

ARTICLE

DOI: 10.1038/s41467-018-03276-y

OPEN

Whole-exome sequencing reveals the origin and evolution of hepato-cholangiocarcinoma

Anqiang Wang^{1,2}, Liangcai Wu¹, Jianzhen Lin¹, Longzhe Han³, Jin Bian¹, Yan Wu⁴, Simon C. Robson⁴, Lai Xue⁵, Yunxia Ge⁶, Xinting Sang¹, Wenzhe Wang⁷ & Haitao Zhao¹

Hepatocellular-cholangiocarcinoma (H-ChC) is a rare subtype of liver cancer with clinicopathological features of both hepatocellular carcinoma (HCC) and intrahepatic cholangiocarcinoma (iCCA). To date, molecular mechanisms underlying the co-existence of HCC and iCCA components in a single tumor remain elusive. Here, we show that H-ChC samples contain substantial private mutations from WES analyses, ranging from 33.1 to 86.4%, indicative of substantive intratumor heterogeneity (ITH). However, on the other hand, numerous ubiquitous mutations shared by HCC and iCCA suggest the monoclonal origin of H-ChC. Mutated genes identified herein, e.g., *VCAN*, *ACVR2A*, and *FCGBP*, are speculated to contribute to distinct differentiation of HCC and iCCA within H-ChC. Moreover, immunohistochemistry demonstrates that EpCAM is highly expressed in 80% of H-ChC, implying the stemness of such liver cancer. In summary, our data highlight the monoclonal origin and stemness of H-ChC, as well as substantial intratumoral heterogeneity.

¹Department of Liver Surgery, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100730, China. ²Department of Gastrointestinal Surgery, Key Laboratory of Carcinogenesis and Translational Research (Ministry of Education), Peking University Cancer Hospital & Institute, Beijing 100142, China. ³Department of Pathology, Yanbian University Hospital, Yanji 133000, China. ⁴Liver Center and The Transplant Institute, Department of Medicine, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA 02215, USA. ⁵Department of Surgery, The University of Chicago Medicine, Chicago 60637, USA. ⁶Novogene Bioinformatics Technology Co., Ltd, Beijing 100083, China. ⁷Department of Pathology, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100730, China. Anqiang Wang, Liangcai Wu, Jianzhen Lin and Longzhe Han contributed equally to this work. Correspondence and requests for materials should be addressed to H.Z. (email: ZhaoHT@pumch.cn) or to W.W. (email: wwwzvsxy@126.com) or to X.S. (email: sangxt@pumch.cn)

Primary liver cancer is one of the most common cancers worldwide¹ and the second leading cause of cancer-related mortality (745,000) every year². Hepato-cholangiocarcinoma (H-ChC) is a special type of liver cancer with pathological features of both hepatocellular carcinoma (HCC) and intrahepatic cholangiocarcinoma (iCCA), accounting for 1–14.1% of primary liver cancers^{3,4}. Risk factors for H-ChC include infection with the hepatitis B virus (HBV) or hepatitis C virus (HCV), alcohol consumption, and primary sclerosing cholangitis^{3,5}.

Coexistence of HCC and iCCA within H-ChC is an intriguing phenomenon. H-ChC was first reported by Allen and Lisa in 1949⁶. Accordingly, H-ChC was divided into three subtypes inclusive of separate masses composed of either HCC or iCCA/contiguous but independent masses of HCC and iCCA/an intermingling of hepatocellular and glandular elements. In 1994, WHO formally defined H-ChC as a tumor-containing intimate and unequivocal admixture of both HCC and iCCA⁷. HCC and iCCA are considered as two distinct types of liver cancer with different pathogenesis, pathological features, prognosis, as well as responses to adjuvant therapies. Up to date, the cellular origins of HCC and iCCA in H-ChC (viz. whether HCC and iCCA differentiate from the same cell origin or from distinct clones) and the underlying mechanisms remain largely unknown.

In this study, we performed whole-exome sequencing (WES) analysis using microdissected HCC and iCCA tissue specimens obtained from H-ChC patients to explore the relationships of these two H-ChC components. We find that H-ChC samples contain substantial private mutations as well as private somatic CNVs, indicative of intratumor heterogeneity (ITH) of H-ChC. However, large amount of ubiquitous nonsynonymous mutations and CNVs are overlapped in HCC and iCCA samples, suggesting the monoclonal origin of H-ChC. Mutated genes identified herein, e.g., *VCAN*, *ACVR2A*, and *FCGBP*, are speculated to contribute to distinct differentiation of HCC and iCCA within H-ChC. Moreover, *EpCAM* is highly expressed in H-ChC samples and associated with poor prognosis of patients with liver cancer. Hence, exome sequencing analyses are suggestive of monoclonal origin and stemness of H-ChC, as well as substantial intratumoral heterogeneity.

Results

Re-evaluation and pathological confirmation of H-ChC. Figure 1 outlined the overall approaches that we used for this study. Tumor samples from 15 H-ChC patients were selected for immunohistochemical (IHC) staining. Tumor specimens from 32 HCC patients and 28 iCCA patients were used as controls. The clinical characteristics of all patients and the immunohistochemical profiles are summarized (Supplementary Data 1 and Supplementary Table 1), respectively.

As shown in Fig. 2a, b for expression of hepatocytic markers Hep and GPC3, vast majority of HCC and H-ChC samples showed positivity with Hep or GPC3 (93.8% for HCC and 80% for H-ChC contained in the HCC component). In contrast, only 14.3% of iCCA samples had hepatocytic expression.

In parallel, biliary differentiation was examined using the two standard markers CK7 and CK19. Hundred percent of iCCA samples exhibited biliary differentiation with positive staining for CK7 or CK19, and 93.3% of H-ChC samples showed such staining pattern only in the iCCA component (Fig. 2a, b). On the contrary, 25% of HCC samples showed biliary differentiation.

These pathological results are concordant with the clinical features of such different subtypes of liver cancer confirming the original diagnosis.

Systematic mutation distribution in H-ChC patient samples.

Next, HCC, iCCA, and adjacent noncancerous tissues were

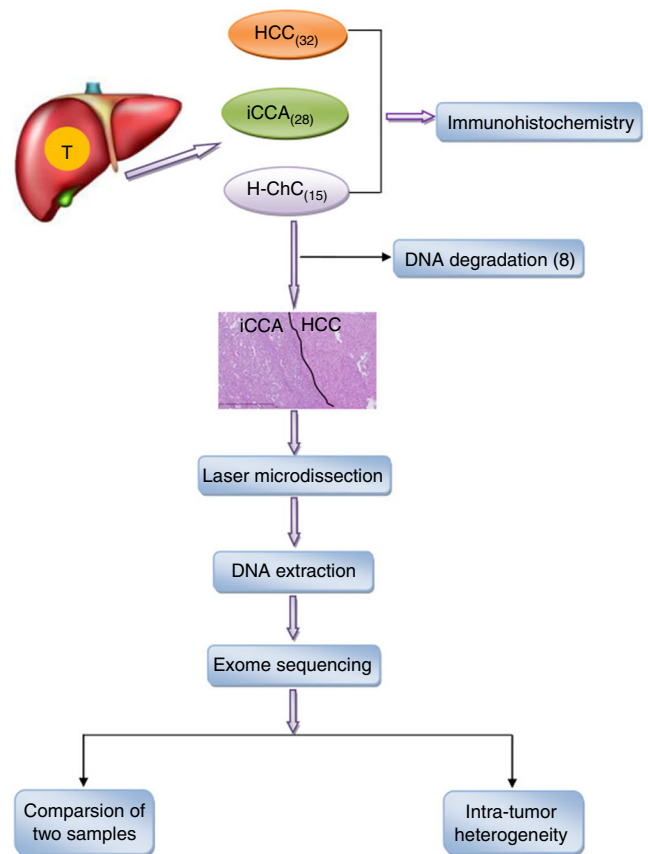


Fig. 1 Flowchart of patient selection and experimental procedure. A total of 32 hepatocellular carcinoma (H-ChC), 28 cholangiocarcinoma (iCCA), and 15 hepatocellular carcinoma (H-ChC) were selected for immunohistochemistry. The blackline in the pathological picture of H-ChC separate HCC component from iCCA component in H-ChC. Finally, seven H-ChC patients were included for microdissection and DNA extraction. Then, we conducted whole-exome sequencing and exome data analysis

isolated from seven H-ChC patients using laser microdissection and subjected for WES. The average depth of WES ranged from 93.01× to 215.6× in all 21 samples (Supplementary Table 2). Somatic nucleotide variants (SNVs), indels, somatic copy number variants (CNVs), and HBV integrations were analyzed to first evaluate distribution of mutations in H-ChC samples.

On average, each H-ChC patient has 159 nonsynonymous SNVs, 6 frameshift indels, 136 CNVs, and 2.6 HBV integration sites (Supplementary Data 2 and 3 and Supplementary Table 3). Six mutational profiles were observed within tumor samples and the predominant mutational profile C:G>T:A is similar to that of liver cancer summarized by Alexandrov et al.⁸. Moreover, 70 known driver mutation genes inclusive of TP53, MTOR, and ARID2 were validated, all of which have been previously linked to liver cancer carcinogenesis (Supplementary Table 4). Among 359 significantly mutation genes (SMGs) identified in all samples, many genes, e.g., *FCGBP*, *ARID1B*, *PTPG13*, *GRM1*, and *MUC12*, have not been well studied in hepatocarcinogenesis (Supplementary Table 5). In addition, we also found 85 genes that may predispose people to liver cancer (Supplementary Table 6).

Substantial intratumor heterogeneity in H-ChC. To explore the relationship between HCC and iCCA components in the same H-ChC tumor, we performed whole-exome mutation and CNV analyses for all tumor samples. We classified somatic

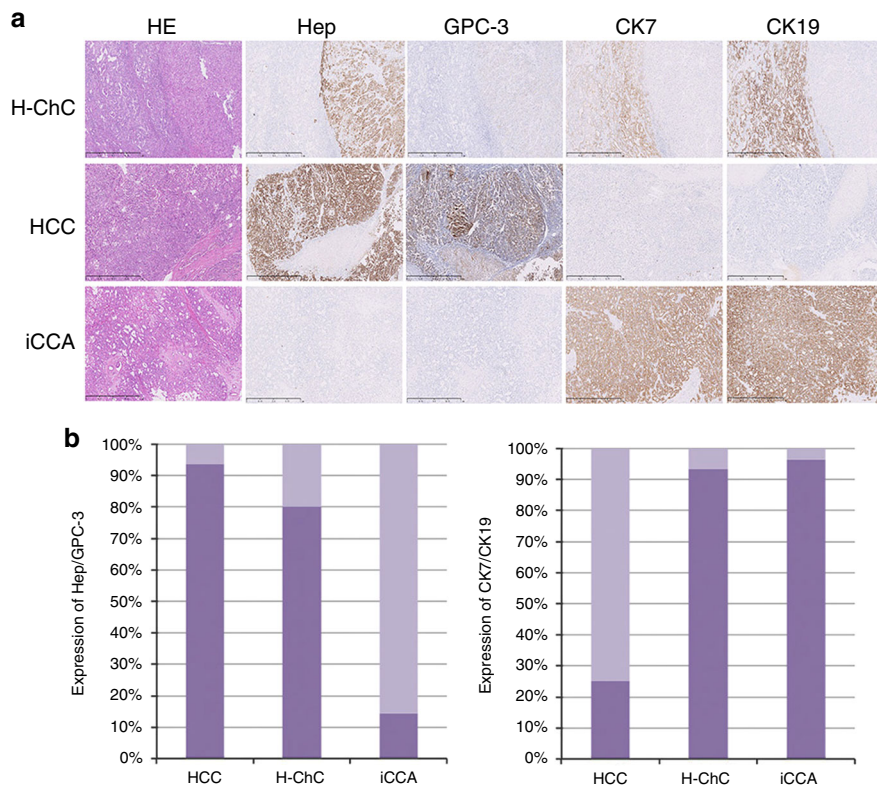


Fig. 2 Immunohistochemical profiles of liver cancer. **a** The tumor cells show immunoreactivity for Hepocyte (Hep), GPC-3, CK7, and CK19 in H-ChC (P6). Immunoreactivity for Hep and GPC-3 are observed in HCC. Tumor cells are positive for CK7, CK19 in iCCA. **b** The columns show the expression level of Hep/GPC3 and CK7/19 in H-ChC, HCC, and iCCA. The scale bar represents 1 mm

nonsynonymous mutations as ubiquitous (present in both tumor components) or private (exclusively present either in HCC or iCCA). As shown in Fig. 3, all H-ChC samples had substantial amount of private nonsynonymous SNVs, accounting for 86.4, 48.1, 77.3, 33.1, 56.6, 48.7, and 53.1% in patient 1 to patient 7 (P1–7), respectively.

Similarly, frequent private somatic CNVs were also noted in HCC or iCCA components ranging from 88.9 to 100% (Supplementary Table 7). These results suggest that substantive genetic heterogeneity exists in H-ChC, which is in keeping with the obvious morphologic differences between HCC and iCCA components seen in the clinic.

HCC and iCCA components share common clonal origin in H-ChC. On the other hand, ubiquitous nonsynonymous SNVs that are overlapped by HCC and iCCA samples range from 29 to 108 (Fig. 3). Moreover, mutation spectrum of iCCA is near identical to that of HCC within a given H-ChC (Fig. 4a). These data indicate that it is unlikely that HCC or iCCA component independently accumulated such substantial amount of ubiquitous nonsynonymous mutations.

Next, nonnegative matrix factorization approach was applied to identify mutational signatures. After cosine similarity analysis with 30 known signatures, three newly identified mutational signatures (A, B, and C) were found in H-ChC (Supplementary Fig. 1–3). For a given H-ChC sample, the percentage of mutations contained in each signature is consistent between HCC and iCCA (Fig. 4a). Additionally, HCC and iCCA share the same mutations on SMGs (Fig. 4d).

In parallel, ubiquitous somatic CNVs that are shared by HCC and iCCA were also identified. Number of ubiquitous somatic CNVs from P1 to P4 are 17, 2, 42, 36, and 5 for P7, respectively (Supplementary Table 9). However, no ubiquitous somatic CNVs

was found in P5 and P6. Within the same H-ChC sample, distribution of somatic CNVs in HCC is similar to that in iCCA (Fig. 4b, c and Supplementary Figures 4 and 5).

Furthermore, HBV integrations were also assessed. We observed that, in P5 samples, HCC and iCCA contain a common HBV integration site on chromosome 15 (Fig. 4c), suggesting that HBV integration is acquired before tumor differentiation into the two components. However, in P1, P4, and P6 samples, private HBV integrations were noted in either HCC or iCCA, indicating that HBV integrations could also occur after tumor differentiation into distinct phenotypes (Supplementary Figure 5). No HBV integration sites were identified in P2, P3, and P7 samples (Supplementary Figure 5).

To further explore the clonal architecture of H-ChC, we constructed phylogenetic trees (select driver genes were marked in the trunk and branch of evolutionary trees; Fig. 5a and Supplementary Table 8) using somatic SNVs and somatic mutations that take the CNVs in consideration. The results revealed different length of trunk in HCC and iCCA components of all H-ChC, which suggested the monoclonal origin of H-ChC (Fig. 5a, c). In addition, clonal analyses suggested common mutation clusters for all seven H-ChC patients after adjusting for cancer cell fraction (CCF) for each sample using Pylone (Fig. 5b)^{9–11}.

The findings above obtained by analyzing systematic mutations, somatic CNVs, clonal analyses as well as HBV integrations all imply that H-ChC originates from a common progenitor.

Genomic dynamics of two distinct differentiation phenotypes.

To further explore the genomic mechanisms of differentiation of monoclonal H-ChC into two distinct phenotypes, we screened all SMGs identified by WES. From 51 genes mutated in at least two H-ChC patients, we selected 17 genes associated with regulation

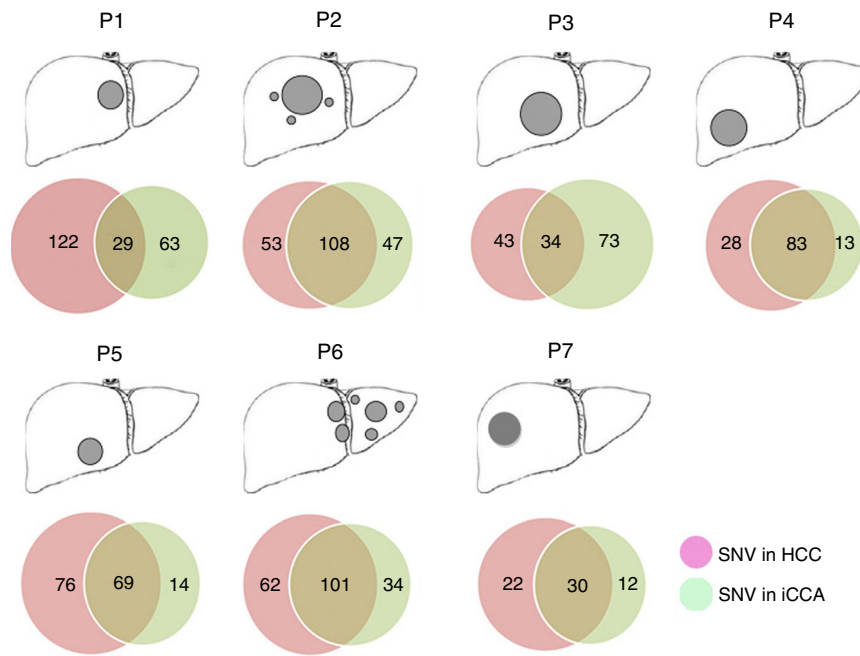


Fig. 3 Distribution of nonsynonymous SNVs. **a** The cartoons of liver show the tumor sites in patients. **b** Venn diagrams show the relationship of nonsynonymous somatic mutations between H-ChC component (red circle) and iCCA component (green circle) in every H-ChC patient. The different number represents the number of nonsynonymous somatic mutations for corresponding samples and the overlapped regions are ubiquitous nonsynonymous somatic mutations between two samples of same H-ChC patients

of cell differentiation for subsequent pathway enrichment analysis (Supplementary Table 9). Mutations of nine genes, such as FCGBP and VCAN, that have been related to the maintenance of stem state were found in all seven H-ChC patients (Supplementary Table 9). Twelve genes, e.g., ACVR2A and APC, that constitute signaling pathways modulating pluripotency of stem cells were mutated in six H-ChC patients (P1–P6) (Supplementary Table 10). Moreover, 12 mutated genes in Wnt and Notch pathways including TP53 and NFATC2/3, which regulate the differentiation of hepatocyte and biliary epithelium, were noted in five H-ChC patients (P1–P4 and P6) (Supplementary Table 11). These data suggest the involvement of such key genes and pathway in the differentiation of HCC and iCCA in H-ChC.

Expression of stem cell markers and survival analyses. To further validate the stem features of H-ChC, we performed immunohistochemical staining with stem cell markers including c-kit and EpCAM in FFPE tissue specimens obtained from 75 liver cancer patients. Except for one HCC sample, all samples showed negative staining for c-kit (Fig. 6a). In contrast, vast majority of specimens exhibited EpCAM immunoreactivity (86.7% in H-ChC and 71.4% in iCCA) (Fig. 6b). Moreover, in HCC samples, EpCAM staining pattern appears to correlate with CK19 positivity (indicative of biliary differentiation), that is 66.7% in CK19 (+) samples vs. 17.2% in CK19 (–) samples (Fig. 6b).

We then conducted survival analysis by protein levels of EpCAM regardless of the cancer subtypes. Univariate analysis with log-rank test showed that EpCAM positivity was significantly associated with poor prognosis of liver cancer patients ($P = 0.028$; Table 1 and Fig. 6c). Yet, the prognostic value was not statistically significant as determined by multivariate Cox regression analysis ($P = 0.087$; Table 1). CA199 and margin status were found to be associated with prognosis of liver cancer by multivariate Cox regression analysis ($P = 0.001$; $P = 0.015$; Table 2).

Discussion

H-ChC is a special type of liver cancer that contains both HCC and iCCA components. Herein, we showed that hepatocytic markers (Hep and GPC3) were strongly expressed in the HCC component of H-ChC, whereas biliary epithelial markers (CK7 and CK19) were strongly expressed in the iCCA component of H-ChC. We then conducted WES to explore the relationship between the two distinct components of H-ChC. We found that HCC and iCCA components possessed many private nonsynonymous mutations and somatic CNVs. Although we conducted CNV analysis using exome data, this method proved to be highly accurate compared to the results of CNV analysis using low-depth whole genome data¹². Private somatic CNVs occurred more frequently than private somatic mutations. These findings support the notion that substantial intratumor heterogeneity exists within H-ChC. Intratumor heterogeneity in H-ChC poses a great challenge for liver cancer-targeted therapy^{12–14}. Currently, it is a big challenge to choose the target drugs from multiple mutated driver genes. However, the common mutations within H-ChC may be a better choice because of their more important role in carcinogenesis. Therefore, it is necessary to conduct microscopic multiregional tissue selection in order to understand the complete genetic landscape of H-ChC, which can aid in the development of targeted molecular therapies.

On the other hand, the prevalence of ubiquitous somatic SNVs (29–108) and CNVs (2–42) noted in H-ChC suggests that HCC and iCCA components of H-ChC share a common progenitor. The relatively narrow coverage for genomic of WES makes it difficult to acquire all of the CNVs. However, the identification of ubiquitous somatic CNVs could also support the monoclonal origin of H-ChC to some extent. And the lack of ubiquitous somatic CNVs could not rule out the possibility of H-ChC monoclonal origin. Because the likelihood of HCC and iCCA component independently accumulating such substantial amount of ubiquitous nonsynonymous mutation is exceedingly low. Clonal analyses using Pyclone identified common mutation

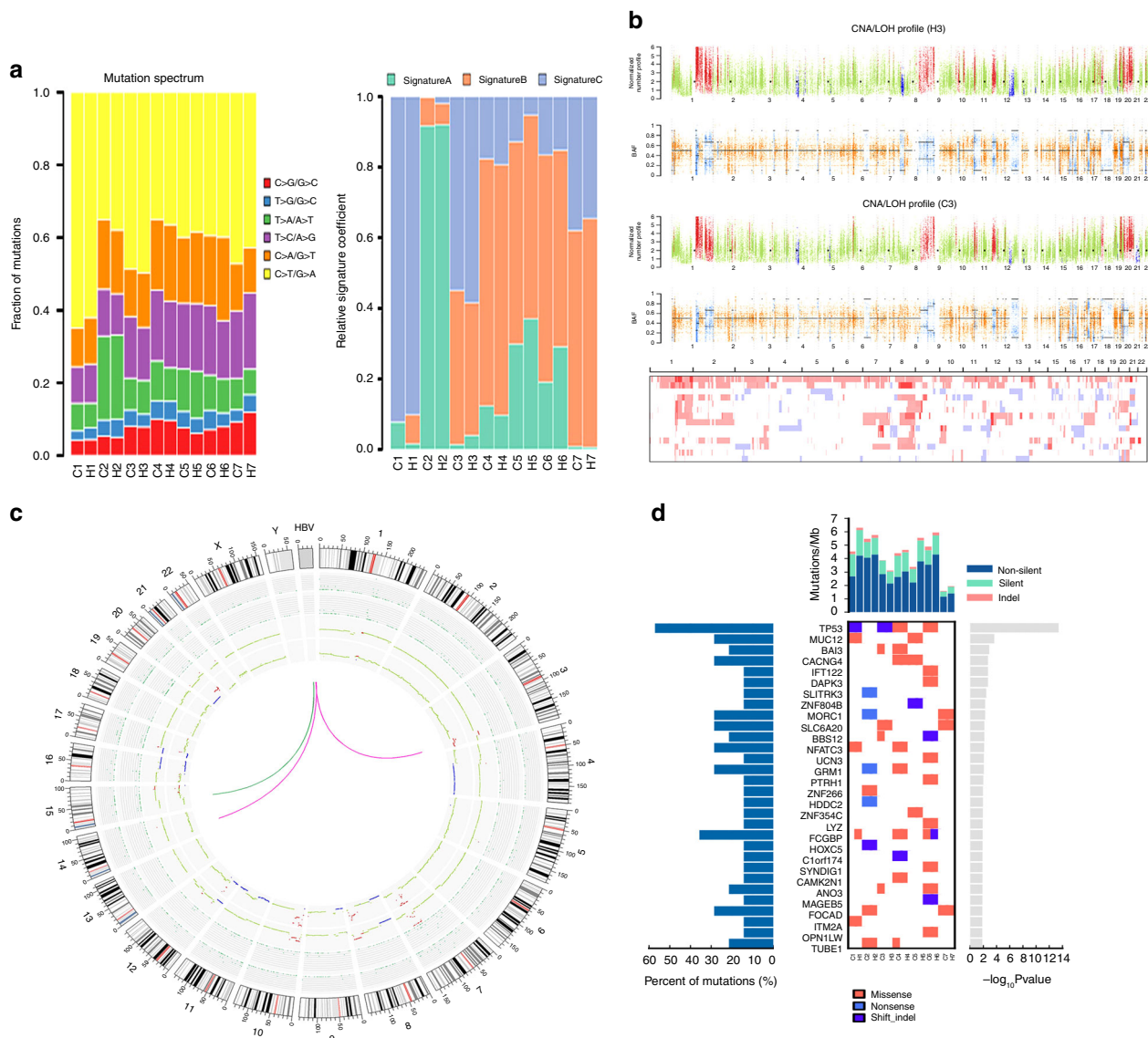


Fig. 4 Mutation spectra and mutation signatures among H-ChC samples. **a** The upper column depicts mutation spectrum within all tumor samples. Red represents mutation type C>G/G>C, blue T>G/A>C, green T>A/A>T, purple T>C/A>G, orange C>A/G>T, and yellow C>T/G>A. Y axis indicates fraction of matations. The below column shows mutation signatures of all tumor samples. **b** The upper chart depicts the distribution of somatic copy number variation (CNV) of patient 3 (P3). Red is for CNV gain, green for normal CNV, and blue for CNV loss. The below chart depicts the distribution of B allele frequency (BAF). Orange represents consistent distribution of allele and blue indicates loss of heterozygosity (LOH). Heat maps show the distribution of CNV for all included H-ChC patients. Red is for CNV gain and blue for CNV loss. **c** Circos plot depicts the relationship between HCC and iCCA components of P5 on the terms of somatic nucleotide variant (SNV), somatic insertions and deletions, somatic copy number variation (CNV), and HBV integration. The first and second circles represent CNV for iCCA and HCC. Red indicates CNV gain, green normal CNV, and blue CNV loss. The third and fourth circles represent SNV and indel for iCCA and HCC. Green dot indicates SNV and indel. Green curve indicates HBV integration sites for iCCA and red curve for HCC. **d** Significantly mutated genes (SMG) landscape shows the distribution of some SMGs between samples in H-ChC. The column on top shows the mutational rate of every sample. Heat map shows the SMGs and mutation type including missense mutation (red block), nonsense mutation (light blue block), and shift indel (dark blue block). H1 and C1 represent the HCC and iCCA components of patient 1 (P1). Similar label was used for other patients. The column on the left stands for the percent of mutations for SMGs. The picture on the right shows the P value for SMGs

usters in both of tumor components, which suggested monoclonal origin of H-ChC. For each H-ChC sample, the probability of monoclonal and bicolonal is equal, just as the two sides of coin. The probability of same clonal features for different samples decreases 50% with one sample increase, especially all samples were chosen randomly. Therefore, although seven H-ChC samples are small, the conclusion is also reliable. HBV integrations were found in four out of seven patients. A unique HBV integration site common to both HCC and iCCA component was found in P5. Considering the random features of HBV integration

and tens of millions of probable integration sites, the probability of common HBV integration sites in two different origin liver cancer samples could hardly be ignored. Therefore, the common integration supports the monoclonal origin of HCC and iCCA components. However, a lack of common HBV integration site in other samples cannot rule out the possibility of common origin of these two tumor components, especially in light of the data from WES analyses, because HBV could integrate in any sites of tumor genome. Furthermore, the HBV integration randomly occurs in any stage of carcinogenesis, either before or after differentiating

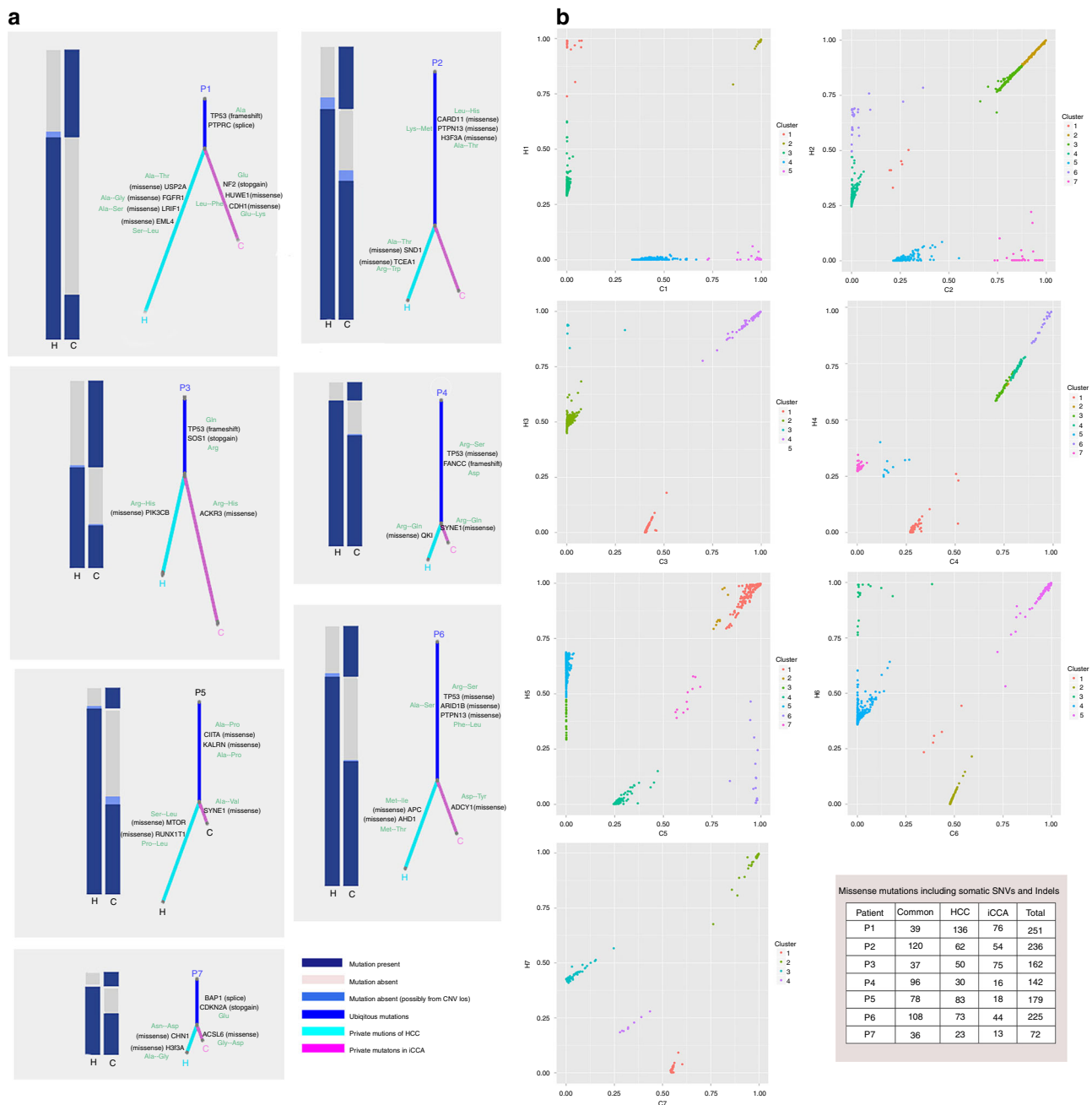


Fig. 5 Missense mutations and cancer cell fraction comparisons within each H-ChC. **a** Heatplot with different bars represents various distributions of missense mutations including somatic SNVs and Indels within H-ChC. Fraction of ubiquitous nonsynonymous somatic mutations (trunk) (blue bar) and unique nonsynonymous somatic mutations (branch) (green bar for HCC and pink bar for iCCA) reveal the relationship of two tumor samples within a single H-ChC. Part of driver mutations was marked on trunk and branch of evolutionary trees. **b** Two-dimensional scatter plots show the cancer cell fraction (CCF) of the mutations in HCC and iCCA components of tumors. Different clusters were calculated from each H-ChC sample. Clusters off the axes indicate mutations in both of tumor components. Clusters on the axes reveal mutations in either HCC or iCCA components. **c** The table shows missense mutations in different tumor components

into distinct phenotypes. Indeed, it is not rigorous to identify HBV integration using WES on FFPE samples. However, we just regard the HBV integration as a supplementary evidence to explore the clonality of H-ChC.

Another puzzle is that how monoclonal H-ChC could become a tumor with unambiguous HCC and iCCA components. We hypothesize that some key genes or pathways may contribute to the loss of the undifferentiated state of stem cells and/or hepatocyte or biliary epithelial differentiation. WES analyses showed that many genes participated in cell stemness and differentiation

are mutated. For example, VCAN is known to favor homeostasis of extracellular matrix, thereby modulating cell proliferation and differentiation¹⁵. It is also essential for the maintenance of stem state of cancer stem cells in many types of cancer¹⁶. ACVR2A is overexpressed in undifferentiated states of mouse and human embryonic stem cell lines and decreased upon differentiation. All seven H-ChC patients possess mutated genes involved in cell stemness and differentiation. Mutated genes identified in six patients are the key components of signaling pathways regulating pluripotency of stem cells. In addition, Wnt and Notch pathways

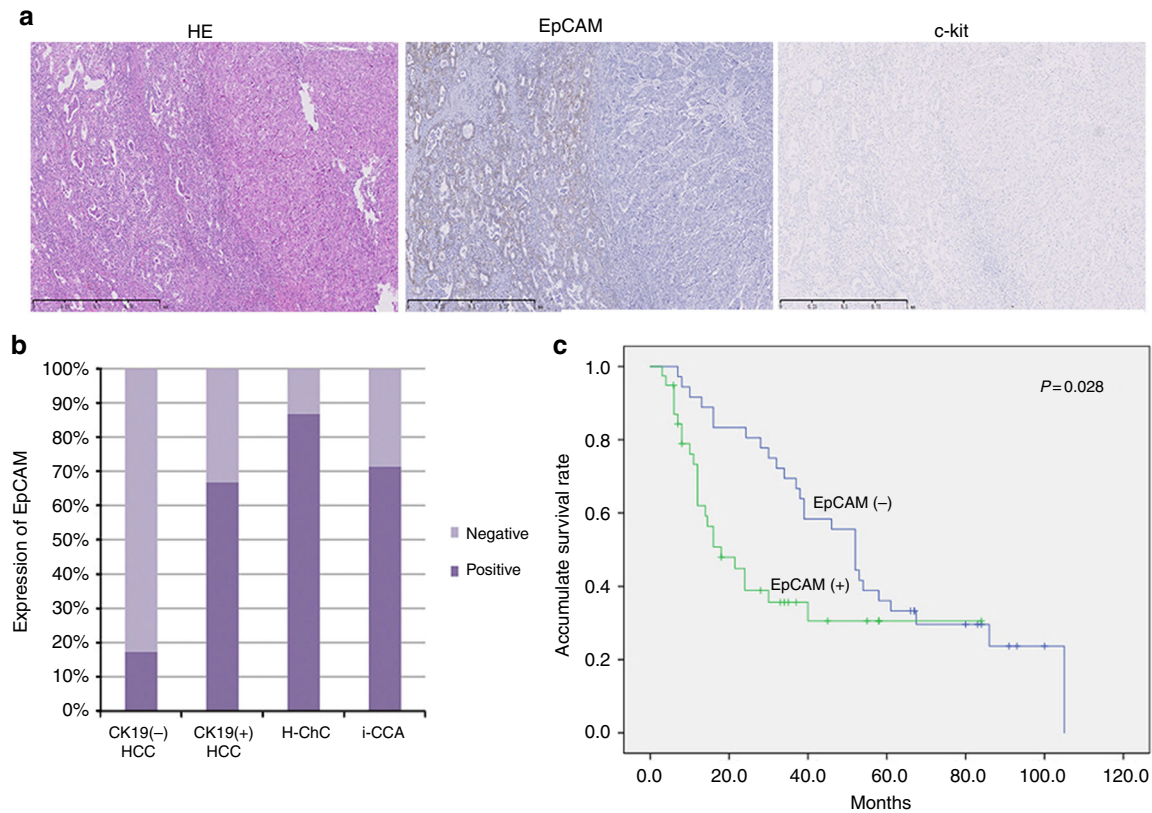


Fig. 6 Expression of stem cell markers and survival analysis. **a** The tumor cells show immunoreactivity for EpCAM in H-ChC. **b** The columns show the expression level of EpCAM in H-ChC, HCC, and iCCA. **c** The survival curve shows the relationship between EpCAM expression and survival. The scale bar represents 1 mm

have been known to modulate the differentiation of hepatocytes or biliary epithelial cells¹⁷. In this study, we also found mutated genes in these two pathways in five out of seven H-ChC patient samples. Overall, we demonstrated that some gene mutations might lead to the instability of stem state, whereas others contributed to distinct differentiation of H-ChC. However, whether there are any corresponding changes in the protein levels of such mutation genes remain unknown, which could be exploited using next-generation RNA sequencing.

Furthermore, we searched NCBI database to examine the relationships between mutation genes and well-known stem cell markers. However, no gene mutations in stem cell markers, e.g., EpCAM, NCAM, c-kit, CD133, and CD90, were found in any samples used for this study. We then analyzed protein levels of select stem cell markers inclusive of EpCAM and C-kit by immunohistochemistry using samples from 75 liver cancer patients. C-kit, considered as a marker for pluripotent stem cell in the liver, has previously been shown to characterize H-ChC, denoting the stem cell origin of H-ChC^{18–20}. C-kit protein expression, however, was absent in nearly all liver cancer samples in our study. The other stem cell marker, EpCAM, is originally identified on the surface of embryonic stem cells²¹ and has been shown to play a role in the maintenance of the undifferentiated state of stem cells^{22,23}. Overexpression of EpCAM has been previously noted in H-ChC patients¹⁸, which is in accordance with our results herein. Nevertheless, we are still not clear with regard to the origin of H-ChC, as stem cell-like features could derive either directly from stem cells or through reprogramming from mature differentiated liver cancer cells^{24–26}.

It is widely accepted that HCC can be classified into two subtypes according to the expression of biliary differentiation marker CK19 and CK19⁺ HCC has a worse prognosis than CK19⁻

HCC²⁷. In this study, we observed that expression of EpCAM is markedly increased in CK19⁺ HCC samples when compared to CK19⁻ samples, although CK19⁺ subtype only accounts for 25% of HCC. In parallel, majority of iCCA and H-ChC samples, those are chiefly CK19⁺, exhibited positivity for EpCAM. These results are in keeping with the previous studies, suggesting that CK19⁺ HCC and iCCA could originate from progenitor cells or biliary tree stem cells^{26–29}. Moreover, by performing survival analyses, we demonstrated that liver cancer patients with high levels of EpCAM expression have poor prognosis using the univariate Cox analysis model. However, when using multivariate Cox analysis, the prognostic value of EpCAM expression was not statistically significant, possibly due to the limited sample size in our study.

Collectively, we found that the HCC and iCCA components of H-ChC originate from a common cell with stem cell-like features. This notion is supported by many other studies showing the stem cell origin of H-ChC^{19,20,30}. Indeed in 2010, WHO updated the classification criteria for H-ChC: H-ChC is divided into the classical type and three subtypes with stem cell features (typical subtype, intermediate cell subtype, and cholangiolocellular subtype)³¹. Accordingly, all seven H-ChC patients subjected for exome analyses herein can be classified as the classical type. However, such WHO classification method is suboptimal having many issues that need to be urgently addressed. Many newly recognized different types of liver cancer have been suggested to be classified as a new subtype of H-ChC. For example, in 2013, Terada et al. reported an H-ChC case with ductal plate malformation features that was characterized by formation of markedly irregular tubules of iCCA cells with intraluminal cell projections, bridge formations, and intraluminal tumor biliary cells³². Jung et al. reported cholangiocellular carcinoma with satellite nodules showing intermediate differentiation¹⁷. In

Table 1 Results of univariate analyses in liver cancer

Variables	No. of pat.	OS			
		No. of event	MST	χ^2	P
Gender					
M	57	38	34.379	1.362	0.243
F	18	13	49.279		
HBV					
Yes	47	31	50.225	0.955	0.328
No	28	20	37.103		
AFP					
Normal	41	27	46.332	0.167	0.683
Elevated	28	21	43.324		
CA199					
Normal	38	24	49.987	8.404	0.004
Elevated	23	20	25.283		
NA					
Histology					
HCC	32	20	61.16	39.516	0.009
ICC	28	21	32.858		
CHC	15	10	29.038		
Size					
<5 cm	43	29	51.091	1.095	0.295
>5 cm	32	22	40.582		
Number					
1	52	32	51.091	1.095	0.295
>2	23	19	37.885		
Differentiation					
High	11	5	65.519	2.642	0.104
Low/moderate	53	38	44.269		
Vascular invasion					
Yes	9	7	36.820	0.778	0.378
No	66	44	46.773		
Margin					
Positive	15	13	51.995	7.732	0.005
Negative	60	38	23.267		
TACE					
Yes	44	31	45.507	0.073	0.788
No	31	20	47.191		
Hepatocyte					
Positive	39	25	54.548	3.711	0.054
Negative	36	26	35.337		
GPC3					
Positive	32	23	45.197	0.104	0.747
Negative	43	28	47.373		
Hepatocyte/GPC3					
Positive	47	30	52.426	2.918	0.088
Negative	28	21	34.572		
CK7					
Positive	43	29	39.425	2.3	0.129
Negative	22	22	54.523		
CK19					
Positive	42	28	35.361	4.698	0.030
Negative	33	23	56.665		
CK7/CK19					
Positive	49	33	39.721	2.670	0.102
Negative	26	18	56.772		
EpCAM					
Positive	39	24	36.206	4.841	0.028
Negative	36	27	56.132		

No. of pat. number of patient; OS, overall survival; MST, mean survival time; TACE, transcatheter arterial chemoembolization

addition, a single H-ChC tumor often contains more than three morphologically distinct tumor components and classification of these atypical tumors remains problematic. Some research¹⁸ made diagnosis by the predominant (50%) components when

Table 2 Results of multivariate analyses for liver cancer

Variables	OR	95% CI	P
CA199	4.156	1.769–9.766	0.001
Histology	1.430	0.816–2.503	0.211
Size	1.291	0.602–2.767	0.512
Number	1.214	0.585–2.518	0.603
Vascular invasion	1.100	0.382–3.169	0.860
Margin	2.874	1.227–6.773	0.015
TACE	2.221	1.029–4.796	0.042
CK19	2.212	0.717–6.826	0.618
EpCAM	2.151	0.896–5.165	0.087

OR, odds ratio; 95% CI, 95% confidence interval

tumor consists of plural components, which we also adopted. Up to now, however, no widely accepted standard is adopted. Furthermore, the classical type of H-ChC is proved to be monoclonal. Therefore, even if they are not positive for widely used stem markers, they may also originate from cells with stem features. In this case, we assumed that classical type of H-ChC also belong to H-ChC with stem cell features. Herein, we suggest a subgroup classification strategy based on the cell origin rather than morphology features, e.g., EpCAM⁺ H-ChC. We speculate that other types of H-ChC, CK19 (+) HCC, and part of iCCA all belong to this type of cancer. Of course, further studies are needed (e.g., by using next-generation sequencing analyses) to validate the clinical utility of such novel classification recommendation.

In conclusion, exome sequencing analyses are suggestive of the monoclonal origin of H-ChC, which may promote the molecule classification of primary liver cancer on the basis of cell origin. In addition, the substantial intratumor heterogeneity noted in H-ChC urges multiregional sequencing analysis to find the common driver mutations that playing more important role in carcinogenesis, thus make target drugs selection more precision and effective.

Methods

Patients and tissue samples. We enrolled 75 patients who underwent surgery at Peking Union Medical College Hospital from April 2008 to December 2015. Fifteen patients were pathologically diagnosed with combined hepatocellular cholangiocarcinoma (H-ChC), 32 were diagnosed with hepatocellular carcinoma (HCC), and 28 with intrahepatic cholangiocarcinoma (iCCA). All histological specimens were reevaluated by two experienced pathologists. HCC, iCCA components of H-ChC samples, and adjacent non-cancerous liver tissue were collected from seven eligible H-ChC patients. The flowchart of patients selected for DNA extraction and whole-exome sequencing was presented (Fig. 1). Fifteen H-ChC were performed immunohistochemistry. Thirty-two HCC and 28 iCCA were also performed immunohistochemistry as control group. Informed consent was obtained from all human participants. This research was also approved by the Ethical Committee of Peking Union Medical University Hospital. All clinical data were obtained from hospital record for each case (Supplementary Data 1).

Immunohistochemistry. We cut 5- μ m-thick sections from representative paraffin blocks of liver cancer for immunohistochemistry (Supplementary Figure 6). The primary antibodies were hepatocyte (IR624, 1:50, Dako), GPC-3 (ZM-0146, 1:1, ZSGB-Bio), CK7 (1:1, Dako), CK19 (1:50, Dako), EpCAM (HEA125, 1:50, abcam, Cambridge, UK), and c-kit (A4502, 1:1, Dako). Immunoreactivity was evaluated according to the percentage of positive cells on sections regardless of intensity of staining, with grading form 0 to 4+ as follows: 0, no staining or equivocal reaction; 1+, 1–5% positive cells; 2+, 6–25% positive cells; 3+, 26–50% positive cells; and 4+, >50% positive cells. Tumors with a grade of 2+ were regarded as positive for antigen expression.

Microdissection and DNA extraction. Deparaffinized tumor specimens were cut into serial consecutive 10 μ m slides, stained with hematoxylin. Two pathologists confirmed the different cancer components independently. Then, HCC and iCCA components were separately microdissected from H-ChC slides using Leica CTR 6000 Microsystem (Wetzlar, Germany). We also dissected adjacent nonmalignant

tissue as control tissue. Then, DNA was extracted using the QIAamp DNA FFPE Tissue Kit Print (Qiagen, Germany) according to the manufacturer's protocol.

Library preparation and sequencing. Paired-end DNA library was prepared according to the manufacturer's instructions (Agilent). The adapter-modified gDNA fragments were enriched by six cycles of PCR. Whole exome capture was carried out using Agilent's SureSelect Human All Exon V5 Kit. Finally, 50 Mb of DNA sequences of 33,4378 exons from 20,965 genes were captured. After DNA quality evaluation, pooled samples were sequenced on Illumina HiSeq 4000 according to the manufacturer's instructions for paired-end 150 bp reads. The average sequencing depth of target region and coverage of target region were summarized in Supplementary Table 2.

Exome sequencing data analysis for SNVs and INDELS calling. Raw data (stored as FastQ format) obtained from HiSeq4000 contains adapter contamination, low-quality nucleotide, and undetected nucleotide (N), which can pose significant influence on downstream processing analysis. Hence, reads with adapter contamination, reads containing uncertain nucleotides more than 10 percentage, and paired reads when single reads have more than 50 percentage low-quality (<5) nucleotides are discarded. After these steps, high-quality clean data are obtained. Finally, QC statistics including total reads number, sequencing error rate, percentage of reads with average quality >Q20, percentage of reads with average quality >Q30, and GC content distribution can be calculated. Paired-end clean reads are aligned to the reference genome (UCSC hg19) using Burrows-Wheeler Aligner (BWA) software³³. If a read or reads pair is mapped to multiple positions, BWA will choose the most likely placement. While if two or more most-likely placements are present, BWA will choose any one randomly. Aligned reads were realigned to the genome. Genome Analysis Toolkit (GATK)³⁴ was used to ignore those duplicates resulted from PCR amplification with Picard-tool. We utilized the Indelrealigner and RealignerTargetCreator in GATK do realignment around the indels according to GATK best practice. Furthermore, we performed base quality score recalibration with GATK to avoid system bias. After realignment to genome, we identified and filtered variants (SNP, INDELS) using GATK HaplotypeCaller and variantFiltration to guarantee meaningful analysis. Variants obtained from previous steps were compared based on the dbSNP³⁵ and 1000 Genomes database³⁶ and annotated with ANNOVAR³⁷. SNVs and somatic INDELS were identified using MuTech³⁸ and Strelka³⁹ with matched adjacent non-cancerous samples, respectively (Supplementary Data 2). The comparison of somatic mutations was presented in circus plot within two tumor samples in same patient⁴⁰.

Mutation spectrum and mutation signature analysis. We performed mutation spectrum and signature analysis to explore the relationship within tumor samples in each same patient. We drew mutation spectrum barplot to present mutation spectrum of each sample. We also conducted clustering analysis on mutation spectrum to draw mutation spectrum heatmap to observe the similarity and difference within tumor samples. We conducted cluster analyses on 96 somatic mutation types using nonnegative matrix factorization^{41,42} and acquired three different mutation signatures. Then, mutation spectra were clustered with 30 known signatures on COSMIC⁴³ to explain mutation process of samples⁴⁴. The similarity of mutation signatures was evaluated with cosine similarity >0.9, which suggested common signatures. Signatures A, B, and C were identified in tumor samples and the distribution of them was presented.

Estimation of tumor purity and cellularity. We used ABSOLUTE to estimate the tumor purity and cellularity from analysis of somatic DNA alteration⁴⁵ (Supplementary Table 12).

Phylogenetic and clonal analyses. We conducted phylogenetic analysis based on the exome sequencing data. Branch and trunk lengths are proportional to the number of nonsynonymous mutations acquired on the corresponding branch and trunk⁴⁶. Driver mutational genes were marked on trunk and branch. In order to evaluate the clonality of tumor more precisely, we conducted clonal analysis using Pyclone⁴⁷.

Identification of potential driver mutations. In order to identify potential driver mutations playing significant role in carcinogenesis, we compared sample mutations with known driver mutations with in-house software. Four driver mutation databases were utilized for the comparison, including Cancer Gene Census (CG513)⁴⁸, Bert Vogelstein125⁴⁹, SMG127⁵⁰, and Comprehensive 435⁵¹ database. Potential driver mutations were analyzed to determine common driver mutations for each patient (Supplementary Table 4).

Determination of significantly mutated genes. Significantly mutated genes (SMGs) were defined as the somatic mutations with higher mutated frequency than background mutation rate. We analyzed mutations with SMG test⁵² based on tumor samples. Significantly mutated genes landscape heatmap was presented. For SMGs, we conducted pathway enrichment analyses with PathScan software⁵³.

Databases such as KEGG⁵⁴, Biocarta, PID, and Teactome were utilized to perform this analysis (Supplementary Table 5).

Screening of predisposing genes. Potential predisposing genes could be identified by detecting germline mutations and comparing with CGC⁴⁶ database with in-house software, which were presented in Supplementary Table 6.

Copy number analysis using exome-sequencing data. We identified copy number variants (CNVs) with control-FREEC⁵⁵ based on exome-sequencing data to analyze the copy number state of each tumor. Control-FREEC could construct copy number profiles with aligned BAM data. Then, the profiles were normalized, segmented, and analyzed to obtain the copy number state of each genomic region. We compared cancer tissues with matched non-cancerous tissues with control-FREEC. Tumor samples with overdioidy could also be analyzed. We also performed comparison of CNVs between two tumor samples in same patient. We defined a common CNV if 100% of CNV regions of two samples overlap⁵⁶ (Supplementary Data 3 and Supplementary Table 7). The comparison of CNVs was presented with circular visualization³⁸ within two tumor samples in same patient.

HBV integration analyses based on exome-sequencing data. To identify the HBV integration sites based on exome-sequencing data, we constructed an HBV virus genome library. First, we conducted reads alignment to human reference genome and some soft-clipped reads were extracted. Then, mapping to human and virus reference genome were conducted. The HBV integration sites would be detected by clipping reveals structure (CTEST)⁵⁷ using an assembly-mapping-searching-assembly-alignment procedure. We explored the common HBV integration sites between HCC and ICCA components within same H-ChC samples.

Data availability. Sequence data have been deposited at the European Genome-phenome Archive (EGA), which is hosted by the EBI and the CRG, under accession number EGAS00001002783. All the data are available within the article, supplementary information, and supplementary data file, or available from the authors on request.

Received: 25 December 2016 Accepted: 2 February 2018

Published online: 01 March 2018

References

1. Fitzmaurice, C. et al. The global burden of cancer 2013. *JAMA Oncol.* **1**, 505–527 (2015).
2. Ferlay, J. et al. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *Int. J. Cancer* **136**, E359–E386 (2015).
3. Wang, A. Q. et al. Combined hepatocellular cholangiocarcinoma: controversies to be addressed. *World J. Gastroenterol.* **22**, 4459–4465 (2016).
4. Cazals-Hatem, D. et al. Clinical and molecular analysis of combined hepatocellular-cholangiocarcinomas. *J. Hepatol.* **41**, 292–298 (2004).
5. Kassahun, W. T. & Hauss, J. Management of combined hepatocellular and cholangiocarcinoma. *Int. J. Clin. Pract.* **62**, 1271–1278 (2008).
6. Allen, R. A. & Lisa, J. R. Combined liver cell and bile duct carcinoma. *Am. J. Pathol.* **25**, 647–655 (1949).
7. Ishak, K. G., Anthony, P. P. & Sobin, L. H. *International Histological Classification of Tumours* 2nd edn (Springer, Hong Kong, 1994).
8. Alexandrov, L. B. et al. Signatures of mutational processes in human cancer. *Nature* **500**, 415–421 (2013).
9. Roth, A. et al. PyClone: statistical inference of clonal population structure in cancer. *Nat. Methods* **11**, 396–398 (2014).
10. Eirew, P. et al. Dynamics of genomic clones in breast cancer patient xenografts at single-cell resolution. *Nature* **518**, 422–426 (2015).
11. Murugaesu, N. et al. Tracking the genomic evolution of esophageal adenocarcinoma through neoadjuvant chemotherapy. *Cancer Discov.* **5**, 821–831 (2015).
12. Xue, R. et al. Variable intra-tumor genomic heterogeneity of multiple lesions in patients with hepatocellular carcinoma. *Gastroenterology* **150**, 998–1008 (2016).
13. Gerlinger, M. et al. Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *N. Engl. J. Med.* **366**, 883–892 (2012).
14. Miao, R. et al. Identification of prognostic biomarkers in hepatitis B virus-related hepatocellular carcinoma and stratification by integrative multi-omics analysis. *J. Hepatol.* **61**, 840–849 (2014).
15. Zhang, Z. et al. Interleukin-11 promotes the progress of gastric carcinoma via abnormally expressed versican. *Int. J. Biol. Sci.* **8**, 383–393 (2012).

16. Oktem, G. et al. Cancer stem cell differentiation: TGFβ1 and versican may trigger molecules for the organization of tumor spheroids. *Oncol. Rep.* **32**, 641–649 (2014).
17. Jung, W. & Kim, B. H. Cholangiolocellular carcinoma with satellite nodules showing intermediate differentiation. *Clin. Mol. Hepatol.* **21**, 183–186 (2015).
18. Akiba, J. et al. Clinicopathologic analysis of combined hepatocellular-cholangiocarcinoma according to the latest WHO classification. *Am. J. Surg. Pathol.* **37**, 496–505 (2013).
19. Zhang, F. et al. Combined hepatocellular cholangiocarcinoma originating from hepatic progenitor cells: immunohistochemical and double-fluorescence immunostaining evidence. *Histopathology* **52**, 224–232 (2008).
20. Kim, H. et al. Primary liver carcinoma of intermediate (hepatocyte-cholangiocyte) phenotype. *J. Hepatol.* **40**, 298–304 (2004).
21. Dolle, L. et al. EpCAM and the biology of hepatic stem/progenitor cells. *Am. J. Physiol. Gastrointest. Liver Physiol.* **308**, G233–G250 (2015).
22. Schnell, U., Cirulli, V. & Giepmans, B. N. EpCAM: structure and function in health and disease. *Biochim. Biophys. Acta* **1828**, 1989–2001 (2013).
23. Zhi, X. S., Xiong, J., Zi, X. Y. & Hu, Y. P. The potential role of liver stem cells in initiation of primary liver cancer. *Hepatol. Int.* **10**, 893–901 (2016).
24. Carpino, G. et al. Stem/progenitor cell niches involved in hepatic and biliary regeneration. *Stem Cells Int.* **2016**, 3658013 (2016).
25. Kohn-Gaone, J., Gogoi-Tiwari, J., Ramm, G. A., Olynyk, J. K. & Tirnitz-Parker, J. E. The role of liver progenitor cells during liver regeneration, fibrogenesis, and carcinogenesis. *Am. J. Physiol. Gastrointest. Liver Physiol.* **310**, G143–G154 (2016).
26. Oikawa, T. Cancer stem cells and their cellular origins in primary liver and biliary tract cancers. *Hepatology* **64**, 645–651 (2016).
27. Lee, J. S. et al. A novel prognostic subtype of human hepatocellular carcinoma derived from hepatic progenitor cells. *Nat. Med.* **12**, 410–416 (2006).
28. Kokuryo, T., Yokoyama, Y. & Nagino, M. Recent advances in cancer stem cell research for cholangiocarcinoma. *J. Hepatobiliary Pancreat. Sci.* **19**, 606–613 (2012).
29. Iwahashi, S. et al. High expression of cancer stem cell markers in cholangiolocellular carcinoma. *Surg. Today* **43**, 654–660 (2013).
30. Komuta, M. et al. Clinicopathological study on cholangiolocellular carcinoma suggesting hepatic progenitor cell origin. *Hepatology* **47**, 1544–1556 (2008).
31. Theise, N. D., Nakashima, O., Park, Y. N., Nakanuma, Y. in: *WHO Classification of Tumours of the Digestive System* 4th edn (eds Bosman, F. T., Carneiro, F., Hruban, R. H. & Theise, N. D.) 225–227 (IARC Press, Lyon, 2010).
32. Terada, T. Combined hepatocellular-cholangiocarcinoma with stem cell features, ductal plate malformation subtype: a case report and proposal of a new subtype. *Int. J. Clin. Exp. Pathol.* **6**, 737–748 (2013).
33. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
34. DePristo, M. A. et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **43**, 491–498 (2011).
35. Sherry, S. T. et al. dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.* **29**, 308–311 (2001).
36. Genomes Project, C. et al. An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56–65 (2012).
37. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **38**, e164 (2010).
38. Cibulskis, K. et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat. Biotechnol.* **31**, 213–219 (2013).
39. Saunders, C. T. et al. in Strelka: accurate somatic small-variant calling from sequenced tumor-normal sample pairs. *Bioinformatics* **28**, 1811–1817 (2012).
40. Ledford, H. Big science: the cancer genome challenge. *Nature* **464**, 972–974 (2010).
41. Gaujoux, R. & Seoighe, C. A flexible R package for nonnegative matrix factorization. *BMC Bioinform.* **11**, 367 (2010).
42. Alexandrov, L. B., Nik-Zainal, S., Wedge, D. C., Campbell, P. J. & Stratton, M. R. Deciphering signatures of mutational processes operative in human cancer. *Cell Rep.* **3**, 246–259 (2013).
43. Forbes, S. A. et al. COSMIC: mining complete cancer genomes in the catalogue of somatic mutations in cancer. *Nucleic Acids Res.* **39**, D945–D950 (2011).
44. Alexandrov, L. B. et al. Signatures of mutational processes in human cancer. *Nature* **500**, 415–421 (2013).
45. Carter, S. L. et al. Absolute quantification of somatic DNA alterations in human cancer. *Nat. Biotechnol.* **30**, 413–421 (2012).
46. Gerlinger, M. et al. Genomic architecture and evolution of clear cell renal cell carcinomas defined by multiregion sequencing. *Nat. Genet.* **46**, 225–233 (2014).
47. Roth, A. et al. PyClone: statistical inference of clonal population structure in cancer. *Nat. Methods* **11**, 396–398 (2014).
48. Futreal, P. A. et al. A census of human cancer genes. *Nat. Rev. Cancer* **4**, 177–183 (2004).
49. Vogelstein, B. et al. Cancer genome landscapes. *Science* **339**, 1546–1558 (2013).
50. Kandath, C. et al. Mutational landscape and significance across 12 major cancer types. *Nature* **502**, 333–339 (2013).
51. Tamborero, D. et al. Comprehensive identification of mutational cancer driver genes across 12 tumor types. *Sci. Rep.* **3**, 2650 (2013).
52. Dees, N. D. et al. MuSiC: identifying mutational significance in cancer genomes. *Genome Res.* **22**, 1589–1598 (2012).
53. Wendt, M. C. et al. PathScan: a tool for discerning mutational significance in groups of putative cancer genes. *Bioinformatics* **27**, 1595–1602 (2011).
54. Kanehisa, M. & Goto, S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**, 27–30 (2000).
55. Boeva, V. et al. Control-FREEC: a tool for assessing copy number and allelic content using next-generation sequencing data. *Bioinformatics* **28**, 423–425 (2012).
56. MacDonald, J. R., Ziman, R., Yuen, R. K., Feuk, L. & Scherer, S. W. The database of genomic variants: a curated collection of structural variation in the human genome. *Nucleic Acids Res.* **42**, D986–D992 (2014).
57. Wang, J. et al. CREST maps somatic structural variation in cancer genomes with base-pair resolution. *Nat. Methods* **8**, 652–654 (2011).

Acknowledgements

We thank Elsevier Language Editing Services for their help in English language revision of this manuscript. We also thank for the support of CAMS Innovation Fund for Medical Science (CIFMS) (2017-12M-4-003), International Science and Technology Cooperation Projects (2015DFA30650 and 2016YFE0107100), Capital Special Research Project for Health Development (2014-2-4012) and Beijing Nature Science Foundation for Young Scholars Project (7164293).

Author contributions

A.W. designed the study and wrote the manuscript. L.W. and J.L. analyzed exome-sequencing data. J.B. collected all samples. L.X. conducted the manuscript revision. L.H. and W.W. conducted pathological diagnosis and immunohistochemistry. Y.W. and S.C.R. reviewed and edited this manuscript. Y.G. helped further analyze the sequencing data. X.S. and H.Z. co-ordinated and provided financial support for this work.

Additional information

Supplementary Information accompanies this paper at <https://doi.org/10.1038/s41467-018-03276-y>.

Competing interests: The authors declare no competing interests.

Reprints and permission information is available online at <http://npq.nature.com/reprintsandpermissions/>

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018