

In-situ multiomics analysis reveals only minor genetic and epigenetic changes in human liver cancer stem cells

Dan Liu

National Center for Liver Cancer, Eastern Hepatobiliary Surgery Hospital

Hong Li

Bio-Med Big Data Center, CAS Key Laboratory of Computational Biology, CAS-MPG Partner Institute for Computational Biology, Shanghai Institute of Nutrition and Health, Shanghai Institutes for Biological Sciences
<https://orcid.org/0000-0001-5679-9980>

Hui Dong

Eastern Hepatobiliary Surgery Institute/Hospital

Min-Cheng Qu

National Center for Liver Cancer, Eastern Hepatobiliary Surgery Hospital

Liguang Yang

Chinese Academy of Sciences <https://orcid.org/0000-0002-2449-5891>

Li-Na Chen

CAS Key Laboratory of Computational Biology, Shanghai Institute of Nutrition and Health

Yixue Li

Shanghai Institute of Nutrition and Health, Shanghai Institutes for Biological Sciences, University of Chinese Academy of Sciences, Chinese Academy of Sciences <https://orcid.org/0000-0002-1198-7176>

Hongyang Wang

Eastern Hepatobiliary Surgery Hospital, Second Military Medical University <https://orcid.org/0000-0002-4709-3334>

Yufei He (✉ yfhe@sibcb.ac.cn)

National Center for Liver Cancer, Eastern Hepatobiliary Surgery Hospital

Article

Keywords: Cancer stem cells (CSCs), liver cancer stem cells

Posted Date: February 16th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-244209/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Cancer stem cells (CSCs) play pivotal roles in human cancer but their genetic and epigenetic changes are still largely unknown. Here we show that human CSCs are more similar to paratumor cells in liver cancer. Using single-cell-based in-situ multiomics analysis, we found that although liver CSCs (LCSCs) from patients had some potential key genetic and epigenetic changes in their genome, they had substantially fewer somatic single nucleotide variants (SNVs), copy number alterations (CNAs) and differentially methylated regions, and less change in the global levels of DNA cytosine modifications, than other tumor parenchymal cells. The cluster analysis of SNVs, CNAs and DNA methylation patterns and spatial transcriptomes all clearly showed that the LCSCs were clustered with the paratumor liver cells. Thus, only minor genetic and epigenetic changes occurred in human LCSCs. Targeting key changes in CSCs, not just changes in bulk tumor cells, should be more effective for human cancer therapy.

Introduction

Human cancer development is a complex evolutionary process. To date, it is universally acknowledged that cancer cells possess many significant genetic and epigenetic changes caused by multiple factors compared with normal cells. Many cancers are thought to originate from cancer stem cells (CSCs), which pose a high risk of therapy resistance and cancer relapse¹. Understanding the genetic and epigenetic changes in human CSCs should shed light on the design of new better cancer therapeutic approaches. However, it is always difficult to study human CSCs, as they generally account for a very small proportion of tumor specimens from clinical patients. Consequently, the genetic and epigenetic changes that occur in human CSCs and whether these changes are the same as those in other tumor parenchymal cells of the same clinical tumor are still poorly understood.

In the current study, we developed a method combining immunohistochemistry (IHC), laser capture microdissection (LCM) and single-cell sequencing techniques (CIL-seq) to uncover the genetic and DNA methylation changes related to human CSCs. We first selected specimens with a stem cell phenotype from mixed hepatocellular cholangiocarcinoma (HCC-CCA) patients to investigate liver CSCs (LCSCs). We found that LCSCs had only minor genetic and epigenetic changes and were more similar to paratumor tissues than to tumor tissues in both genetic and epigenetic patterns. We further confirmed these results in many other liver cancer samples using in-situ multiomics analysis.

Results

Development of CIL-seq method

We selected epithelial cell adhesion molecule (EpCAM), which was used to label liver stem cells^{2,3} or LCSCs⁴⁻⁶, as a marker molecule for liver stem cells in our study. To explore genetic and epigenetic differences between LCSCs and other EpCAM-stained cells in both tumor and paratumor tissues, we developed the CIL-seq method, which combines IHC, LCM and single-cell sequencing techniques (Fig. 1a

and Methods). First, to better obtain LCSCs for further study, we carried out EpCAM IHC staining on frozen sections. Second, we used LCM to capture needed cells more accurately and visually in multiple spatial regions, which is helpful for the investigation of tumor biology and heterogeneity^{7,8}. Then, to test whether the IHC process damages the integrity of genomic DNA, we selected approximately 40 pure HCC or paratumor liver (PL) cells, which were clustered together within the same component in the tumor or paratumor, for whole genome sequencing (WGS) library analysis based on single-cell technology. Simultaneous sequencing of several same-type cells could reduce the inherent single-base false positives caused by whole-genome amplification of single cells as much as possible. Last, the WGS results showed satisfactory coverage; at a 15× mean sequencing depth, WGS covered approximately 78% to 80% of the genome at $\geq 1\times$ coverage (Supplementary Table 1), confirming the integrity of genomic DNA.

Characteristics of LCSCs in liver cancer patients

Mixed HCC-CCA is a rare type (less than 5%) of primary liver cancer (PLC) with mixed differentiation (Supplementary Fig. 1a), and its diagnosis depends on postoperative histopathology⁹. Some cancer cells in mixed HCC-CCAs have stem cell characteristics^{9,10}. Therefore, this type of PLC can provide us with a perfect tumor object directly from clinical patients to help investigate human CSCs and compare the similarities and differences between LCSCs and HCC or CCA cells within the same liver cancer sample. We collected liver cancer samples that were resected and diagnosed with PLC before surgery. According to the clinicopathological diagnosis, one mixed HCC-CCA case (patient P2, Supplementary Table 2) among a total of 35 cases (Supplementary Table 2) collected was identified. Fortunately, this mixed HCC-CCA specimen is the stem-cell feature type, which is the exact type of specimen that we aimed to research. Pathological histocytology (Supplementary Fig. 1b) showed three distinct types of parenchymal cells coexisting in the tumor tissues: HCC cells, CCA cells, and LCSCs, which express liver stem cell markers, including EpCAM, cytokeratin 19 (CK19), OV6 and Sal-like protein 4 (SALL4), and account for approximately 10% of the overall tumor tissues. Some LCSCs grew in clonal clusters, and the cell morphology within each clone was not distinctly different under the microscope (Supplementary Fig. 1b).

By integrating EpCAM staining and cell morphology, the liver parenchymal cells in the paratumor tissue could be further subdivided into two types: the most common type was EpCAM-negative cells, which are ordinary PL cells, and the other type was EpCAM-positive cells, which correspond to paratumor ductular reaction (PDR) cells (Supplementary Fig. 1b). In PDR cells, each clone is generally believed to contain a portion of liver stem cells¹¹. Therefore, EpCAM immunostaining combined with cell morphology features could clearly distinguish the two types of liver parenchymal cells in paratumor tissues and the three types of parenchymal cells in tumor tissues. At the junction between LCSCs and HCC cells or CCA cells, some transitional areas could be clearly observed (Fig. 1b), supporting that HCC and CCA cells may be derived from LCSCs.

LCSCs and PDR cells had similar characteristic markers, such as EpCAM, CK19, OV6, and SALL4 (Supplementary Fig. 1b). Therefore, LCSCs closely resembled PDR cells in both histomorphology and key markers of liver stem cells. We investigated the pathological sections of 12603 PLC patients who received

surgical resection from 2017 to 2019 at our hospital. Among them, 278 cases (2.2%) were pathologically diagnosed as HCC-CCA and 58 cases of them (Supplementary Table 2) could be observed such LCSCs as in patient P2. We also investigated 265 HCC patients and found that 35 cases of HCC (Supplementary Table 2) could be observed such LCSCs, whereas the number of LCSC clones was much less than that in HCC-CCA. Notably, however, LCSCs had significantly more Ki-67-positive staining than PDR cells (Fig. 1c and Supplementary Fig. 1b), indicating that LCSCs have greater proliferative activity.

Whole-exome sequencing (WES) analysis of LCSCs reveals only minor key genetic changes

We performed a WES analysis of specimen P2 based on CIL-seq. According to the aforementioned EpCAM staining, we isolated 5 types of cell samples (Fig. 1d and Supplementary Fig. 2a, b) from frozen sections, including three components in the tumor: LCSCs (3 samples), CCA cells (4 samples) and HCC cells (2 samples), and two components in the paratumor: PDR cells (3 samples) and PL cells (4 samples). In addition, we used bulk immune infiltrating cells (~ 500 cells) from the paratumor tissue (Fig. 1d) in this patient as the normal genomic reference. Taken together, WES libraries were prepared from a total of 17 samples and sequenced. One of the 17 samples, E3A, was sequenced repeatedly, and the results showed good sequencing concordance (Supplementary Fig. 2c). At least two samples of each type had good coverage (average target depth > 62×) (Supplementary Table 1).

We utilized two algorithms to identify somatic single nucleotide variants (SNVs) and removed false positives by commonly used filters (Methods). Consequently, we detected 125 SNVs in all samples, and Sanger sequencing further confirmed the accuracy of SNV detection (Supplementary Table 3). They were predominantly C to T transitions (Fig. 2a), similar to those previously reported¹²⁻¹⁴. Through cluster analysis of the 125 SNVs, it was clearly seen that different samples from the same cell types could be clustered together. Intriguingly, the LCSC samples were clustered with the paratumor samples instead of the HCC or CCA samples (Fig. 2b).

Most SNVs were detected in HCC or CCA samples (73/125 and 70/125, respectively, Supplementary Table 3), and the number of SNVs was consistent with that of previous findings in HCC or mixed HCC-CCA samples¹⁵⁻¹⁸. Notably, only minor SNVs (28 in total) were detected in LCSCs. There were considerably more independent SNVs among LCSC samples (Fig. 2c, Supplementary Fig. 3a and Supplementary Table 3), suggesting that LCSCs harbor strong heterogeneity. HCC and CCA cells shared 52 SNVs (Fig. 2d and Supplementary Table 3), indicating a monoclonal origin, which is in agreement with results from previous reports on the monoclonal origin of mixed HCC-CCA^{17,18}. More than half of the mutations in LCSCs (17/28) coincided with those in HCC and CCA cells (Fig. 2d and Supplementary Table 3), supporting the same origin of LCSCs and HCC or CCA cells. Previous reports showed extensive SNVs in normal or cirrhotic liver cells^{14,19}. Consistent with these findings, we found many SNVs (43 in total, Supplementary Table 3) in PL cells. However, the number of SNVs in PL cells was much greater than that in LCSCs, and less overlapped with that in LCSCs (Fig. 2d and Supplementary Table 3), suggesting that LCSCs are less likely to originate from PL cells.

Of the 69 genes with missense mutations or stop codon changes, only a small portion of these SNVs were detected in LCSCs (Fig. 2e and Supplementary Table 3). We found that three genes (*DHX9*, *ERBB4*, *FAM46D*) were among the 299 reported driver genes²⁰, and one (*ZFH4*) was among the 161 driver genes of HCC¹⁵. These driver gene mutations were detected only in HCC or CCA cells (Supplementary Table 3). Taken together, these results suggest that crucial gene mutations that may promote tumor growth occurred in HCC and CCA cells; however, substantially fewer mutations occurred in LCSCs.

We next analyzed the copy number alteration (CNA) using the WES data. Our analysis focused only on the amplified loci because of the low coverage of some samples. We found many CNA amplification regions in the tumor, but most of them were centralized in HCC and CCA cells, including the amplification of chromosome 8q, which is often detected in liver cancer^{8,18,21}, while only very few CNA regions were detected in LCSCs (Fig. 2f and Supplementary Table 4). We also analyzed the amplified genes. Overall, consistent with the CNA amplification regions, at the gene level, LCSCs also had apparently fewer amplification genes than HCC and CCA cells (Supplementary Fig. 3b and Supplementary Table 4). A phylogenetic tree constructed using the amplified genes clearly showed that similar to SNVs, LCSCs were more similar to paratumor cells than to HCC and CCA cells (Fig. 2g), and this analysis also showed that HCC and CCA cells might be derived from LCSCs, which is consistent with the aforementioned histological characteristics between these cells (Fig. 1b). Interestingly, some of these amplified genes are oncogenes, such as *MYC*, which was also amplified in one sample of LCSCs (Supplementary Fig. 4 and Supplementary Table 4). Activation of *MYC* is thought to be related to the formation of CSCs in the liver^{12,22}. Taken together, WES data showed that LCSCs have only a few SNVs and CNAs compared to those in HCC and CCA cells, but LCSCs might possess a few crucial changes, such as amplification of the oncogene *MYC*.

Whole-genome bisulfite sequencing (WGBS) analysis of LCSCs reveals minor key epigenetic changes

Epigenetic aberrations play an important role in cancer development^{7,23,24}, and our WES data showed few striking genetic changes in LCSCs. Thus, we are curious about what epigenetic changes hidden in LCSCs changed them to malignant cells instead of normal cells. Considering the actual situation of the samples we picked, we mainly examined the DNA methylation landscape of these samples. DNA methylation is the most essential DNA modification and is closely associated with other epigenetic contents²⁵. We performed WGBS analysis of captured different cell clones (Fig. 1d and Supplementary Fig. 2) from specimen P2 using CIL-seq. The bisulfite conversion efficiency of most samples was greater than 98%, and CpG coverage in most samples (Supplementary Table 1) was higher than that reported for single-cell WGBS sequencing²⁶⁻²⁸.

In general, HCC and CCA cells have undergone broad DNA demethylation compared with paratumor cells (Fig. 3a). However, the global methylation level of LCSCs changed much less (Fig. 3a), indicating a closer relationship to paratumor cells. In addition, there was extensive DNA methylation heterogeneity in tumor cells (Supplementary Fig. 5a), as reported^{16,29-31}. Notably, there was also strong DNA methylation heterogeneity among LCSC samples (Supplementary Fig. 5b). The cluster analysis of the DNA

methylation pattern (Fig. 3b) clearly illustrated that HCC and CCA cells were clustered together, and different samples in the paratumor were clustered together. Intriguingly, LCSCs were clustered with paratumor cells. Thus, these results suggest that, consistent with our WES analysis shown above (Fig. 2b, g), for DNA methylation, LCSCs were also more similar to paratumor cells than to HCC and CCA cells.

By comparing the differentially methylated regions (DMRs) of distinct sample types, we found that LCSCs displayed limited variance with paratumor cells but had greater variance with HCC and CCA cells, which is consistent with the global DNA methylation pattern (Fig. 3b). Specifically, 10.1% DMRs were found between LCSCs and PDR cells, while 27.1% or 32.3% DMRs were found between LCSCs and HCC or CCA cells (Fig. 3c and Supplementary Table 5).

Formalin-fixed, paraffin-embedded (FFPE) sample resources are far richer than frozen tissues. We chose an FFPE HCC-CCA sample (P4, Supplementary Table 2) and conducted WGBS analysis based on the CIL-seq method (Fig. 3d). Although the quality of the sequencing data was much lower than that of frozen tissues (Supplementary Table 1), we still observed that the overall methylation level and DNA methylation pattern of LCSCs were similar to those of PDR cells (Fig. 3e, f), consistent with the results of patient P2. Taken together, these results suggest that in the global pattern of DNA methylation, very large methylation changes occurred in HCC and CCA cells compared with paratumor cells, whereas there were fewer differences in DNA methylation patterns between LCSCs and paratumor cells.

DNA methylation changes in LCSCs correspond to a chromatin restriction state

We next examined the distribution region of DMRs in the genome of P2 samples. Overall, there were significantly fewer regions with **prominent** DNA methylation changes in LCSCs than in HCC and CCA cells (Supplementary Table 6). According to the 15-state model of chromatin²⁵, the significant methylation changes in HCC and CCA cells were closely related to the methylation level of the 15 chromatin states in normal cells. Chromatin states with significant demethylation occurred on those initially exhibiting a hypermethylated state in normal cells²⁵, including the 9_Het and 15_Quies chromatin states, and chromatin states with significant hypermethylation occurred on those initially exhibiting a hypomethylated state in normal cells²⁵, including the 1_TssA, 10_TssBiv, 11_BivFlnk and 12_EnhBiv states (Fig. 4a, b and Supplementary Fig. 5c). However, substantially fewer methylation changes occurred in these chromatin states in LCSCs, which is consistent with the minor change in the global methylation level in LCSCs (Fig. 3a).

Remarkably, we found that a few crucial chromatin regions in LCSCs had the same DNA methylation tendency as in HCC and CCA cells. Significant hypermethylation changes in repressive Polycomb group (PcG) protein-marked regions (including the chromatin states 10_TssBiv, 11_BivFlnk, 12_EnhBiv and 13_ReprPC) occurred in HCC and/or CCA cells (Fig. 4a, b), which is consistent with previous findings that PcG-marked genes in both embryonic and adult progenitor systems have a higher chance of becoming hypermethylated in cancers³². In addition, we also found that two of these PcG-marked regions had already undergone significant changes in LCSCs compared with PDR cells, including in the bivalent state

10_TssBiv and the chromatin state 13_ReprPC (Fig. 4b). The functions of those genes corresponding to these hypermethylated regions in LCSCs are mainly involved in cell differentiation and fate determination (Fig. 4c and Supplementary Table 7), and more than half of them (20/32) also correspond to bivalent genes in the human fetal liver³³, such as APC2 (Supplementary Table 7), a homologue of the APC tumor suppressor³⁴. Therefore, hypermethylation changes in these regions in LCSCs might result in a restrictive chromatin state that can block a differentiation program²⁴. Taken together, these results suggest that although there were only a few DNA methylation changes in LCSCs, which were much less than the significant changes observed in HCC and CCA cells, the changes in LCSCs might result in a restrictive chromatin state, which may be important for cancer initiation.

Minor genetic and epigenetic changes in LCSCs of HCC patients

HCC is a major type of PLC. We collected frozen tissues and prepared sections from 23 PLC cases (Supplementary Table 2). One of them (patient P7) was pathologically diagnosed with HCC and had obvious EpCAM+ LCSCs. We performed WES and WGBS analysis of this case with CIL-seq (Fig. 5a and Supplementary Table 1). The results showed that the genetic (SNV and CNV) difference or DNA methylation pattern difference between LCSCs and PDR cells was small (Fig. 5b-f), consistent with the analysis results of patients P2 and P4. Interestingly, among the minor genetic changes in LCSCs, we found a stop gain mutation in the driver gene *MGA*²⁰, which also appeared in an HCC sample (Supplementary Table 3). *MGA* is significantly mutated in lung adenocarcinoma and participates in the negative regulation of *MYC*³⁵.

An important epigenetic change during cancer development is the global demethylation of DNA. We and others have found that 5-methylcytosine (5mC) can be oxidized to the demethylation intermediates 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC) in mammalian cells^{36,37}. Recently, we found that in the early stage of HCC, the global content of 5hmC and 5fC was decreased³⁸, but it is not clear whether there is such a change in LCSCs. We immunostained the FFPE section samples of P7 as well as P2 and P4 for 5mC, 5hmC and 5fC (Supplementary Fig. 6). The results showed that most LCSCs were different from other tumor parenchymal cells which had obviously decreased 5hmC and 5fC staining. There was little difference in 5mC, 5hmC or 5fC staining between LCSCs and PDR cells. Examination of 65 samples with obvious LCSCs (Supplementary Table 3), including those from 35 HCC patients, showed similar results (Fig. 5g, h), suggesting that LCSCs do not undergo extensive DNA demethylation. Taken together, these results are consistent with the results of WGBS analysis showing that minor DNA methylation changes occur in LCSCs.

Finally, we aimed to determine whether the RNA expression profiles of LCSCs are similar to those of paratumor liver cells, since the genetic and epigenetic changes in LCSCs are not significant. We performed a spatial transcriptomic (ST) analysis of P7 frozen tissues. We generated 4493 spot transcriptomes at a medium depth of 14455 UMIS/spot and 4064 genes/spot (Supplementary Fig. 7a), which can clearly distinguish tumor and paratumor cells (Supplementary Fig. 7b). In general, there were large RNA expression differences between the tumor and paratumor cells (Supplementary Fig. 7c).

Further cluster analysis showed that the tumor cells had extensive heterogeneity (Fig. 5i). Because of the small proportion of LCSCs, we manually selected all LCSC spots in the tumor tissues and PDR cell spots in the paratumor tissues (Supplementary Fig. 7d, e). In addition, we randomly selected some spots of tumor or paratumor parenchymal cells. Cluster analysis of the whole expression profiles showed that LCSCs and PDR cells were clustered together (Supplementary Fig. 7f). The cluster analysis of the expression profiles of 11 liver-related genes of these spots showed similar results (Fig. 5j), indicating that there was only a small difference in RNA expression patterns between LCSCs and PDR cells, consistent with the results of the genetic and epigenetic changes.

Discussion

In the present study, using single-cell-based in-situ multiomics analysis, we provide evidence that only a few genetic and epigenetic changes occur in human LCSCs. Our data showed that HCC and CCA cells have undergone considerable genetic and epigenetic changes that we usually see in human cancers, but only minor genetic and epigenetic changes have occurred in LCSCs.

Because of the possible great genetic and epigenetic differences between CSCs and other tumor parenchymal cells, the detection of genetic and epigenetic changes in whole tumor tissue could not reflect what has happened in CSCs. The sequencing signals from bulk tumor samples would be dominated by major tumor parenchymal cells, rendering rare CSCs undetectable. Therefore, treatments that target most of the significant genetic and epigenetic changes (and hence the changes in RNA level and protein level) in these tumors may not be effective for CSCs because CSCs may not possess those changes.

The LCSCs in the patients in this study harbor CSC characteristics, including the expression of stem cell markers and a higher proliferative capacity than PDR cells, and histopathological and genetic evidence suggests that LCSCs may be the common origin of HCC and CCA cells. Nevertheless, LCSCs have significant genetic and epigenetic differences compared with HCC and CCA cells in the same patient. The proportion of CSCs in some mixed HCC-CCAs is usually much higher than that in other types of tumors; therefore, such mixed HCC-CCAs can provide us with an ideal tumor object to help investigate human CSCs. Furthermore, because LCSCs in mixed HCC-CCAs can differentiate into two kinds of tumor parenchymal cells, it is easier to identify them.

The CIL-seq method that we developed in this study, which combines IHC, LCM and single-cell sequencing techniques, is suitable for the investigation of rare CSCs, especially when spatial heterogeneity is taken into consideration. To obtain monoclonal CSCs, in-situ capture, such as LCM, is a direct and effective method. Information on cell spatial positions could not be obtained if using methods such as cell sorting. Because of the breakage of DNA in FFPE tissue cells, such samples are often not suitable for single-cell based genetic and epigenetic sequencing research. Frozen sections are much better, as illustrated in this study.

In conclusion, we found only minor genetic and epigenetic changes in CSCs in liver cancer patients. Further studies are needed to determine whether such changes are common in CSCs in patients with other types of cancer. Identifying few but key genetic and epigenetic changes in both CSCs and other tumor parenchymal cells of each specific cancer patient, such as *MYC* amplification in patient P2 or *MGA* mutation in patient P7 in this study, could help design more effective methods to treat human cancer in the future.

Methods

Patients and collection of clinical specimens

We investigated pathological sections of 12603 PLC patients who received surgical resection from 2017 to 2019 at the Eastern Hepatobiliary Surgery Hospital (EHBH) in Shanghai, China. In addition, 265 HCC patients who underwent surgical resection from 2009 to 2020 at EHBH were randomly selected and investigated for LCSCs (Supplementary Table 2).

Fresh primary liver cancer tissues and paratumor liver tissues were obtained from 58 patients who underwent surgical resection at EHBH (Supplementary Table 2). Each tissue specimen was embedded in optimal cutting temperature (OCT) compound medium, immediately frozen in an isopentane slurry made with liquid nitrogen, and finally stored at -80 °C until further processing. Each sample was embedded within 30 minutes for frozen sectioning after surgical removal.

The archives of all patients were collected by the EHBH archive system. Informed consent was obtained from the patients, and all procedures were approved by the ethical committee of EHBH.

Immunohistochemistry (IHC)

For frozen section IHC, briefly, OCT-embedded frozen tissues were cut into 8 µm thick sections, and endogenous peroxidases were inactivated by incubation in 0.3% H₂O₂. Nonspecific signals were blocked using 5% bovine serum albumin. The primary antibodies used were as follows: anti-EpCAM (1:500, Abcam, ab7504), anti-OV6 (1:200, R&D Systems, MAB2020) and anti-GPC3 (1:100, Abcam, ab207080). HRP-conjugated antibodies were used as the secondary antibodies. Diaminobenzidine colorimetric reagent solution was used for staining followed by hematoxylin counterstaining. The slides were finally scanned, and representative images were illustrated.

For formalin-fixed, paraffin-embedded (FFPE) IHC, the tissues were fixed overnight in 4% PFA, embedded in paraffin and cut into 4 µm serial consecutive sections. After deparaffinization, antigen retrieval was performed with sodium citrate buffer (10 mM sodium citrate, 0.05% Tween-20, pH 6) (for H&E, omit this step, subsequently by hematoxylin and eosin standard protocols). The following steps were the same as those in the aforementioned frozen section. The primary antibodies used here were as follows: anti-SALL4 (1:100, Santa Cruz Biotechnology, sc-101147), anti-CK19 (1:100, Abcam, ab9221), and anti-Ki-67 (1:1000, Abcam, ab15580).

For 5mC, 5hmC or 5fC immunostaining, DNA was denatured with 2 N HCl for 15 minutes at room temperature after antigen retrieval treatment, followed by neutralization with 100 mM Tris-HCl (pH 8.0) for 10 minutes at room temperature. The primary antibodies used were as follows: anti-5mC (1:5000, Cell Signaling Technology, Cat#: 28692), anti-5hmC (1:10000, Active Motif, Cat#: 39769); and anti-5fC (1:1000, Active Motif, Cat#: 61227). For EpCAM and 5hmC double-staining, primary anti-5hmC antibody (rabbit) was first incubated for 16 hours at 4 °C, followed by alkaline phosphatase-based streptavidin/biotinylated link system to visualize dark purple cell nuclei. Anti-EpCAM antibody (mouse) was then incubated for 2 hours at room temperature, followed by detection using a horseradish peroxidase/AEC system to visualize the red cell membrane. Specifically, the slides were sealed with glycerol mounting medium and without hematoxylin counterstaining. We could not obtain good results for EpCAM and 5fC double staining, possibly because of the low 5fC content in cells and thus usually weak staining for 5fC.

CIL-seq

IHC and LCM

OCT-embedded frozen tissues were cut into 10 µm thick sections and mounted on PEN Membrane Glass Slides (Leica). After IHC of EpCAM (see above), tissues were dehydrated through rising ethanol concentrations (50, 75, 85, 95, 100 and 100% ethanol, 60 s each) in 50 mL sterile centrifuge tubes. When membrane slides dried completely, the LCM procedure was initiated by a Leica LMD 7000 laser microdissection system according to the manufacturer's protocol. All captured samples were collected in 0.2 ml PCR tubes. Finally, the membrane slides were cleared with xylene and sealed for long-term preservation. Certified RNase/DNase-free materials were used whenever available.

Genetic libraries

Ultralow DNA from LCM samples was extracted by a Quick-DNA Microprep Plus Kit (Zymo Research) before single-cell multiple displacement amplification (MDA) using a REPLI-g Single Cell Kit (Qiagen) for trace DNA amplification. After obtaining high yields of high-quality whole-genome amplified DNA, whole-exome capture and library construction were carried out by an Agilent SureSelect Human All Exon V6 kit (Agilent) according to the manufacturer's instructions. Whole-genome libraries were constructed by the QIAseq FX Single Cell DNA Library kit (Qiagen).

Whole-genome bisulfite sequencing (WGBS) libraries

Tissue samples from LCM were digested for up to 4 hours with 1 mg/ml proteinase K according to the EZ DNA Methylation-Direct Kit (Zymo Research). After thorough digestion, LCM cells were immediately sent to the Pico Methyl-Seq Library Prep Kit (Zymo Research) to accommodate ultralow DNA input according to the manufacturer's instructions with minor adjustment: bisulfite conversion time was extended to 90 minutes to ensure complete conversion; for the sample M1BULK, the library only needed 6 cycles in Section 4, while other libraries needed 11 cycles.

Sequencing

All libraries were sequenced with the Illumina Xten platform to generate 150 bp paired-end reads.

Whole-genome sequencing (WGS) and whole-exome sequencing (WES) data analysis

Data quality and filtering

Raw reads were filtered with Trimmomatic to remove adaptor sequences and low-quality bases³⁹. The remaining reads were aligned to the human genome (hg38) by the BWA algorithm⁴⁰. After mapping, the duplicate reads were removed. Germline variants were called using GATK⁴¹.

Somatic mutations

Taking bulk immune infiltrating cells as a control, somatic mutations were identified from BAM files using muTect and varScan^{42,43}. High-quality somatic single nucleotide variants (SNVs) were obtained by applying the following filters: 1) Removal of potential germline variants that are recorded in dbSNP or variants with allele frequency larger than 1% in population^{44,45}; 2) Removal of variants located within 10 bp distance, which are more possible to be errors introduced during library preparation; 3) Removal of mutations that are identified by only one algorithm; 4) Depth of the mutation sites should be larger than 10 in sample and control sample; 5) There are at least 2 reads supporting the alternative allele and the allele frequency is larger than 5% in sample, and no reads have the alternative allele in control sample; 6) Since there are more samples in P2 than in P7, only mutations that occur in more than one of P2 samples are analyzed to improve quality. This filter is not suitable for P7.

Functional consequences of somatic SNVs were annotated by ANNOVAR⁴⁶. Somatic SNVs of all samples were converted to a 0-1 matrix (M), in which 1 indicates having a mutation. To reduce the effects of allele dropout, when site i had low coverage ($\leq 5\times$) in the sample, M_{ij} was adjusted based only on other samples of the same pathology group. M_{ij} if more samples carry the mutation or M_{ij} if more samples are wild type. Euclidean distances between paired samples were calculated from M and hierarchical clustering was performed to cluster samples.

Copy number alterations

Copy number alterations (CNAs) were identified from WES data by CNVkit⁴⁷. Samples covering $< 50\%$ target regions were removed from the CNA analysis. Deletions were ignored to avoid the effects of uncovered regions. Segments and genes with $\log_2(\text{CN}/2) \geq 1$ were regarded as amplifications. To decrease false positives of amplified genes, we only analyzed genes that were amplified in $\geq 5\%$ TCGA liver cancer patients (<https://www.cancer.gov/tcga>) and genes with three more amplifications in tumor samples than in paratumor samples. Euclidean distances were calculated from amplified genes, and hierarchical clustering was used to plot the clustering tree.

WGBS data analysis

Data quality and filtering

Raw reads were filtered with Trimmomatic to remove adaptor sequences and low-quality bases³⁹. The remaining reads were aligned to the human genome (hg38) by Bismark⁴⁸. Potential PCR duplicates were removed. The bisulfite conversion rate was estimated as the ratio between the number of methylated non-CpGs and total non-CpG sites.

Data processing

The R package methylKit was used to generate and analyze the methylation matrix⁴⁹. Bases whose coverage was larger than 500× or lower than 3× were discarded. The genome was split into 300-bp windows with 300-bp step size for fresh-frozen samples. For FFPE samples, the genome was split into 1 Mb windows with a 1 Mb step size. Windows that were covered by less than 3 bases were discarded. Similarity between paired samples was measured by Pearson correlation coefficients. Hierarchical clustering was used to explore the clusters of samples. For differential methylation analysis between two groups, the chi-square test was used to compare windows covered by at least two samples in each group. Differentially methylated regions (DMRs) were windows that satisfied Q -value < 0.05 and absolute difference > 0.25 . DMRs were annotated with genomic contexts and 15 chromatin states of embryonic stem cell line H9 (downloaded from Roadmap Epigenomics Project)²⁵. Annotation enrichment analysis was performed by Fisher's exact test. Annotations with Q -values < 0.01 were regarded as significantly enriched.

Gene Ontology (GO) enrichment analysis

The Metascape web-based tool⁵⁰ was used for GO analysis of the genes of interest. The Metascape analysis was performed using the default settings. A value of $P < 0.05$ was considered statistically significant.

DNA modification analysis

For statistical analysis of 5mC, 5hmC and 5fC staining, we set a standard assessment rule: within one patient, the strongest stained areas were marked with 3 scores. Specifically, strong positive, score = 3; medium positive, score = 2; weak positive, score = 1; no signal, score = 0. Five randomly selected equal fields (400×) of IHC images of each cell type were analyzed.

Spatial transcriptome analysis

ST experiments

Spatial transcriptomics (ST) experiments were performed according to the user guide of Visium Spatial Gene Expression Reagent Kits (10x Genomics). Briefly, HCC tissues were gently washed with cold PBS and filled with OCT in a proper mold and then snap frozen in chilled isopentane. Cryosections were mounted onto a spatially barcoded array of ST with 10- μ m thickness, and serial adjacent cyosections

were mounted onto regular glass slides for IHC staining of EpCAM or other markers with 8- μ m thickness. For processing, the tissue was fixed for 30 min with prechilled methanol at -20 °C, followed by H&E staining. Slides were finally taken on a Leica SCN400 F whole-slide scanner at 40 \times resolution. After capturing ideal tissue morphology information and ensuring that RNA was not degraded (RIN \geq 7), tissue permeabilization and reverse transcription were immediately conducted by a Visium Spatial Tissue Optimization Kit (10x Genomics). Finally, the ST library was prepared with second strand synthesis and denaturation and sequenced by NovaSeq 6000 (Illumina). Each of the spots printed onto the array is 50 μ m in diameter and 100 μ m from center to center, covering an area of 6.5 \times 6.5 mm² (ref. 51).

Data processing and cluster analysis

Raw sequenced ST reads were processed using Space Ranger analysis software (version 1.0.0, 10 \times Genomics) mapped to the GRCH38 genome assembly following the standardized analysis rules. UMI counts in each spot were normalized and scaled by the median number transcript count across all spots. Cluster analysis was based on the K-Means algorithm in Space Ranger software.

Data and materials availability

The raw sequencing data were deposited in the National Omics Data Encyclopedia under accession number OEP000991 (<https://www.biosino.org/node/project/detail/OEP000991>).

Declarations

Acknowledgments

We thank L. Guo for assistance with patient specimen acquisition and D. Cao for experimental help. This work was supported by the National Natural Science Foundation of China (81972644 to Y.H., 91859205, 81830054 and 81988101 to H.W., 31771472 to H.L.) and the National Science and Technology Major Project of China (2018ZX10723204 to Y.H.). H.L. was supported by the CAS Youth Innovation Promotion Association, SA-SIBS Scholarship Program.

Author contributions

Y.H. and H.W. designed and supervised the research work. Y.H. wrote the manuscript with assistance from H.L. and D.L. H.L. performed the bioinformatics analyses with assistance from L.Y., L.C. and Y.L. D.L. developed techniques and performed experiments with assistance from H.D. and M.Q. H.D. screened and analyzed all pathological sections.

Competing interests

The authors declare no potential conflicts of interest.

References

- 1 Lytle, N. K., Barber, A. G. & Reya, T. Stem cell fate in cancer growth, progression and therapy resistance. *Nat Rev Cancer* **18**, 669-680, doi:10.1038/s41568-018-0056-x (2018).
- 2 Huch, M. *et al.* Long-term culture of genome-stable bipotent stem cells from adult human liver. *Cell* **160**, 299-312, doi:10.1016/j.cell.2014.11.050 (2015).
- 3 Aizarani, N. *et al.* A human liver cell atlas reveals heterogeneity and epithelial progenitors. *Nature* **572**, 199-204, doi:10.1038/s41586-019-1373-2 (2019).
- 4 Yamashita, T. *et al.* EpCAM-positive hepatocellular carcinoma cells are tumor-initiating cells with stem/progenitor cell features. *Gastroenterology* **136**, 1012-1024, doi:10.1053/j.gastro.2008.12.004 (2009).
- 5 Yamashita, T. *et al.* Discrete nature of EpCAM+ and CD90+ cancer stem cells in human hepatocellular carcinoma. *Hepatology* **57**, 1484-1497, doi:10.1002/hep.26168 (2013).
- 6 Matsumoto, T. *et al.* Proliferating EpCAM-Positive Ductal Cells in the Inflamed Liver Give Rise to Hepatocellular Carcinoma. *Cancer research* **77**, 6131-6143, doi:10.1158/0008-5472.Can-17-1800 (2017).
- 7 Rozenblatt-Rosen, O. *et al.* The Human Tumor Atlas Network: Charting Tumor Transitions across Space and Time at Single-Cell Resolution. *Cell* **181**, 236-249, doi:10.1016/j.cell.2020.03.053 (2020).
- 8 Zhai, W. *et al.* The spatial organization of intra-tumour heterogeneity and evolutionary trajectories of metastases in hepatocellular carcinoma. *Nat Commun* **8**, 4565, doi:10.1038/ncomms14565 (2017).
- 9 Brunt, E. *et al.* cHCC-CCA: Consensus terminology for primary liver carcinomas with both hepatocytic and cholangiocytic differentiation. *Hepatology* **68**, 113-126, doi:10.1002/hep.29789 (2018).
- 10 Munoz-Garrido, P. & Rodrigues, P. M. The jigsaw of dual hepatocellular-intrahepatic cholangiocarcinoma tumours. *Nat Rev Gastroenterol Hepatol* **16**, 653-655, doi:10.1038/s41575-019-0185-z (2019).
- 11 Sato, K. *et al.* Ductular Reaction in Liver Diseases: Pathological Mechanisms and Translational Significances. *Hepatology* **69**, 420-430, doi:10.1002/hep.30150 (2019).
- 12 Marquardt, J. U., Andersen, J. B. & Thorgeirsson, S. S. Functional and genetic deconstruction of the cellular origin in liver cancer. *Nat. Rev. Cancer* **15**, 653-667, doi:10.1038/nrc4017 (2015).
- 13 Blokzijl, F. *et al.* Tissue-specific mutation accumulation in human adult stem cells during life. *Nature* **538**, 260-264, doi:10.1038/nature19768 (2016).
- 14 Zhu, M. *et al.* Somatic Mutations Increase Hepatic Clonal Fitness and Regeneration in Chronic Liver Disease. *Cell* **177**, 608-621.e612, doi:10.1016/j.cell.2019.03.026 (2019).

- 15 Schulze, K. *et al.* Exome sequencing of hepatocellular carcinomas identifies new mutational signatures and potential therapeutic targets. *Nat Genet* **47**, 505-511, doi:10.1038/ng.3252 (2015).
- 16 Lin, D. C. *et al.* Genomic and Epigenomic Heterogeneity of Hepatocellular Carcinoma. *Cancer research* **77**, 2255-2265, doi:10.1158/0008-5472.Can-16-2822 (2017).
- 17 Wang, A. *et al.* Whole-exome sequencing reveals the origin and evolution of hepatocholangiocarcinoma. *Nat Commun* **9**, 894, doi:10.1038/s41467-018-03276-y (2018).
- 18 Xue, R. *et al.* Genomic and Transcriptomic Profiling of Combined Hepatocellular and Intrahepatic Cholangiocarcinoma Reveals Distinct Molecular Subtypes. *Cancer cell* **35**, 932-947.e938, doi:10.1016/j.ccell.2019.04.007 (2019).
- 19 Brunner, S. F. *et al.* Somatic mutations and clonal dynamics in healthy and cirrhotic human liver. *Nature* **574**, 538-542, doi:10.1038/s41586-019-1670-9 (2019).
- 20 Bailey, M. H. *et al.* Comprehensive Characterization of Cancer Driver Genes and Mutations. *Cell* **173**, 371-385.e318, doi:10.1016/j.cell.2018.02.060 (2018).
- 21 The Cancer Genome Atlas Research Network. Comprehensive and Integrative Genomic Characterization of Hepatocellular Carcinoma. *Cell* **169**, 1327-1341.e1323, doi:10.1016/j.cell.2017.05.046 (2017).
- 22 Chow, E. K., Fan, L. L., Chen, X. & Bishop, J. M. Oncogene-specific formation of chemoresistant murine hepatic cancer stem cells. *Hepatology* **56**, 1331-1341, doi:10.1002/hep.25776 (2012).
- 23 Vicente-Dueñas, C., Hauer, J., Cobaleda, C., Borkhardt, A. & Sánchez-García, I. Epigenetic Priming in Cancer Initiation. *Trends in cancer* **4**, 408-417, doi:10.1016/j.trecan.2018.04.007 (2018).
- 24 Flavahan, W. A., Gaskell, E. & Bernstein, B. E. Epigenetic plasticity and the hallmarks of cancer. *Science (New York, N.Y.)* **357**, doi:10.1126/science.aal2380 (2017).
- 25 Kundaje, A. *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317-330, doi:10.1038/nature14248 (2015).
- 26 Smallwood, S. A. *et al.* Single-cell genome-wide bisulfite sequencing for assessing epigenetic heterogeneity. *Nat Methods* **11**, 817-820, doi:10.1038/nmeth.3035 (2014).
- 27 Gravina, S., Dong, X., Yu, B. & Vijg, J. Single-cell genome-wide bisulfite sequencing uncovers extensive heterogeneity in the mouse liver methylome. *Genome Biol* **17**, 150, doi:10.1186/s13059-016-1011-3 (2016).
- 28 Linker, S. M. *et al.* Combined single-cell profiling of expression and DNA methylation reveals splicing regulation and heterogeneity. *Genome Biol* **20**, 30, doi:10.1186/s13059-019-1644-0 (2019).

- 29 Mazor, T., Pankov, A., Song, J. S. & Costello, J. F. Intratumoral Heterogeneity of the Epigenome. *Cancer cell* **29**, 440-451, doi:10.1016/j.ccell.2016.03.009 (2016).
- 30 Zhang, Q. *et al.* Integrated multiomic analysis reveals comprehensive tumour heterogeneity and novel immunophenotypic classification in hepatocellular carcinomas. *Gut* **68**, 2019-2031, doi:10.1136/gutjnl-2019-318912 (2019).
- 31 Marusyk, A., Janiszewska, M. & Polyak, K. Intratumor Heterogeneity: The Rosetta Stone of Therapy Resistance. *Cancer cell* **37**, 471-484, doi:10.1016/j.ccell.2020.03.007 (2020).
- 32 Easwaran, H. *et al.* A DNA hypermethylation module for the stem/progenitor cell signature of cancer. *Genome Res* **22**, 837-849, doi:10.1101/gr.131169.111 (2012).
- 33 Yan, L. *et al.* Epigenomic Landscape of Human Fetal Brain, Heart, and Liver. *J Biol Chem* **291**, 4386-4398, doi:10.1074/jbc.M115.672931 (2016).
- 34 van Es, J. H. *et al.* Identification of APC2, a homologue of the adenomatous polyposis coli tumour suppressor. *Current biology : CB* **9**, 105-108, doi:10.1016/s0960-9822(99)80024-4 (1999).
- 35 Llabata, P. *et al.* Multi-Omics Analysis Identifies MGA as a Negative Regulator of the MYC Pathway in Lung Adenocarcinoma. *Molecular cancer research : MCR* **18**, 574-584, doi:10.1158/1541-7786.Mcr-19-0657 (2020).
- 36 He, Y. F. *et al.* Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science (New York, N.Y.)* **333**, 1303-1307, doi:10.1126/science.1210944 (2011).
- 37 Xu, G. L. & Bochtler, M. Reversal of nucleobase methylation by dioxygenases. *Nature chemical biology* **16**, 1160-1169, doi:10.1038/s41589-020-00675-5 (2020).
- 38 Liu, J. *et al.* Global DNA 5-Hydroxymethylcytosine and 5-Formylcytosine Contents Are Decreased in the Early Stage of Hepatocellular Carcinoma. *Hepatology* **69**, 196-208, doi:10.1002/hep.30146 (2019).
- 39 Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114-2120, doi:10.1093/bioinformatics/btu170 (2014).
- 40 Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-1760, doi:10.1093/bioinformatics/btp324 (2009).
- 41 Van der Auwera, G. A. *et al.* From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr Protoc Bioinformatics* **43**, 11 10 11-11 10 33, doi:10.1002/0471250953.bi1110s43 (2013).
- 42 Cibulskis, K. *et al.* Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat Biotechnol* **31**, 213-219, doi:10.1038/nbt.2514 (2013).

- 43 Koboldt, D. C. *et al.* VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res* **22**, 568-576, doi:10.1101/gr.129684.111 (2012).
- 44 Sherry, S. T. *et al.* dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res* **29**, 308-311, doi:10.1093/nar/29.1.308 (2001).
- 45 Genomes Project, C. *et al.* A global reference for human genetic variation. *Nature* **526**, 68-74, doi:10.1038/nature15393 (2015).
- 46 Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* **38**, e164, doi:10.1093/nar/gkq603 (2010).
- 47 Talevich, E., Shain, A. H., Botton, T. & Bastian, B. C. CNVkit: Genome-Wide Copy Number Detection and Visualization from Targeted DNA Sequencing. *PLoS Comput Biol* **12**, e1004873, doi:10.1371/journal.pcbi.1004873 (2016).
- 48 Krueger, F. & Andrews, S. R. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* **27**, 1571-1572, doi:10.1093/bioinformatics/btr167 (2011).
- 49 Akalin, A. *et al.* methylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles. *Genome Biol* **13**, R87, doi:10.1186/gb-2012-13-10-r87 (2012).
- 50 Zhou, Y. *et al.* Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat Commun* **10**, 1523, doi:10.1038/s41467-019-09234-6 (2019).
- 51 Ståhl, P. L. *et al.* Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science (New York, N.Y.)* **353**, 78-82, doi:10.1126/science.aaf2403 (2016).

Figures

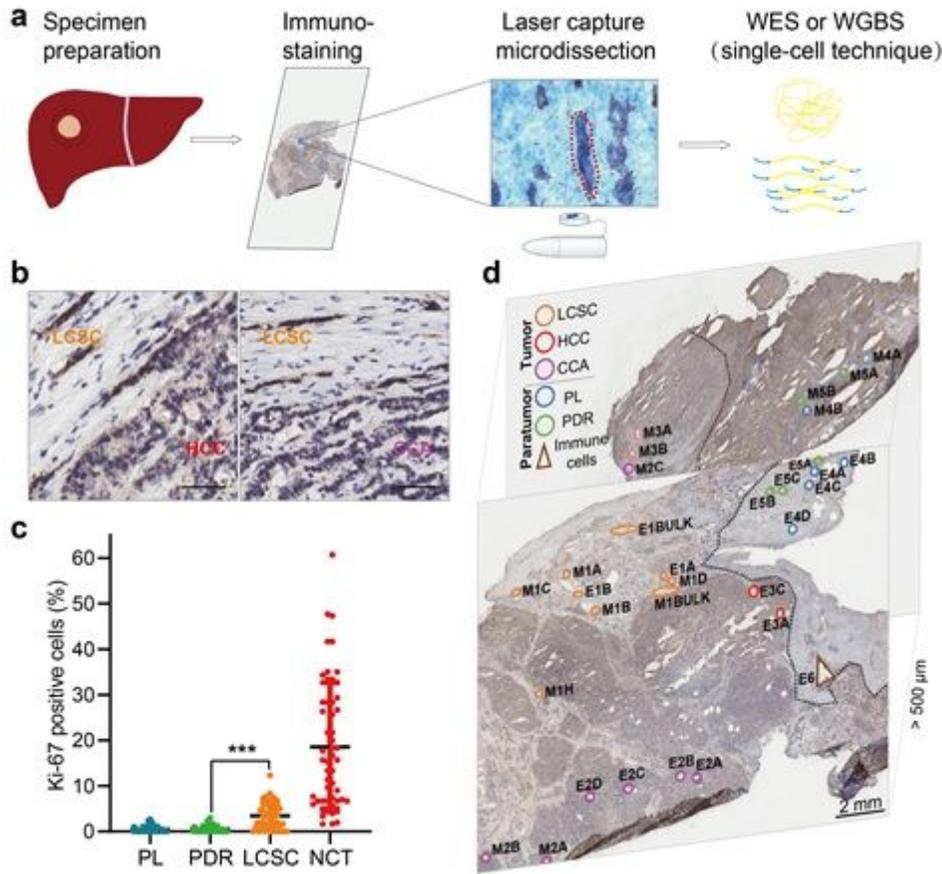


Figure 1

Development of the CIL-seq method and characterization of LCSCs in liver cancer patients. a, Experimental design and the CIL-seq method developed in this study (see also Methods). b, Representative transition regions at the junction of LCSCs and HCC cells or CCA cells (EpCAM immunostaining) in sample P2. Scale bar, 50 μm . c, Dot plot shows the percentages of Ki-67-positive cells in the four cell types of 50 PLC patients with obvious LCSCs (Supplementary Table 2). Each dot represents the mean value of five randomly selected equal fields of IHC images from one patient. NCT: non-CSC tumor cells. Data are presented as the mean \pm SD. ***, $P < 0.001$ by Wilcoxon matched-pairs signed rank test. d, Locations and names of the samples of patient P2 used for WES and WGBS analysis. The samples were collected from two frozen sections that were stained with anti-EpCAM and had a spatial distance greater than 500 μm from each other. The dotted lines indicate the boundaries between the tumor and the paratumor.

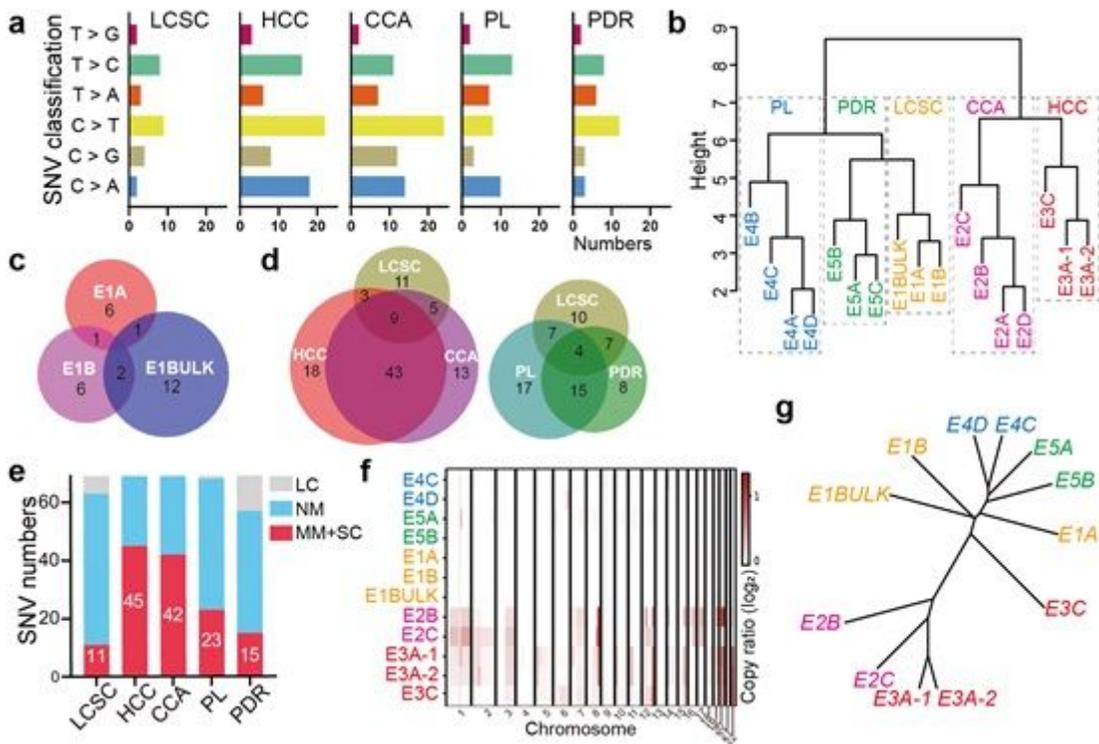


Figure 2

Substantially fewer genetic changes in LCSCs than in HCC and CCA cells. a, Mutation characteristics of different cell types from P2. b, SNV cluster analysis of the 16 sequencing samples from P2. c, Venn diagrams show the SNVs shared among different samples of LCSCs. d, Venn diagrams show the SNV overlap between LCSCs and other types of cells. e, The number of missense mutations (MM) and stop codon changes (SC) identified in different cell types. LC, low coverage; NM, no mutation. f, Heatmap shows the distribution pattern of amplified CNA regions ($\log_2R > 0$) on chromosomes in different samples. g, Clustering evolution analysis of gene amplification in CNAs in the indicated samples. Samples with low WES coverage are not shown in f and g.

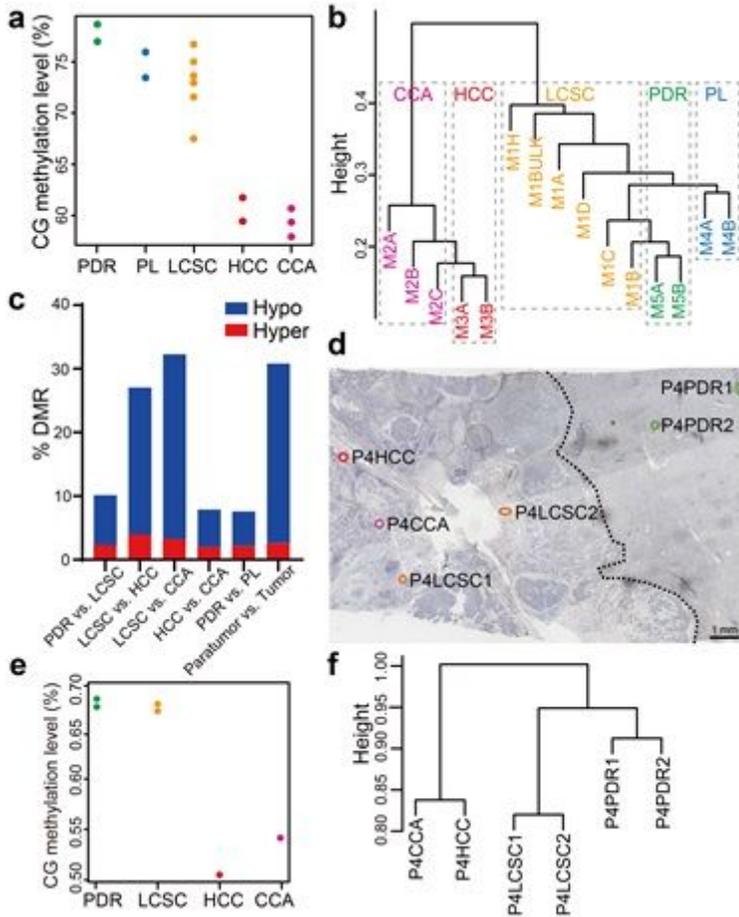


Figure 3

Substantially fewer DNA methylation changes in LCSCs than in HCC and CCA cells. a-c, WGBS analysis of different cell samples from P2. a, The global DNA methylation level of 5 different types of cells. Each dot represents one sample. b, Cluster analysis of the DNA methylation pattern of all WGBS samples. c, DMR frequency between different cell types. d-f, WGBS analysis of the FFPE HCC-CCA sample P4. Different types of cells were captured from FFPE sections of sample P4 and WGBS was performed by CIL-seq. d, Locations and names of the samples. The section was immunostained with anti-EpCAM before LCM. The dotted lines in the figure indicate the boundaries between the tumor and the paratumor. e, The global DNA methylation levels of 6 different types of cells. Each dot represents one sample. f, Cluster analysis of the DNA methylation patterns of all WGBS samples.

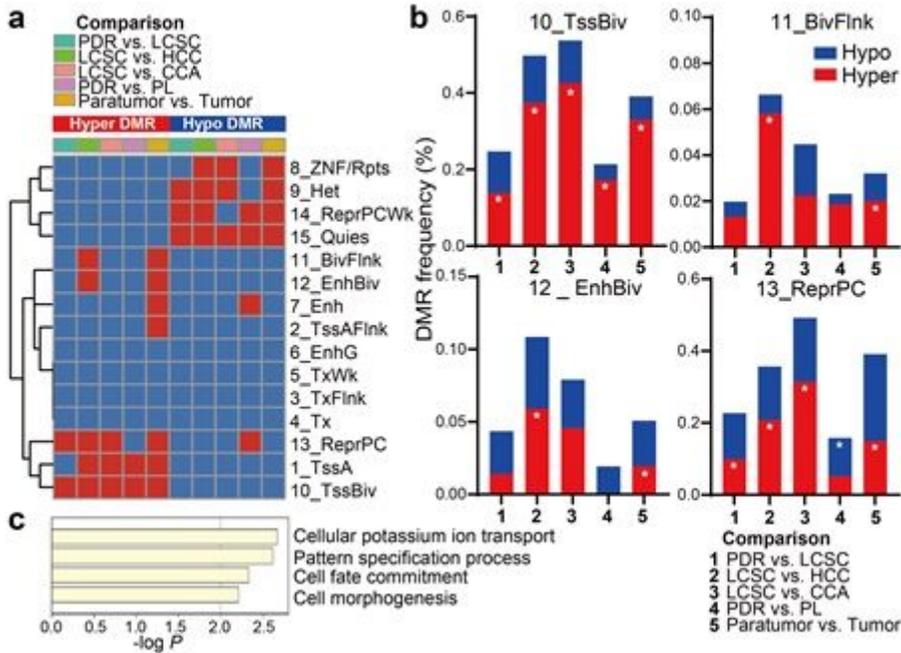


Figure 4

DNA methylation changes in LCSCs correspond to a chromatin restriction state. a, Significant DMR distribution in different chromatin states. We used a 15-state model consisting of 8 active states and 7 repressed states, as suggested in reference 25. The red color represents significant enrichments of hypermethylated or hypomethylated DMRs in the latter cell type (Q-value < 0.01). b, DMR frequency between different cell types in different chromatin states. Both hypomethylation and hypermethylation DMR frequencies between the two indicated cell types in the chromatin states 10_TssBiv, 11_BivFlnk, 12_EnhBiv and 13_ReprPC are shown. DMR frequency = DMR tiles in the chromatin state/all common identified tiles between the two cell types \times 100%. The symbol * represents significant DMR enrichment of hypomethylation or hypermethylation in the latter cell type, as shown in a. c, The functions of genes corresponding to the hypermethylated DMRs between PDR and LCSCs in the chromatin states 10_TssBiv and 13_ReprPC.

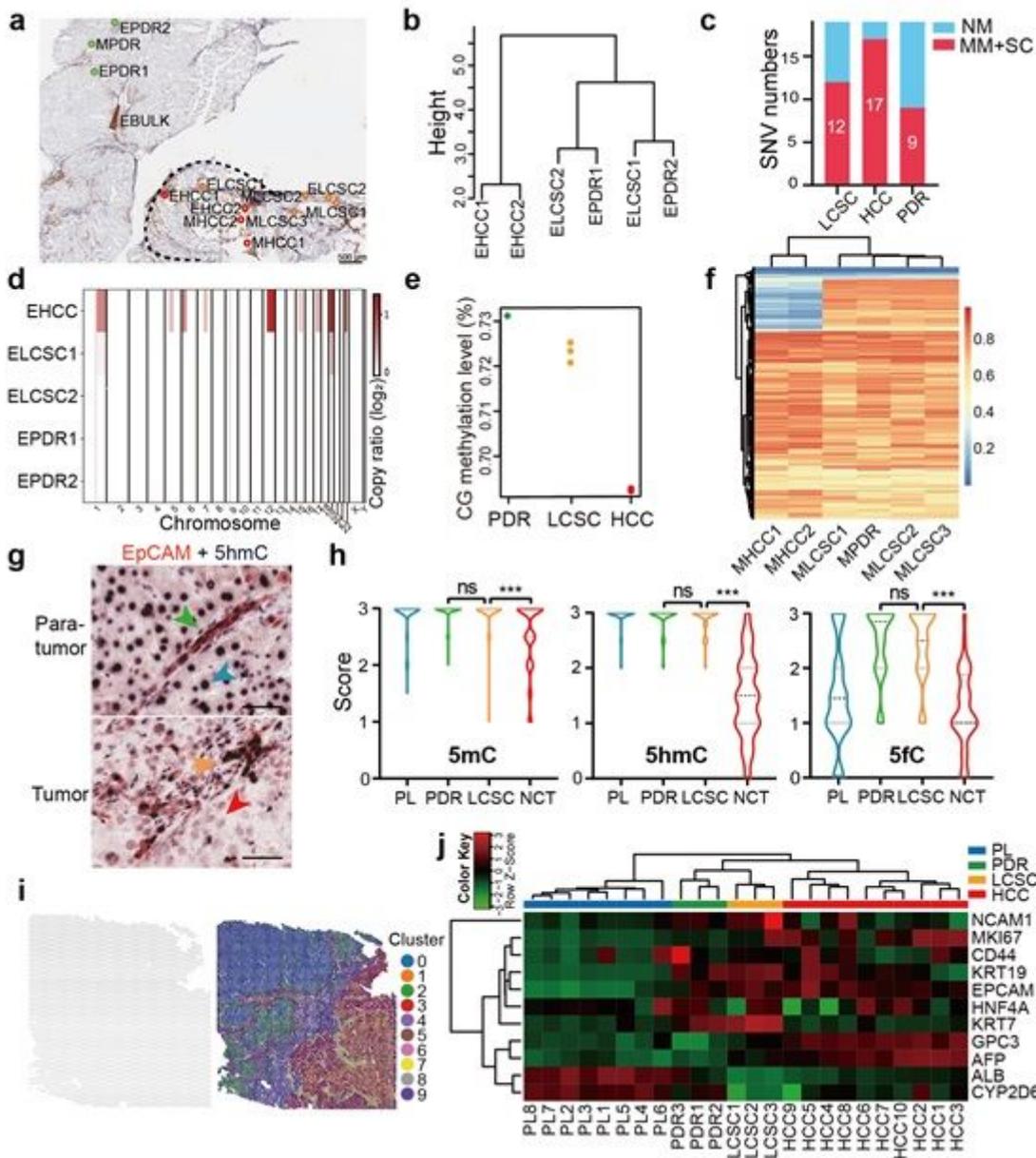


Figure 5

Only minor genetic and epigenetic changes in HCC LCSCs. **a**, Locations and names of the samples used for CIL-seq. Different types of cells were captured by LCM from a frozen section of sample P7 after the section was immunostained with anti-EpCAM. The dotted lines in the figure indicate the boundaries between the tumor and the paratumor. **b**, SNV cluster analysis of the different types of cell samples. **c**, The number of missense mutations (MM) and stop codon changes (SC) identified in different cell types. NM, no mutation. **d**, Amplified CNVs in different cell samples. One sample (EHCC2) with low WES coverage is not shown. **e**, The global DNA methylation levels of 6 different types of cells. Each dot represents one sample. **f**, DNA methylation profiles of 6 cell samples. Each row is a 300-bp window that is covered by at least 3 bases (coverage ≥ 3) in all samples. **g**, Representative double immunostaining of EpCAM (red) and 5hmC (dark purple) in one HCC sample with LCSCs. The green, blue, yellow and red arrowheads indicate PDR, PL, LCSC and HCC cells, respectively. Scale bar, 50 μ m. **h**, Statistical analysis of

5mC, 5hmC and 5fC staining of liver cancer samples. NCT: non-CSC tumor cells. For 5mC and 5hmC, n = 65; for 5fC, n = 34. Data are presented as violin plots with horizontal bars indicating the mean \pm SD. ***, P < 0.001 by Wilcoxon matched-pairs signed-rank test; ns, not significant. i, Unbiased clustering of all ST spots of P7 sample. K-Means cluster analysis, K = 10. j, Cluster analysis of the expression patterns of the selected genes in different spots, as shown in Supplementary Fig. 7d.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryTables.zip](#)
- [SupplementaryMaterialsLiuetal.pdf](#)