0.1mM 2-bertamercaptoethanol and 1% DMSO in Knockout DMEM for 4 days to generate day-7 liver progenitors (Figures 3A-B). Later, following Avior et al., 2015, the liver progenitors were further treated with 20ng/mL OSM, 500nM Dexamethasone, 4ng/mL FGF2, and 1% ITS (Sigma, I13146) in IMDM for 5 days and 500nM Dex, 10ng/mL HGF, ITS-3 (Sigma, I27771), 10µM Lithocholic Acid, and Vitamin K2 (Sigma, V9378) (Figure S5B). Based on Chen et al., the day-7 liver progenitors were treated with 20ng/mL OSM, 500nM Dexamethasone and 1% ITS (Sigma, I3146) in IMDM for 5 days (Figure S5B).

Based on Carpentier et al., 2016, human ESCs were first differentiated into endoderm cells for 4 days using the STEMdiff definitive endoderm differentiation kit (StemCell Technologies Inc.). Later, the endoderm cells were dissociated as a single cell suspension and seeded at a density of 275k cells per well of a 12-well plate in the presence of 10% KOSR, 1mM Glutamax, 1% NEAA, 1% PS, 1% DMSO and 100ng/mL HGF in DMEM/F12 for 3 to 4 days to generate day-7 liver progenitors (Figures 3A-B) or day-8 liver progenitors. 10µM Y27623 was only added on the day of cell plating. The day-8 liver progenitors were further differentiated into hepatocyte-like cells using 100ng/mL HGF, 10% KOSR, 1mM Glutamax, 1% NEAA, 1% PS, and 1% DMSO in DMEM/F12 (Figure S5B).

**Design and construction of Cas9 plasmids**

Primers used for constructing the plasmids are listed in Table S2. All restriction enzymes were purchased from NEB, unless stated otherwise. PCR reactions were conducted using Q5® Hot Start High-Fidelity 2X Master Mix (NEB, M0494L) or Phusion High-Fidelity PCR Master Mix with HF Buffer (ThermoFisher Scientific, F5315). Ligations were conducted using isothermal assembly with NEBuilder® HiFi DNA Assembly Master Mix (NEB, E2621L) or In-Fusion® HD EcoDry™ Cloning Plus (Clontech, 638915). Primers and dsDNA fragments were ordered from Integrated DNA Technologies (IDT). EF1α promoter was PCR amplified from N205 plasmid (Addgene plasmid # 44017), a gift from Jerry Crabtree (Hathaway et al. 2012). Amp pUC fragment was obtained from pCMV-Bsd (Thermo, V51020). Plasmid pX330 (Addgene plasmid # 42230), a gift from Feng Zhang (Cong et al. 2013), was digested with XbaI & AarI (Thermo, ER1582) and ligated with EF1α promoter. The modified plasmid was digested with SpeI and XbaI and then ligated with the Amp pUC fragment. The BsmBI chimeric gRNA cassette was amplified from a gBlock based on pX335 (Addgene plasmid # 42335), a gift from Feng Zhang (Cong et al., 2013). The cassette was subsequently ligated into XbaI cut site of our modified Cas9 plasmid. To add a 2A-linked mRuby2, the plasmid was first digested with PmlI and EcoRI, and the Cas9 3’ fragment was amplified from px330 and 2AmRuby2 (Lam et al., 2012) fragment was amplified from a gBlock. Next, these two PCR fragments were ligated with the digested plasmid. Enhanced specificity of Cas9 was attained by specific mutations of the Cas9-protein sequence (Kleinstiver et al., 2016; Slaymaker et al., 2015). Such mutations (K848A/K1003A/R1060A) were introduced to our Cas9 plasmid by cloning in a gBlock. 5’ fragment of Cas9 and 3’ fragment plus 2AmRuby2 were amplified from the plasmid, and mutated sites from gBlock. The plasmid was digested with FseI and EcoRI and fragments ligated in the digest. hPSCs are very sensitive to DNA damage (Momcilovic et al., 2009; Momcilovic et al., 2010) and we found that Cas9 targeted hPSCs had low survival and low number of correctly targeted clones. Inhibiting the TP53 checkpoint could increase survival of targeted hPSCs (Ihy et al., 2017). mtp53 dominant negative fragment was amplified from gBlock based on pCE-mp53DD (Addgene plasmid # 41856), a gift from Shinya Yamanaka (Okita et al. 2013) and ligated it into our Cas9 plasmid digested with EcoRI to generate our final construct pMIA3 1sg-eSpCas9-2AmRuby2-2Amp53DD.
Design of gRNA
The genomic sequence of the end of human FAH CDS (chr15:80186000-80187000) was uploaded to Benchling (https://benchling.com/) and single gRNAs were designed using the online search algorithm. One gRNA (GAGCAGAGAAAATCTCATGA, negative strand) that overlaps the FAH-stop codon was selected. Oligos with the gRNA sequence and the complementary sequence (TCATGAGATTTCCTGCTC), with CACC- and AAAC- added to the 5’ end of each oligonucleotide were purchased from IDT. Finally, the oligos were annealed and ligated into BsmBI-digested pMIA3 1sg-eSpCas9-2AmRuby2-2Amp53DD.

GFP reconstitution assay
The gRNA cutting efficiency was confirmed by GFP reconstitution assay using pCAG-EGxxFP plasmid (Addgene plasmid # 50716), a gift from Masahito Ikawa (Mashiko et al., 2013), as previously described. Briefly, the target sequence was amplified from H1 hPSCs genomic DNA and then cloned into the SaII cut site on pCAG-eGxxFP (for primers see Table S2). The plasmid was then transfected into HEK293T cells with or without FAH-pMIA3. 48h later, strong GFP signal was observed when both plasmids were transfected indicating Cas9 cleavage activity (Figures SSD-F).

Design and construction of donor plasmids
Our homology-directed repair donor plasmid was derived from OCT4-2A-eGFP-PGK-Puro (Addgene plasmid # 31938), a gift from Rudolf Jaenisch (Hockemeyer et al., 2011). 3’ homology arm for human FAH was amplified from gDNA from H1 hPSCs. The amplified product was then ligated into the OCT4-2A-eGFP-PGK-Puro plasmid that had been double digested with Ascl & FseI. To swap the resistance gene from puromycin to blasticidin, the mpGk1 promoter & polyA fragments were first amplified from the OCT4-2A-eGFP-PGK-Puro plasmid and Blasticidin (Bsd) resistance gene from pCMV-Bsd plasmid. These fragments were ligated into BsrGI & Ascl double-digested donor plasmid to generate a construct containing OCT4-2A-eGFP-PGK-BsdFAH. The modified plasmid was then digested with Nhel & PacI to allow ligation of P2A-linked Clover fluorophore (Lam et al., 2012), that had been amplified from a gBlock. The PGK promoter was amplified from OCT4-2A-eGFP-PGK-Puro to make OCT4-2A-Clover-PGK-BsdFAH. A Gly-Ser-Gly sequence was added to the start of the P2A sequence as this has been shown to increase the 2A-peptide cleavage efficiency (Kim et al., 2011; Szmyczak et al., 2004). The Bsd CDS was amplified from pCMV-Bsd and the negative selection agent thymidine kinase (TK) CDS from pLOX-TERT-iresTK (Addgene plasmid # 12245), a gift from Didier Trono (Salmon et al. 2000). The two PRC fragments were then ligated into PacI & AflIIdigested donor plasmid, to generate the construct OCT4-2AClover-PGK-Bsd2ATK-FAH. Noti was then used to digest this plasmid and a PGK-DTA negative selection cassette was ligated outside the homology arms, amplified from OCT4-2A-eGFP-PGK-Puro and a gBlock. Finally the 5’ homology arm of human FAH was amplified from H1 hPSC gDNA and then ligated into the SbfI- & Nhel-digested plasmid to complete the FAH-2AClover-PGK-Bsd2ATK-FAH-PGK-DTA donor plasmid.

Generation of FAH-2AClover H1 hES reporter cell line
1.5x10^5 H1 hPSCs were used in one reaction of an Amaxa nucleofection (Lonza) targeting. 5 µg of both FAH-pMIA3 and FAH-2AClover-PGK-Bsd2ATK-FAH-PGK-DTA donor plasmid were nucleasefected with the P3 Primary Cell kit (V4XP-3024) using CM-113 program. Prior to nucleasefection the cells were pretreated with ROCK-inhibitor (Y-27632, Tocris) for at least 1 hour and then dissociated into a single-cell suspension with StemPro Accutase (A1110501, ThermoFisher). After nucleasefection, the cells were seeded in mTeSR supplemented by CloneR (05888, Stem Cell Technologies). After 24h, the media was changed to mTeSR1 medium and the cells were left to recover. At approximately 48-72 hours, Bsd (A1113903, ThermoFisher) was added to the media at 10µg/ml to select for targeted cells. Individual colonies were manually picked from wells after 10-14 days of selection and further expanded for screening.

12 clones were picked from the targeted wells. 11 clones survived the expansion and gDNA was isolated for screening. Primers for amplifying the FAH target region in the eGxxFP plasmid (also annotated “eGxxFP primers“) were used for screening the targeted clones alongside FAH-clover forward and reverse primers (Table S2). Expected PCR product size for wild-type cells is 422 bp with eGxxFP primers and for targeted cells 2.3 kbp with “FAH-clo“ primers. 10 out of 11 clones picked amplified PCR bands at both WT and targeted sizes, suggesting they were heterozygous clones (Figures S5G-I). 3 out of the 10 clones had no mutations on the WT allele as confirmed by Sanger sequencing (Figure S5J).

Heterozygous clones that were confirmed were further expanded and targeted with 10µg of pCAG-Cre:GFP (Matsuda & Cepko 2007) (Addgene plasmid # 13776), a gift from Connie Cepko. These cells were allowed to recover for 48h after which they were sorted for GFP. The GFP+ cells were plated in CloneR supplemented mTeSR1 medium at a density of 100 to 1000 cells per well in a 6-well tissue culture plate. After 48h, the media
was changed to mTeSR1 medium and clones were allowed to expand for 10-14 days. Clones were then manually picked for expansion and screening. PCR screening and Sanger sequencing was done with eGxxFP primers as above. Expected PCR amplicon size for correctly targeted allele is 1.3 kbp (Figure S51). Next, confirmed clones were expanded and used in downstream experiments.

Intracellular FACS

AFP antibodies (DakoCytomation, A000829) were conjugated with R-phycocerythrin (rPE) using rPE labeling kit (abcam, ab102918). Cells (either undifferentiated hPSCs or hPSC-derived liver cells) were dissociated into single cells using TrypLE Express (Gibco, 1260413) and centrifuged at 2000 rpm for 3 minutes. Each cell pellet was washed with 1x PBS (Gibco, 14190235), strained using a 100µm strainer (Falcon, 08-771-19) and counted using a hemocytometer before fixation in BD Cytofix/Cytoperm buffer (BD Biosciences, 554714) on ice for 20 minutes. Next, the fixed cells were washed 2 times with 2 mL 1x BD Perm/Wash buffer (pre-warmed to room temperature) (BD Biosciences, 554723) at room temperature and pelleted. The cell pellet was then resuspended in 1x BD Perm/Wash buffer and 100µL of cell suspension was aliquoted into individual tubes for separate stains. Stained or unstained controls were included. Either anti-ALB-APC (R&D, IC1455A, 0.4µL per 150,000 cells) or anti-AFP-PE (0.33µL per 150,000 cells) was used. The anti-AFP-PE antibody/cell mixture was incubated at room temperature for 30 minutes in the dark while anti-ALB-APC antibody/cell mixture was incubated at room temperature for 20 minutes in the dark. Subsequently, the unstained or stained cells were washed 2 times with 2 mL 1x BD perm/wash buffer at room temperature and pelleted. The pellet was then resuspended in 300µL 1x BD perm/wash. FACS run was performed using BD LSR Fortessa X20 and FACS data were analyzed using FlowJo.

Surface marker FACS

Cells were dissociated using TrypLE Express, pelleted, resuspended in FACS buffer (0.5% BSA Fraction V + 5mM EDTA in 1 x PBS) and strained through a 40µm filter (BD Biosciences) to generate a single-cell suspension. The cell suspension was aliquoted into individual tubes and stained with α-CD99 PE, α-CD99-APC, α-CD10 APC, α-CD201-APC or α-CXCR4-PE-Cy7 (Table S2) for 30 minutes on ice in the dark. Subsequently, the cells were washed 3 times with 1-2 mL of FACS buffer, resuspended in 300 µL of FACS buffer containing 100 ng/mL DAPI into a FACS tube and analysed by FACS using BD LSR Fortessa X20.

Sorting of Clover+ and Clover− cells by FACS

Cells were dissociated as single cells and stained with DAPI prior to FACS. Separate samples of FAH-Clover+ and Clover− populations were gated and collected and then harvested for gene expression analyses.

High-throughput antibody screens

BD Lyoplate Screening panel (BD Biosciences, 560747) antibodies (Table S1) were reconstituted with deionized water before their use to stain cells following manufacturer’s instructions. Cell suspension was filtered through 100 µm strainer to remove clumps. 75 µL cell suspension was added to each antibody-containing well, pipette-mixed 3 times and incubated in the dark at 4°C for 20-30mins. The cells were pelleted at 500g for 6 minutes and supernatant was discarded by plate inversion. The cells were then washed twice with 200 µL cell staining buffer by pipette mixing, resuspended in DAPI-containing cell staining buffer and analysed with BD LSR Fortessa X20.

Immunostaining and image analyses

Cells were washed once gently with 1x PBS, fixed with 4% formaldehyde in PBS for 15-20 minutes at room temperature, washed 3 times with 1x PBS and blocked with blocking buffer (10% Donkey Serum + 0.1% Triton X in 1x PBS) for 1 hour at room temperature. Next, the cells were stained with antibodies diluted in 1% Donkey Serum + 0.1% Triton X in 1x PBS at 4°C overnight (see Table S3 for all antibodies used). The next day, the cells were washed 3 times with washing buffer (0.1% Triton X in 1x PBS) and then stained with fluorophore-conjugated secondary antibodies diluted in at 1:1000 (1% Donkey Serum + 0.1% Triton X in 1x PBS) in the dark at room temperature for 1 hour. Finally, the cells were washed once with 100 ng/mL DAPI/PBS once and twice with 1x PBS before visualization using Zeiss Axio Vision. “No primary antibody staining” or undifferentiated hPSC negative controls were used to adjust exposure times for minimal background fluorescence detection. The stained cells were counted using Image J (Abràmoff and Magalhães, 2004; Collins, 2007). Image preprocessing including gray scale conversion, threshold setting, image segmentation and noise reduction were performed.

Western blotting (WB)

Cell samples were lysed using radioimmunoprecipitation assay (RIPA) buffer supplemented with protease inhibitor cocktail (Nacalai Tesque, 25955-11). Protein concentration was determined using Pierce BCA Protein
Assay (ThermoScientific, 23225). 40 μg of protein lysates were separated by SDS-PAGE and transferred onto a PVDF membrane (100V at 4°C, for approximately 1 hour). Membranes were blocked with 5% milk in TBS for 1 hour at room temperature. The membranes were briefly washed with Tris-buffered saline (TBS) and after this incubated with the primary antibody (see Table S3 for all antibodies used) at 4°C overnight. The following day, membranes were washed 3 times with TBS and incubated with the appropriate HRP-conjugated IgG secondary antibody for 2 hours at room temperature. All antibodies, primary and secondary, were used with 5% milk in TBS. Following incubation with the secondary antibody, the membranes were washed 3 times in TBS. Washed membranes were developed using SuperSignal West Pico ECL (Thermo Scientific, 34077) and imaged with BioRad ChemiDoc MP Imager.

Immunohistochemistry (paraffin)

Immunohistochemistry was performed on paraffinized murine tissue that was embedded onto glass microscope slides and sectioned using a microtome. The sections were deparaffinized through a series of 5-minute washes with xylene, 100% ethanol, 95% ethanol, 70% ethanol and lastly milli-Q water. Endogenous peroxidase activity in sections were blocked using blocking buffer (65 mL of 100% methanol, 3.5 mL of 30% hydrogen peroxidase (Sigma) and 31.5 mL of milli-Q water) for 30-minute at room temperature. Finally, antigen retrieval was performed at boiling temperature for 30 minutes in a 10% pH 6-citrate buffer. The tissues were subsequently blocked with donkey serum for 1 hour at room temperature in a hydration chamber. The sections were stained with primary antibodies (Vector Laboratories, Vectastain ABC kit; PK-6101) at 4°C overnight, later washed 3 times with 0.1% Triton in 1x PBS done in 5-minute incubation intervals (with fresh buffer applied during each wash) and stained with secondary biotinylated antibody at room temperature for 30 minutes. Thereafter, the samples were lightly rinsed with 1x PBS for 5 minutes on an orbital shaker. 250 μL of ABC reagent was added to each sample and left to incubate at room temperature for 1 hour. The slides were then washed 3 times with 1x PBS for 5 minutes. For peroxidase detection, DAB kit (Vector Laboratory, SK-4100) was used for substrate binding. The substrate concoction (5 mL of milli-Q water, 2 drops of buffer stock, 4 drops of DAB and 2 drops of hydrogen peroxide) was applied onto the samples and left to incubate for 5 minutes at room temperature. After the 5 minutes incubation, the samples were rinsed 3 times with milli-Q water, changing fresh water each time. The samples were then placed into a vertical plastic chamber containing hematoxylin and incubated for 20 minutes. After the hematoxylin counterstaining step, the samples were rinsed 3 times with milli-Q water before proceeding to the dehydration steps. Dehydration steps comprises: 70% ethanol, 95% ethanol, 100% ethanol and xylene wash steps, all with 5 minutes incubation. The counterstained samples were permanently mounted using Vectamount reagent (H-5000) from Vector Laboratory.

Immunohistochemistry (cryoblock)

Mouse liver tissue was fixed in 4% formaldehyde diluted in 1x PBS at 4°C overnight and then immersed in 30% sucrose solution for cryoprotection. The liver tissue was then embedded in Optimal Cutting Temperature (OCT), sectioned and deposited onto microscope slides (A*STAR IMCB histology lab). The microscope slides were thawed in room temperature and stained following the immunostaining method described above.

Gene expression profiling by RT-qPCR

RNA was extracted using Zymo RNA extraction kit. Reverse transcription was performed with random primers (Applied Biosystems RT kit) to generate cDNA. Gene expression was quantified using gene-specific primers and Quantitect SYBR Green master mix (Qiagen) or KAPA SYBR® Fast qPCR kit (KAPA Biosystems, KK-4602). NCBI primer designing tool was used to design gene-specific primer sequences. The primers used were validated for their specificity and those with efficiency between 90-110% were used (Table S4). Gene expression profile was analysed using Microsoft excel and heatmaps were generated using GenePattern version 2.0 (Reich et al., 2006).

Microarray

Total RNA was extracted from hPSCs and hepatic derivatives using the RNeasy Micro kit (Qiagen) and profiled for RNA integrity using the 2100 Bioanalyzer (Agilent). 500ng of high purity and high integrity samples with 260/280 and 260/230 absorbance ratios >1.9 and RNA integrity numbers (RIN) >8.0 were reverse transcribed into cDNA and in vitro transcribed into biotin-labelled cRNA using the TargetAmp-Nano labeling kit for Illumina Expression BeadChip (Epicentre). The cRNA was hybridised on HumanHT-12 v4 Expression BeadChips and scanned on the HiScan system (Illumina) according to the manufacturer's specifications. The raw microarray data was background subtracted using the BeadStudio Data Analysis Software v3.1.3.0 (Illumina) and normalised using the cross-correlation method (Chua et al., 2006). Differential gene expression was defined based on a fold-change cutoff of >1.5 compared to the average of the undifferentiated hESC baseline controls (Table S5). Heatmaps of the fold-change in gene expression on a Log 2 scale were generated using Cluster and TreeView. Gene ontology analyses on the genes downregulated upon removal of DAPT, Forskolin, Ascorbic
acid and Insulin were conducted using DAVID/EASE (Huang et al., 2008; Huang et al., 2009). Background subtraction using Illumina HT-12 v3 database was chosen in the settings.

Animal husbandry and blood sampling
All animal experiments were performed pursuant to experimental protocols approved by the Institutional Animal Care and Use Committee (IACUC). *Fah^−/Rag2^−/Ilr2g^−* (FRG) mice on NOD or C57BL6 background (Yecuris, 10-0008 or 10-0001) were handled and housed under aseptic conditions. Each mouse had its ears notched for long-term tracking. The mice were fed ad libitum with irradiated Lab Diet 5LJ5 chow formulated with high fat and low protein content to avoid excessive tyrosine levels, which can lead to liver damage. NTBC (Yecuris, 20-0026) was dissolved in sodium carbonate to generate 1 mg/mL stock solution. A final dose of 16 mg/L was given to breeders or non-experimental mice, whereas 0 to 8 mg/L NTBC was provided to experimental mice on selected days. All experimental mice were treated with the same NTBC cycling condition. 3% Dextrose was added to the drinking water to offset the bitter taste of NTBC. Antibiotics such as Sulfamethoxazole (SMX) and Trimethoprin (TMP) (Yecuris, 20-0037) were added to the water once every other week for 2-4 days to prevent infection in the immune-compromised mice. To aid in caloric intake, each cage was supplemented with a dish of liquid diet, prepared by adding 1 volume of STAT high caloric liquid (PRN Pharmaceutical, G8270) to 1 volume of 3% Dextrose (Sigma) drinking water. Masses of the FRG mice were measured weekly. Blood was collected using lancets for submandibular bleeding and allowed to coagulate in 4°C overnight. The next day, serum was harvested by centrifugation and removal of red blood cells.

Animal handling and transplantation
The FRG mice on C57BL6 or NOD backgrounds (Yecuris) were fed *ab libitum* with low-protein, high-fat irradiated Lab Diet 5LJ5 and water containing 16 mg/L NTBC (2-(2-nitro-4-fluoromethylbenzoyl)-1,3-cyclohexanedione). 1 day prior to transplantation, 1.25 x10^7 pfu/mL of *urokinase type Plasminogen Activator (uPA)*-expressing adenovirus particles (Yecuris) were retroorbitally injected into mice and 0.5-1.5x10^6 hPSC-derived day 18 hepatocytes were intrasplenically transplanted. 1 day before transplantation, liver injury was induced by cyclical withdrawal of NTBC for 5 to 7 days and then provision of 8 mg/L NTBC for 2 to 3 days.

Retro-orbital injections, intrasplenic and intrahepatic injections
Adenovirus particles expressing urokinase Plasminogen activator (uPA) (Yecuris, #20-0029) were injected at a dose of 1.25*10^9 pfu at a volume of 100 µL per 25g mice approximately 24 hours before transplantation. Pertaining to intrasplenic or intrahepatic injections, 20-300 µL of cells was injected into the spleen or liver of 4-6 week-old FRG mice using 26G-31G needles.

Intrahepatic injections into neonate and adult FRG mouse livers
For neonatal intrahepatic transplantations, 0-72h old FRG pups were injected with approximately 200,000 cells directly into the liver using 31G needles. The pups were then rubbed with bedding and returned to the cages. Pertaining to intrahepatic transplantation into adult livers, 50 µL of cells were injected into multiple sites within the liver using 26G-31G needles. Bleeding was stopped by gently applying pressure on the puncture site upon withdrawal of needle.

Bioluminescence imaging
The FRG mice were pre-transplanted with hPSC-derived hepatocyte-like cells, which overexpress *luciferase* gene. Prior to bioluminescence imaging, the transplanted mice were anesthetized with 1 to 3% isoflurane-mixed oxygen. To minimize imaging signal interference, mice were depilated at abdominal region to fully expose skin surface as hair will absorb and scatter light, which may result in lower signal output. The mice were then weighed to determine their masses. D-luciferin solution was reconstituted by adding 30 mL of saline or 1 x PBS to 1g of D-luciferin stock (Promega, InvivoGlo) and then filtered. 167µg/g firefly D-luciferin was injected intraperitoneally into each mouse. After 10mins of incubation, the mice were positioned in the Perkin Elmer IVIS Spectrum facing upright (0% relative to supine position) so that the liver was exposed to the camera’s field of view. The parameters of bioluminescence imaging were as follow: Exposure – 40s; binning – medium; excitation – block; emission – open; structure – no; FOV – D; Fstop – 1; and height – 1.5.

Enzyme linked immunosorbent assay (ELISA)
ELISA Accessory kit (Bethyl Laboratories, Inc, E101) was used to quantify human Serum ALBUMIN levels. The assay was performed as per manufacturer’ instructions. Absorbance was measured using an ELISA plate reader at 450 nm.

Bilirubin quantification
Bilirubin levels in mouse serum were measured using a colorimetric Bilirubin assay kit (Sigma, MAK126). Total, direct or blank working reagents were prepared as per manufacturer’s instructions. 20 µL serum was added per well and total, direct or blank working reagents were added to each sample. Colorimetric product was measured at 530 nm using the Sunrise™ microplate reader. The concentration of bilirubin was then calculated using the following formula \( [(A_{530\text{sample}} - A_{530\text{blank}}) \times 5 \text{ mg/dL}] / [A_{530\text{ calibrator}} - A_{530\text{ water}}] \).

**Survival curve analyses**

The FRG mice were checked for survival daily and Kaplan-Meier survival analysis was conducted using GraphPad Prism v7.00 for Mac (GraphPad Software, La Jolla, CA, USA). Statistical analyses were performed using the one-sided Mantel-Cox log-rank test. Data from 3 independent transplantation experiments were analyzed.

**Statistics**

No statistical method was used to pre-determine sample size for *in vitro* or *in vivo* experiments. Experiments were not randomized. The investigators were not blinded to the allocation during experiments or outcome assessment.
Figure S1. Related to Figure 1: Regulation of early foregut competence. A) Experimental strategy to treat DE with RA or TGFβ modulators on the day 2-3 interval and evaluate its subsequent impact on day-18 hepatocyte gene expression as shown in subpanels B and C. B) qPCR gene expression of day-18 hepatocyte markers after inhibition (BMS) or activation of retinoid signaling (using ATRA, 2 µM or various doses of TTNPB, 10-100 nM) in the presence of base media condition A83BF (A83BF: A8301, 1 µM; BMP4, 30 ng/mL; FGF2, 10 ng/mL) on the day 2-3 interval. C) qPCR gene expression of day-18 hepatocyte markers after inhibition (A8301, 1 µM or SB505124, 1 µM) or activation of TGFβ signaling (ACTIVIN, 10 ng/mL) in the presence of base media condition BF (BF: BMP4, 30 ng/mL; FGF2 10, ng/mL) on the day 2-3 interval. D) Experimental strategy to treat definitive endoderm (DE) with FGF2 at 10 ng/mL on the day 2-3 interval to produce day 3 posterior foregut (PFG) and assay subsequent effects on hepatic gene expression by day 8, as shown in subpanel E. E) ALBUMIN qPCR of day-8 hepatic progenitors cells generated from endoderm treated on the day 2-3 interval with FGF2. F) Experimental strategy to treat posterior foregut (PFG) or liver bud progenitors (LB) in the presence of base media condition A83BF (ACTIVIN, 10ng/mL; BMP4, 30ng/mL; 8-Br-A8301).
CAMP, 1mM), with a WNT inhibitor (C59, 1 µM) or R-SPONDIN3 (R100, 100 ng/mL) and WNT3A at varying doses of 50 or 100 ng/mL on the day 3-4 or day 4-5 interval to produce day-4 or day-5 liver bud progenitors, as shown in subpanel G. G) CDX2 qPCR of day 5 liver bud progenitors cells generated from PFG treated on the day 3-4 interval with WNT modulators. H) qPCR gene expression of hPSC, day-3 hPSC-derived PFG and day-5 hPSC-derived liver bud (LB) and midgut/hindgut (MHG) progenitors. Error bars represent ± SE.
Figure S2. Related to Figure 2: Lineage bifurcation between liver bud and pancreatic fates. A) Gene expression of day-6 hPSC-derived liver bud (LB) and midgut/hindgut (MHG) cells derived, as shown by qPCR. B) qPCR gene expression of day-18 hepatocytes generated from PFG cells treated on the day 4-6 interval with WNT inhibitors (C59, 1 µM or XAV939, 1 µM) or CHIR99201 (CHIR) of varying doses (1 µM, 2 µM or 3 µM) in the presence of base media condition ABBr (ACTIVIN, 10ng/mL; BMP4, 30ng/mL; 8-Br-CAMP, 1mM). C) qPCR gene expression of day-18 hepatocytes generated from PFG cells treated on the day 5-6 interval with WNT inhibitors (C59, 1 µM) or CHIR99201 (CHIR, 1 µM) in the presence of base media condition ABBr (ACTIVIN, 10ng/mL; BMP4, 30ng/mL; 8-Br-CAMP, 1mM). D) qPCR gene expression of day-6 cells after treatment of PFG cells on the day 4-6 interval a BMP inhibitor (DM: DM3189, 250 nM) or varying doses of BMP4 (3-25 ng/mL) on top of pancreatic inducing base media condition (ACPRS = ACTIVIN + C59 + PD0325901 + ATRA + SANT1). A10B10: ACTIVIN, 10 ng/mL; BMP4, 10 ng/mL. E) qPCR gene expression of day-6 cells after treatment of PFG cells with a FGF/ERK inhibitor (PD032: PD0325901, 500 nM) or varying doses of FGF2 (10-20 ng/mL) on top of pancreatic-inducing base media condition (ACPRS = ACTIVIN + C59 + PD0325901 + ATRA + SANT1) on the day 4-6 interval.
Figure S3. Related to Figure 3: Identification of hPSC-, endoderm- and liver bud progenitor-specific cell surface markers using BD Lyoplate antibody screens. A) Day-6 H1 hPSC-derived liver progenitors generated using SR2 or other methods (Si Tayeb et al., 2010; Zhao et al., 2013) were immunostained for AFP, HNF4A and TBX3 with a DAPI nuclear counterstain; scale bar = 500 µm. B) Percentages of H1, HES2 and BJC1 hPSC-derived endoderm (blue) and liver bud progenitors (red) expressing surface marker CD99, as shown by FACS. C) qPCR gene expression of H9 hPSC-derived liver bud progenitors generated from SR2 described in the present study and methods previously described in the literature (Si Tayeb et al., 2010; Zhao et al., 2013). Error bars represent ± SE.
Experiment strategy to treat day-6 liver bud progenitors (LB) with NOTCH modulators on the day 7-12 interval and assay effects on hepatic gene expression by day 12, as shown in subpanel A. B) qPCR gene expression of day-16 hepatocytes generated from liver bud progenitors treated on the day 7-12 interval with a NOTCH inhibitor (DAPT, 10 µM). Each qPCR heatmap is representative of 4 independent experiments with technical duplicates. C) qPCR gene expression of day-16 hepatocytes generated from liver bud progenitors treated on the day 7-12 interval with INSULIN (INS: 0.7, 1, 3, 10 and 30 µg/mL). D) Experimental strategy to treat day-6 liver bud progenitors (LB) with a TGFβ inhibitor (SB505124, 1 µM) on the day 7-8, 9-10, 11-12 or 7-12 intervals and assay downstream effects on hepatic gene expression by day 16, as shown in subpanel E. E) qPCR gene expression of day-16 hepatocytes generated from liver bud progenitors treated on the day 7-12, 12-16 or 16-17 intervals with a TGFβ inhibitor (SB505124, 1µM). Each qPCR heatmap is representative of 3 independent experiments with duplicates. F) qPCR gene expression of day-16 hepatocytes generated from liver bud progenitors treated on the day 7-12, 12-16 or 16-17 intervals with Ascorbic acid 2-phosphate (AAP, 200 µg/mL). G) qPCR gene expression of day 12 liver cells after treatment with 8-BromocAMP (cAMP) and Forskolin (Fsk) in the presence of Dexamethasone (Dex). H) Immunostaining of CK7 and CK19 expression in day-13 hPSC-derived biliary progenitors, with DAPI nuclear counterstain DAPI, scale = 100 µm. I) Gene ontology terms enriched amongst the genes regulated by DAPT, AAP or INS during the liver bud → hepatic progenitor differentiation step. J) Immunostaining of ALB (ALBUMIN) and ASGR1 expression in day-18 hPSC-derived hepatocyte-like cells, with DAPI nuclear counterstain DAPI, scale = 50 µm. K) Protein expression of ALBUMIN, HGD and GAPDH in hPSCs (undifferentiated H1, H7, H9), day-18 H1-, H7- and H9-derived hepatocytes and adult human hepatocytes (AHH) as shown by western blot. L) FIBRINOGEN levels detected in culture medium grown with hPSC-derived hepatocytes as measured by ELISA; “Base” denotes the culture
medium alone. M) Day-18 hPSC-derived hepatocytes stained with Periodic Acid-Schiffs (PAS) stain; scale = 400µm. Each western blot is a representative of 2 independent experiments. Error bars represent ± SE.
Figure S5. Related to Figure 5: Generation of FAH-Clover hPSC reporter line by CRISPR/Cas9 mediated gene editing. A) qPCR gene expression of day-18 H1 or H9 hPSC-derived hepatocytes generated by 3 methods: SR2 and previously-reported methods (Si Tayeb et al., 2010; Zhao et al., 2013). B) Day-18 hepatocyte-like cells generated by 4 methods: SR2 (the method described in the present work) and previously-reported methods (Carpentier et al., 2016; Avior et al., 2015; Chen et al., 2012) were stained for ALBUMIN (ALB) and Alpha1 anti-trypsin (AAT) with a DAPI nuclear counterstain and multiple fields were stitched into 1 image. C) In vitro secretion of human serum ALBUMIN into cell culture media by hPSC-derived hepatocytes as measured by ELISA. D) Schematic of FAH-pMIA3 plasmid. E) Schematic of eGxxFP plasmid. F) Guide RNA and eSpCas9 cleavage efficacy test. HEK293T cells transfected with the pCAG-eGxxFP plasmid, FAH-pMIA3 plasmid and both plasmids; scale = 1000 µm. G) PCR screening of H1 FAH-2AClover clones with eGxxFP primers after Cre-excision. H) PCR screening of H1 FAH-2AClover clones with FAH-Clo junction primers. Arrows illustrates DNA regions that PCR screening primers bind to. I) PCR screening of H1 FAH-2AClover clones with eGxxFP primers. J) Sequencing of PCR products from WT allele of clones with no targeting on second allele with donor vector-specific primers. Error bars represent ± SE.
Figure S6. Related to Figure 6: Tracking hPSC-derived liver cells after transplantation into neonatal and adult FRG mice. 

A) Experimental strategy to inject hPSC-derived hepatocytes into the livers of neonatal FRG mice (<72 hrs old) or adult FRG mice (4-6 weeks old) and to subsequently induce liver injury; H9 $E{F}_{14}$-BCL2-2A-GFP;UBC-tdTomato-2A-Luciferase hPSCs (Loh et al., 2016) were used to generate the hepatocytes for injections. 

B) In vivo tracking of 2 FRG mice that were intrahepatically injected as neonates with hPSC-derived liver cells; bioluminescence imaging of the FRG mice 8 or 16 weeks (8 wks, 16 wks) after inducing liver injury. 5 out of 7 mice exhibited an increase in bioluminescence signals over time. 

C) FRG mouse with detected bioluminescence before liver extraction (left) and the correspondingly extracted mouse liver (right); scale bar of signal intensity, in p/sec/cm$^2$/sr. 

D) Mouse liver sections stained for human ALBUMIN (green), as shown by immunostaining; tdTomato-expressing cells appeared red (arrowheads); scale bar =150µm. 

E) Bioluminescence imaging of the FRG mice injected with media as negative control, adult primary human hepatocytes and Luciferase+$+$ hPSC-derived liver cells (1, 4 and 8 weeks after injections). 

F) Kaplan-Meier’s survival curve depicting the percentage survival of adult FRG mice that were either injected with day 18 hPSC-derived (n=6) or media-only control (blue line) (n=5).
**Table S1. Related to Figure 3: Surface marker expression on hPSC, hPSC-derived endoderm and hPSC-derived liver bud progenitors.**

| Description | H7 UD | H7 D2 DE | H7 D6 LB | Description | H7 UD | H7 D2 DE | H7 D6 LB |
|-------------|-------|----------|----------|-------------|-------|----------|----------|
| abTCR       | 0.22  | 0.00882  | 0.57     | CD314(NKG2D)| 11    | 0.00901  | 0.69     |
| B2-aGlob    | 97.3  | 62.9     | 20       | CD32        | 0.064 | 0.00897  | 0.32     |
| BLTR-1      | 0.011 | 0        | 0.64     | CD321(F11 Repr)| 99.9 | 99.8     | 99       |
| CD10        | 87.6  | 64.3     | 2.84     | CD326       | 99.8  | 99.6     | 83.5     |
| CD100       | 60.2  | 0.38     | 0.71     | CD33        | 0.011 | 0.045    | 0.22     |
| CD102       | 19.2  | 0.91     | 1.28     | CD335(NKP46)| 0.074 | 0.027    | 1.48     |
| CD103       | 0     | 0.00957  | 0.23     | CD336       | 0.011 | 0.037    | 0.74     |
| CD104       | 6.72  | 0.00898  | 4.8      | CD337       | 0.096 | 0.00894  | 0.69     |
| CD105       | 17.9  | 5.09     | 0.66     | CD338(ABCG2)| 21.6 | 2.77     | 5.5      |
| CD106       | 0     | 0.00955  | 58.5     | CD34        | 0.011 | 0.019    | 30.2     |
| CD107a      | 4.85  | 48.2     | 16.4     | CD340(Her2)| 100   | 100      | 99.9     |
| CD107b      | 0.1   | 0.37     | 1.73     | CD35        | 1.75  | 0.13     | 0.35     |
| CD108       | 91.1  | 80.6     | 45.5     | CD36        | 0.043 | 0.045    | 0.78     |
| CD109       | 20.1  | 0        | 0.082    | CD37        | 0.011 | 0.027    | 0.24     |
| CD112       | 3.01  | 95.9     | 1.9      | CD38        | 0.31  | 0.063    | 0.55     |
| CD114       | 0     | 0.02     | 0.3      | CD39        | 0.032 | 0        | 0.42     |
| CD116       | 0.021 | 0.00999  | 0.36     | CD4         | 16    | 0.00959  | 0.13     |
| CD117       | 0.032 | 96.6     | 6.5      | CD40        | 60.5  | 57.5     | 1.75     |
| CD118(LIF R)| 0.096 | 3.59     | 24.5     | CD41a       | 0     | 0.00887  | 0.19     |
| CD119       | 99.9  | 68.2     | 82.9     | CD41b       | 0     | 0.045    | 0.19     |
| CD11a       | 0     | 0.011    | 0.27     | CD42a       | 0.011 | 0.02     | 0.14     |
| CD11b       | 0.011 | 0.00925  | 0.18     | CD42b       | 0     | 0.027    | 0.29     |
| CD11c       | 0.032 | 0.028    | 0.42     | CD43        | 43    | 0.13     | 0.92     |
| CD120a      | 5.81  | 2.49     | 0.85     | CD44        | 95.8  | 48.8     | 13       |
| CD120b      | 4.64  | 0.036    | 0.25     | CD45        | 0     | 0.00893  | 0.19     |
| CD121a      | 0.011 | 0.00998  | 0.33     | CD45RA      | 0.064 | 0.018    | 0.2      |
| CD121b      | 0.022 | 0.04     | 0.54     | CD45RB      | 0.05  | 0.018    | 0.34     |
| CD122       | 0.011 | 0.00992  | 0.23     | CD45RO      | 8.77  | 0        | 0.074    |
| CD123       | 0.043 | 0.00971  | 0.25     | CD46        | 99.9  | 100      | 99.9     |
| CD124       | 0.15  | 0.00977  | 4        | CD47        | 100   | 99.6     | 98.4     |
| CD126       | 0.17  | 0.00976  | 2.52     | CD48        | 0.014 | 0.036    | 0.18     |
| CD127       | 0.032 | 0.04     | 0.44     | CD49a       | 32.8  | 0.22     | 27.2     |
| CD128b      | 2.01  | 0.02     | 0.32     | CD49b       | 100   | 76.1     | 94.5     |
| CD13        | 4.37  | 23.1     | 33       | CD49c       | 99.9  | 66.1     | 99.7     |
| CD130       | 2.48  | 36       | 60.4     | CD49d       | 40.2  | 42       | 0.55     |
| CD132       | 0.011 | 0.054    | 0.14     | CD49e       | 100   | 99.9     | 99.5     |
| CD134       | 0     | 0.02     | 0.79     | CD49f       | 99.8  | 44.3     | 97.9     |
| CD135       | 0.043 | 0.049    | 0.3      | CD4V4       | 11.3  | 0.27     | 0.28     |
| CD137       | 0.011 | 0.03     | 5.49     | CD5         | 0.032 | 0.00959  | 0.3      |
| CD138       | 2.2   | 0.03     | 0.28     | CD50        | 96.8  | 38.8     | 0.96     |
| CD138       | 0     | 0.02     | 0.31     | CD51/61     | 0.036 | 10.9     | 4.69     |
| CD14 | 0.032 | 0.009 | 0.23 | CD53 | 0.036 | 0 | 0.27 |
|------|-------|-------|------|------|-------|---|-----|
| CD140a | 0.022 | 36.3 | 37.8 | CD54 | 90.6 | 68.9 | 54.1 |
| CD140b | 62.5 | 46.9 | 1.87 | CD55 | 100 | 100 | 97.1 |
| CD141 | 3.27 | 6.41 | 34.3 | CD56 | 99.9 | 83.2 | 55.9 |
| CD142 | 98.9 | 94.5 | 98.4 | CD57 | 99.4 | 99.7 | 93 |
| CD144 | 1.4 | 4.34 | 0.52 | CD58 | 100 | 100 | 99.6 |
| CD146 | 99.8 | 86.5 | 80.4 | CD59 | 99.9 | 100 | 99.9 |
| CD147 | 99.4 | 68.6 | 61.9 | CD6 | 0 | 0 | 0.19 |
| CD15 | 0.38 | 58.6 | 1.11 | CD61 | 0.014 | 13 | 4.92 |
| CD150 | 0.032 | 2.16 | 0.2 | CD62E | 0 | 0.00912 | 0.18 |
| CD151 | 26.9 | 99.6 | 98.5 | CD62L | 0.021 | 0.018 | 0.31 |
| CD152 | 0.011 | 0.029 | 0.47 | CD62P | 2.85 | 0.00891 | 0.27 |
| CD153 | 0.022 | 0.039 | 0.24 | CD63 | 99 | 99.8 | 96.7 |
| CD154 | 0.011 | 0.02 | 0.3 | CD64 | 0.014 | 0.00895 | 0.081 |
| CD158a | 0.032 | 0.021 | 0.25 | CD66 (a,c,d,e) | 63.4 | 23.3 | 35.7 |
| CD158b | 0 | 0.00983 | 0.16 | CD66b | 0.028 | 0 | 0.063 |
| CD15s | 0 | 0.018 | 0.18 | CD66f | 0.021 | 0.00911 | 0.21 |
| CD16 | 0.043 | 0 | 0.083 | CD69 | 0 | 0.0091 | 0.18 |
| CD161 | 0.032 | 0.019 | 0.21 | CD7 | 1.41 | 0.46 | 1.23 |
| CD162 | 0.021 | 0.039 | 0.34 | CD70 | 0.028 | 0.036 | 0.076 |
| CD163 | 0.032 | 0.02 | 0.4 | CD71 | 40.4 | 81.3 | 15.1 |
| CD164 | 66.1 | 77.9 | 70.8 | CD72 | 0.38 | 0.082 | 0.46 |
| CD165 | 99.8 | 99.9 | 91.9 | CD73 | 0.38 | 0.67 | 2.13 |
| CD166 | 100 | 30 | 94.9 | CD74 | 0.014 | 0.00934 | 0.12 |
| CD171 | 96.7 | 96.4 | 4.09 | CD75 | 0.16 | 0.064 | 18.3 |
| CD172b | 0.011 | 1.13 | 0.68 | CD77 | 0.21 | 0.11 | 0.45 |
| CD177 | 0 | 14.6 | 8.58 | cd79B | 0.021 | 0.00925 | 0.15 |
| CD178 | 7.9 | 0.02 | 0.39 | CD80 | 9.02 | 0.064 | 0.22 |
| CD18 | 0.011 | 0.041 | 0.17 | CD81 | 99.8 | 99.9 | 99.5 |
| CD180 | 0 | 0.00968 | 0.22 | CD83 | 0.029 | 0.037 | 1.1 |
| CD181 | 0 | 0.14 | 0.48 | CD84 | 0.028 | 0 | 0.21 |
| CD183 | 0.011 | 0.029 | 0.19 | CD85 | 0.016 | 0.00933 | 0.17 |
| CD184 | 0 | 99.9 | 5.62 | CD86 | 7.03 | 0.11 | 0.44 |
| CD19 | 0.021 | 0 | 0.12 | CD87 | 0.011 | 0 | 0.46 |
| CD193 | 0.085 | 0.27 | 0.52 | CD88 | 0.021 | 0.00944 | 0.25 |
| CD195 | 0.032 | 0.00998 | 0.37 | CD89 | 1.38 | 0 | 0.26 |
| CD196 | 0.053 | 0.059 | 0.36 | CD8a | 0.22 | 0.037 | 0.15 |
| CD197 | 7.41 | 0.03 | 0.34 | CD8b | 9.6 | 0.019 | 0.35 |
| CD1a | 0.021 | 0 | 0.34 | CD9 | 99.9 | 90.8 | 27.6 |
| CD1b | 0.021 | 0 | 0.2 | CD90 | 99.6 | 99.7 | 85.2 |
| CD1d | 0 | 0.05 | 0.16 | CD91 | 0.011 | 16 | 18.7 |
| CD2 | 0.032 | 0.01 | 0.15 | CD94 | 0.011 | 0 | 0.11 |
| CD20 | 0 | 0.036 | 0.18 | CD95 | 37.1 | 4.22 | 33 |
| CD200 | 100 | 99.9 | 99.3 | CD97 | 11.3 | 0.019 | 0.52 |
| CD201 | 0  | 0.39 | 53.1 | CD98 | 96.9 | 61.3 | 73.7 |
|-------|----|------|------|-----|------|------|------|
| CD205 | 55.3 | 0.17 | 4.85 | CD99 | 1.88 | 2.16 | 87   |
| CD206 | 0.011 | 0.04 | 0.23 | CD99R | 0.011 | 0.00952 | 1.37 |
| CD209 | 0.13 | 0.12 | 0.53 | CDw327 | 0.075 | 0.48 | 0.41 |
| CD21  | 58.2 | 0.095 | 0.38 | CDw328 | 0.011 | 0.00904 | 1.11 |
| CD210 | 0.011 | 0.019 | 0.26 | CDw329 | 0.16 | 0 | 0.59 |
| CD212 | 0 | 0.064 | 0.22 | CDw93 | 18 | 0.03 | 0.45 |
| CD22  | 0.032 | 0.00895 | 0.16 | CLIP | 0.011 | 0.036 | 0.3  |
| CD220 | 61 | 8.1 | 50.2 | CMRF-44 | 0.042 | 0.027 | 71.9 |
| CD221 | 99.9 | 100 | 92.6 | CMRF-56 | 0.011 | 0.00888 | 0.59 |
| CD226 | 0.074 | 0.02 | 0.34 | Cutaneous Lymph. Antigen | 0.011 | 0.046 | 0.78 |
| CD227 | 28 | 1.84 | 10.3 | Disialoganglioside GD2 | 0.18 | 9.34 | 1.21 |
| CD229 | 0.021 | 0.02 | 0.21 | EGF-r | 39.8 | 1.8 | 72.8 |
| CD23  | 23.1 | 0 | 0.31 | Fmly-p-r | 0.085 | 0.027 | 1.79 |
| CD231 | 0 | 0 | 0.31 | gd TCR | 0.053 | 0.027 | 0.53 |
| CD235a | 0.011 | 0.18 | 0.41 | Hem. Prog. Cell | 27 | 1.79 | 20.5 |
| CD24  | 99.8 | 100 | 98 | HLA-A,B,C | 99.9 | 99.8 | 96.4 |
| CD243 | 0.021 | 0 | 0.55 | HLA-A2 | 91 | 20.2 | 26.4 |
| CD244 | 0.011 | 0.00994 | 0.73 | HLA-DQ | 14.4 | 8.09 | 10.9 |
| CD25  | 0.021 | 0 | 0.19 | HLA-DR | 2.41 | 0.22 | 0.69 |
| CD255 | 0.021 | 0.13 | 0.32 | HLA-DR,DQ,DO | 0.095 | 0.00904 | 2.34 |
| CD26  | 0.19 | 1.71 | 37 | INT B7 | 0 | 0.00921 | 0.76 |
| CD267 | 0 | 0.046 | 0.27 | Invariant NKT | 0.053 | 0.00903 | 0.67 |
| CD268 | 0 | 0.02 | 0.41 | MIC A/B | 99.3 | 94.1 | 1.93 |
| CD27  | 0.011 | 0.0089 | 0.29 | mIgG1 | 0.032 | 0.045 | 0.52 |
| CD271 | 93 | 10.2 | 1.39 | mIgG2a | 0.053 | 0.018 | 0.66 |
| CD273 | 0.011 | 0.03 | 0.51 | mIgG2b | 0.053 | 0.036 | 0.38 |
| CD274 | 0 | 0.049 | 1.82 | mIgG3 | 0.064 | 0.00905 | 0.54 |
| CD275 | 0 | 1.06 | 1.03 | mIgM | 0.043 | 0.027 | 0.34 |
| CD278 | 0 | 0.02 | 0.51 | NKB1 | 0.011 | 0.0091 | 0.59 |
| CD279 | 5.89 | 0.098 | 1.11 | rIgG1 | 0 | 0.063 | 0.25 |
| CD28  | 0 | 0.018 | 0.13 | rIgG2a | 0.011 | 0.045 | 0.22 |
| CD282 | 0.075 | 0 | 0.55 | rIgG2b | 0.011 | 0.027 | 0.41 |
| CD29  | 99.5 | 93.4 | 70.7 | rIgM | 0 | 0.046 | 0.5 |
| CD294 | 0.011 | 0.036 | 0.32 | SSEA-1 | 17.8 | 86.9 | 13.6 |
| CD3  | 0.011 | 0.00958 | 0.2 | SSEA-3 | 0 | 7.7 | 0.55 |
| CD30  | 100 | 40.7 | 1.35 | SSEA-4 | 99.6 | 57.1 | 96.1 |
| CD305 | 0.13 | 0.027 | 0.38 | TRA-1-60 | 97.4 | 25.8 | 48.4 |
| CD309 | 91 | 42.9 | 1.41 | TRA-1-81 | 95.8 | 21 | 45.7 |
| CD31  | 0.27 | 0.046 | 1.34 | Vb 23 | 0.032 | 0.00891 | 0.4 |
|       |     |     |     | Vb 8 | 30.9 | 0.00889 | 0.29 |
| Primers for pMIA3 1sg-eSpCas9-2AmRuby2- | Primers for FAH-2AClover-PGK-Bsd2ATK-FAH-PGK |
|------------------------------------------|-----------------------------------------------|
| EF1a_F                                   | gaagaggttgacataggGATTTTCTCTGCTTCTTGGAAC        |
| EF1a_F                                   | cgccctttttgcccgtTGCCCTAGAGACATGCT             |
| Amp pUC_F                                | ATGGACGAGCTgtacaagtaaaGAATTCCGATCATATTCAAT    |
| Amp pUC_R                                | CTCTAGGATTTCTATTACTTTGTACAGCTGCTCCAT          |
| sgRNA cassette_F                         | gacctgatggGATAAGGGGCCGGGATTTTG                |
| sgRNA cassette_F                         | cggctagccggtgccctgtggaattccggtccctgtgaatattc |
| Cas9 3' _F                              | GCCCTGAGCTGCTGCTGGATTTTG                      |
| Cas9 5' _F                              | CGCGCTGGGGATCACCATCATGGAAAGAGCAGCTTGAGAAG    |
| Cas9 3' _R                              | TGAATCTTGGCCATATTTGAAGTCAGGTTAGCTGCTGCTCC   |
| Cas9 5' _R                              | CTCTAGGATTTCTATTACTTTGTACAGCTGCTCCAT        |
| eSp mutations_F                          | GCCCTGAGCTGCTGCTGGATTTTG                      |
| eSp mutations_R                          | GCCCTGAGCTGCTGCTGGATTTTG                      |
| hFAH 3' homo F                           | ggagaggttgacataggGATTTTCTCTGCTTCTTGGAAC      |
| hFAH 3' homo R                           | cgccctttttgcccgtTGCCCTAGAGACATGCT             |
| mPgk prom_F                              | ATGGACGAGCTgtacaagtaaaGAATTCCGATCATATTCAAT    |
| mPgk prom_R                              | CTCTAGGATTTCTATTACTTTGTACAGCTGCTCCAT        |
| Blistaticidin_F                          | GCCCTGAGCTGCTGCTGGATTTTG                      |
| Blistaticidin_R                          | GCCCTGAGCTGCTGCTGGATTTTG                      |
| mPgk1 polyA F II                         | GCCCTGAGCTGCTGCTGGATTTTG                      |
| mPgk1 polyA_R                            | GCCCTGAGCTGCTGCTGGATTTTG                      |
| Bsd II F                                 | GCCCTGAGCTGCTGCTGGATTTTG                      |
| Bsd II R                                 | GCCCTGAGCTGCTGCTGGATTTTG                      |
| P2A TK F                                 | GCCCTGAGCTGCTGCTGGATTTTG                      |
| P2A TK R                                 | GCCCTGAGCTGCTGCTGGATTTTG                      |
| mPgk DTA_F                               | GCCCTGAGCTGCTGCTGGATTTTG                      |
| mPgk DTA_R                               | GCCCTGAGCTGCTGCTGGATTTTG                      |
| DTA bR_F                                 | GCCCTGAGCTGCTGCTGGATTTTG                      |
| hFAH 5' homo F                           | GCCCTGAGCTGCTGCTGGATTTTG                      |
| hFAH 5' homo R                           | GCCCTGAGCTGCTGCTGGATTTTG                      |
| DTA end of CDS target region             | GCCCTGAGCTGCTGCTGGATTTTG                      |
| FAH eGxxFP_F                             | GCCCTGAGCTGCTGCTGGATTTTG                      |
| FAH eGxxFP_R                             | GCCCTGAGCTGCTGCTGGATTTTG                      |

PCR screening
Table S3. Related to Figures 2, 3 and 5: List of antibodies for immunostaining or FACS

| Antibody          | Catalog No. | Stock conc. | Vendor          | App<sup>1</sup> | Host   | Dilution factor                     |
|-------------------|-------------|-------------|-----------------|-----------------|--------|-------------------------------------|
| α-ALBUMIN         | ab2406      | 1-5 mg/mL   | Abcam           | IF              | Rabbit | 1 in 100 to 1 in 500 10µg/mL       |
| α-TBX3            | AF4509      | 200µg/mL    | R&D             | IF              | Goat   | 1 in 100 2µg/mL                    |
| α-HNF4A           | SC8987      | 200µg/mL    | Santa Cruz      | IF              | Rabbit | 1 in 100 4.67µg/mL                 |
| α-AFP             | A00829      | NA          | Dako Cytomation | IF, FCS         | Rabbit | 1 in 300 FACS: 1.8µg/mL            |
| α-AAT             | A001202     | NA          | Dako Cytomation | IF              | Rabbit | 1 in 500                           |
| α-CPS1            | Ab45956     | 1mg/mL      | Abcam           | IF              | Rabbit | 1 in 100 10µg/mL                  |
| α-KRT19           | HPA0024 65  | 200µg/mL    | Sigma           | IF              | Rabbit | 1 in 100 2µg/mL                    |
| α-KRT7            | HPA0072 72  | 200µg/mL    | Sigma           | IF              | Rabbit | 1 in 150                           |
| α-ALB-APC         | IC1455A     | NA          | R&D             | IF              | Mouse  | 0.5µL/150k cells                   |
| α-CD10-APC        | 340923      | 25µg/mL     | BD Biosciences  | FCS             | Mouse  | 5µL per test                       |
| α-CD10-APC        | 312209      | NA          | Biolegend       | FCS             | Mouse  | 1 in 10                            |
| α-CD99-PE         | 555689      | NA          | BD Biosciences  | FCS             | Mouse  | 20µL per test                      |
| α-CD99-APC        | 318010      | NA          | Biolegend       | FCS             | Mouse  | 5µL per test or 1 in 10           |
| α-CXCR4-PeRCP-Cy 7| 560669      | NA          | Biolegend       | FCS             | Mouse  | 1 in 5                             |
| α-CD201-APC       | 351905      | NA          | Biolegend       | FCS             | Rat    | 1 in 10                            |
| α-Mouse IgG_Alexa Fluor 546 | A-10036   | 2mg/mL      | Thermofisher Scientific | IF | Donkey | 1 in 1000                         |
| α-Goat IgG Alexa Fluor® 546 | A-11055   | 2mg/mL      | Thermofisher Scientific | IF | Donkey | 1 in 1000                         |
| α-Rabbit (H+L)_Alexa Fluor 546 | A10040    | 2mg/mL      | Thermofisher Scientific | IF | Donkey | 1 in 1000                         |
| α-Rabbit IgG Alexa Fluor 488 | A21206    | 2mg/mL      | Thermofisher Scientific | IF | Donkey | 1 in 1000                         |
| α-Mouse IgG Alexa Fluor 488 | A21202    | 2mg/mL      | Thermofisher Scientific | IF | Donkey | 1 in 1000                         |
| α-Goat IgG Alexa Fluor® 488 | A11055    | 2mg/mL      | Thermofisher Scientific | IF | Donkey | 1 in 1000                         |
| anti-ALB          | ab2406      | 5mg/mL      | Abcam           | WB              | Rabbit | 1 in 1000                         |
| anti-GAPDH        | ab8245      | 2mg/mL      | Abcam           | WB              | Mouse  | 1 in 1000                         |
| anti-Mouse IgG HRP | NA931       | NA          | GE Healthcare   | WB              | Sheep  | 1 in 10000                        |
| Forward Primer | Reverse Primer |
|----------------|----------------|
| YWHAZ          | CCT TGC TCA GTT ACA GAC TTC ATG CA |
| FAH            | GGA CAG GAC CAG TAA AGA GGT G |
| ALB            | CAT CCT CTG CAT TGA AGG GA |
| TAT            | CAT CCA GAA TTT AGG GGA GGT T |
| TBX3           | CAT CCT CTG CAT TCC GAG G |
| ASS1           | CTT CGT GTA GAG ACC TGG AGG C |
| ASL            | CCA TCC TGC TGA GAT GGG TCA T |
| GS             | TTG CTC ACC ATG GCC CAT TCA C |
| CYP3A4         | GGT GAA GGT TGG AGA CAG CAA TG |
| CYP3A7         | AAC GTC CAA TAG CCC TTA CGG A |
| CPS1           | CAA TGG TGT CTG CCT GGT CAT CTG |
| ARG1           | TTC TTC ACA TCA CAC TCT TGT TGC T |
| AAT            | CGG CAT TGA CGA TCT GGT GAC G |
| FGG            | ACG GTC TTT TAA AGC TGC CCA GC |
| FGB            | AGC AAA GCC TCT TCT TCT TTT |
| FGA            | TGG CCA CTT GCT CCA GCC A |
| MAI            | TAG GGG GTG AGA TCC ACC TTA G |
| HPD            | TTT TCT GTC GTA GTC AGC CAG G |
| OTC            | GCC ATG CTA GAC AAT ACA CGG G |
| Thrombinogen   | TCT CAC AAG CTG TGT AGG G |
| PAH            | AGT AAA CCA GTA AAT TGG GCC GAG C |
| HGD            | GCG GAA CAA TCA AGA GGG G |
| CEBPA          | ATT GTC ACT GGG TCT CAG GAC |
| PROX1          | GGT GAA GGT TGG AGA CAG CAA |
| HNF1B          | GAT GAC AGG GAC ACT TGT TGC T |
| HNF6           | GCC CTC AAT TAC TTA CAG TGC T |
| SOX9           | GCC CTC TCT GCT CGC CTG TTC GAA GT |
| HNF1A          | GGT GAT CTA CGT TCA CCA ACA C |
| CK19           | GTA ACC TGC GCA TGC TGC TGC |
| AFP*           | TAA TGT CAG CGC TCT CTC CTG |
| HNF4A*         | CAT ACT GGC GGT CGT TGA TG |
| HEX            | ATC GCC CTC AAT GTC GAC C |
| PDX1           | AGC TCC CCC GCT GTG TGT TGG AGG |
| SOX17          | GGA TCA GGG ACC TGT CAC AC |
| CDX2           | CCT TGG CTC TGG GGT GCT TCG |
| AFP            | GCA TGT TGA TTT AAG CTG CTG |
| OTC*           | TGG TAG TCA GGT GTG GTG TGT GAC A |
| HNF4A*         | AGT CAT TGC CTA GGA GCA GCA C |

1 IF: Immunofluorescence; WB: Western blot; FCS: Flow cytometry

Table S4. Related to Figures 1, 2, 4 and 5: List of primers for qRT-PCR
Unless indicated, primers are compatible with QuantiTect Fast SYBR green kit and KAPA SYBR green kit; *Compatible with KAPA Sybr green kit.

SUPPLEMENTAL REFERENCES

Abrámoff, M.D., and Magalhães, P.J. (2004). Image processing with ImageJ. Biophotonics International 11, 36-42.

Collins, L. (2007). Image processing with ImageJ. Biophotonics International 11, 36-42.

Cong, L., et al. (2013). Multiplex Genome Engineering Using CRISPR/Cas Systems. Science 339, 819-823.

Hockemeyer, D., et al. (2011). Genetic engineering of human pluripotent cells using TALE nucleases. Nature Biotechnology 29, 731-734.

Huang, D.W., Sherman, B.T., and Lempicki, R.A. (2008). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nature Protocols 4, 44-57.

Huang, D.W., Sherman, B.T., and Lempicki, R.A. (2009). Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic acids research 37, 1-13.

Ihry, R.J., et al. (2017). P53 toxicity is a hurdle to CRISPR/CAS9 screening and engineering in human pluripotent stem cells. bioRxiv.

Kim, J.H., Lee, S.-R., Li, L.-H., Park, H.-J., Park, J.-H., Lee, K.Y., Kim, M.-K., Shin, B.A., and Choi, S.-Y. (2011). High Cleavage Efficiency of a 2A Peptide Derived from Porcine Teschovirus-1 in Human Cell Lines, Zebrafish and Mice. PLoS One 6, e18556.

Kleinstiver, B.P., Pattanayak, V., Prew, M.S., Tsai, S.Q., Nguyen, N.T., Zheng, Z., and Joung, J.K. (2016). High-fidelity CRISPR–Cas9 nucleases with no detectable genome-wide off-target effects. Nature 529, 490-495.

Lam, A.J., et al. (2012). Improving FRET dynamic range with bright green and red fluorescent proteins. Nature Methods 9, 1005-1012.

Loh, K.M., et al. (2014). Efficient Endoderm Induction from Human Pluripotent Stem Cells by Logically Directing Signals Controlling Lineage Bifurcations. Cell Stem Cell 14, 237-252.

Loh, K.M., et al. (2016). Mapping the Pairwise Choices Leading from Pluripotency to Human Bone, Heart, and Other Mesoderm Cell Types. Cell 166, 451-467.

Mashiko, D., Fujihara, Y., Satouh, Y., Miyata, H., Isotani, A., and Ikawa, M. (2013). Generation of mutant mice by pronuclear injection of circular plasmid expressing Cas9 and single guided RNA. Scientific Reports 3, 73.

Momcilovic, O., Choi, S., Varum, S., Bakkenist, C., Schatten, G., and Navara, C. (2009). Ionizing Radiation Induces Ataxia Telangiectasia Mutated-Dependent Checkpoint Signaling and G 2But Not G 1Cell Cycle Arrest in Pluripotent Human Embryonic Stem Cells. Stem Cells 27, 1822-1835.

Momcilovic, O., Knobloch, L., Fornalsaglio, J., Varum, S., Easley, C., and Schatten, G. (2010). DNA Damage Responses in Human Induced Pluripotent Stem Cells and Embryonic Stem Cells. PloS One 5, e13410.

Reich, M., Liefeld, T., Gould, J., Lerner, J., Tamayo, P., and Mesirov, J.P. (2006). GenePattern 2.0. Nature Genetics 38, 500-501.

Si Tayeb, K., Noto, F.K., Nagaoka, M., Li, J., Battle, M.A., Duris, C., North, P.E., Dalton, S., and Duncan, S.A. (2010). Highly efficient generation of human hepatocyte–like cells from induced pluripotent stem cells. Hepatology 51, 297-305.

Slaymaker, I.M., Gao, L., Zetsche, B., Scott, D.A., Yan, W.X., and Zhang, F. (2015). Rationally engineered Cas9 nucleases with improved specificity. Science 351, 84-88.

Szymczak, A.L., Workman, C.J., Wang, Y., Vignali, K.M., Diliglou, S., Vanin, E.F., and Vignali, D.A.A. (2004). Correction of multi-gene deficiency in vivo using a single ‘self-cleaving’ 2A peptide–based retroviral vector. Nature Biotechnology 22, 589-594.

Touboul, T., et al. (2010). Generation of functional hepatocytes from human embryonic stem cells under chemically defined conditions that recapitulate liver development. Hepatology 51, 1754-1765.

Zhao, D., Chen, S., Duo, S., Xiang, C., Jia, J., Guo, M., Lai, W., Lu, S., and Deng, H. (2013). Promotion of the efficient metabolic maturation of human pluripotent stem cell-derived hepatocytes by correcting specification defects. Cell Research 23, 157-161.