Cellular and Molecular Life Sciences (2025) 82:139
https://doi.org/10.1007/s00018-025-05677-x

ORIGINAL ARTICLE

Single-cell transcriptomics reveals liver developmental trajectory during lineage reprogramming of human induced hepatocyte-like cells

Nan Jiang[1,2] · Guangya Li[3] · Sen Luo[1,2] · Xi Kong[1,2] · Shigang Yin[1,2] · Jianhua Peng[1,2] · Yong Jiang[1,2] · Wei Tao[6] · Cheng Li[5] · Huangfan Xie[1,2] · Hongkui Deng[3,4] · Bingqing Xie[1,2]

Received: 30 October 2024 / Revised: 11 March 2025 / Accepted: 24 March 2025

Abstract

Hepatocytes are crucial for drug screening, disease modeling, and clinical transplantation, yet generating functional hepatocytes in vitro is challenging due to the difficulty of establishing their authentic gene regulatory networks (GRNs). We have previously developed a two-step lineage reprogramming strategy to generate functionally competent human induced hepatocytes (hiHeps), providing an effective model for studying the establishment of hepatocyte-specific GRNs. In this study, we utilized high-throughput single-cell RNA sequencing (scRNA-seq) to explore the cell-fate transition and the establishment of hepatocyte-specific GRNs involved in the two-step reprogramming process. Our findings revealed that the late stage of the reprogramming process mimics the natural trajectory of liver development, exhibiting similar transcriptional waves of developmental genes. CD24 and DLK1 were identified as surface markers enriching two distinct hepatic progenitor populations respectively. Lipid metabolism emerged as a key enhancer of hiHeps maturation. Furthermore, transcription factors *HNF4A* and *HHEX* were identified as pivotal gatekeepers directing cell fate decisions between hepatocytes and intestinal cells. Collectively, this study provides valuable insights into the establishment of hepatocyte-specific GRNs during hiHeps induction at single-cell resolution, facilitating more efficient production of functional hepatocytes for therapeutic applications.

Keywords Single-cell RNA sequencing · Lineage reprogramming · Human induced hepatocytes · Lipid metabolism · Gene regulatory networks

Abbreviations

hiHeps	Human induced hepatocytes
hHPLCs	Human hepatic progenitor-like cells
PHHs	Primary human hepatocytes
HEM	Hepatic expansion medium
FACS	Fluorescence-activated cell sorting
scRNA-seq	Single-cell RNA sequencing
GEMS	Gel beads in emulsion
UMI	Unique molecular identifiers
TFs	Transcription factors

Nan Jiang, Guangya Li and Sen Luo have contributed equally to this work.

✉ Huangfan Xie
xie1991@swmu.edu.cn

✉ Hongkui Deng
hongkui_deng@pku.edu.cn

✉ Bingqing Xie
bingqingxie@swmu.edu.cn

[1] Laboratory of Neurological Diseases and Brain Function, the Affiliated Hospital, Southwest Medical University, Luzhou, China

[2] Institute of Epigenetics and Brain Science, Southwest Medical University, Luzhou, China

[3] MOE Engineering Research Center of Regenerative Medicine, School of Basic Medical Sciences, State Key Laboratory of Natural and Biomimetic Drugs, Peking University Health Science Center, the MOE Key Laboratory of Cell Proliferation and Differentiation, College of Life Sciences, Peking-Tsinghua Center for Life Sciences, Peking University, Beijing, China

[4] Changping Laboratory, Beijing, China

[5] School of Life Sciences, Center for Bioinformatics, Center for Statistical Science, Peking University, Beijing, China

[6] Key Laboratory of Cell Proliferation and Differentiation, School of Life Sciences, Peking University, Beijing, China

Published online: 06 April 2025

UGT	UDP-glucuronosyltransferase
GO	Gene ontology
GSEA	Gene set enrichment analysis
KEGG	Kyoto Encyclopedia of Genes and Genomes
PCA	Principal component analysis
UMAP	Uniform Manifold Approximation and Projection
SCENIC	Single-Cell Regulatory Network Inference and Clustering
DEGs	Differentially expressed genes
CDL	Chemically defined lipid concentrate
RSS	Regulon Specificity Score
AH	Adult hepatic cells
FH1	Fetal hepatic cells-1
FH2	Fetal hepatic cells-2
HB1	Hepatoblasts1
HB2	Hepatoblasts2
GRN	Gene regulatory network
HEF	Human embryonic fibroblast
BSA	Bovine serum albumin
PS	Penicillin/streptomycin
PBS	Phosphate-buffered saline

Introduction

Hepatocytes hold significant value in drug screening, disease modeling, and clinical cell transplantation [1, 2]. Despite numerous strategies developed to generate hepatocytes in vitro, several challenges persist primarily due to the incomplete establishment of hepatocyte-specific gene regulatory network. Among these strategies, lineage reprogramming, which employs specific combinations of transcription factors (TFs) or chemical compounds to induce cell fate conversion, stands out as a direct method to generate hepatocytes from differentiated cells such as fibroblasts [3]. However, traditional direct reprogramming approaches encounter several limitations, including the retention of residual memory from the initial cells and the limited functionality of the induced hepatocytes, thereby restricting their practical applications [4, 5].

We have previously developed a novel two-step lineage reprogramming strategy to produce functionally competent human induced hepatocytes (hiHeps) from fibroblasts [6]. In this approach, fibroblasts are initially reprogrammed into expandable human hepatic progenitor-like cells (hHPLCs), which are then induced into highly functional hiHeps. Although hiHeps demonstrate molecular identity and functionality closely resembling primary human hepatocytes (PHHs), the molecular mechanisms underlying this two-step reprogramming remain poorly understood.

The construction of hepatocyte-specific GRNs is a prerequisite for generating functional hepatocytes in vitro [7, 8]. These networks encompass the regulatory interactions between multiple TFs and their downstream target genes, controlling the timing, conditions, and levels of gene expression [9–12]. In recent years, significant attention has been given to the study of hepatocyte-specific GRNs in induced hepatocytes [13–16]. Transcriptional components of the hepatocyte-specific GRNs have been continuously uncovered through bulk RNA-seq analysis; for instance, pioneer TFs *FOXA2* and *GATA4* were identified to form feedforward loops with *HNF4A* as well as fibroblast growth factor signaling cascade to drive hepatic specification [14].

Despite the progress made through bulk RNA-seq analyses in studying hepatocyte GRNs, insights at the single-cell level remain limited. Bulk approaches often aggregate data across cell populations, which can mask rare yet significant events during reprogramming and may draw attention to irrelevant biological processes that do not contribute to this process [17]. Considering that only a small subset of initial cells successfully matures into functional hepatocytes, it is vital to reassess the findings from bulk analyses using single-cell techniques [6, 18–21]. Investigating these intricate biological processes at a single-cell resolution is crucial for capturing transcriptional details of infrequent events, leading to a more nuanced and precise comprehension of hepatocyte differentiation [22–25].

In this study, scRNA-seq analysis revealed that the late stage of the two-step reprogramming process mimics liver development, characterized by distinct transcriptional waves of development-related genes across the reprogramming stages. We identified CD24 and DLK1 as surface markers distinguishing two distinct hepatic progenitor cell populations, crucial for the maturation into functional hepatocytes. Lipid metabolism significantly promotes this maturation. Additionally, transcription factors *HNF4A* and *HHEX* regulate cell fate decisions between hepatocytes and intestinal cells during late reprogramming stages. This study provides a detailed single-cell roadmap of the late stage in the two-step reprogramming process, enhancing the understanding of hepatocyte-specific GRN establishment, which is beneficial for optimizing hepatocyte induction.

Results

High-resolution dissection of two-step reprogramming from fibroblast to induced hepatocytes using scRNA-seq

Our two-step lineage reprogramming consists of a reprogramming step from fibroblasts to hHPLCs (hHPLCs reprogramming step) and a maturation step from hHPLCs to hiHeps (hiHeps maturation step) (Fig. 1A). We observed that critical cellular transformations happened at the late stages

of hHPLCs reprogramming step and the following hiHeps maturation step, the critical time points for establishment of hepatocyte-specific GRNs. Therefore, we selected samples from the late stages of hHPLCs reprogramming at day 38 (R-D38), day 54 (R-D54), and day 63 (R-D63), as well as samples from hiHeps maturation step at day 0 (M-D0), day 2 (M-D2), day 4 (M-D4), and day 7 (M-D7) for dissecting cell fate transition (Fig. 1A). Additionally, we included a starting fibroblast sample (R-D0) and a primary human hepatocyte (PHHs) sample as controls for gene expression analysis. These samples were processed to analyze reprogramming trajectories by measuring single-cell gene expression profiles using the Chromium system (10 × Genomics) (Fig. 1A). Sequencing data were collected from 21,304 individual cells across 9 samples, with a median of 6070 genes and 37,291 UMI per cell (Fig. S1A and S1B). Cells with less than 10% mitochondrial content and more than 2000 detected genes (nFeature) were selected for further analysis (Fig. S1C and S1D).

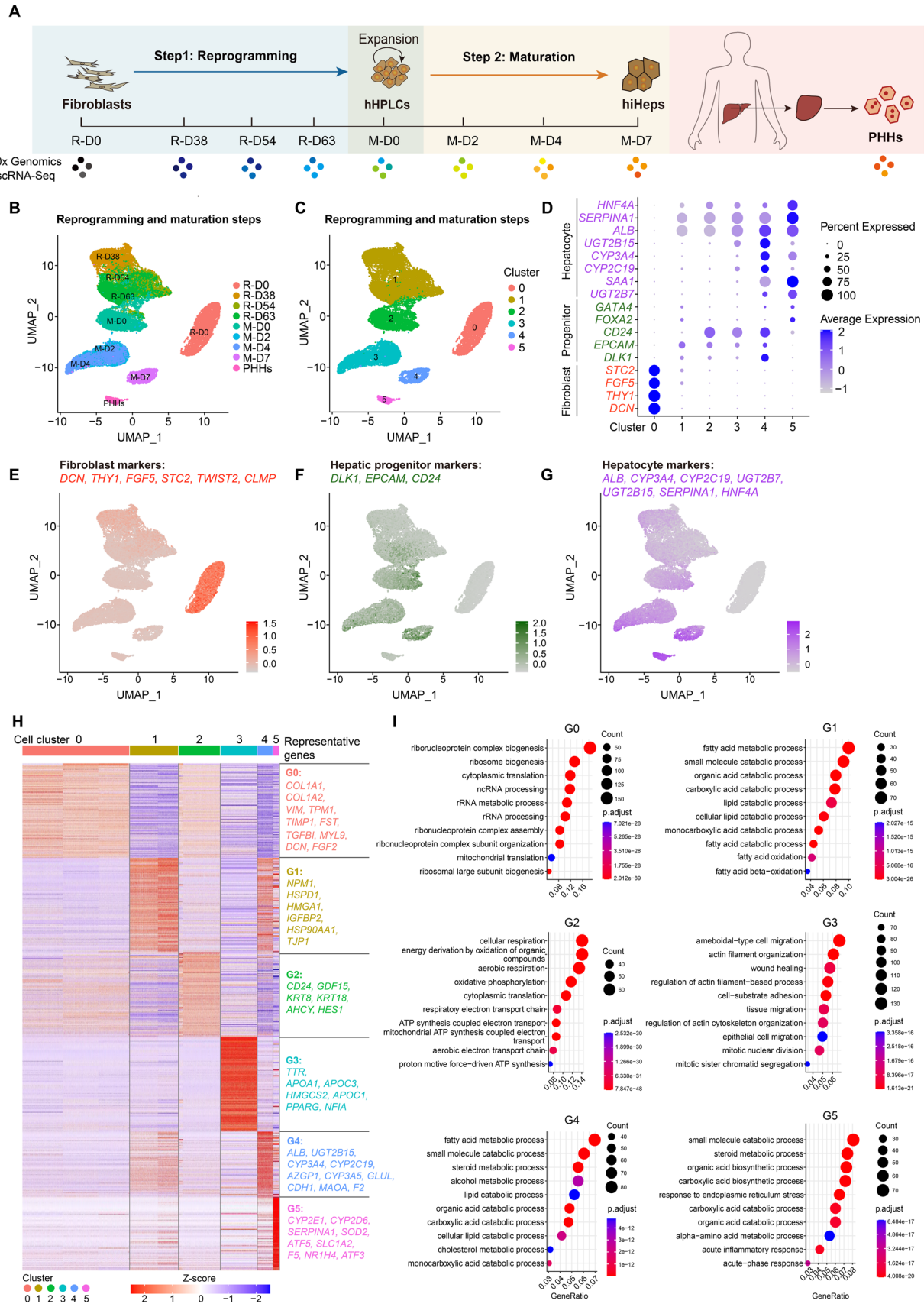
We projected all single cells onto a Uniform Manifold Approximation and Projection (UMAP) plot, identifying 6 transcriptionally distinct clusters (Fig. 1B and C). To classify the main cell types, we annotated each cluster based on marker gene expression (Fig. 1D–G). Sample R-D0 fell into cluster 0 identified as fibroblasts, characterized by the expression of mesenchymal genes such as *DCN*, *THY1*, *FGF5*, and *STC2*. Samples R-D38, R-D54, R-D63 fell into cluster 1, and sample M-D0 into cluster 2. Clusters 1 and 2 were identified as hepatic progenitor-like cells, marked by the expression of *DLK1*, *EPCAM*, and *CD24*. Samples M-D2 and M-D4 in cluster 3 represented immature hepatocytes, showing both progenitor-like markers and low levels of hepatocyte functional genes like *ALB*, *CYP3A4*, *CYP2C19*, *UGT2B7*, and *UGT2B15*. Samples M-D7 fell into cluster 4 corresponding to mature hepatocytes, distinguished by high expression of hepatocyte functional genes. Finally, PHHs were grouped into cluster 5.

To dissect the transcriptomic differences among these samples, we further analyzed the differentially expressed genes (DEGs) (Fig. 1H) and their Gene Ontology (GO) enrichment across different cell clusters (Fig. 1I). We identified 6 distinct groups of DEGs that define specific gene signatures for each cluster (Fig. 1H). Group 0 (G0) was enriched with fibroblast genes such as *COL1A1*, *COL1A2*, and *VIM*, predominantly expressed in cluster 0 (Fig. 1H). Groups 1 and 2 (G1 and G2) contained epithelial and hepatic progenitor markers like *CD24*, *KRT8*, *KRT18*, *TJP1*, and *HES1*, expressed in clusters 1 and 2, and were associated with processes such as lipid metabolism and aerobic respiration (Fig. 1H and I). Group 3 (G3) was enriched with hepatocyte-related genes including *TTR*, *APOA1*, *APOC3*, *HMGCS2*, and *PPARG*, expressed in cluster 3, which involved cell migration and actin filament organization (Fig. 1H and I). Group 4 (G4) included hepatocyte functional genes such as *ALB*, *CYP3A4*, *UGT2B15*, *AZGP1*, *F2*, and *GLUL*, expressed in cluster 4, and was linked to metabolic and catabolic processes like fatty acid metabolic process, small molecule catabolic process and steroid metabolic process (Fig. 1H and 1I). These processes were also similarly represented in Group 5 (G5) genes expressed in PHHs (Fig. 1G). GO enrichment analysis highlighted the significance of metabolic and catabolic pathways, such as fatty acid metabolic processes and small molecule catabolic processes, throughout the reprogramming and maturation stages (Fig. 1I).

Two-step lineage reprogramming recapitulates a trajectory of liver development

Next, we investigated whether the two-step reprogramming process recapitulates a trajectory of liver development in vitro. Wesley et al. recently unveiled a single-cell transcriptome atlas of human liver development, identifying specific liver development stages as hepatoblasts (HB1 and HB2), fetal hepatic cells (FH1 and FH2), and adult hepatic cells (AH), along with key marker genes for each stage [26, 27]. We found that the late stage of our two-step reprogramming process aligns with liver development, as evidenced by corresponding marker gene expression patterns with liver samples across various stages of liver development. Specifically, hHPLCs reprogramming samples expressed hepatoblast (*MAP2K2, NPW, BRI3, BAAT, GSTP1, DLK1, APOM*, and *AMN*) and fetal hepatic markers (*AFP, CYP3A7, F2, AHSG, APOA1, AGT, ALB, ANGPTL3, APOB, GC, VTN*, and *AKR1C1*). This expression profile indicates that hHPLCs exhibit a mixed characteristic of HB and FH identity. hiHeps maturation samples expressed adult hepatic markers (*NNMT, AZGP1, TAT, CES1, ADH1B, HP,* and *SAA1*) (Fig. 2A). Additionally, the expression dynamics of key genes in the late stage of our two-step reprogramming process mirrored those observed in liver development: the reprogramming of hHPLCs from R-D38 to M-D0 resembled the developmental stages from 7 weeks (7W) to 12 weeks (12W) of the human liver, while the maturation of hiHeps from M-D0 to M-D7 paralleled the stages from 13 weeks (13W) to 17 weeks (17W) (Fig. 2B).

To elucidate the trajectory of two-step reprogramming process, we used Monocle 2 to arrange cells in pseudotemporal order (Fig. 2C and D). Samples R-D38, R-D54, R-D63 and M-D0 were positioned in state 1. These cells then branched into two groups, designated as state 2 and 3. State 2 primarily included sample M-D7, while state 3 mainly consisted of samples M-D0 and M-D4 (Fig. 2E and F). According to the pseudotime analysis, state 2 represented the final cell type, illustrating the transition from fibroblasts to hHPLCs and ultimately to mature hepatocytes (Fig. 2F). We then investigated the sequential changes of 3350 DEGs of these cells along pseudotime, which were categorized


A
Step1: Reprogramming
Expansion
Step 2: Maturation
Fibroblasts
hHPLCs
hiHeps
PHHs
R-D0
R-D38
R-D54
R-D63
M-D0
M-D2
M-D4
M-D7
10x Genomics scRNA-Seq
B
Reprogramming and maturation steps
C
Reprogramming and maturation steps
Cluster
D
Hepatocyte
Progenitor
Fibroblast
HNF4A
SERPINA1
ALB
UGT2B15
CYP3A4
CYP2C19
SAA1
UGT2B7
GATA4
FOXA2
CD24
EPCAM
DLK1
STC2
FGF5
THY1
DCN
Percent Expressed
Average Expression
E
Fibroblast markers:
DCN, THY1, FGF5, STC2, TWIST2, CLMP
F
Hepatic progenitor markers:
DLK1, EPCAM, CD24
G
Hepatocyte markers:
ALB, CYP3A4, CYP2C19, UGT2B7, UGT2B15, SERPINA1, HNF4A
UMAP_1
UMAP_2
H
Cell cluster
Representative genes
G0: COL1A1, COL1A2, VIM, TPM1, TIMP1, FST, TGFBI, MYL9, DCN, FGF2
G1: NPM1, HSPD1, HMGA1, IGFBP2, HSP90AA1, TJP1
G2: CD24, GDF15, KRT8, KRT18, AHCY, HES1
G3: TTR, APOA1, APOC3, HMGCS2, APOC1, PPARG, NFIA
G4: ALB, UGT2B15, CYP3A4, CYP2C19, AZGP1, CYP3A5, GLUL, CDH1, MAOA, F2
G5: CYP2E1, CYP2D6, SERPINA1, SOD2, ATF5, SLC1A2, F5, NR1H4, ATF3
Cluster
Z-score
I
G0
G1
G2
G3
G4
G5
Count
p.adjust
GeneRatio

◂**Fig. 1** High-Resolution Dissection of Two-Step Reprogramming from Fibroblast to induced Hepatocytes Using scRNA-Seq. **A** Schematic diagram of the scRNA-seq analysis strategy during the two-step lineage reprogramming process. In the reprogramming step, samples were collected at day 0 (R-D0), day38 (R-D38), day 54 (R-D54), and day 63 (R-D63). In the maturation step, samples were collected at day 0 (M-D0), day 2 (M-D2), day 4 (M-D4), and day 7 (M-D7). Additionally, a primary human hepatocytes (PHHs) sample was collected. All samples were processed using 10 × Genomics single-cell capture and sequencing. **B** UMAP of all 21,304 individual cells during the whole reprogramming process, colored by indicated time points. **C** UMAP of all 21,304 individual cells during the whole reprogramming process, colored by cell clusters. **D** Dot plots showing the expression of marker genes for 6 cell clusters. **E–G** Visualizations of scRNA-seq data, projection the expression score of representative fibroblast markers, hepatic progenitor markers and hepatocyte markers onto a UMAP plot. **H** Heatmap showing 6 groups of DEGs in the 6 cell clusters. Each column represents a single cell. Representative genes from each group of DEGs were listed to the right. **I** GO enrichment analysis of the six groups of DEGs

into 3 clusters (Fig. 2G). We focused on the expression patterns of these 3 clusters as the cells progressed from state 1 to 2 (Fig. 2H and I). Notably, cluster 1 exhibited a stable expression pattern of FH marker genes (*APOA1, AGT, GC*, and *ALB*), cluster 2 showed an up-regulation pattern for AH marker genes *(NNMT, SAA1, TAT*, and *HP*), and cluster 3 demonstrated a down-regulation pattern for HB marker genes (*GSTP1*, *NPW,* and *MAP2K2*), consistent with their expression during liver development (Fig. 2H and I).

We further performed principal component analysis (PCA) on our samples. The cells progressed along the reprogramming timeline in a sequential order of R-D38, R-D54, R-D63, M-D0, M-D2, M-D4 and M-D7 (Fig. S2A). The major differences in gene expression were captured in PC1, which included genes associated with various metabolic processes, highly similar to the related processes in mouse liver development from E17.5 to P60 [28] (Fig. S2B). These findings support the notion that the late stage of two-step lineage reprogramming mirrors the natural development process of liver.

CD24 and DLK1 marks two different hepatic progenitor populations

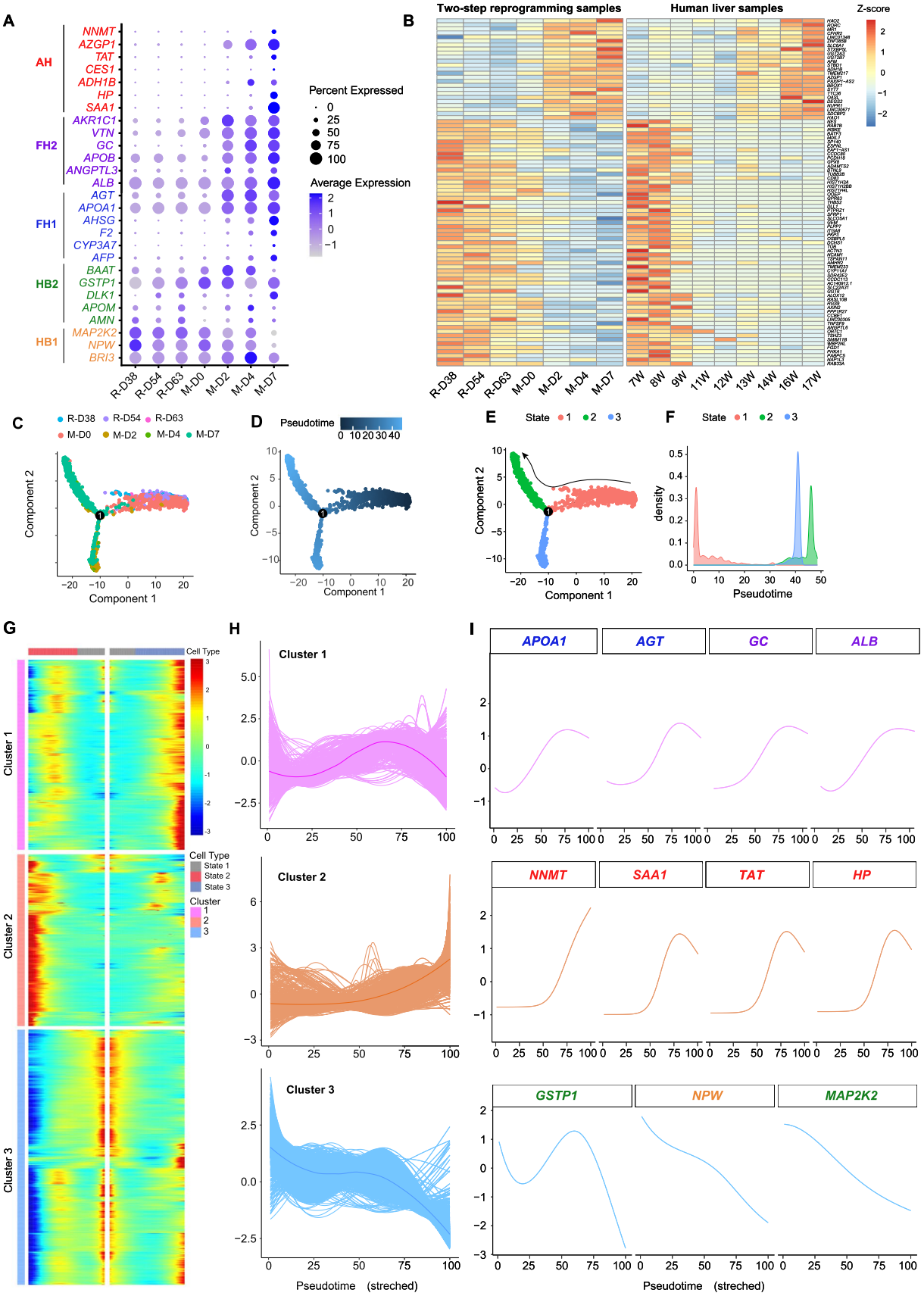
To gain a detailed understanding of the two-step reprogramming process in the late stage, we investigated the hHPLCs reprogramming step and the hiHeps maturation step separately. We first explored the molecular and cellular dynamics of hepatic progenitor reprogramming by visualizing the reprogramming of hHPLCs using UMAP, which revealed 6 distinct clusters (Fig. 3A and B). We focused on the transmembrane proteins *CD24* and *DLK1*, both critical for liver development and the maintenance of progenitor cells [29, 30]. Our analysis revealed that cluster 6, comprising M-D0 cells, exhibited enriched expression of *CD24* as well as other hepatic progenitor markers such as *ITGA6, NQO1, MOK* and *CDH1*. Meanwhile, cluster 2, comprising R-D54 and R-D63 cells, showed enriched expression of *DLK1* and other hepatic progenitor markers such as *ITGB1, CTNNB1, ITGA6* and *ANPEP* [31] (Fig. 3C). Other clusters showed low expression of progenitor markers, suggesting a pre-progenitor stage containing R-D38 cells and subpopulation of R-D54 and R-D63 cells.

To further elucidate the trajectory of progenitor reprogramming, we ordered cells in a pseudotemporal manner using Monocle 2 (Fig. 3D–H). The hHPLCs reprogramming samples were arranged along the pseudotime in the sequence of R-D38, R-D54, R-63, and M-D0 (Fig. 3D–F). These cells were grouped into five states, bifurcating into states 4 and 5, which represented two major progenitor populations at the final progenitor stage (Fig. 3G and H). Cells in the terminal of state 5 expressed key hepatic progenitor genes such as *CD24*, *GSTA1*, *CKB*, and *PATJ*, indicating successful reprogramming into hepatic progenitors (Fig. 3I). In contrast, cells at the alternate terminal exhibited high expression levels of *FABP1*, *SLC35F1*, *SNHG5*, and *KRT19*, indicative of unsuccessful reprogramming with a biased program (Fig. 3I). Notably, *CD24* was enriched in the successful branch, confirming it as a reliable marker for the final hepatic progenitor population (Fig. 3J). Interestingly, *DLK1* was upregulated prior to *CD24* upregulation but was not enriched in the terminal branch, suggesting that *DLK1*-positive cells may represent an earlier stage of progenitor cells (Fig. 3J).

To determine whether CD24 and DLK1 can serve as surface markers for authentic hepatic progenitors during functional hepatocyte maturation, we sorted DLK1-positive and CD24-positive hHPLCs and cultured them in 2C maturation medium for an additional 7 days (Fig. 3K). Both CD24-positive and DLK1-positive cells exhibited significantly enhanced hepatocyte maturation potential compared to unsorted hHPLCs, as evidenced by the upregulation of of key functional hepatocyte markers, including *ALB*, *CYP3A4*, *CYP2C9*, *CYP2C19*, *CYP2C8*, *CPS1*, *OTC*, *ASS*, *ARG1*, *NTCP*, *PXR*, *HNF1A*, *CEBPA*, and *UGT1A1* (Fig. 3L). Notably, while CD24-positive cells exhibited higher expression of *ALB*, most other functional genes were comparably expressed between the two sorted populations (Fig. 3L). These results suggest that *CD24* and *DLK1* define different states of progenitor cells, and both markers may be useful for enriching progenitor cells to enhance the efficiency of hepatocyte maturation.

Lipid metabolism mediates reprogramming of hepatic progenitors into functional hepatocytes

To explore the maturation of hepatic progenitors into functional hepatocytes in 2C maturation medium, we


A
AH
FH2
FH1
HB2
HB1
NNMT
AZGP1
TAT
CES1
ADH1B
HP
SAA1
AKR1C1
VTN
GC
APOB
ANGPTL3
ALB
AGT
APOA1
AHSG
F2
CYP3A7
AFP
BAAT
GSTP1
DLK1
APOM
AMN
MAP2K2
NPW
BRI3
R-D38 R-D54 R-D63 M-D0 M-D2 M-D4 M-D7
Percent Expressed
0 25 50 75 100
Average Expression
2 1 0 −1
B
Two-step reprogramming samples
Human liver samples
Z-score
2 1 0 −1 −2
R-D38 R-D54 R-D63 M-D0 M-D2 M-D4 M-D7 7W 8W 9W 11W 12W 13W 14W 16W 17W
C
R-D38 R-D54 R-D63
M-D0 M-D2 M-D4 M-D7
Component 2
Component 1
D
Pseudotime
0 10 20 30 40
Component 2
Component 1
E
State 1 2 3
Component 2
Component 1
F
State 1 2 3
density
Pseudotime
G
Cell Type
Cluster 1
Cluster 2
Cluster 3
Cell Type
State 1
State 2
State 3
Cluster
1
2
3
H
Cluster 1
Cluster 2
Cluster 3
Pseudotime (streched)
I
APOA1 AGT GC ALB
NNMT SAA1 TAT HP
GSTP1 NPW MAP2K2
Pseudotime (streched)

◂**Fig. 2** Two-Step Lineage Reprogramming Recapitulates a Trajectory of Liver Development. **A** Bubble chart displaying marker gene expression in the late stage of two-step reprogramming samples. Genes are colored by their expression stage, namely hepatoblast stages 1 and 2 (HB1, HB2), foetal hepatocyte stages 1 and 2 (FH1, FH2) and adult hepatocytes (AH), as defined in a previous study of human liver development. **B** Heatmaps displaying marker gene expression in the late stage of two-step reprogramming samples and human liver samples. **C**, **D** Pseudotime trajectory analysis of single cells throughout the late stage of two-step reprogramming, colored by indicated time points. Trajectory reconstruction reveals three branches: pre-branch (before bifurcation) and two terminal branches (after bifurcation). **E**, **F** Pseudotime trajectory analysis of single cells during the late stage of two-step reprogramming, colored by cell state. **G** Gene expression heatmap of the top differentially expressed genes (DEGs), cataloged into 3 clusters in a pseudo-temporal order. Two terminal branches of state 2 and state 3 are shown on the top right and left, respectively. **H** The expression dynamics of the top DEGs categorized into three major clusters in pseudotime trajectory in state 2 cells. Thick lines indicate the average gene expression patterns in each cluster. **I** Gene signatures and expression dynamics of representative liver developmental genes in each gene cluster as shown in **H**

next analyzed maturation samples at different maturation stages. Maturing hepatocytes were grouped into 3 clusters: M-D0 formed cluster 0, M-D2 and M-D4 formed cluster 1, and M-D7 formed cluster 2 (Fig. 4A and B). Differential expression analysis revealed that cluster 0 was enriched with progenitor and proliferative markers such as *CD24*, *CDK1*, *CDK4*, *PCNA*, and *KRT18*. Cluster 1 expressed basic hepatocyte markers including *APOE*, *TTR*, *APOA1*, and *ACSL1*, while cluster 2 was enriched with functional hepatocyte genes like *ALB*, *CYP3A4*, *CYP2C19*, and *UGT2B15* (Fig. 4C and D), reflecting progressive maturation over the 2C culture timeline. GO enrichment analysis further highlighted fatty acid metabolism as a key process in hepatocyte maturation (Fig. 4E).

Furthermore, Monocle pseudotime analysis revealed two distinct cell populations (Fig. 4F–H). M-D0 fell primarily into state 1; M-D2, M-D4 and M-D7 divided into two branches (Fig. 4F–H). One population (state 2) expressed major functional hepatocyte genes such as *UGT2B15*, *SLC7A2*, *CYP3A4*, and *CYP3A5*, indicating successful hepatocyte maturation (Fig. 4I). The other population (state 3) expressed genes such as *AMBP*, *IGFBP2*, *MUC13*, and *AMBRA1*, characteristic of immature hepatocytes and intestinal cells, representing the failed maturation branch (Fig. 4I). Analysis of 2626 DEGs post-bifurcation (Fig. S3A) revealed that cells in the successful branch had elevated levels of genes involved in metabolic processes, particularly fatty acid metabolism, consistent with the GO analysis (Fig. S3B-S3D, and Fig. 4E). These gene expression patterns suggest that lipid metabolism plays a critical role in hepatocyte maturation.

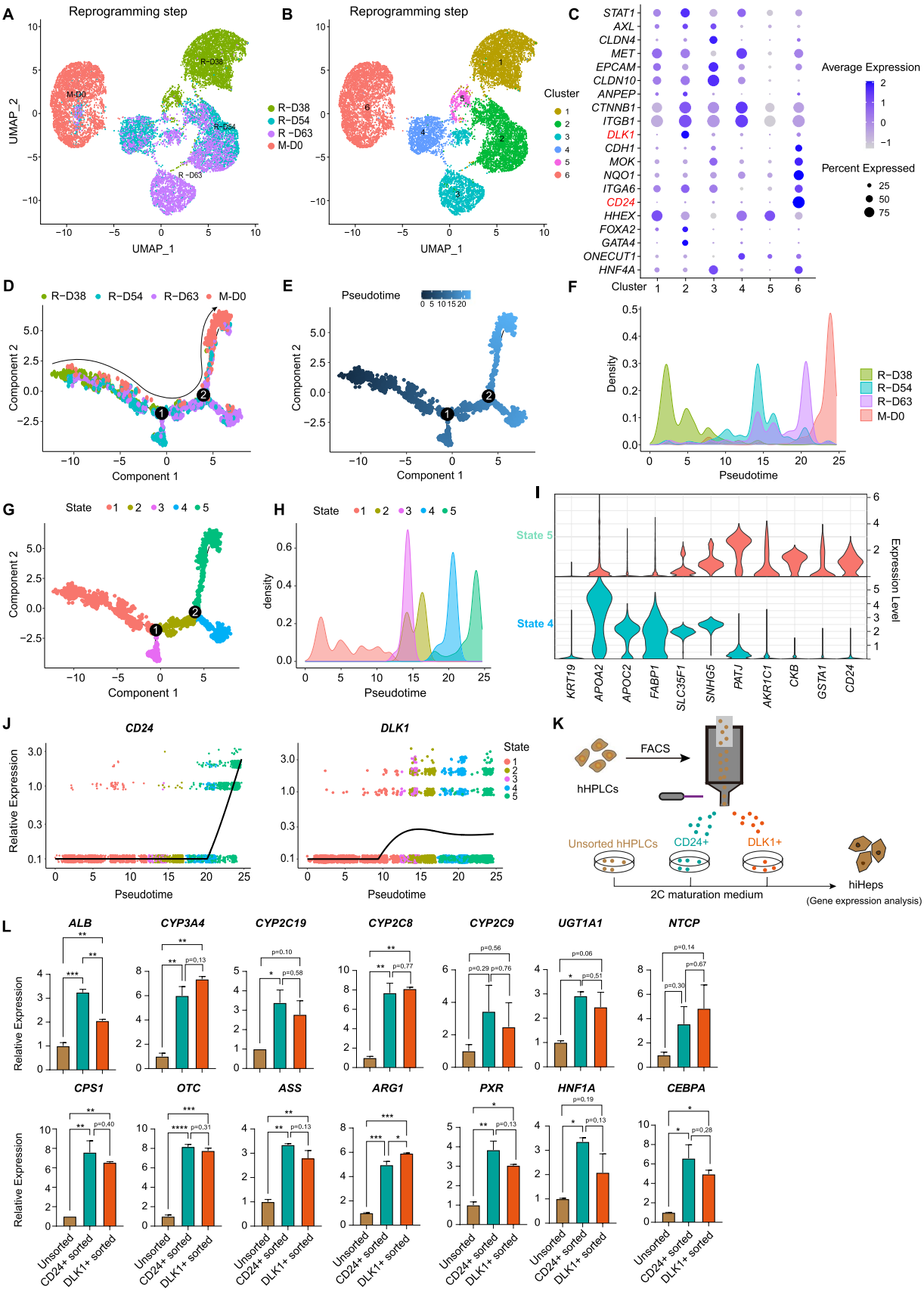
To validate the role of lipids in hepatocyte maturation, we supplemented the 2C maturation medium with chemically defined lipid concentrate (CDL) during hiHeps maturation. CDL supplementation significantly enhanced hiHeps maturation, evidenced by the upregulation of functional hepatocyte genes such as *CYP2C9*, *CYP2C19*, *CYP1A2*, *UGT2B15*, *UGT2B7*, *CPS1*, *MRP2*, and *HNF4A* (Fig. 4J). These findings underscore the essential role of lipid metabolism in promoting hepatocyte maturation.

HNF4A and *HHEX* regulate hepatocyte and intestinal cell fate during maturation step in two-step reprogramming

As the maturation process indicated an intestinal cell fate in the hiHeps, we further investigated the cell identity at the final stage (M-D7). M-D7 cells were divided into two sub-populations (Fig. 5A). Cluster 0 expressed functional hepatocyte genes, including *CYP3A4*, *CYP2C19*, *CYP2C9*, *CYP2D6*, *CYP2C8*, *UGT1A1*, *UGT2B7*, *UGT2B15*, and *PLG* (Fig. 5B-5C and S4A), while cluster 1 expressed intestinal cell genes like *AGR2*, *LGALS3*, *LCN2*, *S100A6*, *S100P*, *MUC5AC*, *MUC5B*, *TTF1*, and *VSTM2L* (Fig. 5B, C and S4B). Both clusters expressed common hepatocyte markers, such as *ALB*, *TTR*, *FABP1*, and *SERPINA1*, indicating a subpopulation of hiHeps acquired a mixed hepatocyte and intestinal cell fate (Fig. 5B, C and S4C). Differential expression and GO enrichment analysis confirmed that cluster 0 was enriched with genes involved in oxidation and respiration processes, while cluster 1 was enriched with metabolic process genes typical of intestinal cells (Fig. 5D and E). CellNet analysis further validated these findings, showing that cluster 0 had a significantly higher liver score, while cluster 1 had a higher intestinal score (Fig. S4D and S4E).

To investigate the biased cell fate determination, we examined the correlation between five reprogramming TFs that are expressed across various endoderm-derived tissues and their influence on the determination of hepatocyte and intestinal lineages. Cluster 0 was closely correlated with *HNF4A*, while cluster 1 was correlated with *HHEX* (Fig. 5C). *HNF4A* is essential for liver development and functional hepatocyte differentiation [32, 33], whereas *HHEX* is a marker of the definitive endoderm, crucial for early liver development [34]. However, the contrasting roles of these TFs in hepatocyte reprogramming remain unclear.

To confirm the roles of *HNF4A* and *HHEX* in hiHeps reprogramming, we knocked down these genes in hHPLCs and matured them in 2C medium for an additional 7 days to generate hiHeps. Knockdown of either *HNF4A* and *HHEX* reduced the expression of *ALB* and other key markers (Fig. 5F and G). Specifically, *HNF4A* knockdown led to a significant reduction in functional hepatocyte markers like *ALB*, *CYP3A4*, *MRP2*, *CYP2C9*, and *CYP2C19*, with minimal effect on intestinal genes like *MUC5AC*, *AGR2*, and


A
Reprogramming step
UMAP_2
UMAP_1
R−D38
M-D0
R−D54
R −D63
B
Reprogramming step
R−D38
R−D54
R −D63
M-D0
Cluster
1
2
3
4
5
6
C
STAT1
AXL
CLDN4
MET
EPCAM
CLDN10
ANPEP
CTNNB1
ITGB1
DLK1
CDH1
MOK
NQO1
ITGA6
CD24
HHEX
FOXA2
GATA4
ONECUT1
HNF4A
Cluster 1 2 3 4 5 6
Average Expression
Percent Expressed
25
50
75
D
R−D38 R−D54 R−D63 M-D0
Component 2
Component 1
E
Pseudotime
0 5 10 15 20
Component 2
Component 1
F
Density
Pseudotime
R−D38
R−D54
R−D63
M-D0
G
State 1 2 3 4 5
Component 2
Component 1
H
State 1 2 3 4 5
density
Pseudotime
I
State 5
State 4
Expression Level
KRT19
APOA2
APOC2
FABP1
SLC35F1
SNHG5
PATJ
AKR1C1
CKB
GSTA1
CD24
J
CD24
DLK1
State 1 2 3 4 5
Relative Expression
Pseudotime
Pseudotime
K
FACS
hHPLCs
Unsorted hHPLCs
CD24+
DLK1+
2C maturation medium
hiHeps
(Gene expression analysis)
L
ALB
CYP3A4
CYP2C19
CYP2C8
CYP2C9
UGT1A1
NTCP
CPS1
OTC
ASS
ARG1
PXR
HNF1A
CEBPA
Relative Expression
Unsorted
CD24+ sorted
DLK1+ sorted

◂**Fig. 3** CD24 and DLK1 Marks Two Different Hepatic Progenitor Populations. **A** UMAP of individual cells during the first step of reprogramming process, colored by indicated time points. **B** UMAP of individual cells during the first step of reprogramming process, colored by cell clusters. **C** Dot plots showing the expression levels of marker genes for each cluster. **D** Pseudotime trajectory analysis of reprogramming cells, colored by indicated time points. **E**, **F**. Pseudotime trajectory analysis of reprogramming cells, colored according to pseudotime progression. **G**, **H** Pseudotime trajectory analysis of reprogramming cells, colored by cell state. **I** Violin plot displaying the expression of representative hepatic progenitor genes in state 4 and state 5 cells. **J** Expression pattern of *CD24* and *DLK1* along pseudotime trajectory. **K** Schematic diagram of sorting CD24 and DLK1 positive hHPLCs followed by their maturation in 2C medium. **L** RT-qPCR analysis of hepatocyte gene expression (n = 2). The data are presented as the mean ± SD. Significance was assessed using one-way ANOVA. ***p < 0.001; **p < 0.01; *p < 0.05

FABP2 (Fig. 5F). Conversely, *HHEX* knockdown decreased the expression of intestinal genes *MUC5AC*, *AQP8*, *EPHB2*, *DLFA*, *FABP2*, *CHGA*, and *CDX2*, while only slightly affecting *ALB* expression (Fig. 5G). These results suggest that *HHEX* promotes an intestinal cell fate in hiHeps during maturation, whereas *HNF4A* is crucial for maintaining hepatocyte function.

Gene regulatory networks in the two-step reprogramming process

To gain insight into specific TF regulatory networks involved in the late stage of two-step reprogramming process, we re-mapped the 7 samples representing different stages of reprogramming and maturation, clustering them into four distinct groups (Fig. 6A and B). Single-cell regulatory network inference and clustering (SCENIC) analysis revealed that cluster 0 of early pre-progenitor samples from day 38 to day 63 (R-D38, R-D54 and R-D63) was regulated by TFs such as *HES7* and *TBX15* (Fig. 6C and D). Cluster 1 of hepatic progenitor samples (M-D0) was regulated by *MYBL2* and *SMARCA4*, etc. (Fig. 6C and D). Cluster 2 representing immature hepatocyte samples (M-D2 and M-D4) was primarily regulated by *HOXA10, HOXB7*, *PPARA,* and *NR1H4*, etc., while cluster 3, which contained mature hepatocyte samples (M-D7) was regulated by *RARA* and *FOXO3*, etc. (Fig. 6C and D). The expression pattern of these TFs was shown in the feature plots (Fig. S5A).

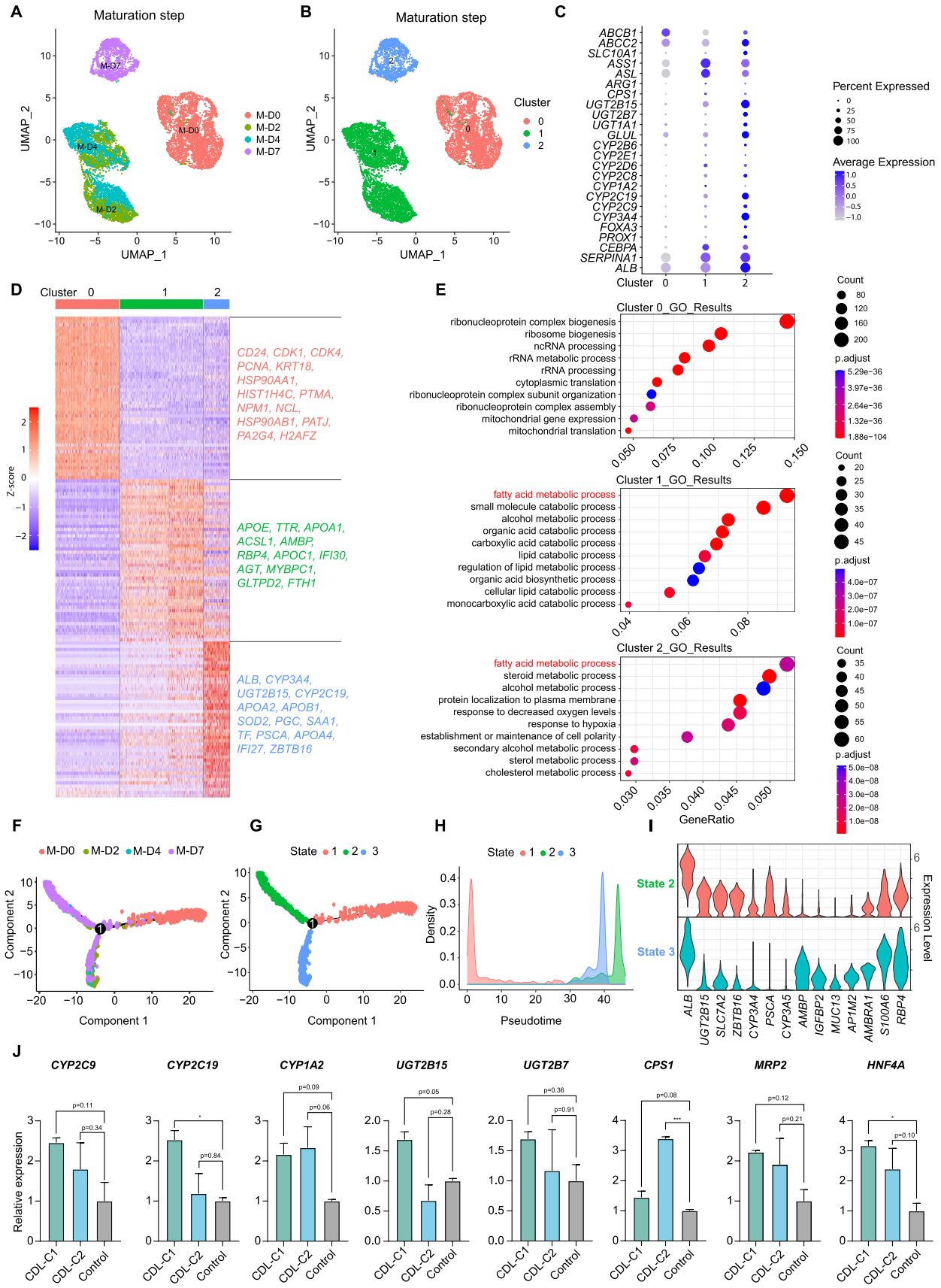
We also constructed a gene network based on pairwise correlations of the top DEG expressions during progressive cell fate transitions from R-38 to M-D7 (Fig. 6E). This analysis revealed four chronological subnetworks. The pre-progenitor subnetwork was closely linked to progenitor related genes, and the immature hepatocyte genes were tightly connected to the mature hepatocyte subnetwork (Fig. 6E). The sequential switch of transcriptional circuits underscored the close relationship between the pre-progenitor and progenitor subnetworks, which acted as a bridge linking immature hepatocytes subnetwork to that of mature hepatocytes, which consisted of key liver functional transcription factors such as *RARA*, *XBP1*, *PPARA*, *CEBPB*, and *NR5A2* (Fig. 6E).

Discussion

Our study presents a comprehensive, high-resolution single-cell roadmap of the two-step reprogramming process converting human fibroblasts to functional hiHeps using scRNA-seq and computational analysis. This roadmap offers valuable insights for optimizing hiHeps induction and enhancing our understanding of hepatocyte fate determination, which is beneficial to hepatocytes-based therapeutic applications.

Our most impressive finding is that the late stage of the two-step lineage reprogramming process accurately recapitulates the human liver developmental trajectory, thus offering an efficient platform to explore hepatocyte fate determination and GRN establishment. Pseudotime analysis reveals a cell-fate transition from pre-progenitors to hHPLCs, immature hepatocytes and ultimately functional hepatocytes, each stage characterized by unique gene expression profiles mirroring its in vivo counterpart (Fig. 2C–I). Additionally, SCENIC analysis unveils a sequential establishment of different GRNs throughout reprogramming stages (Fig. 6C–E). While some factors involved in these GRNs, such as PPARA and NR1H4 [35, 36], have been previously reported, others are newly identified and remain to be further explored and validated.

While our study primarily focuses on the two-step reprogramming process which relies on exogenous transgenes, alternative methods such as chemical reprogramming have also been explored for generating hepatocyte-like cells. These approaches rely on small molecules to modulate signaling pathways and transcriptional networks, bypassing the need for exogenous transcription factors. Both transgenic and chemical reprogramming strategies converge on the activation of HNF4α, a master regulator of hepatocyte identity, to establish hepatocyte-specific GRNs. While transgenic approaches directly introduce exogenous HNF4α, chemical reprogramming achieves this by modulating signaling pathways that activate endogenous HNF4α expression [37, 38]. According to recent studies, chemical reprogramming induces hepatocyte fate determination in a more direct manner, as evidenced by the generation of chemically induced hepatocytes (CiHeps) without clearly identifiable intermediate progenitor stages [38]. This study demonstrated that CiHeps rely on rapid HNF4α activation and chromatin remodeling to bypass pluripotency-associated transitions, further underscoring the role of HNF4α as a linchpin


A
Maturation step
M-D0
M-D2
M-D4
M-D7
UMAP_1
UMAP_2
B
Maturation step
Cluster
0
1
2
C
ABCB1
ABCC2
SLC10A1
ASS1
ASL
ARG1
CPS1
UGT2B15
UGT2B7
UGT1A1
GLUL
CYP2B6
CYP2E1
CYP2D6
CYP2C8
CYP1A2
CYP2C19
CYP2C9
CYP3A4
FOXA3
PROX1
CEBPA
SERPINA1
ALB
Percent Expressed
Average Expression
D
Cluster
Z-score
CD24, CDK1, CDK4, PCNA, KRT18, HSP90AA1, HIST1H4C, PTMA, NPM1, NCL, HSP90AB1, PATJ, PA2G4, H2AFZ
APOE, TTR, APOA1, ACSL1, AMBP, RBP4, APOC1, IFI30, AGT, MYBPC1, GLTPD2, FTH1
ALB, CYP3A4, UGT2B15, CYP2C19, APOA2, APOB1, SOD2, PGC, SAA1, TF, PSCA, APOA4, IFI27, ZBTB16
E
Cluster 0_GO_Results
ribonucleoprotein complex biogenesis
ribosome biogenesis
ncRNA processing
rRNA metabolic process
rRNA processing
cytoplasmic translation
ribonucleoprotein complex subunit organization
ribonucleoprotein complex assembly
mitochondrial gene expression
mitochondrial translation
Cluster 1_GO_Results
fatty acid metabolic process
small molecule catabolic process
alcohol metabolic process
organic acid catabolic process
carboxylic acid catabolic process
lipid catabolic process
regulation of lipid metabolic process
organic acid biosynthetic process
cellular lipid catabolic process
monocarboxylic acid catabolic process
Cluster 2_GO_Results
fatty acid metabolic process
steroid metabolic process
alcohol metabolic process
protein localization to plasma membrane
response to decreased oxygen levels
response to hypoxia
establishment or maintenance of cell polarity
secondary alcohol metabolic process
sterol metabolic process
cholesterol metabolic process
GeneRatio
Count
p.adjust
F
M-D0 M-D2 M-D4 M-D7
Component 1
Component 2
G
State 1 2 3
H
Density
Pseudotime
I
State 2
State 3
Expression Level
ALB
UGT2B15
SLC7A2
ZBTB16
CYP3A4
PSCA
CYP3A5
AMBP
IGFBP2
MUC13
AP1M2
AMBRA1
S100A6
RBP4
J
CYP2C9
CYP2C19
CYP1A2
UGT2B15
UGT2B7
CPS1
MRP2
HNF4A
Relative expression
CDL-C1
CDL-C2
Control

◂**Fig. 4** Lipid Metabolism Mediates Reprogramming of Hepatic Progenitors into Functional Hepatocytes. **A** UMAP of individual cells during the second step of the maturation process, colored by the indicated time points. **B** UMAP of individual cells during the second step of the maturation process, colored by cell clusters. **C** Dot plots showing the expression of marker genes for each cluster. **D** Heatmap showing three groups of DEGs of the 3 cell clusters. Each column represents a single cell. Representative genes for each group of DEGs were listed on the right. **E** GO enrichment analysis of three groups of DEGs. **F** Pseudotime trajectory analysis of reprogramming cells, colored by the indicated time points. **G**, **H** Pseudotime trajectory analysis of reprogramming cells, colored by cell states. **I** Violin plot displaying the expression of representative hepatocyte genes in state 2 and 3 cells. **J** RT-qPCR analysis of functional hepatocyte gene expression (n = 2). The data are presented as the mean ± SD. Significance was assessed using one-way ANOVA. ***p < 0.001; **p < 0.01; *p < 0.05

in hepatocyte GRN establishment [38]. Despite these mechanistic differences, both methods face challenges in suppressing off-target lineages, such as intestinal programs, which may arise due to incomplete GRN specification [37]. In the future, the combination of TFs and chemical compounds, such as CDK8 inhibition [39], may further enhance the efficiency and precision of hepatocyte reprogramming by synergistically reinforcing HNF4α-driven GRNs.

A key aspect of our study is the identification of CD24 and DLK1 as surface markers that enrich distinct hepatic progenitor populations. During reprogramming, DLK1 marked a hepatic progenitor population that emerge earlier than CD24-positive hepatic progenitors, both displaying higher efficiency for hiHeps induction compared to unsorted hHPLCs (Fig. 3J–L). Notably, CD24 has been reported to enrich hepatic progenitors in adult mouse liver [40] and was also expressed in highly proliferative human chemically induced liver progenitors derived from infant liver samples [41]. In our reprogramming samples, CD24-positive hepatic progenitors which may correspond to fetal hepatocytes, enriched the expression of progenitor markers *ITGA6, NQO1, MOK* and *CDH1* (Fig. 3C), suggesting a conservative role of CD24 between mouse and human. Likewise, DLK1 was highly expressed in the E10.5 mouse liver bud [42] and DLK1-positive mouse hepatoblasts exhibit high proliferation and hepatocyte specification potential [43]; however, the role of DLK1 has not been little reported in human hepatic progenitors yet. Our study found that DLK1-positive hepatic progenitors exist in the reprogramming process, enriched the expression of progenitor markers *ITGB1, CTNNB1, ITGA6* and *ANPEP* (Fig. 3C), with capacity of hepatocyte specification (Fig. 3L). These DLK1-positive progenitors align with early hepatoblasts (7–11 post-conception weeks (PCW)). Collectively, CD24 and DLK1 serve as valuable tools for isolating and enhancing progenitor cell populations primed for hepatocyte maturation. Investigating the role of these markers in human hepatic progenitors in vivo would also be valuable.

Another notable finding is that lipid metabolism plays a crucial role in the maturation of hepatic progenitors into functional hepatocytes. GO analysis of the reprogramming samples reveals enrichment of lipid metabolism related genes during cell-fate transition, especially at timepoints around late stages of hHPLCs reprogramming and the subsequent hiHeps maturation steps (Fig. 4E). This includes up-regulation of processes involving fatty acids, steroids, organic acids, carboxylic acids, cholesterol, and monocarboxylic acids, reflecting a metabolic transition conducive to the functionality of hepatocytes [44–46] (Fig. 4E). Intriguingly, our study demonstrates that lipids can promote hiHeps maturation (Fig. 4J). This aligns with previous findings suggesting that hydrocortisone promotes PHH expansion through PPARα signaling [47], highlighting the importance of lipids in hepatocyte cell fate decisions and proliferation.

Our study also reveals a novel role for the transcription factors *HNF4A* and *HHEX* in determining the fate of reprogrammed cells between hepatocyte and intestinal lineages. Knockdown experiments demonstrate the essential role of *HNF4A* in driving the hepatocyte differentiation of hHPLCs (Fig. 5F), aligning with previous findings that highlight its role in liver development and hepatocyte differentiation [48, 49]. While *HHEX* has been previously reported to govern definitive endoderm formation [50], our study uncovers its novel role in intestinal fate determination. Specifically, knockdown of *HHEX* results in the downregulation of intestinal gene expression in hiHeps. These findings suggest that sophisticated modulation of certain transcription factors is required for the authentic production of hiHeps.

In summary, this study provides a detailed single-cell roadmap for generating functional hepatocytes in vitro, uncovering the molecular mechanisms and regulatory networks involved. By addressing the key scientific challenge of constructing authentic hepatocyte-specific GRNs, we pave the way for more effective hepatocyte generation and therapeutic applications.

Materials and methods

Isolation and culture of human fibroblasts

This study received approval from the Clinical Research Ethics Committee of China-Japan Friendship Hospital (Ethical approval No: 2009–50) and the Stem Cell Research Oversight of Peking University (SCRO201103-03), adhering

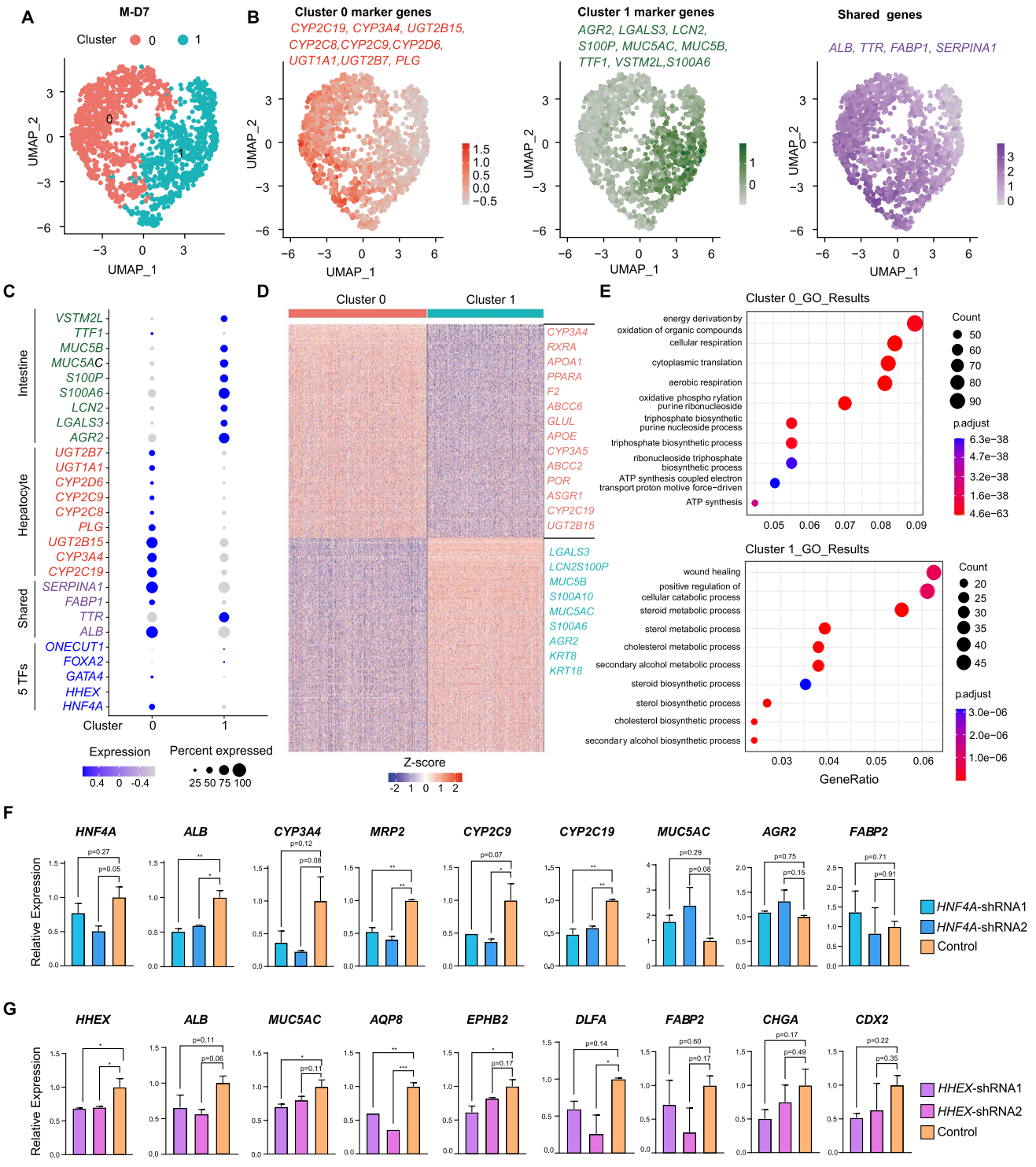


Fig. 5 *HNF4A* and *HHEX* Regulate Hepatocyte and Intestinal Cell Fate During Maturation Step in Two-Step Reprogramming. **A** UMAP of individual cells of the M-D7 sample. **B** UMAP visualizations showing the expression of representative cluster 0 marker genes (left), cluster 1 marker genes (middle), and shared genes of cluster 0 and 1 (right). **C** Dot plots showing the expression of marker genes. **D** Heatmap showing the two groups of DEGs in the 2 cell clusters. Each column represents a single cell. Representative genes for each group of DEGs were listed on the right. **E** GO enrichment analysis of the two groups of DEGs. **F** RT-qPCR analysis of hepatocyte gene expression in *HNF4A* knockdown hiHeps (n = 2). **G** RT-qPCR analyses of intestine gene expression in *HHEX* knockdown hiHeps (n = 2). The data are presented as the mean ± SD in figure **H** and **I**. Significance was assessed using one-way ANOVA. ***p < 0.001; **p < 0.01; *p < 0.05

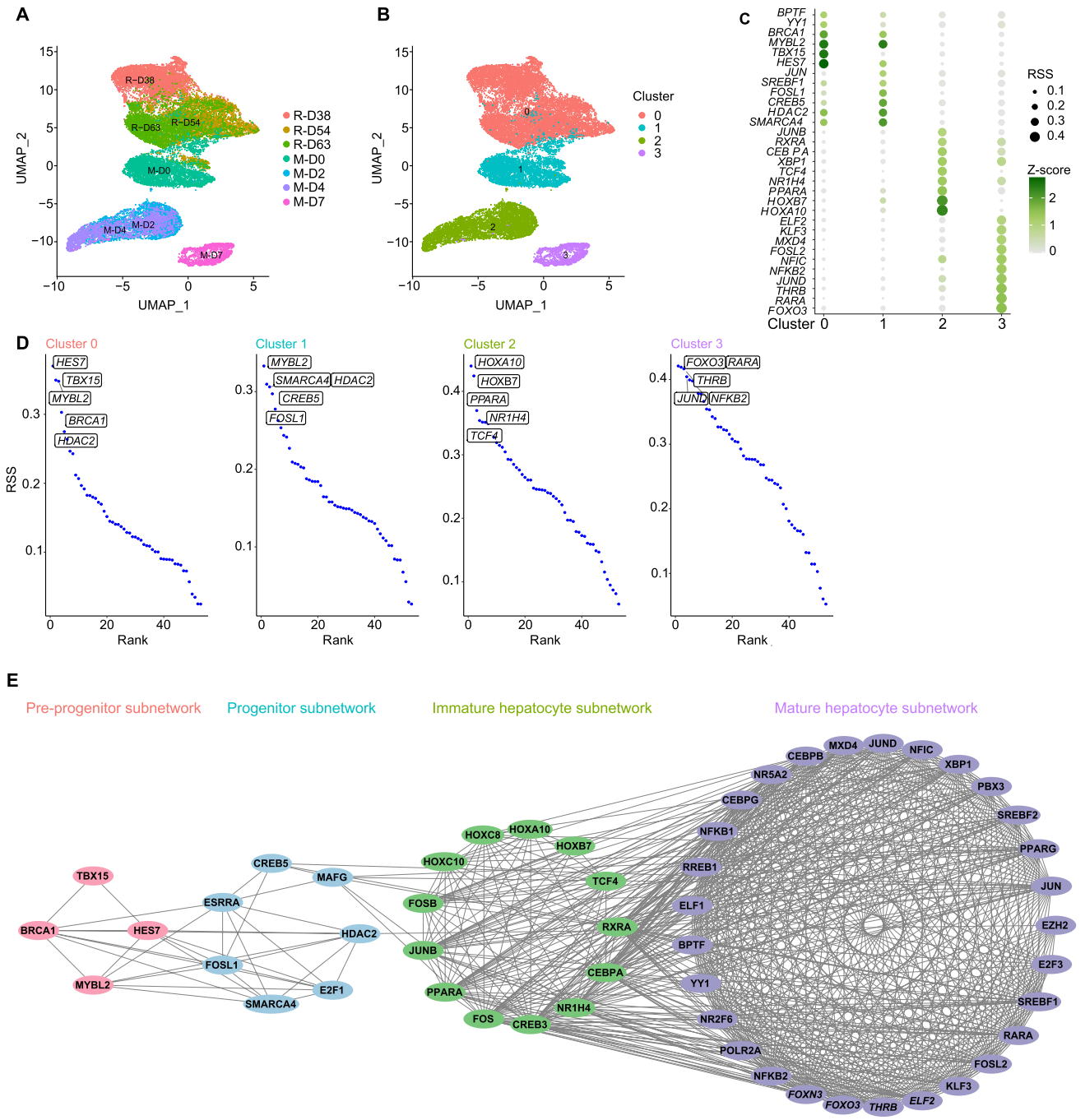


Fig. 6 Gene Regulatory Networks in the Two-step Reprogramming Process. **A** UMAP of individual cells from the late stage of the two-step reprogramming samples, colored by indicated time points. **B** UMAP of individual cells from the late stage of the two-step reprogramming samples, colored by cell cluster. **C** Single-Cell Regulatory Network Inference and Clustering (SENIC) analysis of regulatory TFs of each cluster. Top ranked TFs are shown. **D** Dot plots of representative TFs for each cluster. **E** Transcriptional regulatory network showing the progression from pre-progenitor cells to mature hepatocytes. This network is divided into four subnetworks, each representing distinct cell states during hepatocyte development: pre-progenitor (red), progenitor (blue), immature hepatocyte (green), and mature hepatocyte (purple)

to the principles outlined in the Declaration of Helsinki. Human embryonic skin samples were obtained from aborted tissue with appropriate informed consent provided by the patients. The skin tissue was minced using forceps and subsequently incubated in a solution of 1 mg/mL collagenase IV (Gibco, 17104019) for 1–2 h at 37 °C. Following this enzymatic treatment, the cells were harvested via centrifugation and resuspended in human embryonic

fibroblast (HEF) medium. The HEF medium consisted of Dulbecco's Modified Eagle's Medium (DMEM, Gibco, 11965092), supplemented with 10% fetal bovine serum (FBS, Ausbian, WS500TZ), 1% GlutaMAX (Gibco, 35050061), 1% Non-Essential Amino Acids (NEAA, Gibco, 11140035), and 1% penicillin/streptomycin (PS, Gibco, 15140122). The isolated cells were then plated onto 10 cm tissue culture dishes and maintained in the HEF medium for further culture.

Two-step lineage reprogramming protocol

The two-step lineage reprogramming procedure was adapted with modifications from earlier studies [6]. Human fibroblasts were initially plated at a density of approximately 1×10^4 cells per well in 12-well plates the day prior to viral infection. These fibroblasts were subsequently infected with a lentivirus containing five hepatic transcription factors (HNF4A, ONECUT1, GATA4, FOXA2, and HHEX), along with Tet-on c-MYC, rtTA, and shRNA targeting P53. This infection occurred at a multiplicity of infection (MOI) of around 15 in HEF medium that included 10 μg/mL polybrene (MCE, HY-112735).

After a 12-h incubation, the cells were washed with phosphate-buffered saline (PBS, Gibco, 10010023) and then cultured in fibroblast medium supplemented with 2 μg/mL doxycycline (MCE, HY-N0565) for a duration of 10 days. Following this period, the cells were detached using Accutase (Gibco, 00–4555-56) at a dilution ratio of 1:3–4 based on cell density and were then plated on culture dishes coated with 0.2 mg/mL Matrigel (Corning, 354277).

Post-passage, the cells were transferred to hepatic expansion medium (HEM), which consisted of a 50/50 mixture of DMEM/F12 (Gibco, 11330032) and William's E Medium (Gibco, A1217601). This medium was enriched with 1% penicillin–streptomycin solution (Gibco, 15140122), 2% B27 without vitamin A (Gibco, 12587010), 5 mM nicotinamide (MCE, HY-B0150), 200 μM 2-phospho-L-ascorbic acid (pVc, Sigma-Aldrich, 49752), 3 μM CHIR99021 (MCE, HY-10182), 5 μM SB431542 (MCE, HY-10431), 0.5 μM sphingosine-1-phosphate (S1P, Sigma-Aldrich, S9666), 5 μM lysophosphatidic acid (LPA, Sigma-Aldrich, L7260), 50 ng/mL epidermal growth factor (EGF, Peprotech, AF-100-15), and 2 μg/mL doxycycline (MCE, HY-N0565). The reprogrammed cells were classified as hHPLCs based on their successful epithelial conversion, expression of hepatic progenitor genes, and ability to respond to maturation media for functional hepatocyte production. These hHPLCs were maintained in HEM and passaged every five days at a ratio of 1:5.

To obtain functional human hepatocytes (hiHeps), hHPLCs were cultured until they reached compact confluence on Matrigel-coated plates and then treated with 2C medium for a period of seven days. The 2C medium was formulated by modifying HCM (Lonza, CC-3198) to exclude EGF and included 50 μM forskolin (MCE, HY-15371) and 10 μM SB431542 (MCE, HY-10431). In experiments involving lipid supplementation, a chemically defined lipid concentrate (CDL, Gibco, 11905031) was added at a 1 × concentration.

RT-qPCR analysis

Total RNA was extracted using the Direct-zol RNA Miniprep (ZYMO RESEARCH, R2050) method and subsequently reverse-transcribed with the TransScript FirstStrand cDNA Synthesis SuperMix (TransGen Biotech, AT311-02). Quantitative real-time PCR (RT-qPCR) was carried out with the KAPA SYBR® FAST Universal qPCR Mix (KAPA Biosystems) on a BIO-RAD CFX384TM Real-time System. The resulting quantification values were normalized to the input, determined by housekeeping genes (RPL13A or GAPDH). The primer sequences used for RT-qPCR are listed in Table S1.

Fluorescence-activated cell sorting

hHPLCs were dissociated into single-cell suspension by Accutase at 37 °C for 3–5 min. Thoroughly re-suspended cells were washed twice in PBS. After that cells were conjugated with antibodies including Human DLK1 APC-conjugated Antibody (R&D, FAB1144A-100) and PE Mouse Anti-Human CD24 (BD Biosciences, 560991) in 200 μL PBS supplemented with 0.5% bovine serum albumin (BSA) in 4 °C for 2 h. Then, stained cells were washed twice in PBS and re-suspended in PBS. CD24 and DLK1 positive cells were finally sorted using fluorescence-activated cell sorting (FACS) on CytoFLEX SRT (Beckman-Coulter).

Gene knockdown

The shRNA plasmids targeting *HHEX* and *HNF4A* were acquired from the National Center at Peking University. These oligonucleotides were inserted downstream of the U6 promoter within a pLKO.1 vector, and the specific sequences are detailed in Table S2. The production and collection of lentiviruses have been previously described [51].

Single-cell RNA sequencing

Cells at indicated time points were digested and resuspended at 1×10^6 cells per mL in 1 × PBS with 0.04% BSA. Then, cell suspensions (300–1000 living cells per microliter determined by trypan blue staining) were loaded on a Chromium Single Cell Controller (10 × Genomics) to generate single-cell gel beads in emulsion (GEMs) by using Single Cell 30 Library and Gel Bead Kit (10 × Genomics). Captured cells were lysed and the released RNA were barcoded through reverse transcription in individual GEMs. Barcoded cDNAs were pooled and cleanup by using DynaBeads MyOne Silane Beads (Invitrogen). Single-cell RNA-seq libraries were prepared using Single Cell 30 Library Gel Bead Kit (10 × Genomics) following the manufacture's introduction. Sequencing was performed on an Illumina HiSeq 4000 with pair end 150 bp (PE150).

scRNA-seq data analysis

Raw FASTQ files of scRNA-seq data were processed with CellRanger software (v3.1.0), including alignment, filtering, barcode counting and unique molecular identifiers (UMI) counting. Reads were aligned with GRCh38 genomes. For data preprocessing and integration, we processed each dataset individually using the Seurat package. Cells with fewer than 1000 or more than 10,000 detected genes, or with mitochondrial gene expression levels greater than 10%, were removed to filter out low-quality cells. The datasets were then merged into a single Seurat object for downstream analysis. The merged dataset was normalized using LogNormalize and scaled using ScaleData. We selected 2000 variable features for downstream analysis and performed Principal Component Analysis (PCA) to reduce dimensionality. The top 10 principal components were then used to perform Uniform Manifold Approximation and Projection (UMAP) for visualization of the cellular landscape. All differentially expressed genes (DEGs) and Gene ontology (GO) terms in this study were listed in Table S3.

Single cell trajectory analysis

Single cell trajectory was analyzed using matrix of cells and gene expressions by Monocle 2 [52]. Before monocle analysis for gene selection, we identified highly variable genes or used cluster-specific DEGs from Seurat analysis. Then, we used monocle to construct lineage trajectory and branch points by analyzing DEGs. Dimensionality reduction was performed using the DDRTree algorithm implemented in Monocle. The state representing earlier cell type was chosen as the root. Finally, the trajectory plot was generated, visualizing the progression of cells through pseudotime. Gene expression heatmap of pseudotime is based on lineage trajectory and performed by ''plot_genes_branched_heatmap'' function of monocle 2. The core idea of this function is to merge all cells into total 200 bins in the order of pseudotime.

Gene regulatory network analysis

Gene regulatory network analysis was performed using SCENIC (v1.3.1) [53]. The input to SCENIC is an expression matrix, in which rows correspond to genes and columns correspond to cells. The genes with at least 60 UMIs across all cells and detected in at least 20 cells were selected as the input genes. The expression matrix was loaded into GENIE3 and the co-expressed genes to each TF was constructed. The TF co-expression modules were then analyzed by RcisTarget (v1.18.2). The filtered potential targets by RcisTarget human hg19 database from the co-expression module were used to build the regulons. The regulons activity (Area Under the Curve) was analyzed by AUCell and the active regulons are determined by AUCell default threshold.

To calculate the Regulon Specificity Score (RSS) for each regulon, we used SCENIC's calcRSS function, which computes RSS based on the AUC values of regulons across different cell types. To provide an overview of RSS distribution across different regulons and cell types, we created a heatmap of the RSS matrix using the plotRSS function. The heatmap displays the relative activity of each regulon across all cell types, offering insights into the regulatory activity landscape. We constructed a TF network based on the calculated RSS matrix. First, for each molecule (TF), we assigned it to the cell group where its RSS score was highest using a threshold-based approach. We then computed a Pearson correlation matrix between molecules, which allowed us to identify significant correlations between pairs of molecules (correlation coefficient > 0.5). The final network was visualized by Cytoscape.

Cell fate analysis

We analyzed cell fate in our terminal induced cells by CellNet (v0.1.1) [54]. To classify query samples, we log-transformed the pseudo bulk scRNA data and used a pre-trained classifier model (liver_broadClassifier100.rda) to predict cell types. This model was based on reference expression data from engineered liver cells. The classification result was visualized as a heatmap using the acn_queryClassHm function. To evaluate the gene regulatory network (GRN) status of the samples, we used the pre-processed GRN data (liver_grnAll.rda) and normalized training parameters (liver_trainNormParam.rda) from https://github.com/CahanLab/PACNet. The GRN status

for each sample was calculated and visualized by bar plots illustrating GRN activity across samples.

Ontology annotation

GO term enrichment analyses were performed using "clusterProfiler", "enrichplot", and "ggplot2" packages [55]. Terms that had a P-value < 0.05 was defined as significantly enriched.

Statistical analysis

Statistical analyses were performed using GraphPad Prism (version 10.2.3). P values for the purpose of sample comparisons were calculated using one-way ANOVA. The level of significance in all graphs is represented as follows: *P < 0.05, **P < 0.01, ***P < 0.001. All the error bars represent SD.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00018-025-05677-x.

Acknowledgements We thank the National Center at Peking University in Beijing, China, for providing shRNA plasmids.

Author contributions Nan Jiang and Sen Luo performed scRNA-seq data analysis. Guangya Li, Xi Kong performed the experiments, and interpreted the results. Shigang Yin, Jianhua Peng, and Yong Jiang assisted the research design and manuscript writing. Wei Tao and Cheng Li assisted the research design and scRNA-seq data analysis. Hongkui Deng supervised this project and designed the research. Huangfan Xie and Bingqing Xie supervised this project, designed the research, interpreted the results, and wrote the manuscript.

Funding This work was supported by the National Natural Science Foundation of China (32300675, 32400598, 32288102, 82301658), Sichuan Science and Technology Program (2022YFS0615, 2025ZNSFSC1019), Luzhou Science and Technology Program of China (2023SYF136, 2023JYJ019), Southwest Medical University and Luzhou Government (2021LZXNYD-Z05), Doctoral Research Initiation Fund of Affiliated Hospital of Southwest Medical University (22069, 22070).

Availability of data and materials All data are available in the Article and its Supplementary Information. The scRNA-seq datasets in this study have been deposited to the NCBI Gene Expression Omnibus (GEO) with the accession GSE278249. The public scRNA-seq datasets of PHHs and fibroblasts can be accessed at GEO database under accession numbers GSE261965 and GSE178325. Human liver scRNA-seq data have been deposited in ArrayExpress under accession E-MTAB-7407.

Declarations

Conflict of interest The authors declare that there are no competing interests.

Ethical approval and consent to participate All research using human fibroblasts was approved by the Clinical Research Ethics Committee of China-Japan Friendship Hospital (Ethical approval No: 2009–50) and Stem Cell Research Oversight of Peking University (SCRO201103-03).

Consent for publication Not applicable.

References

1. Nulty J, Anand H, Dhawan A (2024) Human hepatocyte transplantation: three decades of clinical experience and future perspective. Stem Cells Transl Med 13:204–218
2. Jin M, Yi X, Liao W, Chen Q, Yang W, Li Y, Li S, Gao Y, Peng Q, Zhou S (2021) Advancements in stem cell-derived hepatocyte-like cell models for hepatotoxicity testing. Stem Cell Res Ther 12:84
3. Wang H, Yang Y, Liu J, Qian L (2021) Direct cell reprogramming: approaches, mechanisms and progress. Nat Rev Mol Cell Biol 22:410–424
4. Zakikhan K, Pournasr B, Vosough M, Nassiri-Asl M (2017) In vitro generated hepatocyte-like cells: a novel tool in regenerative medicine and drug discovery. Cell J 19:204–217
5. Morris SA, Cahan P, Li H, Zhao AM, San Roman AK, Shivdasani RA, Collins JJ, Daley GQ (2014) Dissecting engineered cell types and enhancing cell fate conversion via Cell Net. Cell 158:889–902
6. Xie B, Sun D, Du Y, Jia J, Sun S, Xu J, Liu Y, Xiang C, Chen S, Xie H, Wang Q, Li G, Lyu X, Shen H, Li S, Wu M, Zhang X, Pu Y, Xiang K, Lai W, Du P, Yuan Z, Li C, Shi Y, Lu S, Deng H (2019) A two-step lineage reprogramming strategy to generate functionally competent human hepatocytes from fibroblasts. Cell Res. 29(90):696–710
7. Godoy P, Widera A, Schmidt-Heck W, Campos G, Meyer C, Cadenas C, Reif R, Stöber R, Hammad S, Pütter L, Gianmoena K, Marchan R, Ghallab A, Edlund K, Nüssler A, Thasler WE, Damm G, Seehofer D, Weiss TS, Dirsch O, Dahmen U, Gebhardt R, Chaudhari U, Meganathan K, Sachinidis A, Kelm J, Hofmann U, Zahedi RP, Guthke R, Blüthgen N, Dooley S, Hengstler JG (2016) Gene network activity in cultivated primary hepatocytes is highly similar to diseased mammalian liver tissue. Arch Toxicol 90:2513–2529
8. Wu H, Yang ASP, Stelloo S, Roos FJM, te Morsche RHM, Verkerk AH, Luna-Velez MV, Wingens L, de Wilt JHW, Sauerwein RW, Mulder KW, van Heeringen SJ, Verstegen MMA, van der Laan LJW, Marks H and Bártfai R. Integrative omics analysis reveals gene regulatory mechanisms distinguishing organoid-derived hepatocytes from primary human hepatocytes. *bioRxiv*. 2023:2023.12.05.570132.

9. Macneil LT, Walhout AJ (2011) Gene regulatory networks and the role of robustness and stochasticity in the control of gene expression. Genome Res 21:645–657
10. Davidson EH, Erwin DH (2006) Gene regulatory networks and the evolution of animal body plans. Science 311:796–800
11. Huang X, Song C, Zhang G, Li Y, Zhao Y, Zhang Q, Zhang Y, Fan S, Zhao J, Xie L, Li C (2024) scGRN: a comprehensive single-cell gene regulatory network platform of human and mouse. Nucleic Acids Res 52:D293-d303
12. Zhang S, Pyne S, Pietrzak S, Halberg S, McCalla SG, Siahpirani AF, Sridharan R, Roy S (2023) Inference of cell type-specific gene regulatory networks on cell lineages from single cell omic datasets. Nat Commun 14:3064
13. Godoy P, Schmidt-Heck W, Natarajan K, Lucendo-Villarin B, Szkolnicka D, Asplund A, Björquist P, Widera A, Stöber R, Campos G, Hammad S, Sachinidis A, Chaudhari U, Damm G, Weiss TS, Nüssler A, Synnergren J, Edlund K, Küppers-Munther B, Hay DC, Hengstler JG (2015) Gene networks and transcription factor motifs defining the differentiation of stem cells into hepatocyte-like cells. J Hepatol 63:934–942
14. Gérard C, Tys J, Lemaigre FP (2017) Gene regulatory networks in differentiation and direct reprogramming of hepatic cells. Semin Cell Dev Biol 66:43–50
15. Velazquez JJ, LeGraw R, Moghadam F, Tan Y, Kilbourne J, Maggiore JC, Hislop J, Liu S, Cats D, de Chuva Sousa Lopes SM, Plaisier C, Cahan P, Kiani S, Ebrahimkhani MR (2021) Gene regulatory network analysis and engineering directs development and vascularization of multilineage human liver organoids. Cell Syst. 12:41-55.e11
16. Kim DS, Ryu JW, Son MY, Oh JH, Chung KS, Lee S, Lee JJ, Ahn JH, Min JS, Ahn J, Kang HM, Kim J, Jung CR, Kim NS, Cho HS (2017) A liver-specific gene expression panel predicts the differentiation status of in vitro hepatocyte models. Hepatology 66:1662–1674
17. Buganim Y, Faddah DA, Jaenisch R (2013) Mechanisms and models of somatic cell reprogramming. Nat Rev Genet 14:427–439
18. Du Y, Wang J, Jia J, Song N, Xiang C, Xu J, Hou Z, Su X, Liu B, Jiang T, Zhao D, Sun Y, Shu J, Guo Q, Yin M, Sun D, Lu S, Shi Y, Deng H (2014) Human hepatocytes with drug metabolic function induced from fibroblasts by lineage reprogramming. Cell Stem Cell 14:394–403
19. Huang P, He Z, Ji S, Sun H, Xiang D, Liu C, Hu Y, Wang X, Hui L (2011) Induction of functional hepatocyte-like cells from mouse fibroblasts by defined factors. Nature 475:386–389
20. Huang P, Zhang L, Gao Y, He Z, Yao D, Wu Z, Cen J, Chen X, Liu C, Hu Y, Lai D, Hu Z, Chen L, Zhang Y, Cheng X, Ma X, Pan G, Wang X, Hui L (2014) Direct reprogramming of human fibroblasts to functional and expandable hepatocytes. Cell Stem Cell 14:370–384
21. Sekiya S, Suzuki A (2011) Direct conversion of mouse fibroblasts to hepatocyte-like cells by defined factors. Nature 475:390–393
22. Butler A, Hoffman P, Smibert P, Papalexi E, Satija R (2018) Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nat Biotechnol 36:411–420
23. Li B, Hon GC (2021) Single-cell genomics: catalyst for cell fate engineering. Front Bioeng Biotechnol 9:748942
24. Guo L, Lin L, Wang X, Gao M, Cao S, Mai Y, Wu F, Kuang J, Liu H, Yang J, Chu S, Song H, Li D, Liu Y, Wu K, Liu J, Wang J, Pan G, Hutchins AP, Liu J, Pei D, Chen J (2019) Resolving cell fate decisions during somatic cell reprogramming by single-Cell RNA-seq. Mol Cell 73:815-829.e7
25. Jindal K, Adil MT, Yamaguchi N, Yang X, Wang HC, Kamimoto K, Rivera-Gonzalez GC, Morris SA (2024) Single-cell lineage capture across genomic modalities with cell tag-multi reveals fate-specific gene regulatory changes. Nat Biotechnol 42:946–959
26. Popescu DM, Botting RA, Stephenson E, Green K, Webb S, Jardine L, Calderbank EF, Polanski K, Goh I, Efremova M, Acres M, Maunder D, Vegh P, Gitton Y, Park JE, Vento-Tormo R, Miao Z, Dixon D, Rowell R, McDonald D, Fletcher J, Poyner E, Reynolds G, Mather M, Moldovan C, Mamanova L, Greig F, Young MD, Meyer KB, Lisgo S, Bacardit J, Fuller A, Millar B, Innes B, Lindsay S, Stubbington MJT, Kowalczyk MS, Li B, Ashenberg O, Tabaka M, Dionne D, Tickle TL, Slyper M, Rozenblatt-Rosen O, Filby A, Carey P, Villani AC, Roy A, Regev A, Chédotal A, Roberts I, Göttgens B, Behjati S, Laurenti E, Teichmann SA, Haniffa M (2019) Decoding human fetal liver haematopoiesis. Nature 574:365–371
27. Wesley BT, Ross ADB, Muraro D, Miao Z, Saxton S, Tomaz RA, Morell CM, Ridley K, Zacharis ED, Petrus-Reurer S, Kraiczy J, Mahbubani KT, Brown S, Garcia-Bernardo J, Alsinet C, Gaffney D, Horsfall D, Tysoe OC, Botting RA, Stephenson E, Popescu DM, MacParland S, Bader G, McGilvray ID, Ortmann D, Sampaziotis F, Saeb-Parsy K, Haniffa M, Stevens KR, Zilbauer M, Teichmann SA, Vallier L (2022) Single-cell atlas of human liver development reveals pathways directing hepatic cell fates. Nat Cell Biol 24:1487–1498
28. Yang L, Wang X, Zheng JX, Xu ZR, Li LC, Xiong YL, Zhou BC, Gao J, Xu CR (2023) Determination of key events in mouse hepatocyte maturation at the single-cell level. Dev Cell 58:1996-2010.e6
29. Ko S, Russell JO, Molina LM, Monga SP (2020) Liver Progenitors and Adult Cell Plasticity in Hepatic Injury and Repair: Knowns and Unknowns. Annu Rev Pathol 15:23–50
30. Miyajima A, Tanaka M, Itoh T (2014) Stem/progenitor cells in liver development, homeostasis, regeneration, and reprogramming. Cell Stem Cell 14:561–574
31. Segal JM, Kent D, Wesche DJ, Ng SS, Serra M, Oulès B, Kar G, Emerton G, Blackford SJI, Darmanis S, Miquel R, Luong TV, Yamamoto R, Bonham A, Jassem W, Heaton N, Vigilante A, King A, Sancho R, Teichmann S, Quake SR, Nakauchi H, Rashid ST (2019) Single cell analysis of human foetal liver captures the transcriptional profile of hepatobiliary hybrid progenitors. Nat Commun 10:3350
32. DeLaForest A, Nagaoka M, Si-Tayeb K, Noto FK, Konopka G, Battle MA, Duncan SA (2011) HNF4A is essential for specification of hepatic progenitors from human pluripotent stem cells. Development 138:4143–4153
33. Parviz F, Matullo C, Garrison WD, Savatski L, Adamson JW, Ning G, Kaestner KH, Rossi JM, Zaret KS, Duncan SA (2003) Hepatocyte nuclear factor 4alpha controls the development of a hepatic epithelium and liver morphogenesis. Nat Genet 34:292–296
34. Jackson JT, Nutt SL, McCormack MP (2023) The Haematopoietically-expressed homeobox transcription factor: roles in development, physiology and disease. Front Immunol 14:1197490
35. Preidis GA, Kim KH, Moore DD (2017) Nutrient-sensing nuclear receptors PPARα and FXR control liver energy balance. J Clin Invest 127:1193–1201
36. Dubois V, Lefebvre P, Staels B, Eeckhoute J (2024) Nuclear receptors: pathophysiological mechanisms and drug targets in liver disease. Gut 73:1562–1569
37. Bai Y, Yang Z, Xu X, Ding W, Qi J, Liu F, Wang X, Zhou B, Zhang W, Zhuang X, Li G, Zhao Y (2023) Direct chemical induction of hepatocyte-like cells with capacity for liver repopulation. Hepatology 77:1550–1565

38. Zhong Z, Du J, Zhu X, Guan L, Hu Y, Zhang P, Wang H (2024) Highly efficient conversion of mouse fibroblasts into functional hepatic cells under chemical induction. J Mol Cell Biol. https://doi.org/10.1093/jmcb/mjad071
39. Li J, Bai Y, Liu Y, Song Z, Yang Y, Zhao Y (2023) Transcriptome-based chemical screens identify CDK8 as a common barrier in multiple cell reprogramming systems. Cell Rep 42:112566
40. Qiu Q, Hernandez JC, Dean AM, Rao PH, Darlington GJ (2011) CD24-positive cells from normal adult mouse liver are hepatocyte progenitor cells. Stem Cells Dev 20:2177–2188
41. Katsuda T, Matsuzaki J, Yamaguchi T, Yamada Y, Prieto-Vila M, Hosaka K, Takeuchi A, Saito Y, Ochiya T (2019) Generation of human hepatic progenitor cells with regenerative and metabolic capacities from primary hepatocytes. Elife. https://doi.org/10.7554/eLife.47313
42. Tanimizu N, Nishikawa M, Saito H, Tsujimura T, Miyajima A (2003) Isolation of hepatoblasts based on the expression of Dlk/Pref-1. J Cell Sci 116:1775–1786
43. Huang J, Zhao X, Wang J, Cheng Y, Wu Q, Wang B, Zhao F, Meng L, Zhang Y, Jin M, Xu H (2019) Distinct roles of Dlk1 isoforms in bi-potential differentiation of hepatic stem cells. Stem Cell Res Ther 10:31
44. Nguyen P, Leray V, Diez M, Serisier S, Le Bloc'h J, Siliart B, Dumon H (2008) Liver lipid metabolism. J Anim Physiol Anim Nutr (Berl) 92:272–283
45. Sheng L, Jiang B, Rui L (2013) Intracellular lipid content is a key intrinsic determinant for hepatocyte viability and metabolic and inflammatory states in mice. Am J Physiol Endocrinol Metab 305:E1115–E1123
46. Enjoji M, Kohjima M, Nakamuta M (2016) Lipid metabolism and the liver. In: Ohira H (ed) The liver in systemic diseases. Springer, Tokyo, pp 105–122
47. Xie H, Li G, Fu Y, Jiang N, Yi S, Kong X, Shi J, Yin S, Peng J, Jiang Y, Lu S, Deng H, Xie B (2024) A two-step strategy to expand primary human hepatocytes in vitro with efficient metabolic and regenerative capacities. Stem Cell Res Ther 15:281
48. Kotulkar M, Robarts DR, Apte U (2023) HNF4α in hepatocyte health and disease. Semin Liver Dis 43:234–244
49. Watt AJ, Garrison WD, Duncan SA (2003) Hnf4: a central regulator of hepatocyte differentiation and function. Hepatology 37:1249–1253
50. Martinez Barbera JP, Clements M, Thomas P, Rodriguez T, Meloy D, Kioussis D, Beddington RS (2000) The homeobox gene Hex is required in definitive endodermal tissues for normal forebrain, liver and thyroid formation. Development 127:2433–2445
51. Zhao Y, Yin X, Qin H, Zhu F, Liu H, Yang W, Zhang Q, Xiang C, Hou P, Song Z, Liu Y, Yong J, Zhang P, Cai J, Liu M, Li H, Li Y, Qu X, Cui K, Zhang W, Xiang T, Wu Y, Zhao Y, Liu C, Yu C, Yuan K, Lou J, Ding M, Deng H (2008) Two supporting factors greatly improve the efficiency of human iPSC generation. Cell Stem Cell 3:475–479
52. Trapnell C, Cacchiarelli D, Grimsby J, Pokharel P, Li S, Morse M, Lennon NJ, Livak KJ, Mikkelsen TS, Rinn JL (2014) The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. Nat Biotechnol 32:381–386
53. Aibar S, Gonzalez-Blas CB, Moerman T, Huynh-Thu VA, Imrichova H, Hulselmans G, Rambow F, Marine JC, Geurts P, Aerts J, van den Oord J, Atak ZK, Wouters J, Aerts S (2017) SCENIC: single-cell regulatory network inference and clustering. Nat Methods 14:1083–1086
54. Cahan P, Li H, Morris SA, da Lummertz Rocha E, Daley GQ, Collins JJ (2014) Cell Net: network biology applied to stem cell engineering. Cell. 158:903–915
55. Yu G, Wang LG, Han Y, He QY (2012) clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS 16:284–287