ORIGINAL ARTICLE

Diverse modes of clonal evolution in HBV-related hepatocellular carcinoma revealed by single-cell genome sequencing

Meng Duan[1, *], Junfeng Hao[2, *], Sijia Cui[3, *], Daniel L Worthley[4], Shu Zhang[1], Zhichao Wang[1], Jieyi Shi[1], Longzi Liu[1], Xiaoying Wang[1], Aiwu Ke[1], Ya Cao[5], Ruibin Xi[6], Xiaoming Zhang[7], Jian Zhou[1, 8], Jia Fan[1, 8], Chong Li[2], Qiang Gao[1, 9]

[1]Department of Liver Surgery and Transplantation, Liver Cancer Institute, Zhongshan Hospital, and Key Laboratory of Carcinogenesis and Cancer Invasion (Ministry of Education), Fudan University, Shanghai 200032, China; [2]Core Facility for Protein Research, Institute of Biophysics, Chinese Academy of Sciences, Beijing 100101, China; [3]Hangzhou Cancer Institute, Hangzhou Cancer Hospital, Hangzhou, Zhejiang 310002, China; [4]Cancer Theme, South Australian Health and Medical Research Institute and Department of Medicine, University of Adelaide, Adelaide, SA, Australia; [5]Cancer Research Institute, Xiangya School of Medicine, Central South University, Changsha, Hunan 410013, China; [6]School of Mathematical Sciences and Center for Statistical Science, Peking University, Beijing 100871, China; [7]Key Laboratory of Molecular Virology & Immunology, Institute Pasteur of Shanghai, Chinese Academy of Sciences, Shanghai 200032, China; [8]Institutes of Biomedical Sciences, Fudan University, Shanghai 200032, China; [9]State Key Laboratory of Genetic Engineering, Fudan University, Shanghai 200433, China

Hepatocellular carcinoma (HCC) is a cancer of substantial morphologic, genetic and phenotypic diversity. Yet we do not understand the relationship between intratumor heterogeneity and the associated morphologic/histological characteristics of the tumor. Using single-cell whole-genome sequencing to profile 96 tumor cells (30-36 each) and 15 normal liver cells (5 each), collected from three male patients with HBV-associated HCC, we confirmed that copy number variations occur early in hepatocarcinogenesis but thereafter remain relatively stable throughout tumor progression. Importantly, we showed that specific HCCs can be of monoclonal or polyclonal origins. Tumors with confluent multinodular morphology are the typical polyclonal tumors and display the highest intratumor heterogeneity. In addition to mutational and copy number profiles, we dissected the clonal origins of HCC using HBV-derived foreign genomic markers. In monoclonal HCC, all the tumor single cells exhibit the same HBV integrations, indicating that HBV integration is an early driver event and remains extremely stable during tumor progression. In addition, our results indicated that both models of metastasis, late dissemination and early seeding, have a role in HCC progression. Notably, early intrahepatic spreading of the initiating clone leads to the formation of synchronous multifocal tumors. Meanwhile, we identified a potential driver gene ZNF717 in HCC, which exhibits a high frequency of mutation at both single-cell and population levels, as a tumor suppressor acting through regulating the IL-6/STAT3 pathway. These findings highlight multiple distinct tumor evolutionary mechanisms in HCC, which suggests the need for specific treatment strategies.

Keywords: liver cancer; single-cell sequencing; HBV integration; tumor morphology; ZNF717

*These three authors contributed equally to this work.
Correspondence: Qiang Gao[a], Chong Li[b], Jia Fan[c]
[a]Tel: +86-21-64037181
E-mail: gaoqiang@fudan.edu.cn
[b]Tel: +86-10-64888424
E-mail: lichong@moon.ibp.ac.cn
[c]E-mail: fan.jia@zs-hospital.sh.cn

Introduction

Hepatocellular carcinoma (HCC) is the third leading cause of cancer-related deaths worldwide, and hepatitis B virus (HBV) infection is one of its leading causes [1]. During the past several years, next-generation sequencing studies using bulk tumor samples have revealed con-

siderable intratumor molecular and genetic heterogeneity in HCC [2-6]. Such intratumor heterogeneity poses a great challenge for tumor characterization and therapeutic management of HCC patients [7]. As is well known, tumor initiation and evolution are mediated by sequential genetic alterations in single cells. Single-cell sequencing has the potential to provide new insights into cancer biological diversity that were difficult to resolve in genomic data from bulk tumor samples [8, 9]. Of note, recent single-cell genome (SCG)/exome sequencing in many solid tumors, including breast [10, 11], renal [12], bladder [13], and colon cancers [14] has provided strong evidence for clonal tumor evolution [15]. All these reports showed that single cells within a tumor shared common founder genetic alterations with intratumor subpopulations occasionally detected, strongly favoring the evolution from a common single cell while arguing against field effects of mutagenesis and multicell origins.

Likewise, single-cell triple omics sequencing applied to 25 tumor cells derived from a human HBV-related HCC identified two subpopulations within these cells [16]. This study sequenced a single-nodular type HCC, providing informative data for deciphering intratumor heterogeneity of HCC at single-cell level, but left some unanswered questions. It is well known that HCC is defined as an extraordinarily heterogeneous disease due to its morphological and histological diversity. For typical examples, HCC with macroscopic portal vein tumor thrombus (PVTT), HCC with multifocal tumors, and the confluent multinodular HCC are very common and of great clinical significance. Previous studies have shown that distinct signatures of gene expression were closely related to morphological heterogeneity in HCC [17-19], supporting the notion that genotypes affect morphological phenotypes. However, the true extent of intratumor heterogeneity and clonal diversity correlated with morphologic and histological features has not been clarified yet in HCC.

The aim of this study was to systematically investigate intratumor heterogeneity in HCC by correlating tumor morphology with single-cell genomic landscape. We applied SCG sequencing using MALBAC [20] to a total of 96 tumor cells (30-36 each) and 15 normal liver cells (5 each), collected from three HBV-related male HCC patients with distinct tumor morphologic features. Our results illustrate that multiple distinct tumor evolutionary models that are closely related to morphological phenotypes may operate in HCC.

Results

Patients and clinical characteristics

Given that HCC shows various morphological and pathological appearances [21], we selected three morphological types that are of great clinical significance. First, macroscopic PVTT, detected in up to 30%-62% of HCC patients, is one of the most unfavorable prognostic factors, and these patients have a median survival of 3 months without treatment [22]. Second, different from other cancers, multifocal tumors are very common and occur in > 50% of HCC, arising as independent tumors (i.e., *de novo* tumor) or as intrahepatic metastases of the primary cancer [3, 4, 23]. However, their molecular pathogenesis remains elusive. Third, the confluent multinodular morphology is characterized by a cluster of confluent nodules, where each nodule has a clear margin or capsule and the overall tumor is lobulated [24]. This morphology is associated with tumor aggressiveness and is most prevalent in HCC with a HBV background compared with other etiologies [17]. Therefore, we selected a single-nodular HCC with PVTT (H-PT), a multifocal HCC (H-MF), and a confluent multinodular HCC (H-CM) for SCG sequencing (Figure 1A; Supplementary information, Figure S1 and Table S1). The three male patients were diagnosed with HBV-related HCC and underwent curative resection without any adjuvant therapy before. To systematically assess tumor spatial heterogeneity, we sliced the tumors into pieces and obtained geographically separated samples from each cut face (Figure 1B). In H-PT, 2 tumor cells from the PVTT, along with 28 cells from the tumor mass, were sorted and sequenced. In H-MF, 12 and 24 tumor cells were sorted from the two synchronous tumors H-MFa and H-MFb, respectively. In H-CM, a total of 30 tumor cells were obtained, covering each intratumor nodule. Cells were sequenced at a mean of $13.2\times$ depth and 76.0% coverage (Supplementary information, Figures S2 and S3), and single-nucleotide variations (SNVs, both in coding and noncoding regions) and copy number variations (CNVs) were detected. HCC cells were clearly different from normal liver cells as shown by neighbor-joining tree analysis of CNVs, indicating that no normal cells were mistakenly regarded as tumor cells (Figure 1C).

HBV integration patterns indicate the monoclonal origin of H-PT and H-MF cases

HBV integration plays a causal role in HCC occurrence and progression [25]. Integrated HBV DNA sequences and episomal HBV genomes have been found in 85%-90% of HBV-related HCC. In this study, H-PT and H-MF were both HBV-related HCC with high load of serum HBV-DNA. We also performed population sequencing of the bulk tumors ($50\times$) and matched blood samples ($30\times$) of the two cases. In H-PT, we identified 5 HBV integration sites by analyzing bulk sequencing data
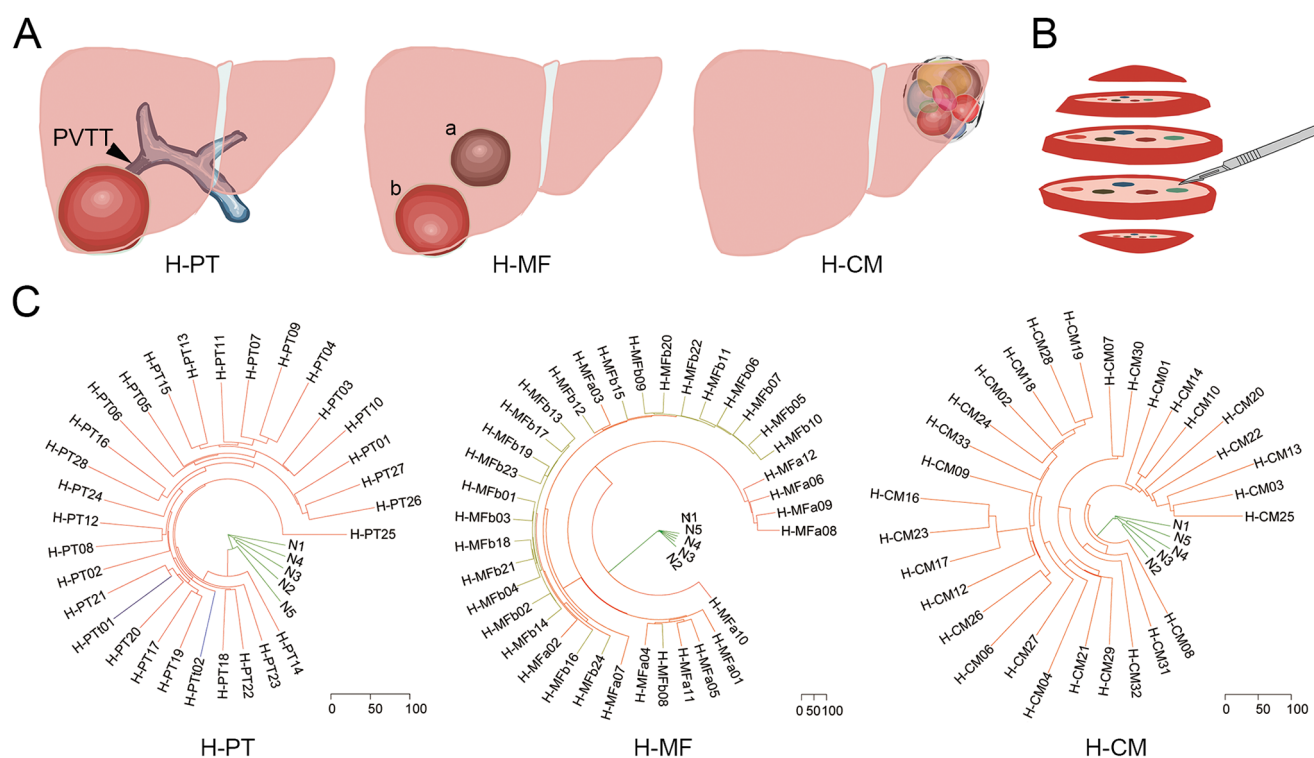
Figure 1 Single-cell whole-genome sequencing of three HBV-related HCC patients. **(A)** Schematic illustration of the tumors. Five normal liver cells were sorted from the adjacent liver tissues in each case. PVTT, portal vein tumor thrombus. **(B)** Geographical sampling approach for each tumor. We evenly sliced the tumors to pieces and obtained geographically separated samples from each cut face. **(C)** Neighbor-joining tree constructed using CNV data by Euclidean distance. Tumor cells were represented in red and normal liver (N1-N5) cells in green. In H-PT, two PVTT cells were in blue. In H-MF, tumor cells from H-MFb were in brown.

(Figure 2A). All the 5 HBV integrations were present in each tumor cell and PVTT cell, but none in normal cells, as revealed by SCG sequencing and Sanger validation (Figure 2A and Supplementary information, Table S2), indicating a monoclonal expansion of HBV-integrated tumor-initiating cells. Consistently, no additional HBV integrations showing common or private patterns were detected in those cells by SCG sequencing. The results also indicated that PVTT, a feature for advanced tumor stages, was derived from tumor clonal dissemination at single-cell level.

In H-MF, the synchronous tumors H-MFa and H-MFb had the exact same 6 HBV integration sites as detected by bulk sequencing (Figure 2B). Likewise, all the tumor cells from H-MFa and H-MFb exhibited the same 6 HBV integrations as revealed by SCG sequencing and Sanger validation, without any new integrations detected (Figure 2B and Supplementary information, Table S2). Thus, the two separated tumors in H-MF most likely originated from a single clone that had gained sufficient growth advantage after early intrahepatic spreading, mimicking

a pair of monozygotic twins. During patient selection process, we assumed that the two separated tumors with comparable size (6 cm and 7 cm in diameter each) in H-MF should represent multicentric carcinogenesis (i.e., *de novo* tumor) under a chronic hepatitis background. However, based on the findings that multicentric tumor exhibited exclusively different HBV integrations and highly different genomic landscape [4, 19], this H-MF case could be a model of early seeding (as opposed to the late dissemination model), where the two separated tumors are from one founding clone whose early intrahepatic spreading led to the formation of two large tumors.

For intratumor diversity, previous studies mainly focused on comparing the host endogenous genomic alterations [4, 5, 16]. Herein, we precisely dissected the clonal origins of human cancer using HBV-derived foreign genomic markers. Annotation of all the 11 HBV integrations showed that they all located in intergenic or intronic areas, where the X and S sequences of HBV may lead to chromosome instability (Supplementary information, Figure S4 and Table S2). The remarkable uniformity of
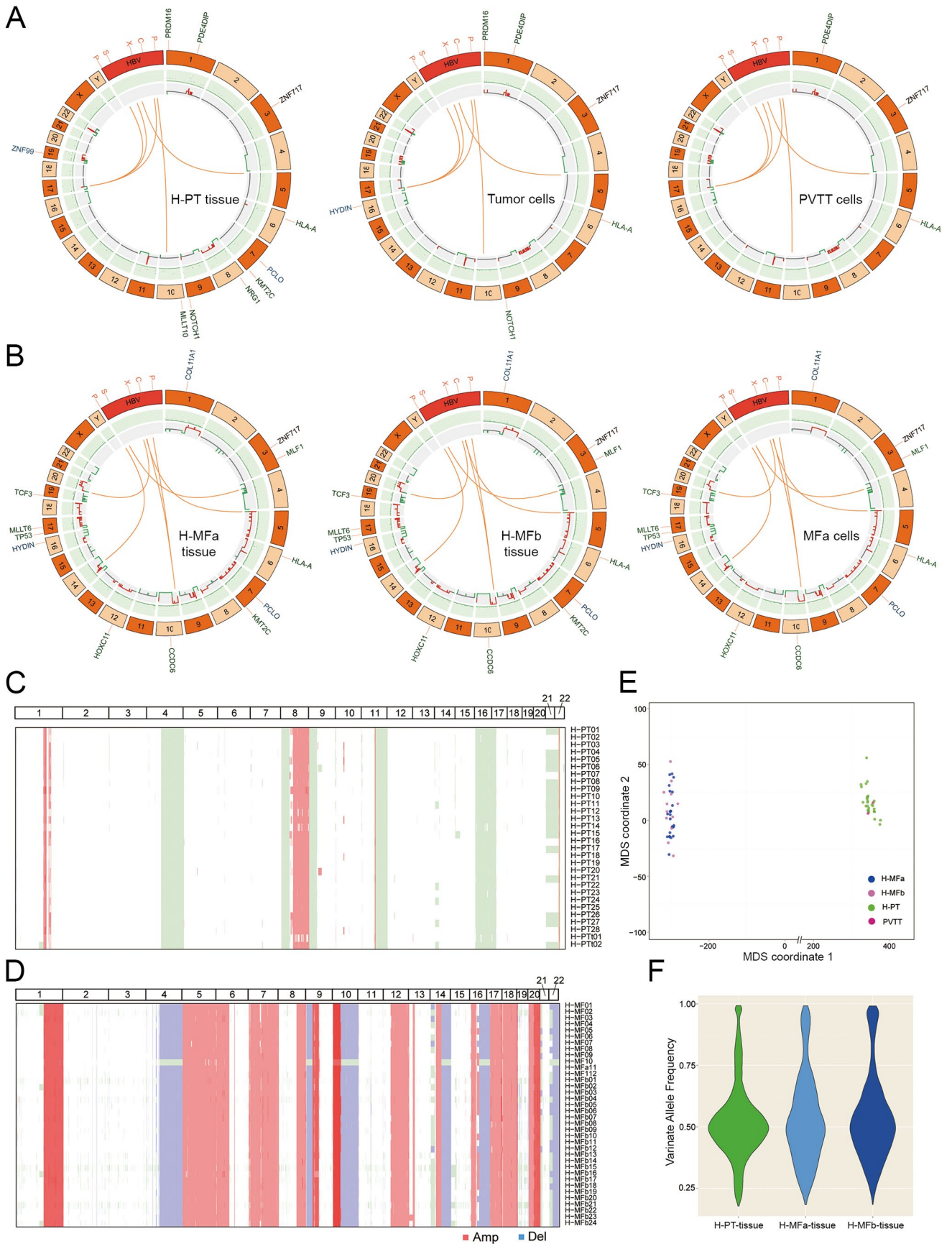
HBV integration among spatially separated tumor cells within a tumor strongly supported that HBV integration was an early event during hepatocarcinogenesis [19] and remained extremely stable during tumor evolution. Comparing the number of HBV integration per individual, the 5 or 6 HBV integrations per individual in our analysis is similar to the 5-7 HBV integrations per individual reported in previous large-scale deep-sequencing studies [26, 27]. In addition to HBV integrations, a large number of reads that are uniquely mapped to the HBV genome were detected in those single cells from H-PT and H-MF, consistent with the presence of high load of serum HBV-DNA.

CNV profiles confirm the monoclonal origin of H-PT and H-MF cases

HBV integrations may not be sufficient to dissect the clonal diversity of HCC, considering that various selective pressure including microenvironment and persistent chronic hepatitis may lead to acquisition of driver alterations and thus emergence of subclones with survival advantage. Thus, single-cell genomic alterations were used to assess intratumor clonal heterogeneity for H-PT and H-MF cases.

MALBAC is well suited for the simultaneous characterization of SNVs and CNVs, in particular for profiling CNVs [28]. Based on our sequencing depth, we can precisely measure the CNV profiles of each cell. The error rates of almost all the single-cell samples were < 0.1% except H-CM22 and the proportion of Q20 and Q30 are all > 80%, which was suggestive of the stable sequencing result (Supplementary information, Figures S5 and S6). The statistically significant CNVs in each patient were illustrated in Supplementary information, Figure S7, revealing large significant CNVs of the single cells in the chromosomes that may play crucial roles in HCC. In particular, both H-PT and H-MF harbored TP53 loss at chromosome 17q and MYC amplification at chromosome 8q. Of note, tumor cells from H-PT and H-MF showed highly reproducible CNV profiles (H-PT mean $R^2 = 0.89$ and H-MF mean $R^2 = 0.88$), representing a monoclonal population in each case (Figure 2C and 2D; Supplementary information, Table S3). These data demonstrated that chromosome rearrangements occurred early during tumor initiation [10, 11], presented as punctuated bursts of copy number evolution, to generate a specific CNV profile per case, which was maintained throughout tumor formation and metastasis. Recent studies have reported substantial intratumor heterogeneity in HCC using multiregional sequencing [3-5]. We assumed that such intratumor heterogeneity in HCC was mainly attributed to somatic mutations that evolved gradually overtime in monoclonal populations. Multidimensional scaling analysis based on CNV profiles showed that tumor cells in each case were tightly clustered without any significant subclusters, further indicating that both H-PT and H-MF may be of monoclonal origin (Figure 2E). Meanwhile, tumor cells from H-PT and H-MF were obviously separated, indicating distinct CNV profiles. Analyzing the bulk exome data for distributions of mutant allele frequencies revealed that mutations for H-PT, H-MFa and H-MFb were all near a frequency of 50%, suggesting a single dominant clone within each tumor (Figure 2F).

Confluent multinodular HCC is of polyclonal origin

Previous studies have reported that confluent multinodular tumors display the highest heterogeneity of DNA ploidy compared to other types of nodular-type HCC [29]. However, the clonal relationship among these confluent nodules within a tumor remains largely unknown. SCG sequencing provides a good opportunity to elucidate the clonal diversity of these clearly separated intratumor nodules. As shown in Figure 3A, CNV heatmaps of H-CM single cells were quite different from those of two

Figure 2 Single-cell and bulk sequencing of two HCCs of monoclonal origin. **(A)** Circos plots of SNVs and CNVs detected by bulk sequencing (left panel) and SCG sequencing of cells from tumor mass (middle panel) or PVTT (right panel) in H-PT. Cancer genes that are documented in the Cancer Gene Census database or have been reported as HCC driver mutations were indicated on the outer ring. A total of 5 HBV integrations were detected by bulk sequencing, which were all reproduced in single cells from tumor mass and PVTT. **(B)** Circos plots of SNVs and CNVs detected by bulk sequencing of H-MFa (left panel) and H-MFb (middle panel) tumors, as well as SCG sequencing of tumor cells from H-MFa (right panel). The exact same 6 HBV integrations were identified in H-MFa and H-MFb by bulk sequencing, which were all reproduced in single cell from each tumor (plot for SCG sequencing of tumor cells within H-MFb was not shown, which was highly similar to that of H-MFa). Cancer genes that are documented in the Cancer Gene Census database or have been reported as HCC driver mutations were indicated on the outer ring. **(C)** Heatmap of 30 tumor single-cell CNV profiles of H-PT. Tumor cells showed highly reproducible CNV profiles (mean $R^2 = 0.89$). **(D)** Heatmap of 36 tumor single-cell CNV profiles of H-MF. Tumor cells showed highly reproducible CNV profiles (mean $R^2 = 0.88$). **(E)** Multidimensional scaling analysis of tumor single cells based on CNV profiles. No obvious subclones could be found in each case. **(F)** Mutation allele frequency distributions for SNV profiles by bulk sequencing. Mutations for H-PT, H-MFa, and H-MFb were all near a frequency of 50%, suggesting a single dominant clone in each tumor.

monoclonal tumors (H-PT and H-MF; Figure 2C and 2D). Each H-CM single cell has a distinct CNV profile and no large copy number segments were shared by all the tumor cells (mean $R^2 = 0.27$; Figure 3A and 3B), indicating a polyclonal origin of these intratumor nodules. Multidimensional scaling analysis roughly clustered those cells into several clones (Figure 3C). Phylogenetic tree analysis indicated that there were almost no trunk mutations (Figure 3D), consistent with the above results. Collectively, the field effect of mutagenesis and multicell origins in hepatocarcinogenesis resulted in independent genetic lineages that hardly share any founder genetic alterations, as exemplified by H-CM. Furthermore, we attempted to provide additional supportive evidence for

its polyclonal origin from the HBV integration patterns. The serum HBV-DNA of H-CM was negative, indicating an occult HBV status. Coincidently, no HBV integrations (Figure 3B) and almost no episomal HBV-DNA reads were detected in tumor cells or normal cells, suggesting that HBV integrations may not be a causal event during hepatocarcinogenesis in this case.

Despite the relatively low coverage, single cells from H-PT and H-MF harbored high-prevalence recurrent nonsynonymous SNVs (12.7% and 5.1% of nonsynonymous SNVs were detected in > 50% tumor cells, respectively), whereas single cells from H-CM did not contain any common SNVs (present in > 50% of cells), strongly supporting its polyclonal origin (Figure 4A). Similar to
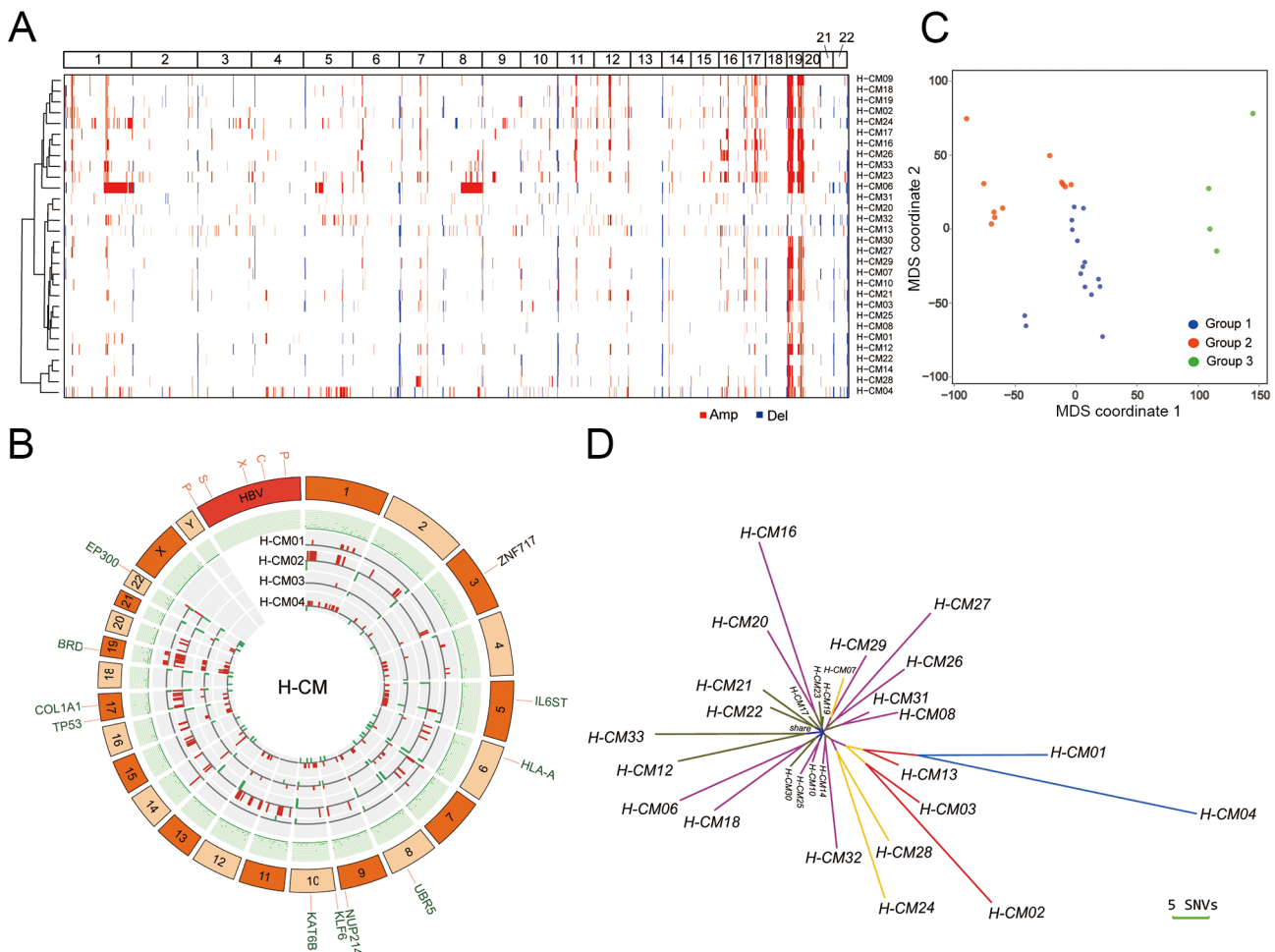

Figure 3 Polyclonal origin in a confluent multinodular HCC. **(A)** Heatmap of 30 tumor single-cell CNV profiles of H-CM. No common CNVs were shared by all the tumor cells. **(B)** Circos plot of SNVs and CNVs detected by SCG sequencing of H-CM. Four representative cells from different intratumor nodules were shown. No HBV integrations were detected in this case. Cancer genes that are documented in the Cancer Gene Census database or have been reported as HCC driver mutations were indicated on the outer ring. **(C)** Multi-dimensional scaling plot of tumor single cells based on CNV profiles. Tumor cells from H-CM were clustered into three groups. **(D)** Phylogenetic tree of H-CM based on mutational profile. The length of each line was proportional to the number of non-synonymous mutations.

coding SNV distribution in tumor single cells, H-CM harbored 0.2% common noncoding SNVs ($\geq$ 50%) and no somatic noncoding SNVs shared by all the tumor single cells. In contrast to H-CM, H-PT harbored 6.1% and H-MF harbored 4.0% common SNVs in the noncoding regions, also supporting the conclusion of the monoclonal origins of these two tumors. In addition, cells from a single clone displayed similar CNV profiles, whereas cells of polyclonal origin showed different CNV profiles which resulted in a much lower frequency of shared CNVs among them. As such, the two monoclonal tumors

reflected a well-known punctuated evolution model of CNV, whereas H-CM reflected a cumulative model that could be explained by its polyclonal origin (Figure 4B). Altogether, we conclude that these confluent nodules within H-CM were of polyclonal origin, each borne from distinct cancer-initiating cells.

ZNF717 acts as a new tumor suppressor in HCC

Next, we sought to identify coding SNVs that potentially had functional impacts. The number of single cells in each case were above 30, yielding a total of > 330×
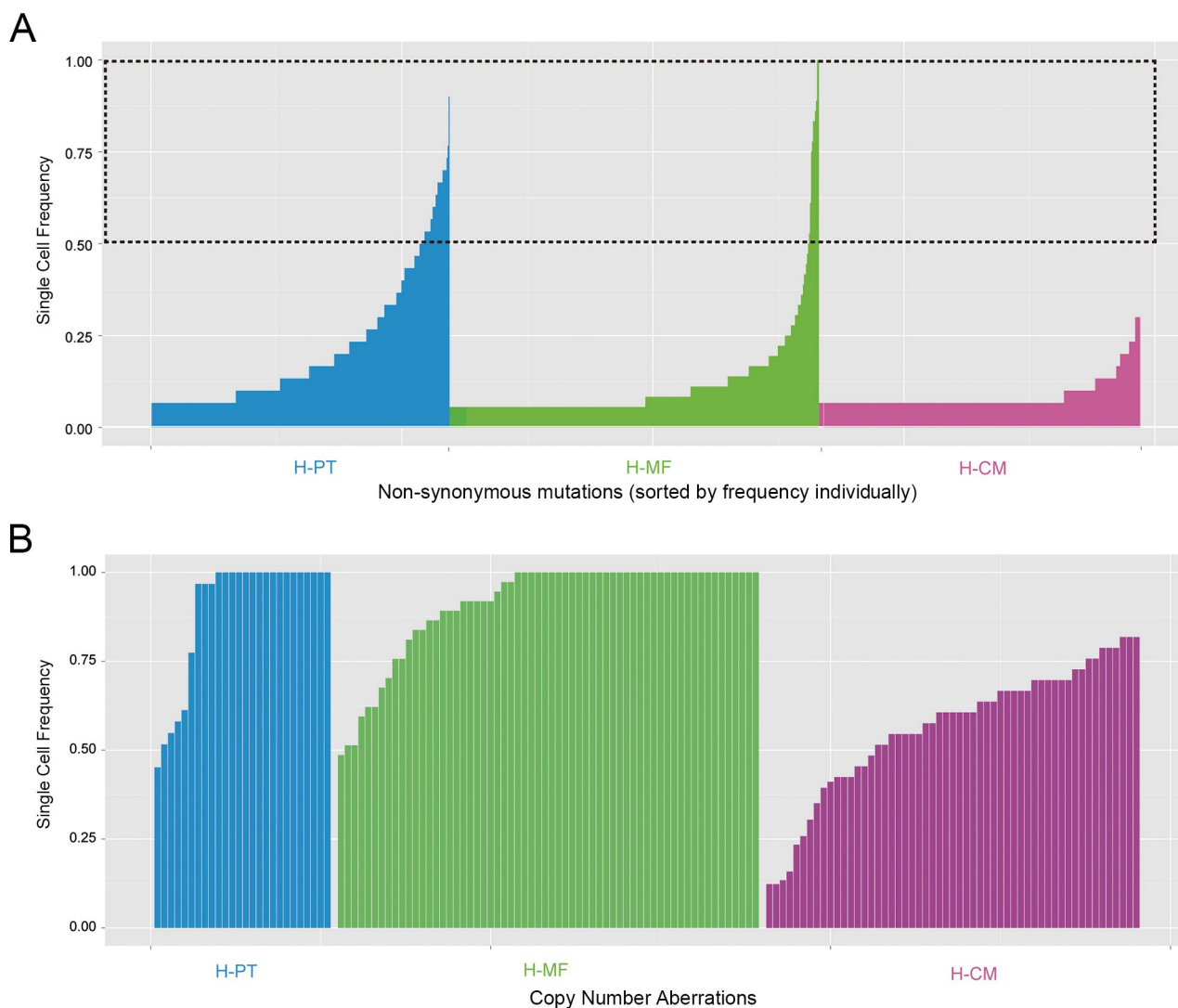

Figure 4 Patterns of mutation and copy number profiles. **(A)** Bar plots showing mutations shared by two or more cells in each case. Common mutations could be detected in monoclonal tumors H-PT and H-MF, but were absent in polyclonal tumor H-CM. Mutations accumulate gradually over time in both monoclonal and polyclonal tumors. **(B)** Bar plots showing CNVs shared by two or more cells in each case. CNV profiles were highly similar and occurred in punctuated burst pattern in monoclonal tumors H-PT and H-MF, whereas CNV profiles were highly different and showed a cumulative pattern in polyclonal tumor H-CM.

depth each, which enabled SNV calling on most exome regions. Using stringent criteria, we identified 275, 349 and 302 coding SNVs in H-PT, H-MF and H-CM, respectively (Supplementary information, Table S4). The mutated genes that are documented in the Cancer Gene Census dataset [30] or have been reported as HCC driver mutations [31-34] were listed in the circos plots (Figures 2 and 3). Using OncodriveCLUST [35], the mountain gene ZNF717 located at 3p12.3 was identified to be of particular interest, in which mutations recurred in the three cases with the highest frequency (Supplementary information, Figure S8A). ZNF717 mutation rate was 90% in H-MF and 72% in H-PT, achieving an average of 80% in the monoclonal origin tumor single cells, but only 17% in the polyclonal origin tumor H-CM. ZNF717 mutations were also detected in 47 out of 88 HCCs in an early study but at different sites (Supplementary information, Figure S8B) [36], indicating that ZNF717 alterations were mountains at both individual and population levels. Protein encoded by ZNF717 belongs to the zinc-finger family, which is known to play key roles in regulating gene expression [36]. However, the exact role of ZNF717 in tumor pathogenesis remains unknown.

Next, we used shRNA to knock down ZNF717 in two HCC cell lines HepG2 and Huh7. Our results showed that ZNF717 knockdown in HCC cells significantly promoted cell growth, adhesion, migration and invasion, while potently inhibiting cell apoptosis *in vitro*, as compared to the control (Supplementary information, Figure S9). We further examined STAT3-regulated adhesion and migration molecules such as MMP2, CD44 and ITGA3, and their expression was significantly increased by ZNF717 knockdown (Supplementary information, Figure S10). In xenograft HCC models, ZNF717 knockdown accelerated both the subcutaneous and orthotropic tumor growth, and significantly correlated with reduced survival time of tumor-bearing mice (Figure 5). Clinically, low expression of ZNF717 was revealed to be significantly correlated with dismal prognosis in two independent patient cohorts (Figure 6). First, using publically available microarray dataset ($n = 289$) [37, 38], we revealed that low expression of ZNF717 mRNA correlated with reduced survival ($P = 0.003$, Figure 6B). Second, using tissue microarrays containing 236 consecutive HCC patients [4], we determined ZNF717 protein expression by immunohistochemistry. We found that ZNF717-positive staining was mainly detected in cell nuclei. Normal liver cells showed uniformly high expression of ZNF717, whereas HCC cells showed significant downregulation of ZNF717 expression (Figure 6A). Likewise, downregulation of ZNF717 expression correlated with poor survival ($P = 0.032$, Figure 6B). Collectively, the results indicated that ZNF717 may act as a tumor suppressor in HCC.

To further identify the underlying mechanism of tumor suppressing function of ZNF717 in HCC, gene expression profiles and the activities of a total of 45 signal transduction pathways were compared between shZNF717 and control HepG2 cells. The results showed that STAT3 pathway was the leading pathway that was significantly activated in shZNF717 HepG2 cells (4.2-fold upregulation vs control; Figure 7A), and similar results were obtained in tumor samples from HCC patients (Supplementary information, Figure S11). IL-6 was found to be one of the top genes that was significantly upregulated in shZNF717 HepG2 cells (17.0-fold upregulation vs control; Figure 7B), suggesting that the IL-6/STAT3 pathway may be downstream of ZNF717 in HCC. Furthermore, increased levels of IL-6 (Figure 7C), upregulation of STAT3 target gene expression (Figure 7D), and elevated expression of STAT3 (Figure 7E) were observed in HCC cell lines following ZNF717 knockdown.

The mechanisms of ZNF717 in regulating STAT3 has not been investigated before. Herein, we found that ZNF717 bound to three regions of the STAT3 promoter, and the binding to the −1 171 to −1 022 region was the most significant (an average of 937% enrichment; Supplementary information, Figure S12A). CRISPR/Cas9 was applied to abrogate the −1 171 to −1 022 segment of the STAT3 promoter. In heterozygous mutant HepG2 or Huh7 cells (Mut), ZNF717 was only associated with a single allele of STAT3 promoter (Supplementary information, Figure S12B and S12C). Similarly, in DNase I digestion assay, abrogation of STAT3 promoter remarkably enhanced the chromatin accessibility of STAT3 locus in Mut HepG2 or Huh7 cells compared to WT HCC cells (Supplementary information, Figure S12D). Furthermore, overexpression of ZNF717 in HepG2 or Huh7 cells (oeZNF717) resulted in an average of 83% decrease in STAT3 expression compared to vector-transfected (vec) HCC cells. Nevertheless, co-overexpression of STAT3 and ZNF717 in HepG2 or Huh7 cells (oeZNF717/STAT3) restored STAT3 expression (Supplementary information, Figure S12E). Moreover, ZNF717 more efficiently bound to the −1 171 to −1 022 region of STAT3 promoter in oeZNF717 HepG2 or Huh7 cells compared to vec HCC cells (Supplementary information, Figure S12F and S12G). These results suggest that the tumor-inhibiting effect of ZNF717 in HCC is likely mediated through regulation of the IL-6/STAT3 pathway.

Discussion
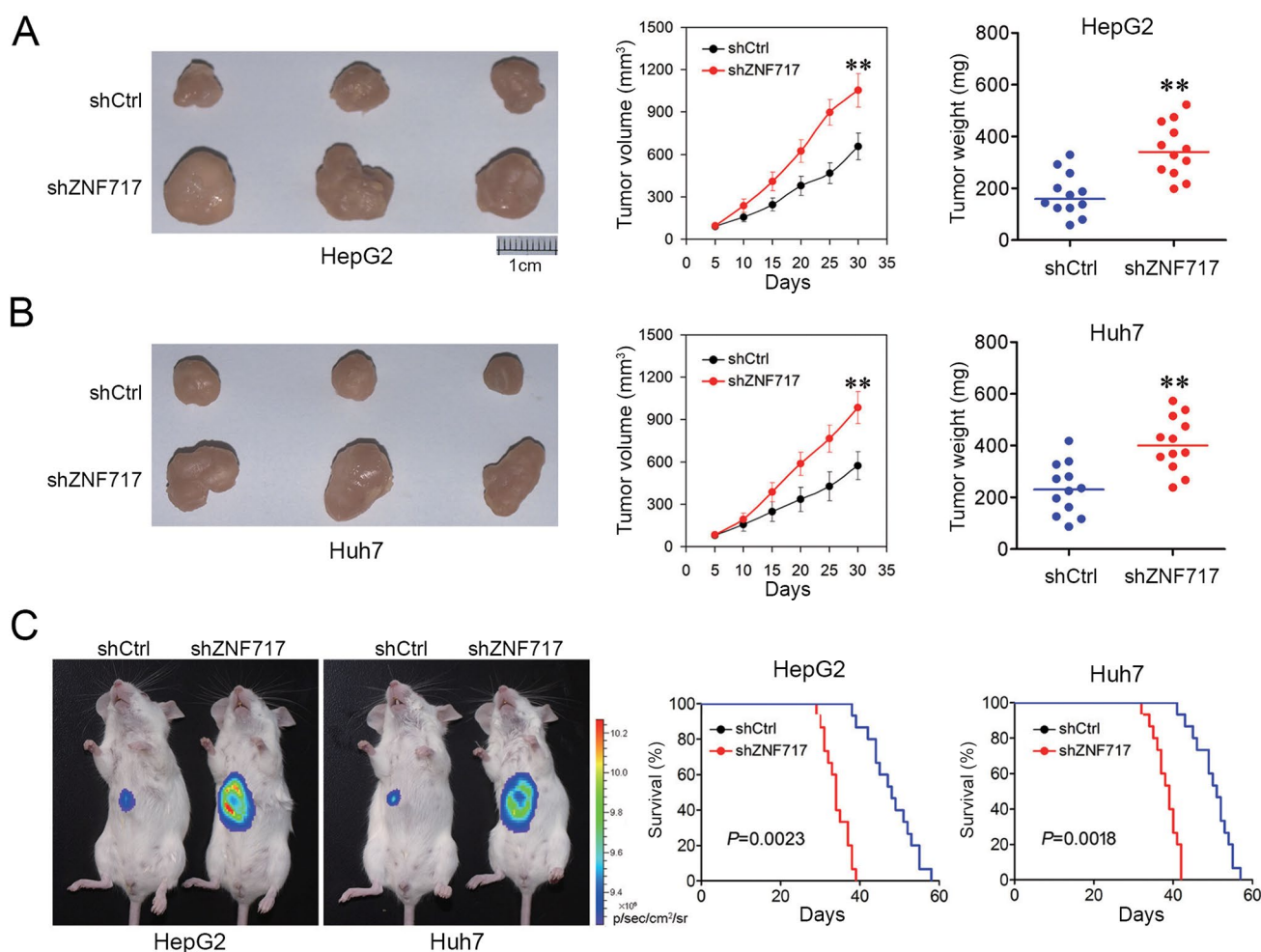
With the advent of next-generation sequencing, there

Figure 5 ZNF717 acts as a tumor suppressor in HCC. **(A**, **B)** Knockdown of ZNF717 significantly promoted subcutaneous tumor growth in HCC. HepG2 **(A)** or Huh7 **(B)** cells transfected with shZNF717 or shCtrl were subcutaneously injected into the back of NOD/SCID mice to establish xenograft HCC models (*n* = 11 each group). Left panel: representative images of xenograft tumors. Scale bar, 1 cm. Middle panel: tumor growth curves. Right panel: dot plot of tumor weight. *******P* < 0.01. **(C)** Knockdown of ZNF717 significantly promoted orthotopic tumor growth and reduced host survival. Luciferase-labeled HepG2 or Huh7 cells were transplanted into NOD/SCID mouse livers to establish orthotopic HCC models.

has been significant interest in studying intratumor genetic heterogeneity in HCC although previous efforts have mainly focused on the spatial heterogeneity within a tumor or the relationship between the primary and metastatic tumors [2-6]. It is well established, however, that HCC frequently displays heterogeneous growth patterns and morphologic features. Over 50% of HCC cases harbor multifocal lesions [3, 4, 23], about 30% of cases show a confluent multinodular morphology featured by a cluster of confluent nodules [17, 24], and 30%-62% of cases ultimately develop macroscopic PVTT upon tumor progression [22]. We sought to establish the relationship between intratumor genetic diversity and distinct morphologic phenotypes, which should be of pivotal impor-

tance in HCC.

In this study, using single-cell whole-genome sequencing, we evaluated three HBV-related HCC cases with typical morphologic phenotypes. Analyzing these data allowed us to delineate a detailed intratumor genetic landscape of HCC at a single-cell level, and relate it to distinct tumor morphologies. Our results supported that both models of metastasis, late dissemination and early seeding [39], have a role in human HCC. The H-PT case with PVTT exemplifies the model of late dissemination, where the thrombotic tumor cells genetically resemble the cells from tumor mass. The H-MF case is a good model of early seeding, where the two separated tumors are from one founding clone whose early intrahepatic
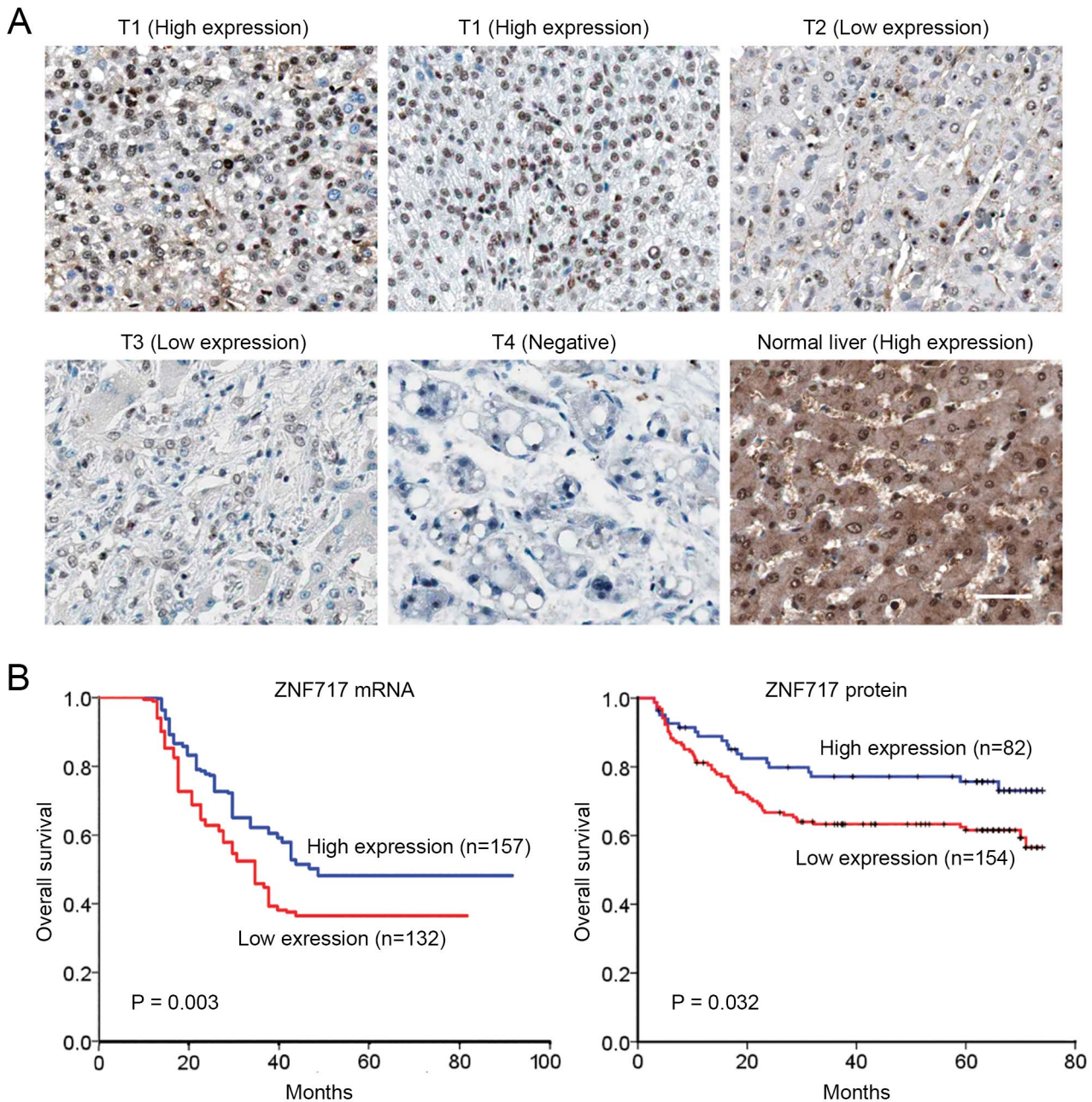
A

T1 (High expression)   T1 (High expression)   T2 (Low expression)

T3 (Low expression)   T4 (Negative)   Normal liver (High expression)

B

ZNF717 mRNA

High expression (n=157)

Low exression (n=132)

P = 0.003

Overall survival / Months

ZNF717 protein

High expression (n=82)

Low expression (n=154)

P = 0.032

Overall survival / Months

Figure 6 Prognostic significance of ZNF717 expression in HCC patients. **(A)** Representative immunostaining images of ZNF717 in HCC patients. ZNF717-positive staining was mainly detected in cell nuclei. Normal liver cells showed uniformly high expression of ZNF717, whereas HCC cells showed significant downregulation of ZNF717 expression. Scale bar, 50 μm. **(B)** Low expression of ZNF717 mRNA or protein significantly correlated with poor survival in HCC patients. ZNF717 mRNA expression data were adopted from [37, 38] (n = 289). ZNF717 protein expression was determined by immunohistochemistry on tissue microarray containing 236 consecutive HCC patients as previously reported [4].

spreading led to the formation of two large tumors with comparable size. Importantly, the H-CM case exemplifies the relevance of the mutagenic field effect of chronic hepatitis and multicellular origins in hepatocarcinogenesis (Figure 8), resulting in independent genetic lineages that hardly share any founder genetic alterations. Consistent with our results, a latest study applying genome and transcriptome sequencing on a nodule-in-nodule HCC also showed no sharing in CNV and expression profiles, and on another case with confluent multiple
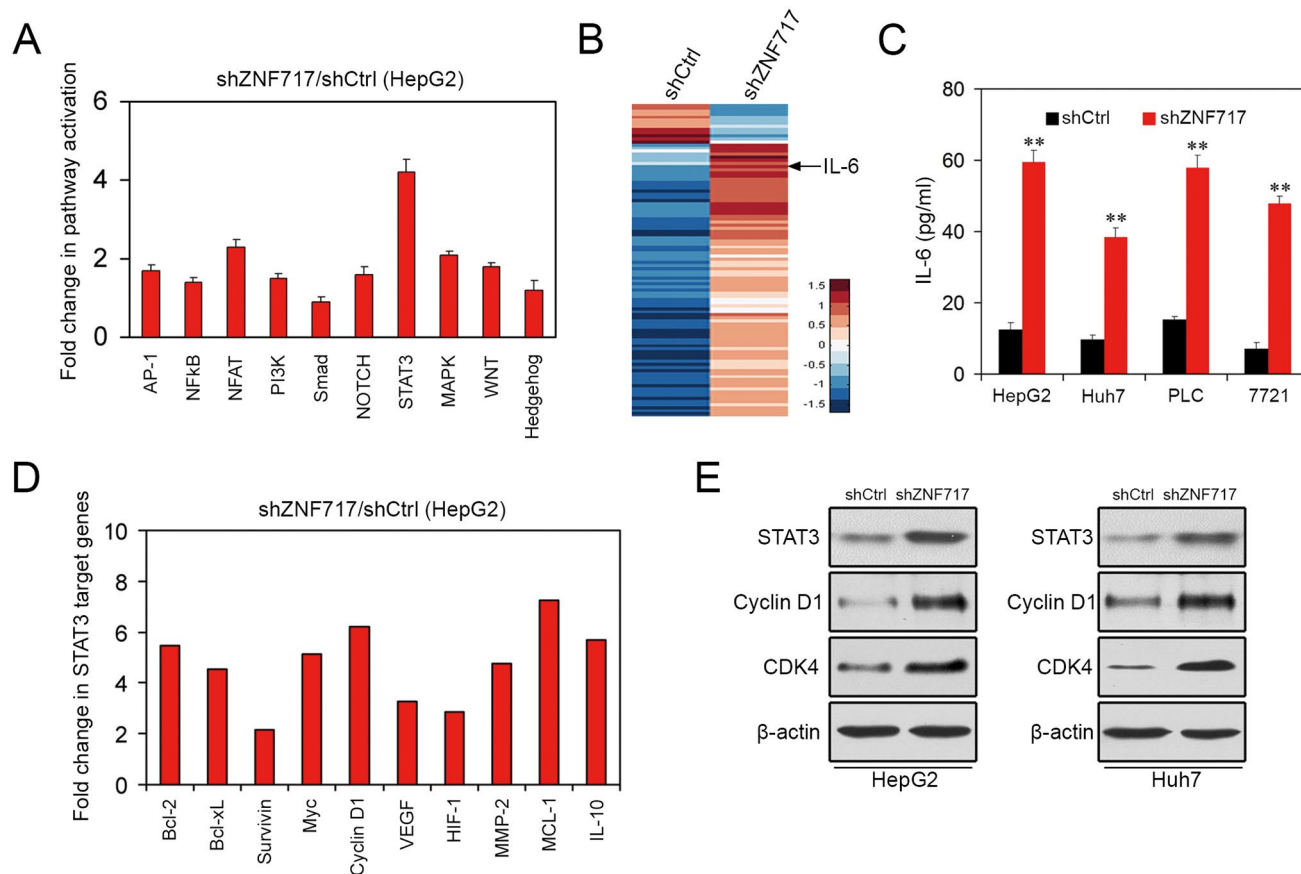
Figure 7 ZNF717 inhibits IL-6 expression via STAT3 signaling in HCC. **(A)** Microarray analysis of gene expression from shCtrl and shZNF717 HepG2 cells in each group. **(B)** Pathway Array analysis of gene expression from shCtrl and shZNF717 HepG2 cells in each group. **(C)** Knockdown of ZNF717 significantly enhanced IL-6 production in a panel of HCC cell lines. Supernatant IL-6 concentrations were determined by ELISA. **$P < 0.01$. **(D)** qRT-PCR analysis of STAT3 target gene expression from shCtrl and shZNF717 HepG2 cells in each group. **(E)** Knockdown of ZNF717 in HCC cells led to the overexpression of STAT3 and its downstream molecules.

nodules showed distinct somatic SNV profiles among these intratumor nodules [40]. Therefore, these intratumor nodules in the H-CM case were likely independently derived from different liver cancer progenitor cells, i.e., polyclonal origin. These findings highlight that multiple distinct tumor evolutionary mechanisms operate in HCC and that these different pathways of evolution result in different HCC-specific morphologies and importantly call for different treatment strategies. Hence, understanding cancer evolution has the potential to guide cancer treatment practice.

HBV integration into the host liver cell genome has been suspected to occur during chronic HBV infection before the initiation of HCC tumorigenesis [19]. During chronic hepatitis/cirrhosis, hepatocytes with HBV integrations undergo certain rounds of expansion, and finally gain growth advantages that lead to clonal expansion and

tumorigenesis. In this study, the HBV integration pattern in monoclonal tumors H-PT and H-MF was consistent with this model. We revealed the remarkable uniformity of HBV integration among spatially separated tumor cells within H-PT and H-MF, which further supports that HBV integration is an early driver event during hepatocarcinogenesis and remains extremely stable during tumor evolution. As such, in addition to the CNV and mutational profiles, HBV integration pattern could be used to aid the illustration of the clonal diversity of HBV-related HCC.

Previous deep-sequencing studies indicated that multifocal tumors in HCC develop as independent tumors or as intrahepatic metastases [3, 4, 23]. However, whether the confluent multinodular HCC derives from multiple carcinogenesis events under the chronic hepatitis/cirrhosis background or intratumor subclonal ex-
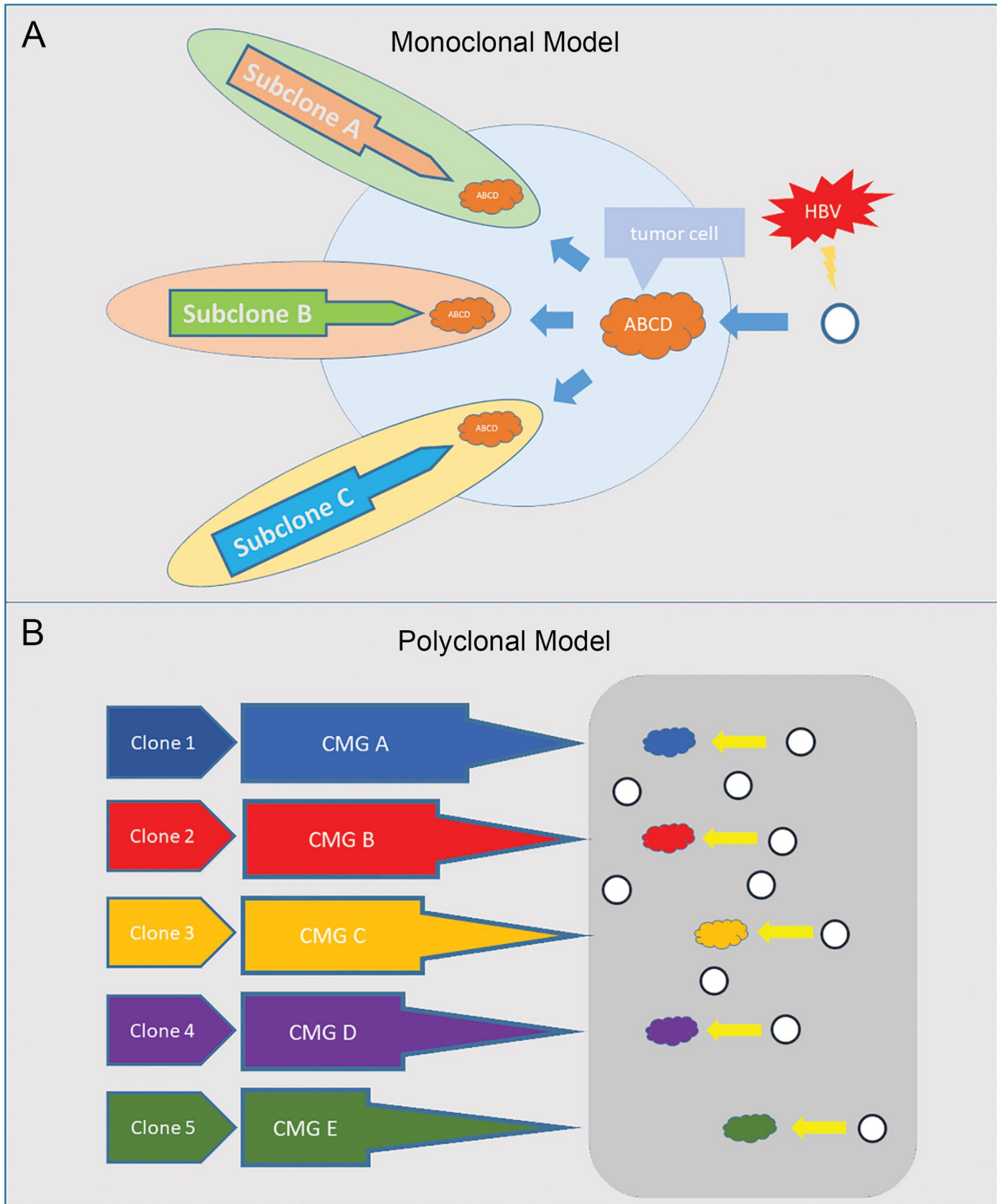
Figure 8 Models of monoclonal and polyclonal evolution in HCC. **(A)** The monoclonal evolution model. Monoclonal HCC arises from a single cell that is stimulated by the HBV integration events. Similar to the punctuated CNV evolution model, HBV integration sites (ABCD) are shared by all the tumor cells during tumor evolution (in the light blue circle). Subclonal tumor cells gradually accumulate private mutations (the part of ellipse outside the light blue circle), but HBV integration sites persist. **(B)** The polyclonal evolution model. Each tumor clone derives from different tumor single cells (cloud shape), harboring distinct CMGs (clone mutation groups, ABCDE).

pansion remains unknown. Our data showed that these intratumor nodules within H-CM were probably derived independently from distinct liver progenitor cells rather than from a common precancerous ancestor. Therefore, although the confluent multinodular HCC is confined within a tumor mass, it should no longer be considered as one tumor but multiple independent tumors in future clinical practice. Moreover, we demonstrate that intratumor heterogeneity derives from subclonal mutations in monoclonal tumors or from the quite different genetic architectures of each independent clone in polyclonal tumors.

In conclusion, these findings in HBV-related HCC reconcile the existing theory of "seed and soil" and raise fascinating new issues for HCC science and treatment practice. Future work with single-cell multi-omic platforms applied to more HCCs and novel experimental models are needed for a more complete understanding of the initiation, progression and therapeutic resistance of this fatal malignancy.

Materials and Methods

Sample collection

HCC demonstrates various morphological and pathological appearances. Herein, we selected a single-nodular HCC with PVTT, a multifocal HCC, and a confluent multinodular HCC for SCG sequencing. The three male patients were diagnosed with HBV-related HCC who underwent curative resection at Zhongshan Hospital without any adjuvant therapy before. Tumor samples were obtained immediately after resection and sent for single cell isolation and histological examination. The study was approved by the Institutional Review Board with written informed consent from each patient.

Single-cell isolation

The paired normal liver and HCC tissue were minced and digested with collagenase I (GIBCO) at 37 °C for 2 h. The normal liver and HCC cell suspension were filtered with screen cloth and stained with CD31 and CD45. The single cell was then sorted into a 96-well plate using a FACSArila II flow cytometry system (BD Biosciences). Moreover, the cells were shipped in drikold and stored at −80 °C until DNA extraction.

Whole-genome sequencing and data analysis

Whole-genome sequencing of lysed single cells and bulk samples, and data analyses including somatic coding and noncoding SNV detection, somatic CNV detection, single-cell classification, detection of HBV integration sites, phylogenetic tree construction, and cancer driver gene prediction were detailed in Supplementary information, Data S1.

Cell lines and animals

Human HCC cell lines HepG2 and Huh7 were obtained from the Chinese Academy of Medical Sciences. Silencing of ZNF717, overexpression of ZNF717 and STAT3, as well as CRISPR/Cas9

manipulation were detailed in Supplementary information, Data S1. NOD/SCID mice were obtained from the Animal Center of the Chinese Academy of Medical Science, Beijing.

Functional experiments in vitro

Cell proliferation, attachment and invasion assays, wound healing assay, apoptosis and cell cycle assays, ELISA and western blot were detailed in Supplementary information, Data S1.

Animal experiments

Six-week-old NOD/SCID female mice weighing ~15 g were used. Animal xenograft models of shCtrl/shZNF717 HepG2/Huh7 cells were created by subcutaneously injecting cells (1×10^6 resuspended in PBS) into the back of mice. Tumor size was measured twice a week. As for the orthotopic NOD/SCID mouse model, cells were labeled with luciferase and transplanted into the murine liver. The photon counts of each mouse were indicated by the pseudo-color scales.

Prognostic value of ZNF717 expression

The prognostic value of ZNF717 expression was evaluated in two independent patient cohorts. First, mRNA expression data of ZNF717 were adopted from Roessler *et al.* (GSE14520) [37, 38], which contained a total of 289 HCC patients with complete clinicopathologic and survival information. The median duration of follow-up of this cohort was 72.5 months. Second, protein expression of ZNF717 and pSTAT3 was investigated using immunohistochemical staining on tissue microarrays containing a total of 236 HCC patients (Supplementary information, Data S1) [4]. The median duration of follow-up of this cohort was 60 months.

Pathway reporter array

Profile of the changes in the activities of 45 signaling pathways in shCtrl and shZNF717 HepG2 cells was evaluated using the Cignal 45-Pathway Reporter Array (QIAGEN). Dual-luciferase reporter assays, with a pathway-specific firefly luciferase reporter and a constitutively expressed Renilla reporter serving as a normalization control, was performed.

RNA extraction and microarray analysis

Total RNA of shCtrl or shZNF717 HepG2 cells was extracted with TRIzol reagent (Invitrogen) using a standard isopropanol/chloroform protocol. Gene expression patterns were analyzed with a human gene chip that contained clones of 35 000 human genes (CapitalBio Corp, Beijing, China). Microarray slides were scanned with a ScanArray 4000 Microarray Analysis System (Packard BioScience), and data were analyzed with data analysis software (Dapple version 0.86 beta).

Chromatin immunoprecipitation

Chromatin immunoprecipitation (ChIP) was performed as previously described [41]. Briefly, 10^7 cells were crosslinked with 1% formaldehyde in lysis buffer, and sonicated on ice resulting in an average DNA fragment length of 500 bp. Immunoprecipitation was performed in ChIP dilution buffer overnight with IgG and ZNF717 antibody with agitation. A protein A agarose/Salmon Sperm DNA (Merck Millipore) slurry was added and incubated for 2 h at 4 °C with agitation. After centrifugation and 5-time washes, the immunoprecipitated fraction was eluted. The crosslinking was

reversed by incubation at 65 °C for 4 h. The DNA was recovered by phenol/chloroform extraction and precipitated, and the abundance of specific sequence was measured by qRT-PCR.

DNase I digestion assay

Cell nuclei were isolated and lysed for DNase I digestion assay. After digestion at 37 °C for 5 min, total DNA was extracted for qRT-PCR.

Data availability

The whole-genome sequencing data have been deposited in the EGA (European Genome-phenome Archive) database (EGAD00001002064). Microarray data have been submitted to the NCBI's Gene Expression Omnibus (GEO; GSE86852).

Acknowledgments

Author Contributions

QG, CL and MD conceived and designed the study. QG, JH, MD and SC contributed to data acquisition. QG, JH and DW analyzed and interpreted the data. SC, JH and MD performed the experiments. LL, XW, ZW and AK contributed to sample acquisition and clinical data collection. QG, MD, XZ and JS drafted the manuscript. CL, YC, JF, JZ and SZ contributed to critical revision of the manuscript. QG, SC and RX performed statistical analysis. All authors reviewed the data and approved the final manuscript.

Competing Financial Interests

The authors declare no competing financial interests.

References

1   Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, Jemal A. Global cancer statistics, 2012. *CA Cancer J Clin* 2015; **65**:87-108.

2   Friemel J, Rechsteiner M, Frick L, *et al*. Intratumor heterogeneity in hepatocellular carcinoma. *Clin Cancer Res* 2015; **21**:1951-1961.

3   Xue R, Li R, Guo H, *et al*. Variable intra-tumor genomic heterogeneity of multiple lesions in patients with hepatocellular carcinoma. *Gastroenterology* 2016; **150**:998-1008.

4   Shi JY, Xing Q, Duan M, *et al*. Inferring the progression of multifocal liver cancer from spatial and temporal genomic heterogeneity. *Oncotarget* 2016; **7**:2867-2877.

5   Ling S, Hu Z, Yang Z, *et al*. Extremely high genetic diversity in a single tumor points to prevalence of non-Darwinian cell evolution. *Proc Natl Acad Sci USA* 2015; **112**:E6496-E6505.

6   Gao Q, Wang ZC, Duan M, *et al*. Cell culture system for analysis of genetic heterogeneity within hepatocellular carcinomas and response to pharmacologic agents. *Gastroenterology* 2017; **152**:232-242.

7   Weiskirchen R. Intratumor heterogeneity, variability and plasticity: questioning the current concepts in classification and treatment of hepatocellular carcinoma. *Hepatobiliary Surg Nutr* 2016; **5**:183-187.

8   Gawad C, Koh W, Quake SR. Single-cell genome sequencing: current state of the science. *Nat Rev Genet* 2016; **17**:175-188.

9   Wang Y, Navin NE. Advances and applications of single-cell sequencing technologies. *Mol Cell* 2015; **58**:598-609.

10 Navin N, Kendall J, Troge J, *et al*. Tumour evolution inferred by single-cell sequencing. *Nature* 2011; **472**:90-94.

11 Wang Y, Waters J, Leung ML, *et al*. Clonal evolution in breast cancer revealed by single nucleus genome sequencing. *Nature* 2014; **512**:155-160.

12 Xu X, Hou Y, Yin X, *et al*. Single-cell exome sequencing reveals single-nucleotide mutation characteristics of a kidney tumor. *Cell* 2012; **148**:886-895.

13 Li Y, Xu X, Song L, *et al*. Single-cell sequencing analysis characterizes common and cell-lineage-specific mutations in a muscle-invasive bladder cancer. *GigaScience* 2012; **1**:12.

14 Yu C, Yu J, Yao X, *et al*. Discovery of biclonal origin and a novel oncogene SLC12A5 in colon cancer by single-cell sequencing. *Cell Res* 2014; **24**:701-712.

15 Martelotto LG, Baslan T, Kendall J, *et al*. Whole-genome single-cell copy number profiling from formalin-fixed paraffin-embedded samples. *Nat Med* 2017; **23**:376-385.

16 Hou Y, Guo H, Cao C, *et al*. Single-cell triple omics sequencing reveals genetic, epigenetic, and transcriptomic heterogeneity in hepatocellular carcinomas. *Cell Res* 2016; **26**:304-319.

17 Murakata A, Tanaka S, Mogushi K, *et al*. Gene expression signature of the gross morphology in hepatocellular carcinoma. *Ann Surg* 2011; **253**:94-100.

18 Budhu A, Forgues M, Ye QH, *et al*. Prediction of venous metastases, recurrence, and prognosis in hepatocellular carcinoma based on a unique immune response signature of the liver microenvironment. *Cancer Cell* 2006; **10**:99-111.

19 Miao R, Luo H, Zhou H, *et al*. Identification of prognostic biomarkers in hepatitis B virus-related hepatocellular carcinoma and stratification by integrative multi-omics analysis. *J Hepatol* 2014; **60**:346-353.

20 Zong C, Lu S, Chapman AR, Xie XS. Genome-wide detection of single-nucleotide and copy-number variations of a single human cell. *Science* 2012; **338**:1622-1626.

21 Li L, Wang H. Heterogeneity of liver cancer and personalized therapy. *Cancer Lett* 2016; **379**:191-197.

22 Pesi B, Ferrero A, Grazi GL, *et al*. Liver resection with thrombectomy as a treatment of hepatocellular carcinoma with major vascular invasion: results from a retrospective multicentric study. *Am J Surg* 2015; **210**:35-44.

23 Gao Q, Wang XY, Zhou J, Fan J. Multiple carcinogenesis contributes to the heterogeneity of HCC. *Nat Rev Gastroenterol Hepatol* 2015; **12**:13.

24 He J, Shi J, Fu X, *et al*. The clinicopathologic and prognostic significance of gross classification on solitary hepatocellular carcinoma after hepatectomy. *Medicine* 2015; **94**:e1331.

25 Ringelhan M, O'Connor T, Protzer U, Heikenwalder M. The direct and indirect roles of HBV in liver cancer: prospective

markers for HCC screening and potential therapeutic targets. *J Pathol* 2015; **235**:355-367.

26 Sung WK, Zheng H, Li S, *et al.* Genome-wide survey of recurrent HBV integration in hepatocellular carcinoma. *Nat Genet* 2012; **44**:765-769.

27 Ding D, Lou X, Hua D, *et al.* Recurrent targeted genes of hepatitis B virus in the liver cancer genomes identified by a next-generation sequencing-based approach. *PLoS Genet* 2012; **8**:e1003065.

28 Chen M, Song P, Zou D, *et al.* Comparison of multiple displacement amplification (MDA) and multiple annealing and looping-based amplification cycles (MALBAC) in single-cell sequencing. *PLoS One* 2014; **9**:e114520.

29 Oriyama T, Yamanaka N, Fujimoto J, Ichikawa N, Okamoto E. Progression of hepatocellular carcinoma as reflected by nuclear DNA ploidy and cellular differentiation. *J Hepatol* 1998; **28**:142-149.

30 Futreal PA, Coin L, Marshall M, *et al.* A census of human cancer genes. *Nat Rev Cancer* 2004; **4**:177-183.

31 Schulze K, Imbeaud S, Letouze E, *et al.* Exome sequencing of hepatocellular carcinomas identifies new mutational signatures and potential therapeutic targets. *Nat Genet* 2015; **47**:505-511.

32 Totoki Y, Tatsuno K, Covington KR, *et al.* Trans-ancestry mutational landscape of hepatocellular carcinoma genomes. *Nat Genet* 2014; **46**:1267-1273.

33 Ahn SM, Jang SJ, Shim JH, *et al.* Genomic portrait of resectable hepatocellular carcinomas: implications of RB1 and FGF19 aberrations for patient stratification. *Hepatology* 2014; **60**:1972-1982.

34 Cleary SP, Jeck WR, Zhao X, *et al.* Identification of driver genes in hepatocellular carcinoma by exome sequencing. *Hepatology* 2013; **58**:1693-1702.

35 Tamborero D, Gonzalez-Perez A, Lopez-Bigas N. Oncodrive-CLUST: exploiting the positional clustering of somatic mutations to identify cancer genes. *Bioinformatics* 2013; **29**:2238-2244.

36 Chen Y, Wang L, Xu H, Liu X, Zhao Y. Exome capture sequencing reveals new insights into hepatitis B virus-induced hepatocellular carcinoma at the early stage of tumorigenesis. *Oncol Rep* 2013; **30**:1906-1912.

37 Roessler S, Long EL, Budhu A, *et al.* Integrative genomic identification of genes on 8p associated with hepatocellular carcinoma progression and patient survival. *Gastroenterology* 2012; **142**:957-966.

38 Roessler S, Jia HL, Budhu A, *et al.* A unique metastasis gene signature enables prediction of tumor relapse in early-stage hepatocellular carcinoma patients. *Cancer Res* 2010; **70**:10202-10212.

39 Klein CA. Parallel progression of primary tumours and metastases. *Nat Rev Cancer* 2009; **9**:302-312.

40 Furuta M, Ueno M, Fujimoto A, *et al.* Whole genome sequencing discriminates hepatocellular carcinoma with intrahepatic metastasis from multi-centric tumors. *J Hepatol* 2017; **66**:363-373.

41 Yang Z, He L, Lin K, *et al.* The KMT1A-GATA3-STAT3 circuit is a novel self-renewal signaling of human bladder cancer stem cells. *Clin Cancer Res* 2017; **23**:6673-6685.

(**Supplementary information** is linked to the online version of the paper on the *Cell Research* website.)