Stem Cell Reports, Volume *9*

Supplemental Information

Distinct Gene Expression and Epigenetic Signatures in Hepatocyte-like Cells Produced by Different Strategies from the Same Donor

Yimeng Gao, Xiaoran Zhang, Ludi Zhang, Jin Cen, Xuan Ni, Xiaoying Liao, Chenxi Yang, Ying Li, Xiaotao Chen, Zhao Zhang, Yajing Shu, Xin Cheng, David C. Hay, Dongmei Lai, Guoyu Pan, Gang Wei, and Lijian Hui

Supplemental Information

Figure S1:

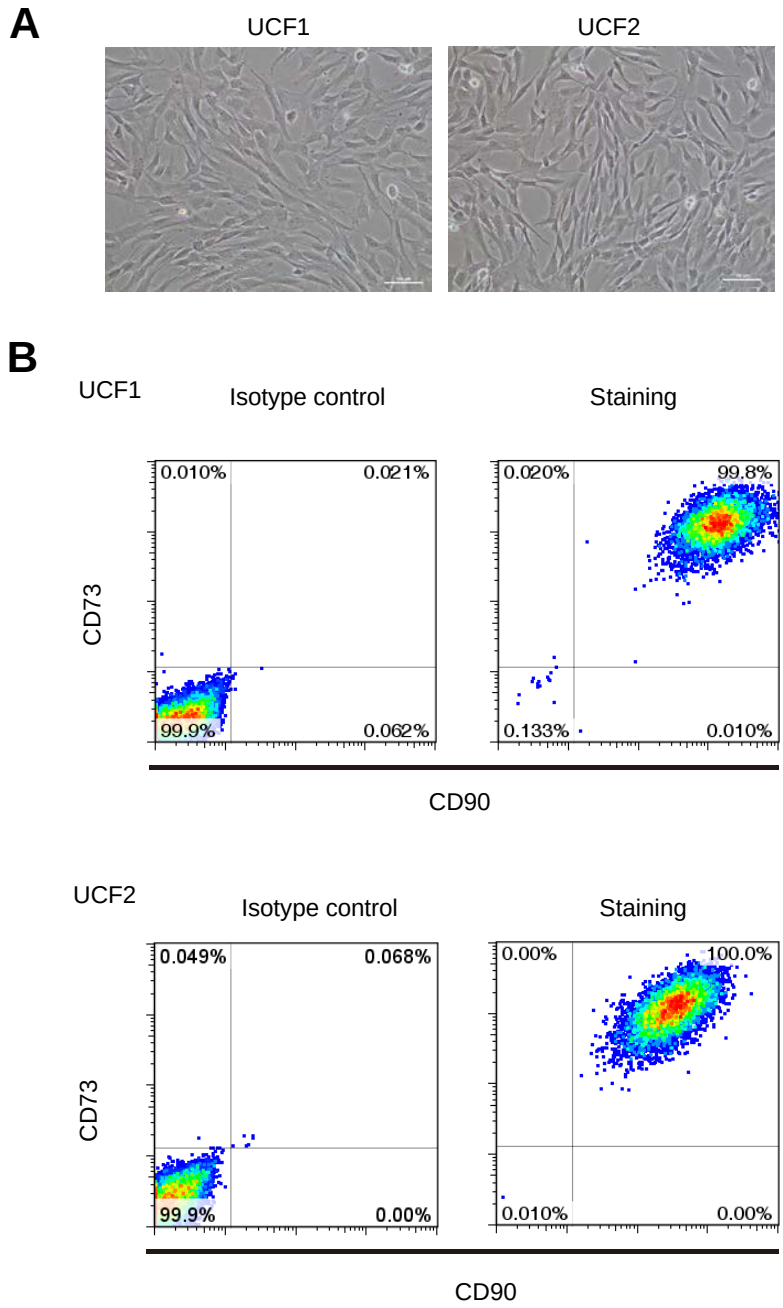


Figure S1. Characterization of umbilical cord-derived fibroblasts (UCF). Related to Figure 1.

(A) Fibroblast-like morphology of two UCFs. Scale bar, 100 μm.

(B) Flow cytometric analysis of surface marker expression on UCFs. Both UCF1 and UCF2 were positive for CD73 and CD90 as measured by flow cytometry.

Figure S2:

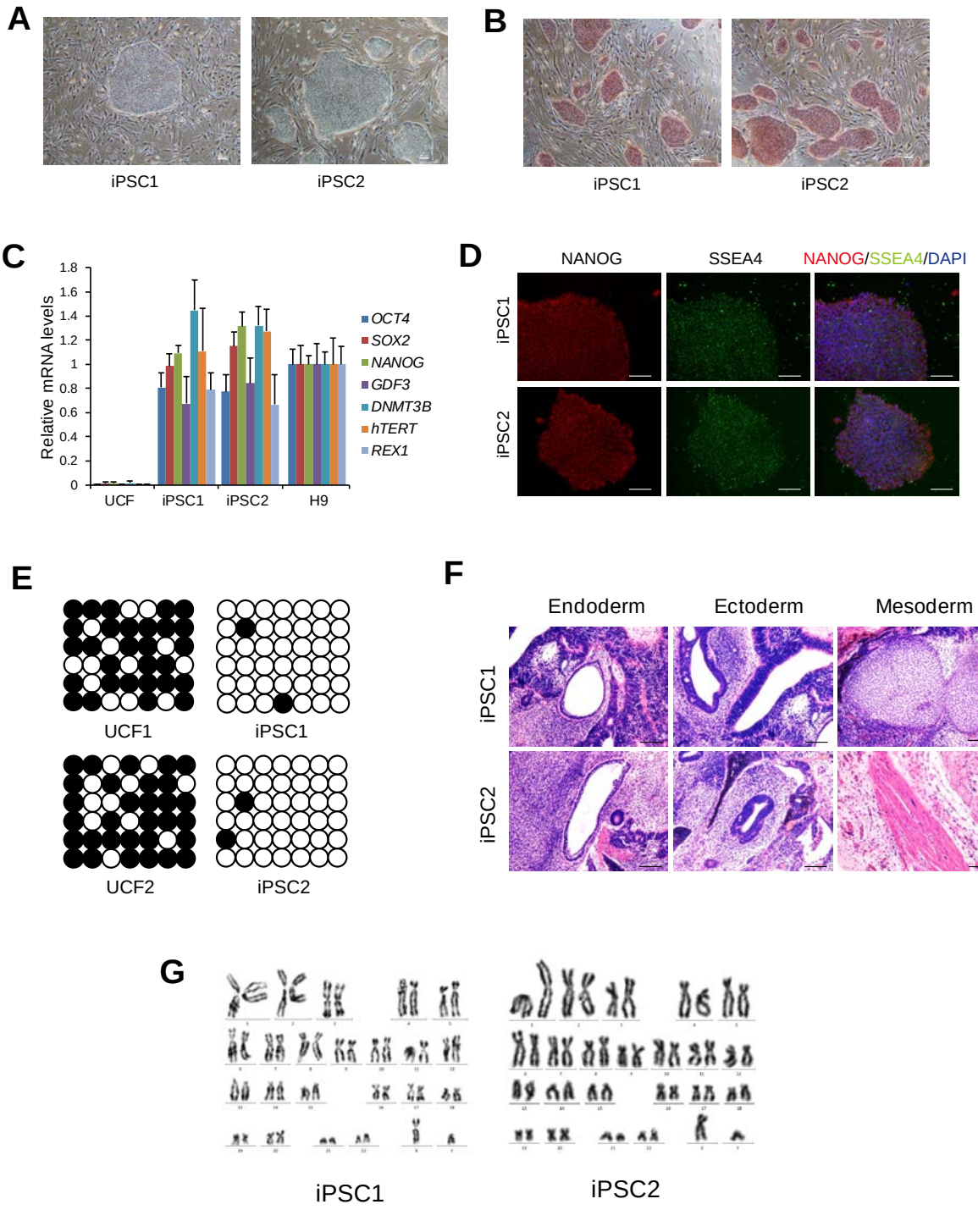

A
iPSC1
iPSC2
B
iPSC1
iPSC2
C
Relative mRNA levels
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
UCF
iPSC1
iPSC2
H9
OCT4
SOX2
NANOG
GDF3
DNMT3B
hTERT
REX1
D
NANOG
SSEA4
NANOG/SSEA4/DAPI
iPSC1
iPSC2
E
UCF1
iPSC1
UCF2
iPSC2
F
Endoderm
Ectoderm
Mesoderm
iPSC1
iPSC2
G
iPSC1
iPSC2

Figure S2. Characterization of human iPSCs. Related to Figure 1.

(A) iPSC lines showed typical morphology of human embryonic stem cells (ESC). Scale bar, 100 μm.

(B) iPSC lines showed alkaline phosphatase activities as measured by AP staining. Scale bar, 100 μm.

(C) iPSCs showed similar gene expression levels of pluripotent stem cell genes compared with human ESCs (H9).

(D) iPSCs were measured by immunofluorescent staining for pluripotent stem cell specific markers, NANOG and SSEA4. Scale bar, 100 μm.

(E) Bisulfite genomic sequencing of the promoter regions of OCT4. Open and closed circles indicated unmethylated and methylated CpGs.

(F) Hematoxylin and eosin (HE) staining of teratoma derived from iPSC lines. Scale bar, 100 μm.

(G) iPSC lines showed normal karyotypes after long-term culture.

Figure S3:

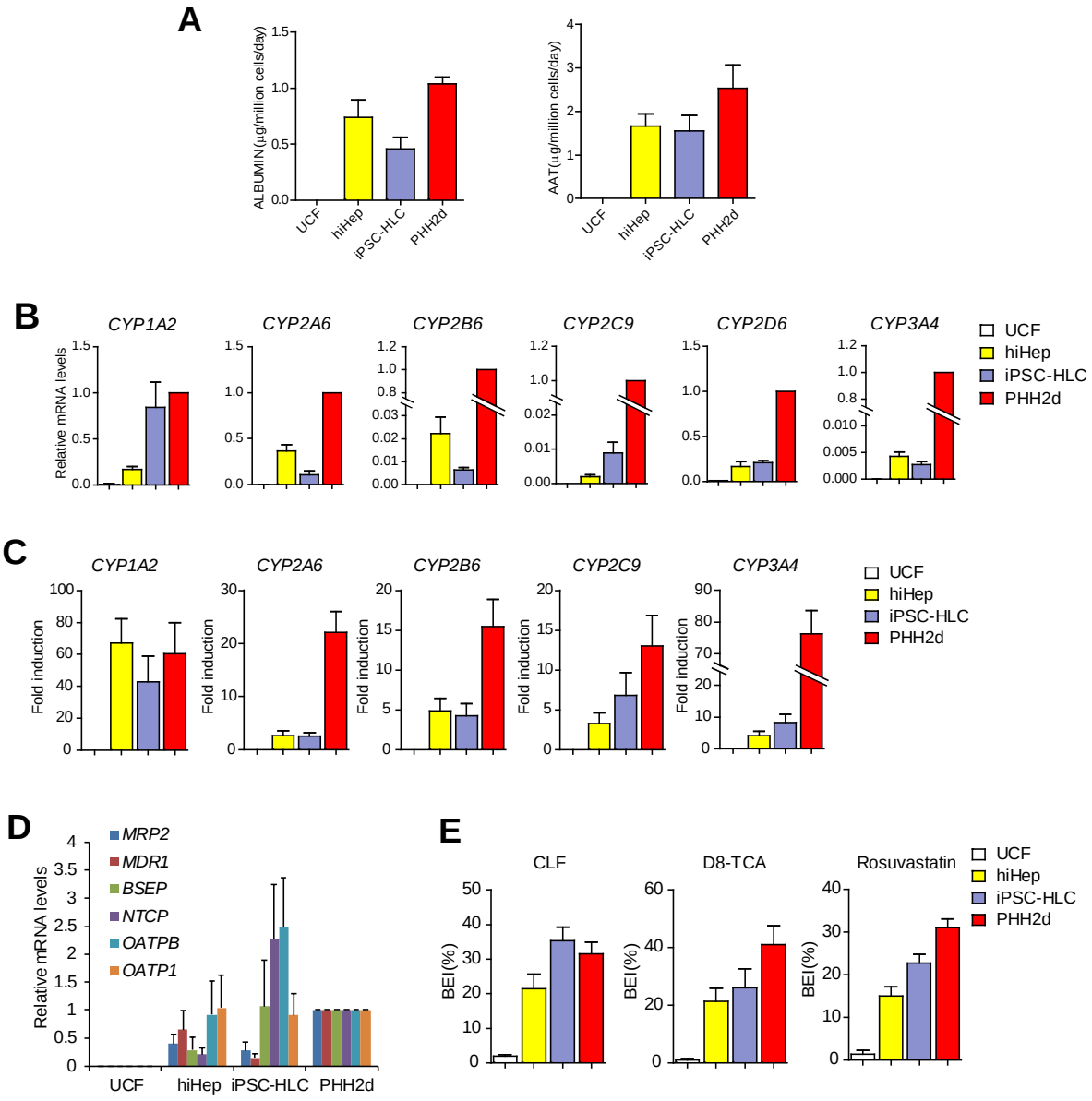

A
ALBUMIN (µg/million cells/day)
AAT (µg/million cells/day)
UCF
hiHep
iPSC-HLC
PHH2d
B
CYP1A2
CYP2A6
CYP2B6
CYP2C9
CYP2D6
CYP3A4
Relative mRNA levels
C
Fold induction
D
MRP2
MDR1
BSEP
NTCP
OATPB
OATP1
Relative mRNA levels
E
CLF
D8-TCA
Rosuvastatin
BEI (%)

Figure S3. *In vitro* characterization of hiHep and iPSC-HLC. Related to Figure 3.

(A) ALBUMIN and α-1-antitrypsin secretion levels were determined by ELISA. UCF included one UCF1 and one UCF2; hiHep included four independent replicates (hiHep1, hiHep2, hiHep3 and hiHep4); iPSC-HLC included four replicates (iPSC-HLC1, iPSC-HLC2, iPSC-HLC3 and iPSC-HLC4); PHHs included two independent replicates which were cultured for two days.

(B) The mRNA levels of Cytochrome P450 (CYP) genes of hiHep and iPSC-HLC were determined by q-PCR. The combination of replicates is the same as Figure S3A.

(C) CYP1A2 can be significantly induced in HLCs, while other CYP genes can be induced slightly. Fold changes were determined by q-PCR. The combination of replicates is the same as Figure S3A.

(D) Transporter expression levels in hiHep and iPSC-HLC were determined by q-PCR. The combination of replicates is the same as Figure S3A.

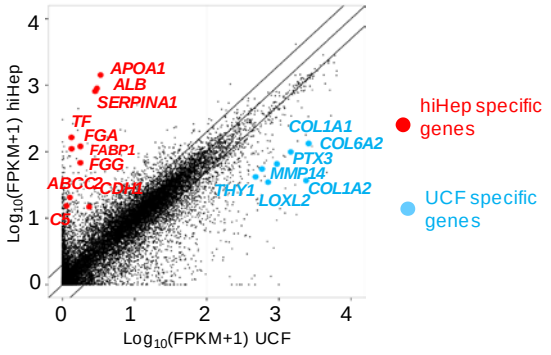
(E) Biliary excretion capabilities of HLCs were measured by clearance of cholyl-lysyl-fluorescein (CLF), D8-taurocholic acid (D8-TCA) and Rosuvastatin. The combination of replicates is the same as Figure S3A.

Figure S4:

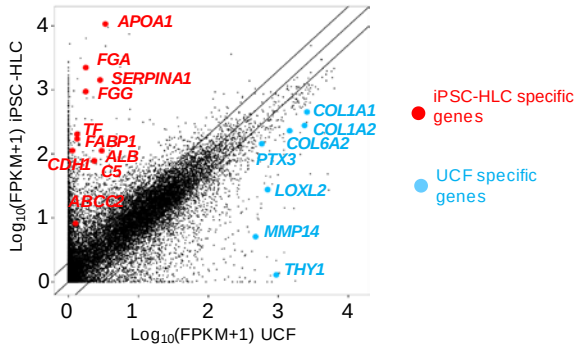
A

Clusters	Enriched Pathways	Representative genes
Cluster V	Regulation of actin cytoskeleton	*FGF7*, *LIMK1*, *PIK3CD*, *TWIST2*, *ITGA11*, *ITGB5*, *MYL9*, *VIM*
	Focal adhesion	*ITGA11*, *ITGB1*, *FLNA*, *VEGFC*, *LAMA4*, *CCND1*, *COL8A1*, *SNAI2*
	Cytokine-cytokine receptor interaction	*CSF3*, *IL6*, *CCL2*, *CXCL8*, *HGF*, *CCL7*, *IL11*, *LIF*, *CXCL1*, *CXCL3*
	Cell adhesion molecules (CAMs)	*NCAM2*, *CD274*, *CNTN1*, *CLDN11*, *ITGA4*, *CDH2*, *JAM2*, *CDH4*
Cluster VI	ECM-receptor interaction	*COL4A2*, *COL4A1*, *ITGB8*, *COL6A3*, *ITGA1*, *TIMP2*, *LOXL2*, *PTX3*
	Focal adhesion	*FLT1*, *ITGA1*, *ITGA3*, *COL5A2*, *LAMA2*, *DOCK1*, *FYN*, *SERPINE2*
	Protein digestion and absorption	*SLC1A5*, *COL7A1*, *COL6A3*, *COL15A1*, *MME*, *COL11A1*, *COL5A2*
	PI3K-Akt signaling pathway	*FGF5*, *ITGB8*, *PDGFRA*, *PDGFRB*, *GNB4*, *COL11A1*, *AKT3*
Cluster VII	ECM-receptor interaction	*LAMA1*, *COL6A2*, *COL1A2*, *COL6A1*, *COL1A1*, *MMP2*, *MMP14*
	Focal adhesion	*SRC*, *LAMB2*, *EGF*, *FN1*, *FLNC*, *COL5A1*, *COL4A6*, *KDR*
	Hippo signaling pathway	*WNT5A*, *BMP4*, *FZD8*, *TEAD2*, *TEAD3*, *WTIP*, *TGFB1*, *BMP6*

B



C



D

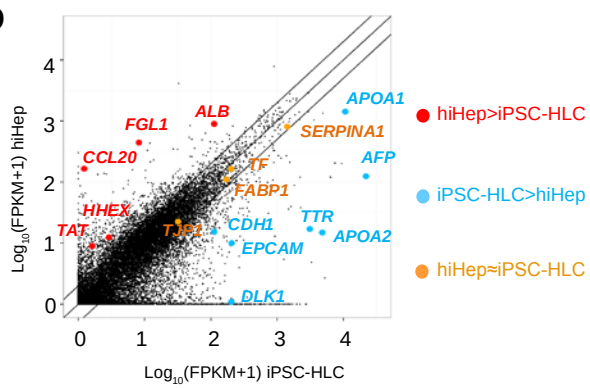


Figure S4. Gene expression analysis of hiHeps and iPSC-HLCs. Related to Figure 2 and Figure 4.

(A) Enriched pathways and representative genes highly expressed in UCFs were summarized.

(B-C) Pair-wise gene expression comparisons showed that most hepatic markers were increased, whereas fibroblast-specific genes were silenced in hiHeps (B) and iPSC-HLCs (C). The combination of replicates is the same as Figure 4A.

(D) A number of hepatic genes expressed differently between hiHeps and iPSC-HLCs. Black lines indicate 2-fold changes between the sample pairs. The combination of replicates is the same as Figure 4A.

Figure S5:

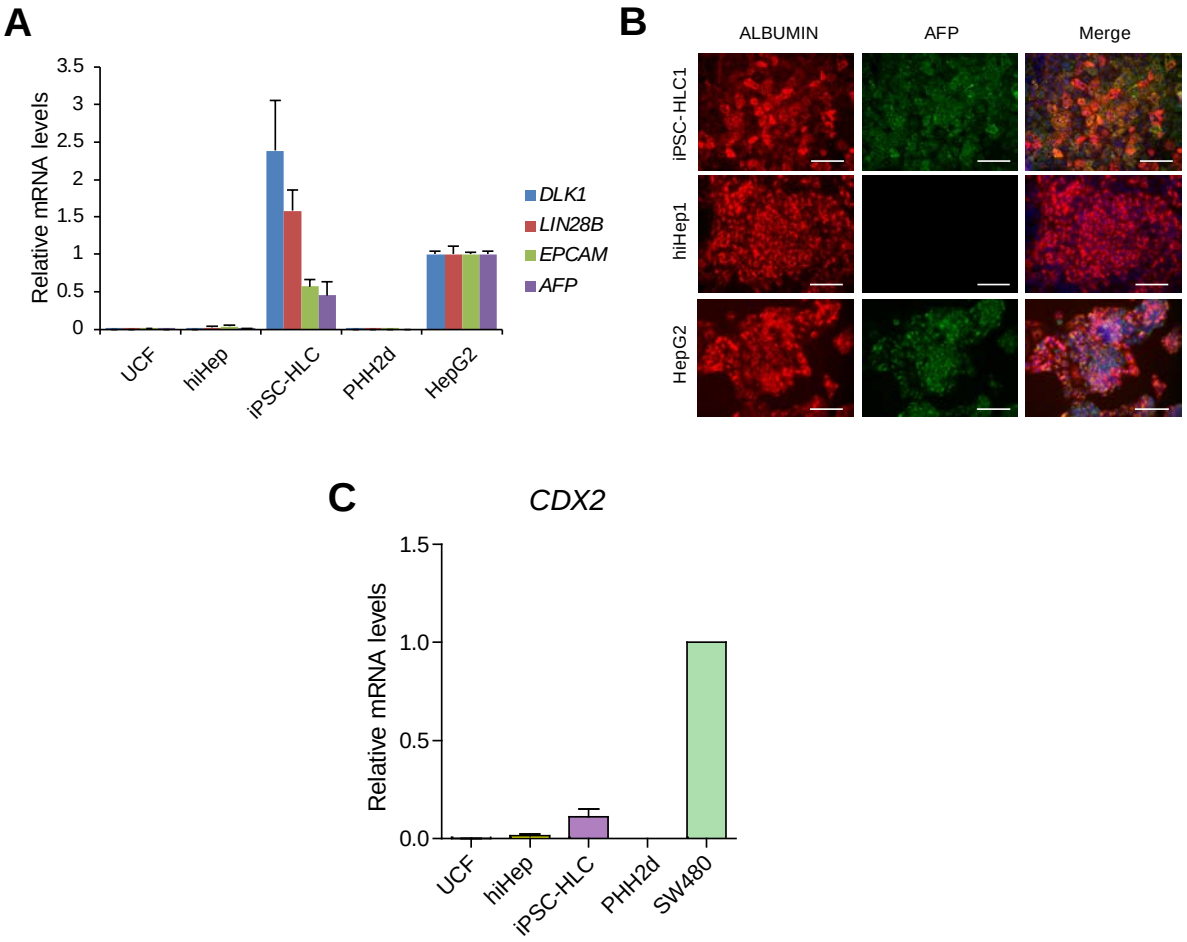

A
Relative mRNA levels
3.5
3
2.5
2
1.5
1
0.5
0
UCF
hiHep
iPSC-HLC
PHH2d
HepG2
DLK1
LIN28B
EPCAM
AFP
B
ALBUMIN
AFP
Merge
iPSC-HLC1
hiHep1
HepG2
C
CDX2
Relative mRNA levels
1.5
1.0
0.5
0.0
UCF
hiHep
iPSC-HLC
PHH2d
SW480

Figure S5. iPSC-HLC expressed immature markers. Related to Figure 4.

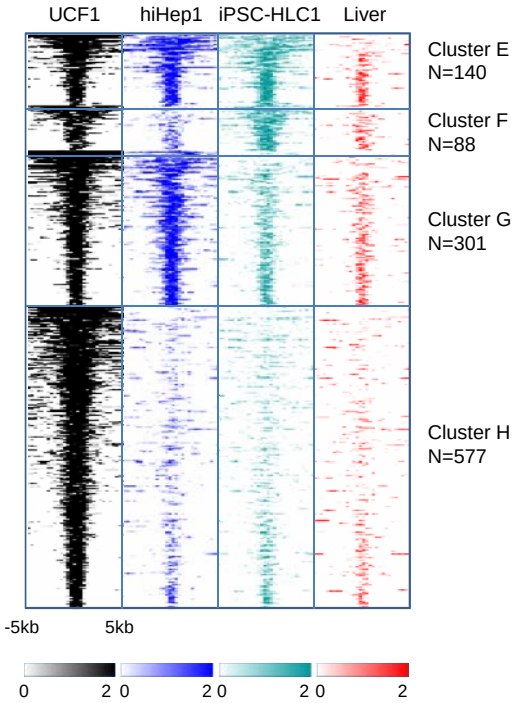
(A) Marker genes of immature hepatocytes were expressed in iPSC-HLC, as determined by q-PCR. HepG2 was used as positive control. UCFs included one UCF1 and one UCF2; hiHeps included four independent replicates (hiHep1, hiHep2, hiHep3 and hiHep4); iPSC-HLCs included four independent replicates (iPSC-HLC1, iPSC-HLC2, iPSC-HLC3 and iPSC-HLC4); HepG2 included two replicates.

(B) ALBUMIN and α-fetoprotein (AFP) were co-expressed in iPSC-HLCs, while hiHep expressed ALBUMIN uniquely. HepG2 was used as positive control. Scale bar, 100 μm.

(C) iPSC-HLC showed low CDX2 gene expression level. SW480 was used as positive control. The combination of replicates is the same as Figure S5A.

Figure S6

A



B

Clusters	Representative genes
Cluster E	*CDH11*, *COL5A2*, *IGF2BP1*, *PRC1*, *SLC7A5*
Cluster F	*ALPK2*, *E2F4*, *RARB*, *VCAN*
Cluster G	*CXCL1*, *FOXL1*, *IL8*, *MED24*
Cluster H	*ABCC1*, *COL6A1*, *COL7A1*, *E2F1*, *FGF2*, *IL6*, *MMP2*, *SLC2A6*, *TWIST2*

C

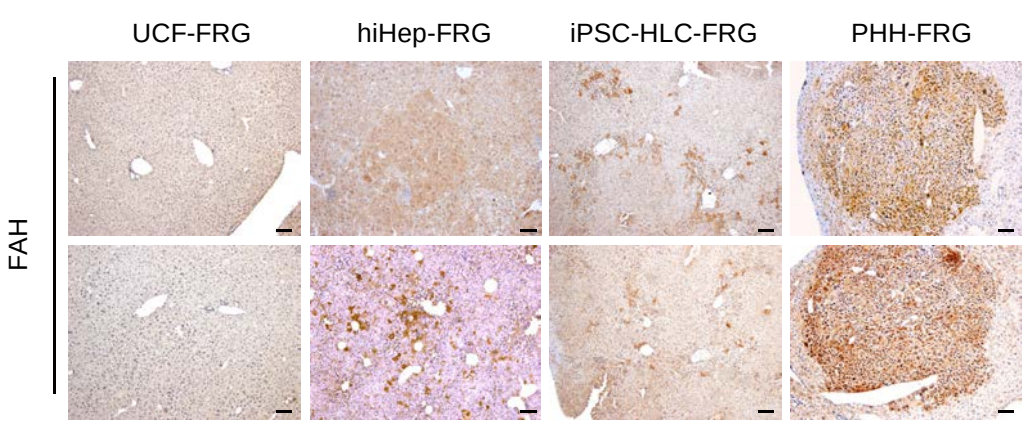


Figure S6. HLCs erased their fibrotic H3K27ac modifications and repopulated in the liver of FRG mice. Related to Figure 5 and Figure 6.

(A) Distribution of the H3K27ac signals in the four samples. Genes which have H3K27ac in UCF1 but not in liver are shown. Each row represents a promoter region of ± 5 kb around the TSS and ranked by the H3K27ac read counts in promoter regions. Peaks were grouped based on presence/absence of peaks comparing the UCF1 and HLCs samples.

(B) Gene Ontology analysis for the genes associated with H3K27ac peaks in different cluster groups.

(C) Representative photos of low magnification showed the engraftment efficiency of HLCs in the livers of FRG mice. Scale bar, 100 μm.

Table S1:

Sample	Donor1			Donor2		
	UCF1	**iPSC-HLC1**	**hiHep1**	**UCF2**	**iPSC-HLC3**	**hiHep3**
SEX	Male	Male	Male	Female	Female	Female
AME	XY	XY	XY	XX	XX	XX
CSF1PO	10,12	10,12	10,12	10,12	10,12	10,12
D5S818	10	10	10	11,12	11,12	11,12
D7S820	10,11	10,11	10,11	10,12	10,12	10,12
D13S317	8,10	8,10	8,10	8	8	8
D16S539	9,11	9,11	9,11	9	9	9
TH01	9,10	9,10	9,10	9	9	9
TPOX	8,9	8,9	8,9	8,10	8,10	8,10
vWA	16,17	16,17	16,17	13,14	13,14	13,14

Table S1. Short tandem repeat (STR) analysis of representative different cell lines. Related to Figure 1.

Table S2:

Dataset	Type	Cell type description	Raw data reads	Uniquely aligned reads(hg19)
UCF1	mRNA-seq	UCF	15,587,063	14,986,937
UCF2	mRNA-seq	UCF	12,886,105	12,351,277
hiHep1	mRNA-seq	hiHep	11,782,612	11,300,955
hiHep2	mRNA-seq	hiHep	11,793,071	11,315,088
hiHep3	mRNA-seq	hiHep	15,387,774	14,678,448
hiHep4	mRNA-seq	hiHep	13,777,553	13,136,450
iPSC-HLC1	mRNA-seq	iPSC-HLC	14,001,840	13,556,128
iPSC-HLC2	mRNA-seq	iPSC-HLC	14,152,516	13,703,578
iPSC-HLC3	mRNA-seq	iPSC-HLC	16,957,633	14,969,686
iPSC-HLC4	mRNA-seq	iPSC-HLC	16,159,800	15,616,574
PHH0d	mRNA-seq	PHH	13,627,191	13,060,672
PHH1d	mRNA-seq	PHH	16,068,468	15,277,617
PHH2d	mRNA-seq	PHH	15,455,998	14,720,381
PHH4d	mRNA-seq	PHH	13,989,274	13,391,254
H3k27ac_UCF1	Histone	UCF	23,634,469	19,500,077
H3k27ac_hiHep1	Histone	hiHep	24,147,889	21,664,019
H3K27ac_iPSC-HLC1	Histone	iPSC-HLC	20,296,814	16,848,508
Input_UCF1	Input	UCF	38,312,544	32,042,769
Input_hiHep1	Input	hiHep	27,152,133	22,888,166
Input_iPSC-HLC1	Input	iPSC-HLC	27,853,342	23,274,017

Table S2. Summary of data sets generated in this study. Related to Figure 2, Figure 4 and Figure 5.

Supplemental Experimental Procedures

Cell culture and generation of hepatocyte-like cells.

UCFs were cultured in human fibroblast medium (HFM), which is DMEM/F12 medium containing 10% FBS, 1% nonessential amino acids, 0.1 mM 2-mercaptoethanol, and 4 ng/ml bFGF. hiHep was generated from immortalized UCFs with transduction of *FOXA3*, *HNF1A* and *HNF4A*, and cultured in hepatocyte maintaining medium (HMM) as previously reported(Huang et al., 2014). iPSC-HLC was generated from iPSCs through a three-step method as previously reported(Szkolnicka et al., 2014). Briefly, iPSCs were maintained in mTeSR (Stemcell Technologies) on Matrigel (BD Biosciences). Before hepatic differentiation, iPSCs were passaged on Matrigel. At day 0-3, iPSCs were treated with RPMI/B27 plus Activin A (100 ng/mL) and Wnt3a (50 ng/mL). Then, in following 6 days, cells were cultured in SR/DMSO (DMEM/F12 containing 20% KOSR, 1 mM glutaMAX, 1% nonessential amino acids, 0.1 mM 2-mercaptoethanol, and 1% dimethyl sulfoxide [DMSO]). The final maturation step was culturing the cells in HepatoZYME medium (Life Technologies) supplemented with 0.5 × GlutaMAX, 10 μM hydrocortisone, 10 ng/ml HGF and 20 ng/ml OSM.

Primary human hepatocytes (PHH) from 3 individuals were purchased from Celsis In Vitro Technologies (Lot number: TLQ, FLO, YJM). Detailed information about the PHHs was provided in the product instructions. PHHs were pooled together at the same number and plated in the density of 1.25×10^5 /cm^2.

RNA extraction Polymerase Chain Reaction, PCR

For most experiments, total RNA was isolated from cells by Trizol (Invitrogen). For RNA extraction from formalin-fixed-paraffin-embedded (FFPE) tissues, RNA was extracted using RNeasy FFPE Kit (Qiagen).

A total of 1 μg RNA was reverse transcribed into cDNA with M-MLV Reverse Transcriptase (Promega) according to the manufacturer's instructions. Quantitative PCR (q-PCR) was performed with SYBR Premix Ex Taq (TaKaRa) on an ABI StepOne Plus real-time PCR system (Applied Biosystems). Primer sequences will be provided upon request. All q-PCR data were performed with at least 3 repeats.

Alkaline phosphatase staining and bisulfite genomic sequencing

The alkaline phosphatase activities of iPSCs were measured by Alkaline Phosphatase Kit (Sigma). For the methylation analysis of *OCT4* promoter regions, the genomic DNA of UCFs and iPSCs were collected and using CpGenome Turbo Bisulfite Modification Kit (Millipore) to measure their methylation levels.

Teratoma formation by human iPSCs

Feeder-free human iPSCs were collected by ReleSR treatment and injected subcutaneously into NOD-SCID mice. Teratoma were observed 1-2 months after injection and were collected typically in 2–3 months and processed for paraffin embedding and hematoxylin and eosin staining following standard procedures.

Flow cytometry analysis

For characterization of UCF by flow cytometry, UCFs were harvested and washed once in PBS solution (Gibco) with 1% BSA, and incubated with PE anti-human CD73 and APC anti-human CD 90 in dark for 30 min at 4 ℃. After incubation, cells were washed with PBS with 1% BSA twice and analyzed by the Calibur Flow Cytometer (Becton Dickinson). For analysis of HLCs, cells were harvested and washed once in PBS solution (Gibco) with 1% BSA, and then fixed by 4% PFA for 15 min and permeabilized by 0.3% Triton

X-100 for 15 min. After that, cells were incubated with ALBUMIN and α-1-antitrypsin (AAT) antibodies for 30 min at room temperature, and then incubated with the secondary antibodies for 15 min in dark at room temperature. After incubation, cells were washed twice and analyzed by the Calibur Flow Cytometer (Becton Dickinson). Antibodies for flow cytometry are as follows: PE anti-human CD73 (Biolegend 344003, 0.5μg/10^6cells), APC anti-human CD90 (Biolegend 328113, 0.5μg/10^6cells), mouse anti-Albumin (R&D mab1455, 2.5μg/10^6 cells), rabbit anti-AAT (NeoMarkers RB-367-A1, 1:200), goat anti-mouse IgG-PE (Santa Cruz sc-3738, 1:500), Alexa Fluor® 647-conjugated donkey anti-rabbit (Jackson Lab 711-605-152, 1:500). Data were analyzed with FlowJo software (Tree Star).

Immunofluorescent staining

For immunofluorescent staining, the cells were fixed with 4% paraformaldehyde for 15 min at room temperature, and then incubated with PBS containing 0.3% Triton X- 100 (Sigma) for 15 min. Cells were then washed three times with PBS. After being blocked by 3% BSA in PBS for 30 min at room temperature, cells were incubated with primary antibodies at 4 ℃ overnight, washed three times with PBS, and then incubated with appropriate fluorescence-conjugated secondary antibody for 30 min at room temperature in dark. Nuclei were stained with DAPI for 5 min.

Antibodies used for immunofluorescent staining are as follows: goat anti-human ALBUMIN (Bethyl Laboratories a80-229a, 1:100), mouse anti-human AFP (eBioscience 14-6583-80, 1:100), mouse anti-human SSEA4 (DSHB MC-813-70, 1:100), goat anti-human NANOG (R&D AF1997, 1:100), Cy3-conjugated donkey anti-goat IgG (Jackson Lab 705-165-003, 1:1000), Alexa Fluor® 647-conjugated donkey anti-mouse (Jackson Lab 715-605-150, 1:1000).

Human ALBUMIN and α-1-Antitrypsin ELISA

To determine the secretion of human ALBUMIN and α-1-Antitrypsin, supernatants of cell culture were collected after 48 hours culture from UCF, hiHep, iPSC-HLC. PHH were seeded on 12-well plates for 12 hours, and then maintained in HMM for 48 hours until collection of supernatants. Human ALBUMIN and α-1-Antitrypsin were measured by the human ALBUMIN ELISA Quantitation Set (Bethyl Laboratory) and the human α-1-Antitrypsin ELISA kit (Bethyl Laboratory) according to the manufacturer's instructions.

Assays for PAS, Oil Red O staining and UGT metabolism

For PAS staining, cells were stained by Periodic-Acid-Schiff (PAS, Sigma). Glycogen storage of UCF, hiHep, iPSC-HLC and PHH were determined by Glycogen Assay Kit (Abnova). For Oil Red O staining, confluent cells were stained by Oil Red O (Sigma-Aldrich), and quantitatively measured by counting oil numbers average to each cell. UGT metabolic abilities were measured by adding substrates to the supernatants and then measured by UGT Glo Assay Kit (Promega)

CYP induction, CYP metabolism and biliary excretion assay

For the measurement of CYP enzyme induction, UCF, hiHep, iPSC-HLC and PHH cells were cultured in HMM for 48 hours and then change to HMM supplemented with 3-methylcholanthrene (25μM), rifampicin (25μM) and sodium phenobarbital (2mM) for additional 48 hours.

For measurement of CYP metabolic activities, UCF, hiHep and iPSC-HLC were cultured for 6 days. On the experimental day, culture medium was removed, and hiHep cells were incubated with 5 μM testosterone (Sigma), the CYP3A substrate in 2 ml incubation medium. Cells were incubated on an orbital shaker for 0, 30,

60, 120, 240 min at 80-120 rpm for a 6-well plate. To stop the reaction, 300 μl cold methanol was added into 100 μl medium containing the substrate. The supernatants were collected for measurement of indicated productions by LC-MS/MS (LCMS-8030; Shimadzu, Kyoto, Japan). Plated cryopreserved hepatocytes were used as a positive control and UCF cells were used as a negative control.

D8-TCA and Rosuvastatin were analyzed by LC/MS/MS (LCMS-8030; Shimadzu, Kyoto, Japan). The amount of CLF was quantified by measuring fluorescence at 492 nm and 536 nm with a Synergy 4 microplate reader (Biotek, Winooski, USA). Biliary Excretion Index (BEI) was calculated as: BEI=(AHBSS-AHBSS(Ca^{2+} free))/AHBSS×100%.

Transplantation of UCF, hiHep, iPSC-HLC and PHH into FRG mice

Cells were intrasplenically transplanted into FRG mice after the withdrawn of NTBC water. Body weight was monitored twice a week post transplantation. Survived recipient mice were sacrificed to collect blood and liver samples 8 weeks after transplantation.

Histology and immunohistochemistry

Tissues were fixed overnight with 4% neutral formalin. Tissue sections were stained with haematoxylin and eosin for pathological evaluation. For immunohistochemical staining, paraffin sections (3-4 μm thick) were used. Slides were treated in 3% H_2O_2 for 15 min, blocked in 5% normal goat or horse serum in 1% BSA-PBS for 20 min, and stained with the indicated antibodies in 1% BSA-PBS overnight. Secondary antibodies were used according to Vectastain ABC kits (Vector Laboratories), followed by DAB staining (DAKO). FAH antibody (Cell Lab Tech, Inc. CLT602-910, Sunnyvale, CA) was used for immunohistochemical staining.

ChIP-seq data analysis

Raw reads were aligned by Bowtie (V1.1.1) (Langmead et al., 2009) against hg19 human reference genome with up to one mismatch. Then, non-uniquely mapped reads and PCR duplicates were removed. To generate the wig files, we extended the mapped reads to 200 bp. For further analysis, we normalized the read counts by computing RPKM (the numbers of reads per kilobase of bin per million of reads). Peaks of ChIP-seq samples were called by MACS(V1.3.7.1, parameter: -nomodel -nolambda) (Zhang et al., 2008) and weak peaks with p-value >=10e-40 were removed. The ChIP-seq data of Liver was downloaded from Roadmap Epigenomics and the GEO number of this data set is GSM1112808.

RNA-seq data processing

RNA-Seq datasets were mapped using TopHat2 (v2.0.13) (Kim et al., 2013) to the hg19 human reference genome. The mapped reads were further analyzed by Cufflinks(Roberts et al., 2011) and annotated using RefSeq database (Pruitt et al., 2012). The expression levels for each transcript were quantified by FPKM (Fragments Per Kilobase of transcript per Million mapped reads). For genes with multiple isoforms, the FPKM values were summed across all isoforms as the expression values for the genes. Differential expression analysis was carried out using the R (v.3.1.1) package DEseq (v.1.18.0) (Anders and Huber, 2010). Genes with Q value less than 0.01 and normalized more than 2 fold change were considered as differentially expressed genes. Heatmaps were constructed using R(v,3.1.1). Gene ontology analysis was performed using DAVID(v.6.7). Total RNA-seq data of RBE, Intestine, Endoderm and Liver progenitors were downloaded from (Forster et al., 2014; Loh et al., 2014; Sampaziotis et al., 2015; Yu et al., 2013) and mapped as described. The k-means clustering of gene expression levels at various samples were conducted using BICskmeans(Liu

et al., 2013).

Original data are available in the NCBI Gene Expression Omnibus (accession number GSE103078).

Supplemental References

Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. Genome biology *11*, R106.

Forster, R., Chiba, K., Schaeffer, L., Regalado, S.G., Lai, C.S., Gao, Q., Kiani, S., Farin, H.F., Clevers, H., Cost, G.J., *et al.* (2014). Human intestinal tissue with adult stem cell properties derived from pluripotent stem cells. Stem cell reports *2*, 838-852.

Huang, P., Zhang, L., Gao, Y., He, Z., Yao, D., Wu, Z., Cen, J., Chen, X., Liu, C., Hu, Y., *et al.* (2014). Direct reprogramming of human fibroblasts to functional and expandable hepatocytes. Cell Stem Cell *14*, 370-384.

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S.L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome biology *14*, R36.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome biology *10*, R25.

Liu, Y., Qiao, N., Zhu, S., Su, M., Sun, N., Boyd-Kirkup, J., and Han, J.D. (2013). A novel Bayesian network inference algorithm for integrative analysis of heterogeneous deep sequencing data. Cell Res *23*, 440-443.

Loh, K.M., Ang, L.T., Zhang, J., Kumar, V., Ang, J., Auyeong, J.Q., Lee, K.L., Choo, S.H., Lim, C.Y., Nichane, 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.

Pruitt, K.D., Tatusova, T., Brown, G.R., and Maglott, D.R. (2012). NCBI Reference Sequences (RefSeq): current status, new features and genome annotation policy. Nucleic Acids Res *40*, D130-135.

Roberts, A., Pimentel, H., Trapnell, C., and Pachter, L. (2011). Identification of novel transcripts in annotated genomes using RNA-Seq. Bioinformatics *27*, 2325-2329.

Sampaziotis, F., Cardoso de Brito, M., Madrigal, P., Bertero, A., Saeb-Parsy, K., Soares, F.A., Schrumpf, E., Melum, E., Karlsen, T.H., Bradley, J.A., *et al.* (2015). Cholangiocytes derived from human induced pluripotent stem cells for disease modeling and drug validation. Nature biotechnology *33*, 845-852.

Szkolnicka, D., Farnworth, S.L., Lucendo-Villarin, B., and Hay, D.C. (2014). Deriving functional hepatocytes from pluripotent stem cells. Current protocols in stem cell biology *30*, 1G 5 1-1G 5 12.

Yu, B., He, Z.Y., You, P., Han, Q.W., Xiang, D., Chen, F., Wang, M.J., Liu, C.C., Lin, X.W., Borjigin, U., *et al.* (2013). Reprogramming fibroblasts into bipotential hepatic stem cells by defined factors. Cell Stem Cell *13*, 328-340.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., *et al.* (2008). Model-based analysis of ChIP-Seq (MACS). Genome biology *9*, R137.