



Whole-exome sequencing reveals the metastatic potential of hepatocellular carcinoma from the perspective of tumor and circulating tumor DNA

Chenhao Zhou^{1,2,3} · Jialei Weng^{1,3} · Shaoqing Liu^{1,3} · Qiang Zhou^{1,3} · Zhiqiu Hu^{3,4} · Yirui Yin⁵ · Peng Lv⁵ · Jialei Sun¹ · Hui Li¹ · Yong Yi¹ · Yinghao Shen¹ · Qinghai Ye¹ · Yi Shi⁶ · Qiongzhu Dong^{3,4} · Chunxiao Liu² · Xiaoqiang Zhu^{7,8} · Ning Ren^{1,3,4} 

Received: 29 January 2023 / Accepted: 15 April 2023 / Published online: 23 May 2023
© Asian Pacific Association for the Study of the Liver 2023

Abstract

Background Relapse of hepatocellular carcinoma (HCC) due to vascular invasion is common, but the genomic mechanisms remain unclear, and molecular determinants of high-risk relapse cases are lacking. We aimed to reveal the evolutionary trajectory of microvascular invasion (MVI) and develop a predictive signature for relapse in HCC.

Methods Whole-exome sequencing was performed on tumor and peritumor tissues, portal vein tumor thrombus (PVTT), and circulating tumor DNA (ctDNA) to compare the genomic profiles between 5 HCC patients with MVI and 5 patients without MVI. We conducted an integrated analysis of exome and transcriptome to develop and validate a prognostic signature in two public cohorts and one cohort from Zhongshan Hospital, Fudan University.

Results Shared genomic landscapes and identical clonal origins among tumor, PVTT, and ctDNA were observed in MVI (+) HCC, suggesting that genomic changes favoring metastasis occur at the primary tumor stage and are inherited in metastatic lesions and ctDNA. There was no clonal relatedness between the primary tumor and ctDNA in MVI (–) HCC. HCC had dynamic mutation alterations during MVI and exhibited genetic heterogeneity between primary and metastatic tumors, which can be comprehensively reflected by ctDNA. A relapse-related gene signature named RGS_{HCC} was developed based on the significantly mutated genes associated with MVI and shown to be a robust classifier of HCC relapse.

Conclusions We characterized the genomic alterations during HCC vascular invasion and revealed a previously undescribed evolution pattern of ctDNA in HCC. A novel multiomics-based signature was developed to identify high-risk relapse populations.

Keywords Hepatocellular carcinoma · Microvascular invasion · Whole-exome sequencing · Circulating tumor DNA · Relapse

Introduction

Liver cancer has had the fastest growing mortality rate over the decades, with a 5-year survival rate of 20%, the second lowest among all cancers [1]. Hepatocellular carcinoma (HCC) is the most common pathological type of liver cancer, accounting for 75–85% of all cases [2]. As technology has improved, so has our understanding of the genetic alterations related to liver carcinogenesis. Genomic analyses have

delineated several characteristic genomic events associated with the etiology of HCC, including *FGF19* gene amplifications and mutations in the *CTNNB1* and *TP53* genes [3–5]. A comprehensive understanding of HCC pathogenesis and progression at the molecular level would help us identify core oncogenic events for developing new therapeutic strategies but currently remains an unmet need.

Although optimal clinical management has improved the survival of HCC patients, the high frequency of postoperative relapse and the presence of metastases limit long-term outcomes [6–8]. The risk of relapse is unpredictable in most HCC patients in current clinical practice. Microvascular invasion (MVI), considered the initial stage of hematogenous metastasis of HCC, is an important risk factor that

Chenhao Zhou, Jialei Weng and Shaoqing Liu authors contributed equally to this work.

Extended author information available on the last page of the article

influences the relapse and prognosis of HCC after hepatectomy [9, 10]. However, the genetic mechanisms underlying the vascular invasion of HCC, especially the clonal evolution between the primary tumor and portal vein tumor thrombus (PVTT), remain poorly understood. Therefore, comprehensive genomic profiling based on emerging high-throughput technologies and a gene-based predictive model highly associated with disease relapse are required to enhance the molecular understanding of HCC progression and promote individualized treatment of HCC.

Another major challenge in achieving precision medicine for HCC is the intratumor heterogeneity and lack of tissue of advanced HCC [11]. In this context, circulating tumor DNA (ctDNA), a key component of liquid biopsy, has received much attention as it integrates genomic information from different lesions and can be obtained in a minimally invasive manner [12]. The application prospects of ctDNA in the diagnosis and treatment of HCC have been widely discussed [13–15]. However, although ctDNA is released throughout malignant progression, from initial growth to metastatic spread, the evolutionary pattern of ctDNA in different stages of HCC remains largely unknown.

Here, we aim to systematically elucidate the genomic landscape during the clonal evolution of HCC from the perspective of tumor and ctDNA and to identify mutations associated with MVI and relapse. We performed whole-exome sequencing (WES) of tumor and peritumor tissues, PVTT, ctDNA, and peripheral blood mononuclear cells (PBMCs) in 5 HCC patients with MVI and 5 HCC patients without MVI. Using these samples, we compared genomic profiles, including mutational landscapes and clonal evolution patterns, between HCC patients with MVI and those without MVI to investigate the genomic evolution during HCC vascular invasion. Finally, focusing on the significantly mutated genes (SMGs) associated with MVI, we performed an integrative genomic analysis with the transcriptome and developed a prognostic signature for relapse prediction in HCC.

Materials and methods

Patient cohort

Biospecimens for WES were collected from 10 patients diagnosed with pathologically confirmed HCC during curative resection at Zhongshan Hospital, Fudan University between June 2016 and July 2016, including 5 with MVI and 5 without MVI. Peritumor tissues, tumor tissues, PBMCs, and ctDNA from these 10 patients, as well as 5 PVTTs from MVI-positive patients, were subjected to WES (a total of 45 samples). PBMCs were used as the control. Patients were included based on the following criteria: radical resection; no prior intervention, such as transhepatic

artery embolization, chemotherapy, or radiotherapy; and no history of other malignancies or inflammatory disease. The presence of MVI in the tumor was independently judged by two experienced pathologists. In addition, we retrospectively collected 172 tumor samples from HCC patients who received hepatectomy during January 2012 and December 2012 in Zhongshan Hospital, Fudan University, as an external validation for our signature (PCR cohort) with the same inclusion criteria.

Demographic and clinicopathological data, including age, sex, laboratory tests, and pathology reports, were collected from the electronic medical record system. Liver function was assessed based on the Child–Pugh scoring system. Tumors were staged according to the 8th edition of the tumor-node-metastasis (TNM) classification system [16]. The postoperative follow-up scheme was conducted as previously described [17]. Relapse-free survival (RFS) was defined as the interval between surgery and relapse or the study endpoint. The clinicopathological characteristics of patients were summarized in the Supplementary Table 1. All patients were HBV positive and patients with MVI were diagnosed as MVI (+) after surgery. The study design was conducted in accordance with both the Declarations of Helsinki and Istanbul and approved by the Ethics Review Board of Zhongshan Hospital, and all participants provided written consent.

Genomic DNA preparation and WES

Genomic DNA from fresh-frozen specimens and PBMCs was extracted using the QIAamp DNA FFPE Tissue Kit and QIAamp DNA Blood Mini Kit (Qiagen, Germany), respectively, according to the manufacturer's instructions. DNA was then quantified using a Qubit 2.0 Fluorometer (Thermo Fisher Scientific, USA) and assessed for integrity by agarose gel electrophoresis.

Sequencing libraries were prepared using the Agilent SureSelectXT Low Input Reagent Kit (Agilent Technologies). Briefly, qualified genomic DNA was sheared to fragments in the range of 180–280 bp using a Biorupter Ultrasonicator (Diagenode) and then purified using AMPure SPRI beads. Later, DNA fragments were ligated with Illumina paired-end indexed adapters and amplified by 8 cycles of PCR to construct uniquely indexed libraries. After size distribution examination on the 4200 TapeStation High Sensitivity D1000 ScreenTape, the libraries were subjected to exon target capture using the SureSelect Human All Exon V5 Kit (Agilent Technologies). The captured exon-enriched libraries were assessed on an Agilent 2100 Bioanalyzer and quantified using a Qubit 2.0 Fluorometer (Thermo Fisher Scientific, USA). Finally, libraries were sent for sequencing on the Illumina HiSeq 2000 platform using the format of 150 bp paired-end reads.

Sequence alignment and variant calling

Filtered raw sequencing data in FastQ format were aligned to the reference human genome (UCSC GRCh37/hg19) using Burrows–Wheeler Aligner (BWA) software to generate mapping results in BAM format [18]. GATK2 and Picard were then used for local realignment and base recalibration [19]. Somatic single-nucleotide variations (SNVs) were identified using MuTect, and somatic indels were detected by GATK Somatic Indel Detector [20]. High-confidence variants were identified using the following criteria: coverage for all samples was more than $10\times$, the variant allele fraction (VAF) was at least 10% in the tumor samples and less than 2% in peritumoral samples and the number of reads supporting the variant allele was > 3 . ANNOVAR was conducted to annotate Variant Call Format (VCF). Somatic SNVs and indels were then compared against the single nucleotide polymorphisms (SNPs) presented in the 1000 Genomes Project to remove common SNPs and SNVs with minor allele frequencies greater than 1% or located in segmental duplications.

Spectra of somatic mutations and mutation signature analysis

Mutation spectra of six mutation types (C > T, T > C, C > A, T > A, C > G, and T > G) were analyzed in all samples. For mutation signature analysis, the relative contribution of each mutation signature in each sample was statistically quantified based on somatic SNVs using the “deconstructSigs” package in the R project [21]. The extracted signatures were analyzed for similarity with 30 previously depicted COSMIC mutation signatures.

Copy number variation (CNV) analysis

GATK software was applied to detect somatic CNVs and identify loss of heterozygosity (LOH) regions. The recurrently amplified and deleted genomic regions were then inferred using the GISTIC algorithm [22]. G-scores were calculated to evaluate the frequency and magnitude of amplified and deleted genomic fragments. Peak regions with a *p* value less than 0.05 were defined as significant CNV regions.

Identification of SMGs

The MutSigCV method was used to quantify the importance of nonsilent mutations (including SNVs and indels) and estimate the background mutation rate

based on silent mutations. SMGs were defined as genes with a significantly higher mutation frequency than the background mutation rate (false discovery rate (FDR) < 0.05). In HCC patients with MVI, shared SMGs among tumors, PVTs, and ctDNA were identified as SMGs associated with MVI.

Phylogenetic tree construction and driver mutation labeling

A phylogenetic tree was constructed based on nonsynonymous somatic mutations, including SNVs and indels, using PHYLIP to analyze the evolutionary trajectory of HCC [23]. Phylogenetic tree can be used to visualize the affinities between species or genes using tree branching graphs, and to calculate the evolutionary relationships between organisms by performing mathematical and statistical algorithms on gene mutations. The lengths of the branches and trunks were determined according to the number of mutated genes. For driver labeling, we collected a list of potential driver genes for HCC by a literature review of sequencing studies and assessed the mutation types in these genes. Genes were identified as driver genes and labeled to the phylogenetic tree when they had one of the following three variant types: nonsynonymous SNVs or indels, copy number amplification of oncogenes or fragment deletion of tumor suppressor genes and breakpoints in gene structure.

Cancer cell fraction (CCF) and mutation cluster analysis

Dynamic changes in the mutational landscape were analyzed using CCF value as mentioned previously [24]. This allows us to analyze the proportion of cells carrying a particular gene mutation in different samples, which can reflect the dynamic changes in the mutational characteristics of tumors during metastasis. For each somatic mutation, CCF was calculated using the following formula: $VAF = \frac{\alpha \times CCF}{\alpha \times CNT + 1 - \alpha \times CNP}$, where α indicates the tumor purity and CNT and CNP represent the copy number of tumor and peritumor samples, respectively. The tumor purity and copy number of each sample were detected using ABSOLUTE [25]. VAF was calculated using the following formula: $VAF = \frac{N^{var}}{N^{var} + N^{ref}}$, where N^{var} and N^{ref} indicate the number of reads supporting the variant allele and reference allele, respectively. For mutation cluster analysis, mutation clusters were inferred based on somatic SNVs and indels using the PyClone method. Clusters containing only silent mutations were removed before further analysis.

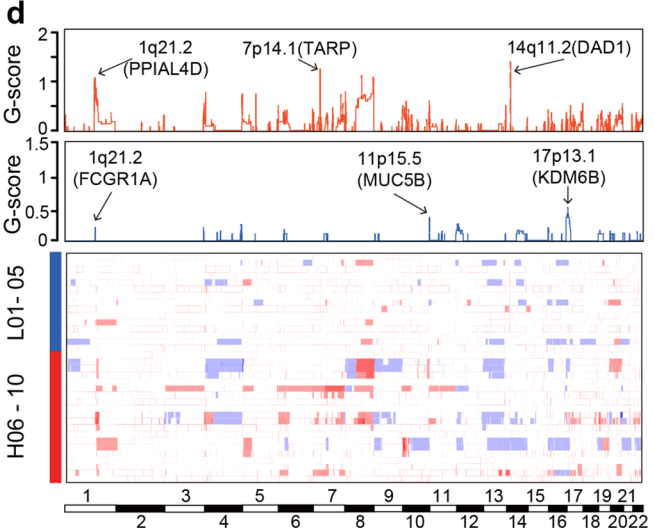
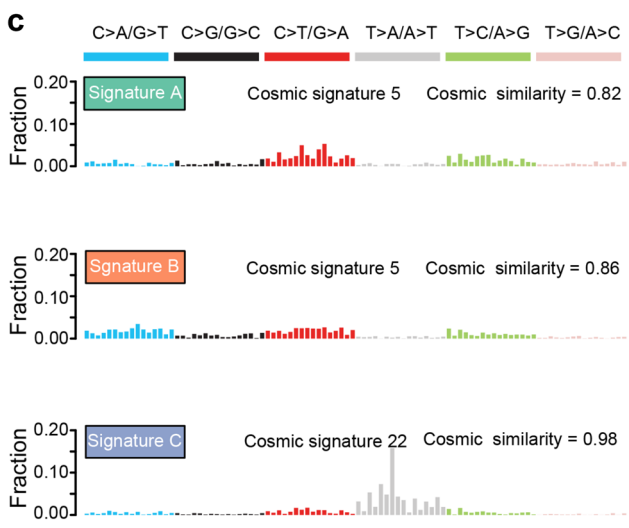
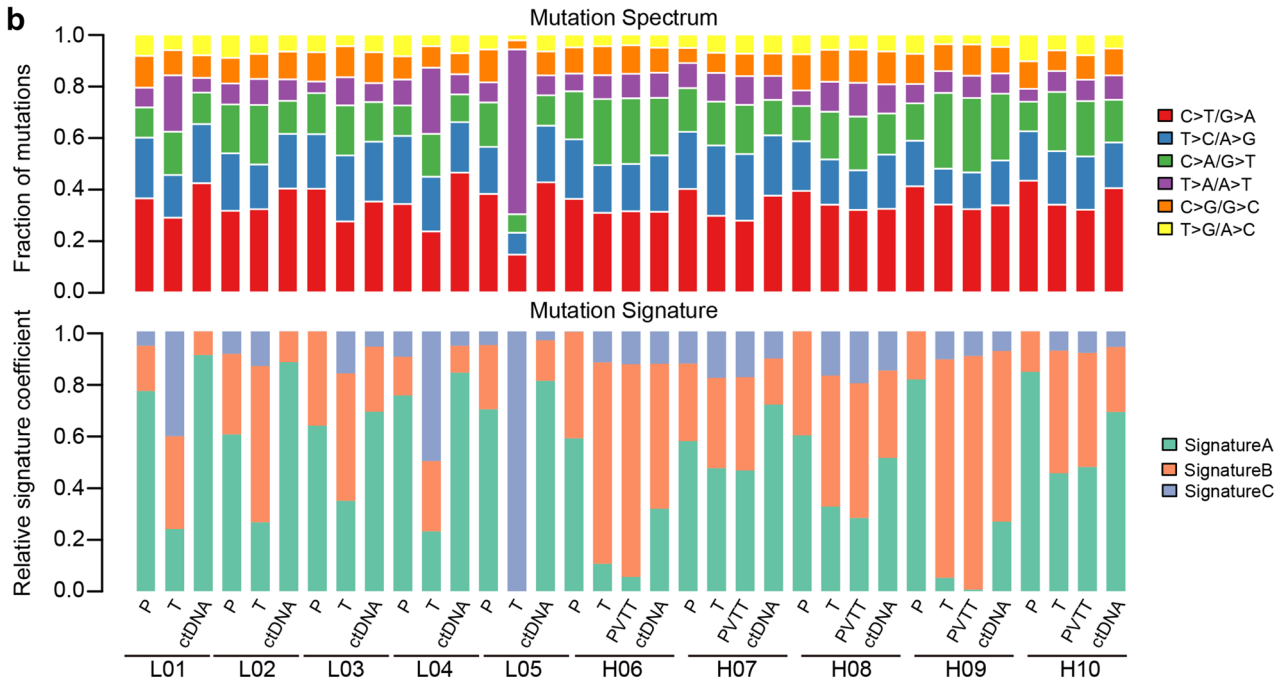
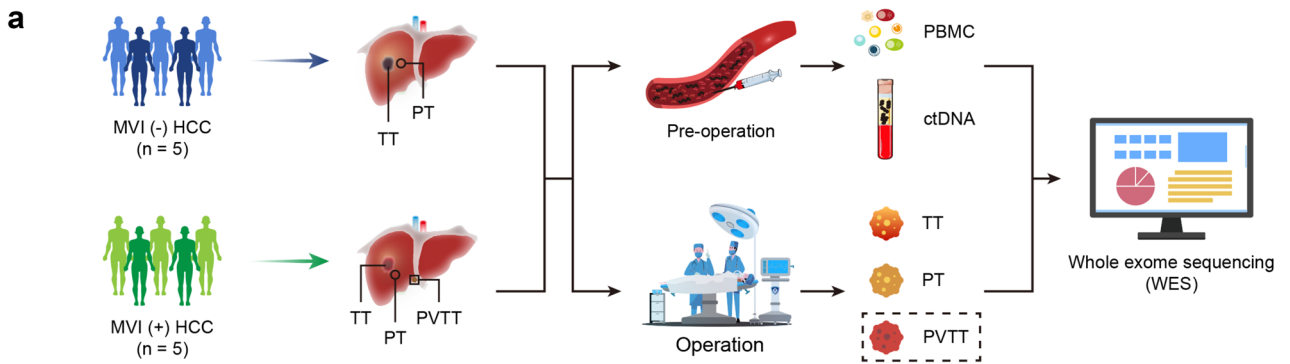


Fig. 1 Overview of the genomic alterations in HCC patients with and without MVI. **A** Flow chart of whole-exome sequencing in the present study. Preoperative PBMCs and ctDNA and postoperative peritumor tissue (PT), tumor tissue (TT), and PVTT from MVI (+) cases of 10 HCC patients were collected for sequencing. **B** Top: The distribution of 6 substitution mutations in all samples. Bottom: The relative coefficients of mutation signatures in all samples. **C** Patterns of 3 identified mutation signatures (Signature A, B, and C) and their most similar COSMIC mutation signatures. **D** Top: The whole-genome distribution of CNVs (upper, red represents amplification; lower, blue represents deletion) inferred using the GISTIC algorithm. The G-score was used to assess the frequency and magnitude of CNVs. Bottom: The distribution of CNVs in samples from the MVI (–) group (L01-05) and the MVI (+) group (H06-10)

Analysis of public databases

Liver cancer cohorts from two public databases were employed in this study: the LIHC cohort from The Cancer Genome Atlas (TCGA) database ($n = 356$) and the GSE14520 cohort in the Gene Expression Omnibus (GEO) database ($n = 221$) from the Liver Cancer Institute (LCI) at Zhongshan Hospital. Patients with pathological types of HCC were included. The LIHC cohort was used as the training dataset, and the GSE14520 cohort was employed for external validation (LCI validation dataset 1). RNA sequencing data and clinical information were downloaded from the corresponding websites. For data processing, transcriptome data from the LIHC cohort were logarithmically transformed and normalized; gene expression profiles from the GSE14520 cohort were processed on the Affymetrix platform. The probe names were converted into the corresponding gene symbols according to the annotation of each cohort. Batch effects were eliminated using the ComBat method. Gene set enrichment analysis was performed with C2 gene sets obtained from the Molecular Signatures Database. The Tumor Immune Dysfunction and Exclusion (TIDE) algorithm, a computational approach to evaluate tumor immune evasion potential based on gene expression profiles, was used to predict the immunotherapy response of HCC samples [26]. The regimen of immunotherapy refers to anti-PD-1 antibodies (pembrolizumab, nivolumab) and anti-CTLA-4 antibody (ipilimumab) based on the TIDE website.

Development of the relapse-related gene signature (RGS_{HCC}) for HCC

Overlapping genes in MVI-associated SMGs and the TCGA-LIHC cohort and GSE14520 cohort were identified as relapse-related genes. Univariate Cox regression analysis of these genes was then performed in the TCGA training dataset to screen for prognosis-related genes. The least absolute shrinkage and selection operator (LASSO) Cox regression model was used to determine the final genes to be included in the signature and to calculate the corresponding gene

coefficients. Analysis was performed using the “glmnet” package in the R project with tenfold cross-validation. A prognostic signature based on the weighted gene expression was finally constructed, and the model score was calculated using the following equation: $RGS_{HCC}Score = \sum_{i=1}^n c_i x_i$, where x and c indicate the expression levels of specific genes and the corresponding coefficient, respectively.

Quantitative RT–PCR (qRT–PCR)

Tumor samples were treated with TRIzol reagent (Invitrogen) to extract total RNA using a general protocol. After quantification by a Qubit 2.0 Fluorometer (Thermo Fisher Scientific, USA), RNA was reverse-transcribed into cDNA using the PrimeScript RT reagent Kit with gDNA Eraser (TaKaRa, Japan). The PCR system was prepared using RT2 SYBR Green qPCR Master Mix (QIAGEN), and qRT–PCR was performed on the Applied Biosystems™ QuantStudio™ 3 platform (Thermo Fisher Scientific, USA). GAPDH was used as the control. The relative mRNA levels of target genes were determined using the $\Delta\Delta Ct$ method. The primers for qRT–PCR are summarized in Supplementary Table 2.

Statistical analyses

The data are presented as the mean \pm standard deviation (SD) and frequencies (proportion) according to the variable type. Quantitative data were compared using Student’s t test or the Mann–Whitney U test, and the composition ratio between groups was compared using Pearson’s chi-square or Fisher’s exact test. The Kaplan–Meier method and log-rank test were used for survival analysis. All statistical analyses were performed using SPSS (version 23.0; IBM) software or R project (version 4.0.0) with a two-sided p value of 0.05 as a statistically significant threshold.

Results

Profiles of the genomic alterations in HCC patients with and without MVI

To investigate the genomic alterations in HCC during the process of MVI, we performed WES of 45 samples, including paired tumor and peritumor tissues, PVTTs, PBMCs, and ctDNA, from 10 HCC patients (5 of whom had no MVI and 5 of whom had MVI) with an average sequencing depth of $232.74 \times$ (ranging from $198 \times$ to $280 \times$) for tissues and $339.6 \times$ (ranging from $302 \times$ to $406 \times$) for ctDNA (Fig. 1A). Approximately 99.6% of the targeted bases had a coverage depth greater than $10 \times$. We performed strict data quality control on each sample by applying filters to exclude sequence artifact features enriched in FFPE samples. Finally,

by mapping sequence reads to the human reference genome, we identified a total of 10,945,468 SNVs and 1,567,536 indels, with a median of 775 (ranging from 587 to 1,067) somatic mutations per tumor.

Our analysis of the mutation spectrum revealed a predominance of C > T and T > C transitions and C > A transitions in all samples (Fig. 1B, top panel), a feature shared by other HCC cohorts [4, 5, 27, 28]. Furthermore, the mutation spectra were fairly consistent between tumors and PVTTs from the same patient with MVI. No significant difference in mutational spectrum was observed between MVI (–) HCCs and MVI (+) HCCs.

We next performed mutation signature analysis of 96 substitution patterns to explore the temporal dynamics of the genomic landscapes in MVI (–) HCCs and MVI (+) HCCs and identified 3 signatures across almost all samples (except tumor tissue of L05) (Fig. 1B, bottom panel). Signatures A and B, predominant signatures in most samples, were mainly characterized by C > T transitions and exhibited high similarity to the previously described Cosmic Signature 5 (A: cosmic similarity = 0.82; B: cosmic similarity = 0.86), which has been reported to be associated with aging (Fig. 1C) [29]. Signature C, characterized by T > A mutations, was only dominant in 3 MVI (–) tumor samples and had a close correlation (cosmic similarity = 0.98) with Cosmic Signature 22, which is known to be a result of aristolochic acid exposure [30]. A significant feature revealed by analyses was the similar mutation signature profile of tumor and PVTT samples in HCC patients with MVI, suggesting that metastatic lesions formed after vascular invasion follow a similar mutational process as the primary tumors. In addition, we found that in HCC patients with MVI (–), the relative contributions of signatures showed substantial variation between tumors and ctDNA; in contrast, in HCC patients with MVI (+), the contribution of signatures was comparable, suggesting that ctDNA can effectively reflect the mutation features of tumors and metastases only when HCC progresses to the MVI stage.

In addition, CNV analysis revealed 56 amplified fragments harboring several genes, such as *PPIAL4D*, *TARP*, and *DADI*, and 17 deleted fragments containing genes, including *FCGRIA*, *MUC5B*, and *KDM6B*, across the entire sample (Fig. 1D, top panel). The number of CNVs in the tumor and PVTT was significantly greater than that in the peritumor tissue, and more CNVs were detected in tumor and PVTT samples from the MVI (+) group than in those from the MVI (–) group, suggesting the involvement of CNVs in hepatocarcinogenesis and vascular invasion (Fig. 1D, bottom panel). In addition, similar numbers of CNVs to those in tumor tissue could be detected in ctDNA in 3 HCC patients with MVI, indicating that the characteristics of genomic CNVs can be reflected from ctDNA to some extent in MVI (+) HCCs. Taken together, these

findings demonstrate that initial hematogenous metastases share similar genomic alterations with the primary tumor, and these features can be captured from ctDNA at the MVI stage of HCC.

Overview of somatic mutations and SMGs in HCC patients with and without MVI

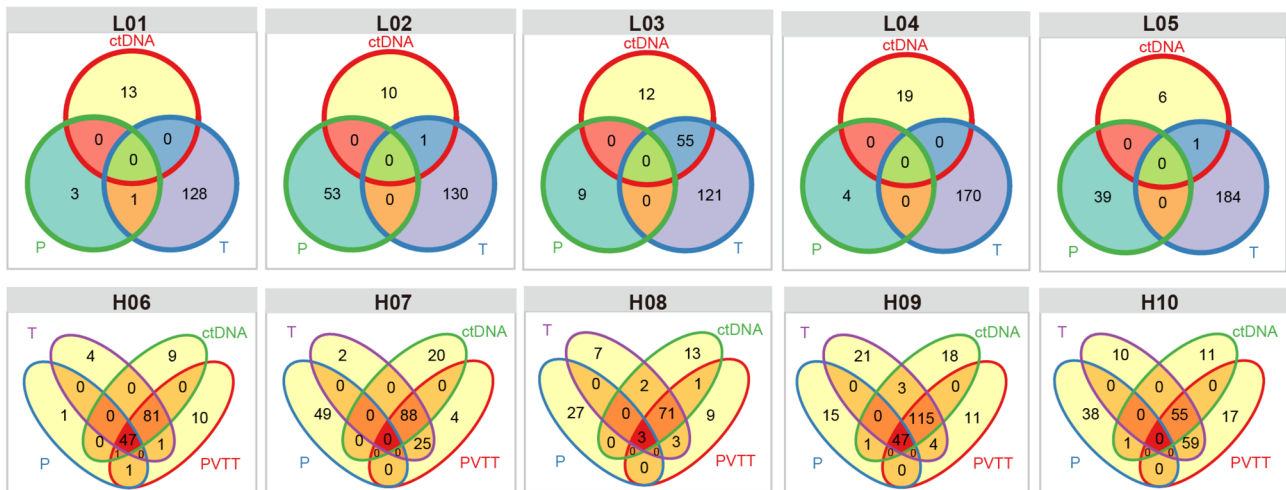
To determine the potential molecular relationships among peritumor tissue, tumor tissue, PVTT, and ctDNA in each patient, we analyzed the nonsynonymous somatic mutations, including SNVs and indels, across all samples. As shown in Fig. 2A, different extents of shared mutations were detected in tumor tissues and PVTTs of MVI (+) HCCs (accounting for 56.6%–83.2% of all mutations in each patient). In addition, tumor and PVTT samples, as well as ctDNA, of a given MVI (+) case harbored fairly overlapping repertoires of somatic mutations, although to varying extents among individuals (ranging from 28.8% to 52.3%). However, in the MVI (–) patients (except L03), there were few shared mutations between tumor tissue and ctDNA (0.0–0.5%). These results suggest a potential clonal relatedness and identical origin between the primary tumor and PVTT, as well as ctDNA, in MVI (+) HCC.

Discovery of key mutation events is an essential step to fully uncover the molecular mechanisms of tumorigenesis and progression [31]. Hence, we next revealed genes that were significantly mutated in the samples (Fig. 2B, Top 30 genes are shown). *TP53* was the most significant SMG and was mutated in all tumor and PVTT samples (100.0%) from the patients with MVI but only 2 of 5 tumor samples from the patients without MVI (40.0%). Moreover, we observed that mutations in 20 of these 30 genes (66.7%) were detected only in the MVI (+) but not MVI (–) HCC patients, suggesting that mutations affecting these genes mainly contribute to vascular invasion in HCC. In addition, somatic mutations in some genes, such as *LPHN2*, *GRIA1*, *SBSPON*, *OSBPL3*, *LGR6*, and *ZNF541*, were detected in the tumor, PVTT, and ctDNA of at least two HCC patients with MVI, indicating that mutations in these genes are more likely to be involved in MVI of HCC rather than incidental occurrence. Moreover, the profile of SMGs in ctDNA was highly consistent with that of the tumor and PVTT samples in the MVI (+) patients. Collectively, we identified some metastasis-promoting gene mutation events that occur at the primary tumor stage and are inherited in metastatic tumors and ctDNA.

The evolutionary trajectory of vascular invasion in HCC

To further determine the evolutionary pattern in vascular invasion of HCC, we constructed a phylogenetic tree for

a



b

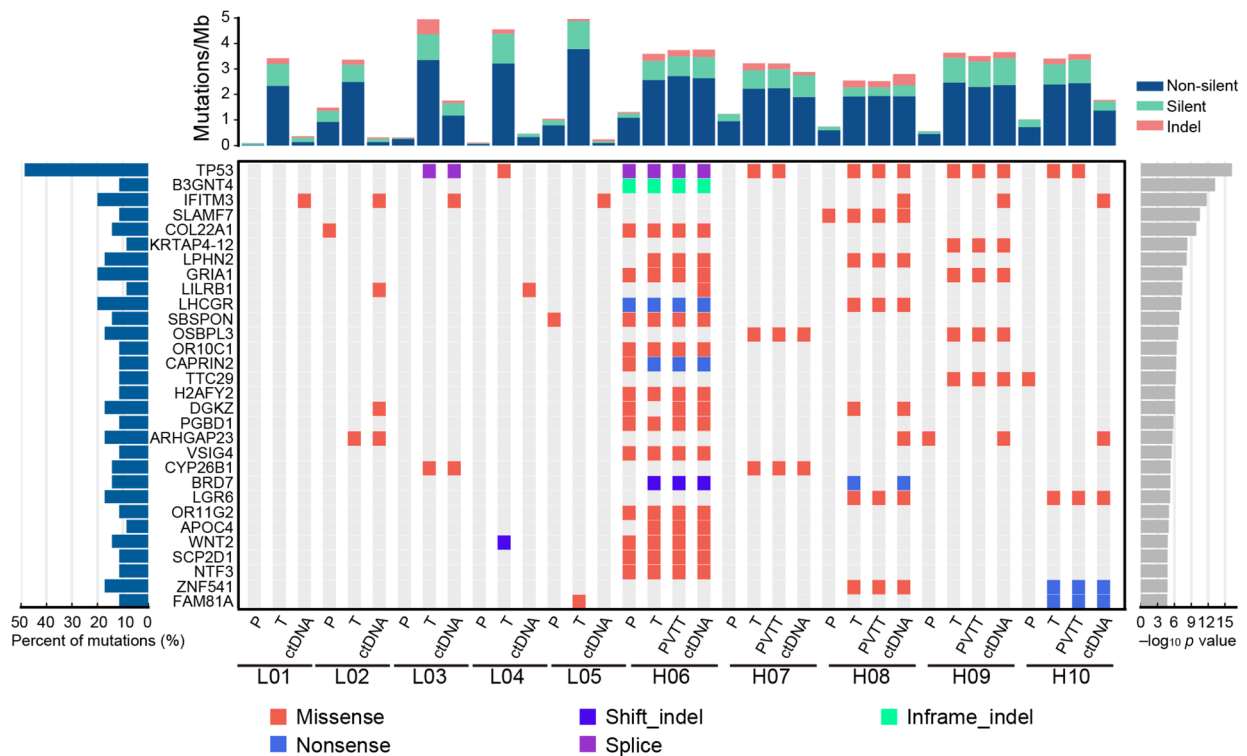


Fig. 2 Profiles of somatic mutations and SMGs in HCC patients with and without MVI. **A** Venn diagrams of shared nonsynonymous somatic mutations among peritumor (P), tumor (T), PVTT, and ctDNA samples from the MVI (–) group (L01–05) and the MVI (+)

group (H06–10). **B** The distribution of SMGs in each sample. The top 30 statistically significant genes are shown. Mutation types are annotated in the lower panel

each patient based on the somatic mutated genes (Fig. 3A). The first noteworthy finding is that the phylogenetic tree varies considerably between different MVI (+) HCC cases, i.e., different branch evolution times between individuals resulted from vascular invasion. Nevertheless, we observed different lengths of trunk genes in the highly metastatic HCC group (H06, H07, H08, H09, and H10), suggesting

a monoclonal origin of metastatic tumors, primary tumors, and ctDNA. No trunk gene length was observed in the low metastatic HCC group (L01, L02, L03, L04, and L05). In addition, as indicated by the divergent evolution of branching clones in different samples, there was significant spatial heterogeneity between geographically separated primary tumors, PVTT and ctDNA, suggesting that different patterns

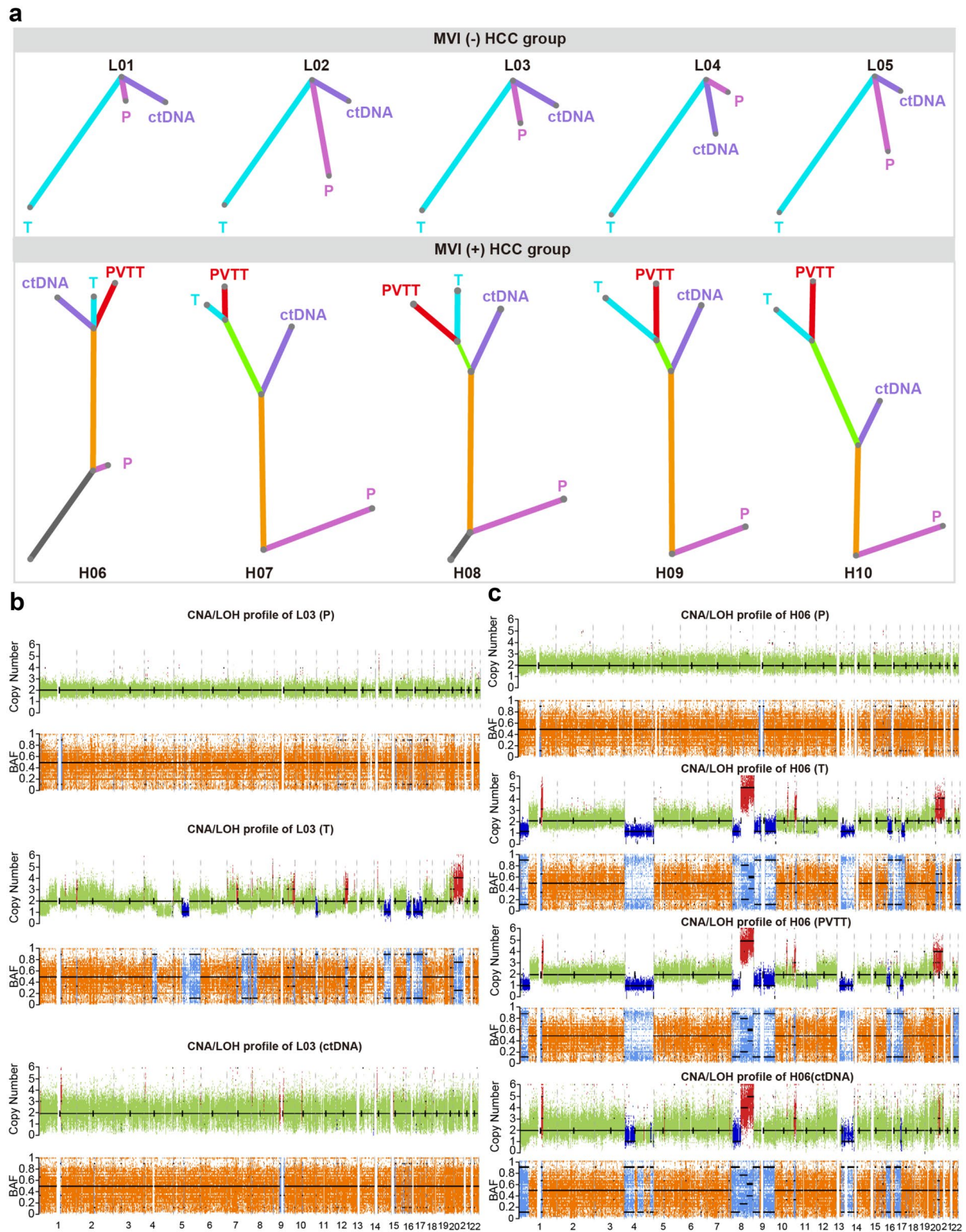


Fig. 3 The evolutionary trajectory of vascular invasion in HCC. **A** Phylogenetic trees showing the clonal relationship among peritumor (P), tumor (T), PVTT, and ctDNA samples from one HCC patient based on somatic mutations. The lengths of the branch and trunk

reflect the number of mutated genes. Driver genes are labeled on the tree. **B–C** The copy number (upper) and LOH (lower, shown as BAF distribution) profiles of peritumor (P), tumor (T), PVTT, and ctDNA samples from L03 and H06

of genetic evolution occur in primary and metastatic HCC, as well as ctDNA, upon isolation.

We next analyzed the distributions of copy number and B-allele frequency (BAF) of samples from the same patient. As illustrated by the representative profiles of the L03 and H06 cases in Fig. 3B and C, tumors, PVTTs, and ctDNA in MVI (+) HCCs shared mutation clusters, again indicating the monoclonal origin of primary/metastatic tumors and ctDNA from another perspective. This phenomenon was not observed in the HCC patients without MVI. These results confirm the clonal relevance of ctDNA to primary and metastatic tumors in MVI (+) HCC.

Dynamic changes in the mutational landscape during vascular invasion of HCC

To better understand the dynamic alterations in the mutational landscape during HCC progression and to identify specific mutations related to vascular invasion, we next performed clonal and subclonal structure analyses of the mutations in MVI (+) HCC based on the CCF value calculated using the PyClone method [32]. Comparison of CNVs harbored by three samples from the H08 cases confirmed the clonal relatedness across samples (Fig. 4A). Interestingly, as illustrated by representative clonal architectures of the H08 case in Fig. 4B, we observed that *TP53*^{R117S} maintained high cellular prevalence both in primary tumors and in PVTT, suggesting that it may be the trunk mutation and acquired at the early stage of carcinogenesis. Furthermore, the cellular prevalence of *SNX6*^{G6C}, *SLC14A2*^{E526K}, and *KIF18A*^{K57N}, which were present in the primary tumor but relatively low, increased in the PVTT samples, while *BRD7*^{Y455X} mutations were eliminated after vascular invasion. This finding indicates that HCC subclones harboring these specific mutations acquired invasive ability and finally generated vascular metastasis. In addition, novel mutations, such as *HS3ST2*^{S47T}, *SCN9A*^{W349X}, and *NEB*^{V196I}, were acquired at the vascular invasion stage in the evolutionary cascade, suggesting that the cancer cells in the tumor embolus continue to mutate to adapt to the metastatic state. We also found that trunk mutations, metastasis-promoting mutations or metastasis-specific mutations were detected in ctDNA to different extents. These results demonstrate that despite its monoclonal origin, HCC has dynamic mutational changes during vascular invasion and mutational heterogeneity between primary and metastatic foci, which can be comprehensively reflected by ctDNA.

Development of an integrated molecular signature to predict relapse for HCC

Given that genes found to be significantly mutated in MVI (+) HCC are likely to play an important role in the invasion

and relapse of HCC and that transcriptomic data can provide additional information to determine whether the expression of mutated genes is altered, we next integrated the exome and transcriptome data to construct a gene expression-based prognostic signature. A concise flowchart of signature development and validation is shown in Fig. 5A. We obtained a total of 220 shared mutated genes from MVI-associated SMGs and two public HCC datasets. By applying univariate and LASSO Cox regression analyses, we finally developed a relapse-related prognostic signature containing 11 genes and named it RGS_{HCC} . These genes, with characteristic amino acid mutation patterns, were detected in tumors, PVTT, and ctDNA of MVI (+) HCC (Fig. 5B). Based on the coefficients for each gene shown in Fig. 5C, the RGS_{HCC} score for each patient was calculated, and the median score was used as the cutoff value to determine the high- RGS_{HCC} and low- RGS_{HCC} groups. Association analyses of the risk score and clinicopathological characteristics of HCC patients indicated that patients with high RGS_{HCC} scores had worse tumor differentiation and tumor staging, larger tumor size, higher α -fetoprotein (AFP) levels, and vascular invasion than those with low RGS_{HCC} scores (Supplementary Table 3). In addition, as indicated in the TCGA training dataset, the RGS_{HCC} score successfully segregated cases with a high risk of relapse from those with a low risk (hazard ratio [HR], 2.34 [95% confidence interval (CI), 1.71–3.21], $p < 0.0001$; Fig. 5D). Furthermore, the signature was externally validated in the GSE14520 cohort (LCI validation dataset 1) and PCR cohort (LCI validation dataset 2), for which HCC patients in the high- RGS_{HCC} group experienced more relapse events than those in the low- RGS_{HCC} group (validation dataset 1, HR, 1.7 [95% CI, 1.19–2.44], $p = 0.0034$; validation dataset 2, HR, 1.78 [95% CI, 1.14–2.79], $p = 0.0097$; Fig. 5E and F).

To examine whether the RGS_{HCC} score is independent of other prognostic factors, we next performed univariate and multivariate Cox regression analyses. The results indicated that the RGS_{HCC} score remained significant for RFS after considering other prognostic variables, including MVI and TNM stage, in the TCGA training dataset (Supplementary Figs. S1A and S1B). Consistently, the prognostic signature exhibited independent predictive value in two validation datasets (Supplementary Figs. S1C–S1F). These results suggest that the RGS_{HCC} score is a robust risk factor for HCC relapse.

Biological features underlying RGS_{HCC} and its predictive value for therapeutic response

We next examined the biological features underlying the signature using gene set enrichment analysis. As illustrated in Fig. 6A, several biological processes, such as tumorigenesis, epigenetic alterations, and tumor-driving transcription,

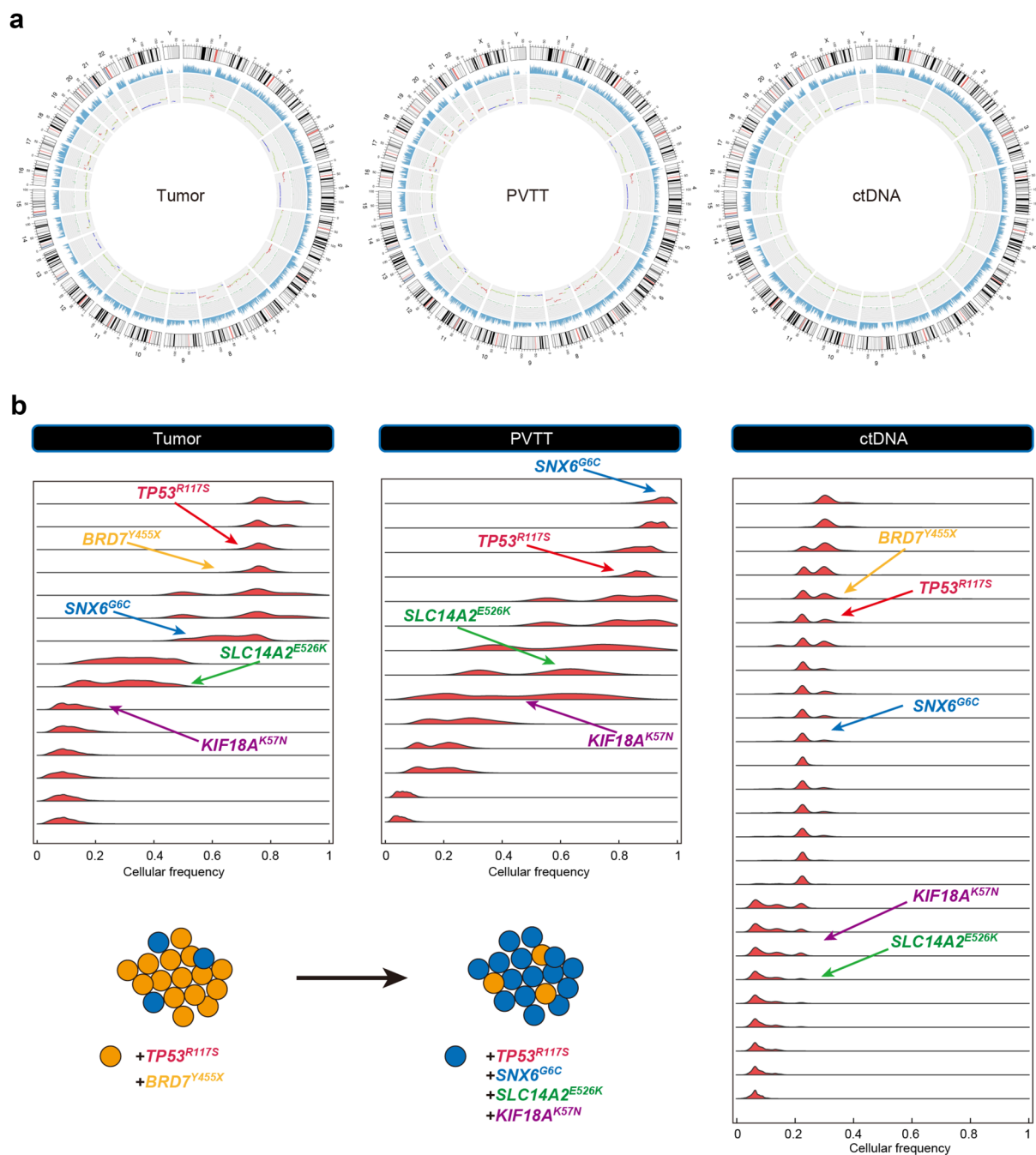


Fig. 4 Monoclonal origin and dynamic mutation alterations during vascular invasion of HCC. **A** Circos plots showing the CNVs in each sample of H08. The first circle represents the coverage of sequencing, the second circle represents the density of SNP indels, and the third circle represents the distribution of CNVs (red indicates amplification, blue indicates deletion, green indicates normal copy number). **B**

The cellular prevalence of specific somatic mutations in tumor (left), PVTT (middle), and ctDNA (right) samples from H08. The peak and distribution of somatic mutations were determined based on the allelic fraction and copy number using PyClone analysis

were enriched in the high-RGS_{HCC} group, while the low-RGS_{HCC} tumors had stronger inflammatory responses and higher chemotherapeutic sensitivity. This result drove us to investigate whether the high and low RGS_{HCC} groups had different chemotherapeutic responses. As expected, based on the IC₅₀ obtained from the GDSC database, we

found that low-RGS_{HCC} tumors tended to be more sensitive to four chemotherapy drugs approved for solid tumor treatment: docetaxel, erlotinib, temsirolimus, and dasatinib (Fig. 6B–E). Considering that immunotherapy is currently a promising treatment for HCC, we also used the TIDE algorithm to predict the immunotherapy response for each

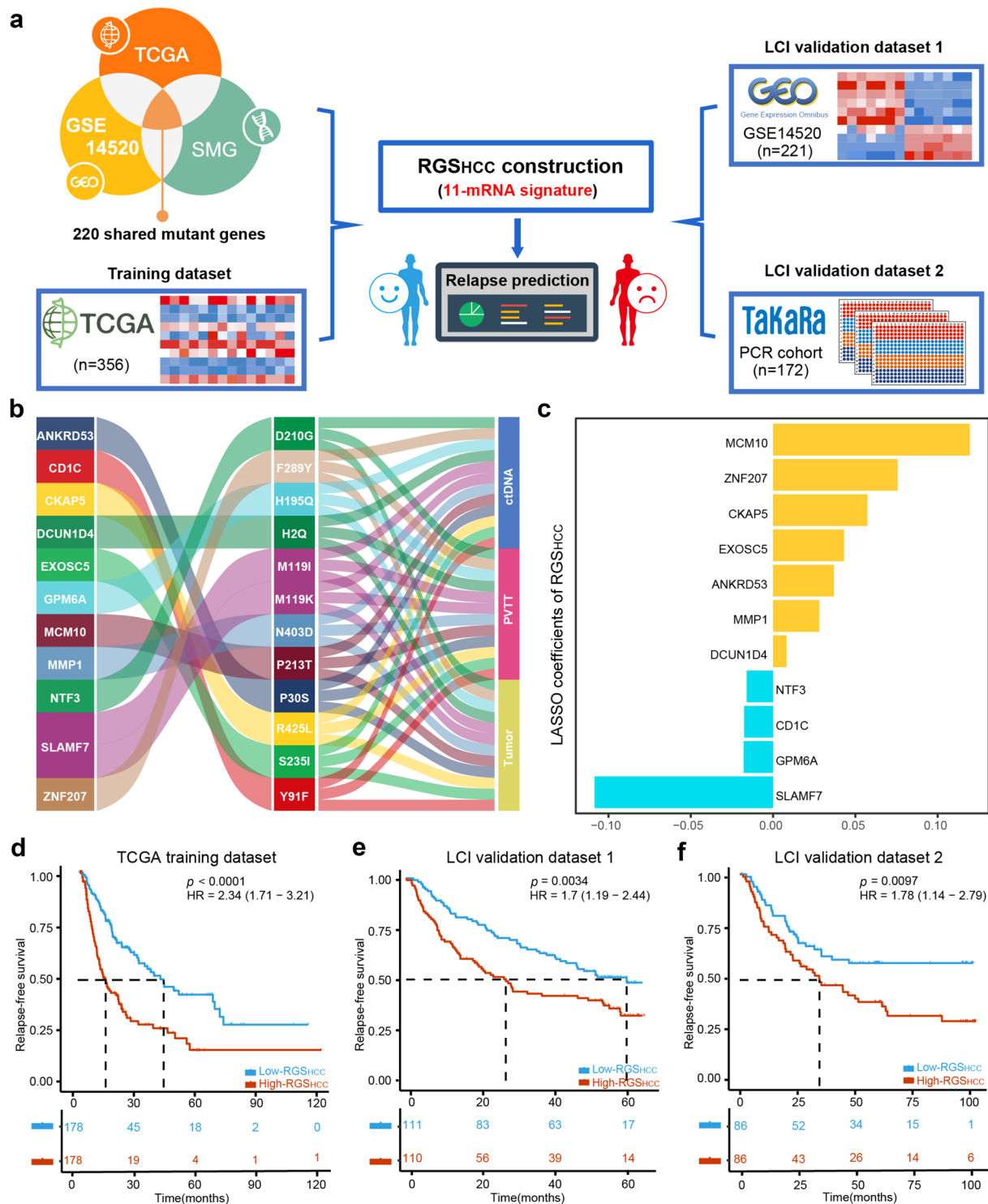


Fig. 5 Development of an 11-mRNA signature named RGS_{HCC} to predict relapse for HCC. **A** The flow chart of the signature development and validation in the present study. A total of 220 shared mutated genes in the MVI-associated SMGs and the TCGA-LIHC cohort and GSE14520 cohort were included for univariate and LASSO Cox regression analyses in the TCGA training dataset ($n=356$). In total, 11 genes were finally selected to construct the signature named RGS_{HCC}. The prognostic value of the signature

was externally validated in the GSE14520 cohort ($n=221$) and PCR cohort ($n=172$). **B** Graphical summary of the sample distribution (right) and the amino acid mutation pattern (middle) of 11 genes included in the signature. **C** The LASSO coefficient of 11 genes in the signature. **D–F** Kaplan–Meier analyses of RFS between the low-RGS_{HCC} and high-RGS_{HCC} groups in the TCGA training dataset (**D**), LCI validation dataset 1 (**E**), and LCI validation dataset 2 (**F**). Log-rank test was used for survival comparisons in Fig. 5D–F

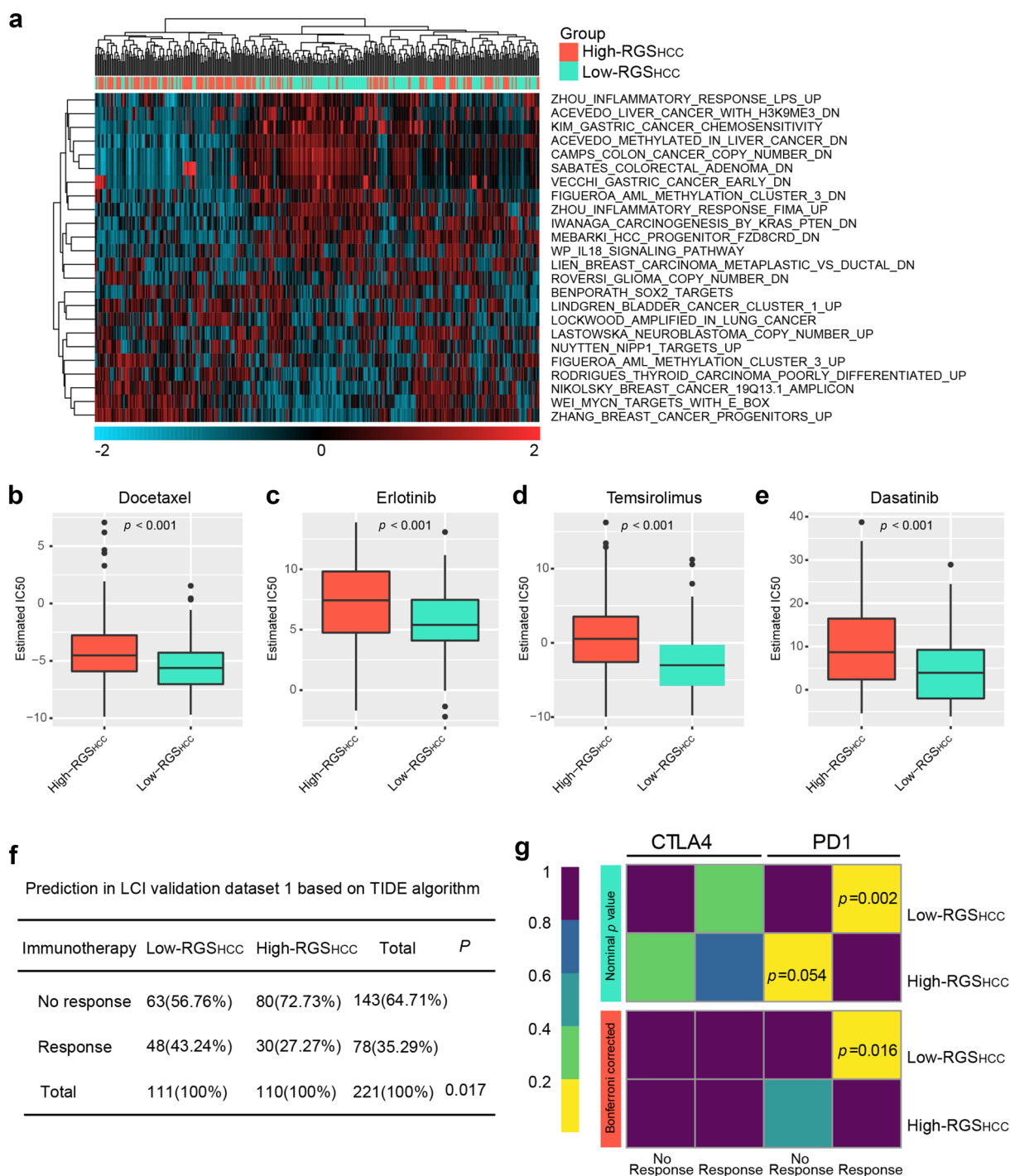


Fig. 6 The biological processes underlying RGS_{HCC} and its predictive value for therapeutic response. **A** Gene set enrichment analysis of differentially enriched gene sets in the low- and high-RGS_{HCC} groups in the TCGA dataset. **B–E** Box plots of the estimated IC₅₀ values for docetaxel, erlotinib, temsirolimus, and dasatinib between the low-

and high-RGS_{HCC} groups. **F** Association between RGS_{HCC} and the predicted immunotherapy response of HCC patients using the TIDE algorithm in LCI validation dataset 1. **G** Submap analysis of the putative immunotherapeutic response in the low- and high-RGS_{HCC} groups

sample in LCI validation dataset 1. As indicated in the four-fold contingency table in Fig. 6F, HCC patients in the low-RGS_{HCC} group had a higher immunotherapy response rate

than those in the high-RGS_{HCC} group (43.24% vs. 27.27%, $p=0.017$). Furthermore, we matched the expression profiles of HCC patients in two groups with a published cohort

of 47 melanoma patients receiving anti-CTLA4 and anti-PD-1 treatments and found that patients with low RGS_{HCC} were likely to have a better response to anti-PD-1 therapy (Fig. 6G). These findings suggest that the dismal outcomes of high- RGS_{HCC} HCC may result from more aggressive biological processes, and HCC patients with low RGS_{HCC} scores may benefit more from chemotherapy and immunotherapy.

Discussion

Vascular invasion of HCC is a spatially and temporally dynamic process that is poorly understood, resulting in very limited therapeutic interventions to effectively inhibit hematogenous metastasis and a lack of risk evaluation systems to accurately predict relapse of HCC. Here, by performing WES and comprehensive bioinformatics analyses, we obtained an explicit view of the clonality of metastatic HCC and the evolution pattern of ctDNA. First, genomic alterations that promote vascular invasion in HCC occur at the primary tumor stage and are inherited by metastatic foci. Geographically separated primary tumors and metastases will continue to evolve independently and present genetic heterogeneity. Furthermore, ctDNA can reflect the landscape of genomic alterations in metastatic HCC but has no clonal correlation with primary tumors in the absence of MVI. Moreover, we developed a practical and reliable molecular signature containing 11 genes for HCC relapse prediction based on the integration analysis of SMGs associated with MVI. This work contributes to a better understanding of the evolutionary process of HCC and to the prognostic stratification of HCC patients.

The high frequency of relapse and poor prognosis of HCC after surgery are largely the result of intrahepatic metastases through invasion of the portal vein [33]. One of our earlier studies found that the gene expression profile of primary tumors was very similar to that of metastases in metastatic HCC, thus pioneering the novel concept that gene expression events favoring metastasis were initiated at the primary tumor stage from a transcriptome perspective [34]. However, no relevant studies have been conducted to comprehensively characterize the evolutionary trajectory of HCC metastasis at the genomic level. In this work, by comparing genomic alterations, including mutation spectrum and signatures, CNVs, SMGs, and clonal evolutionary trajectories, between 5 paired tumor and PVTT samples, we found that primary tumors and metastases have very similar genomic alteration profiles and high clonal correlation, indicating a monoclonal origin, which is consistent with the clonal evolution theory of cancer [35]. In addition, our results further support and complement our previous concept that genomic alterations and expression changes promoting metastasis occur in primary tumors. This concept may have theoretical

implications for the exploration of mechanisms underlying hepatocarcinogenesis and the development of therapeutic strategies but also presents a formidable challenge for early interventions of HCC progression. In addition, we observed that the landscape of SMGs and potential MVI-driven mutations were different among individuals, suggesting that the metastatic drive of HCC involves multiple pathways and is heterogeneous, highlighting the necessity of detecting tumor genetic features before targeted therapy in each HCC patient.

In HCC, although the most prevalent mutations, such as *TP53*, are not actionable, approximately a quarter of tumors have potentially pharmacologically targetable drivers [36]. However, the biggest obstacle to achieving precision treatment for HCC is the high intratumor heterogeneity and the inaccessibility of tumor tissues [11]. In this setting, ctDNA has been proposed to serve as an alternative source of tumor DNA and can provide comprehensive and dynamic insight into the tumor genome in a minimally invasive manner, which may be a way to address the dilemma of current HCC biopsy [13]. Indeed, evidence has demonstrated that HCC patients benefit from corresponding targeted therapies based on the mutations detected in ctDNA [37]. Here, we provide new insights into the evolutionary pattern of ctDNA in HCC: ctDNA can accurately characterize tumor genomic alterations when MVI occurs but has no clear genetic correlation with primary tumors without vascular invasion. From a pathophysiological perspective, this can be explained by more cancer cells entering the circulation after the tumor breaks through the blood vessels, but more detailed molecular mechanisms need to be further investigated. This new concept suggests that ctDNA may have greater application prospects in advanced HCC and can effectively address the issues of tissue unavailability at the advanced stage of HCC, helping guide treatment decisions and dynamic detection of therapeutic efficacy. However, since the mutation frequency of single genes in ctDNA is relatively low, a ctDNA panel that integrates mutational features of multiple genes to more comprehensively, accurately, and adequately reflects the metastatic potential of HCC is warranted in the future.

By identifying SMGs and comparing dynamic changes in mutation landscapes between tumors and PVTT, we identified a series of genes and their specific mutations that may be involved in the invasion and relapse of HCC. Dissection of the subclonal architecture has demonstrated that HCC vascular invasion is associated with a lack of *BRD7*^{Y455X} mutation, enrichment of *SNX6*^{G6C}, *SLC14A2*^{E526K}, and *KIF18A*^{K57N} mutations, and acquisition of *HS3ST2*^{S47T}, *SCN9A*^{W349X}, and *NEB*^{V196I} mutations. The role of some genes, such as *SNX6*, *KIF18A*, and *HS3ST2*, in promoting tumor invasion and metastasis has been reported [38–40]. Therefore, therapies targeting these genes or specific mutations may effectively prevent the postoperative relapse of HCC. We also found that some mutations can be detected in

ctDNA. As mentioned above, whether ctDNA can be used as a new tool to obtain important molecular information in metastatic HCC deserves further investigation. On the other hand, we have to point out that the patients in this study were mainly with HBV-associated HCC, so whether the aforementioned evolutionary trajectories regarding HCC metastasis and ctDNA, as well as the gene mutations related to HCC metastasis, are also present in non-HBV-associated HCC need to be further studied. Another issue is that our sample size is relatively small, so the findings here are yet to be validated by a large-sample cohort in the future.

Finally, by integrating MVI-related SMGs with transcriptomic data from several independent HCC cohorts, we successfully developed and validated an RNA-based signature to robustly distinguish high- and low-risk relapsed HCC patients with good predictive performance. Our multivariate Cox regression analyses revealed that RGS_{HCC} score was an independent prognostic factor in HCC, which may be attributed to its use as a gene-based predictive system that can additionally reflect information on molecular features associated with metastasis and recurrence within the tumor. In addition, we revealed the biological features underlying the signature, which may explain the poor prognosis in the high- RGS_{HCC} group. The association of the RGS_{HCC} score with chemotherapeutic and immunotherapeutic sensitivity was also highlighted. Nevertheless, although we have fully validated the feasibility of applying RGS_{HCC} for relapse prediction in HCC from our institution, prospective multicenter and large-scale cohort studies are still required to thoroughly assess its application potential.

In summary, we characterized the genomic alterations during HCC vascular invasion and demonstrated a previously undescribed evolution pattern of ctDNA in HCC, which provides an important theoretical basis for a deeper understanding of HCC progression. We also developed a novel integrated multiomics-based molecular platform to identify populations with a high risk of relapse, which has significant implications for personalized treatment to prolong HCC patient survival.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s12072-023-10540-x>.

Author contributions C.Z., J.W., and S.L. designed and performed experiments, analyzed data, and wrote the manuscript. H.L., Y.Y., Y.S., and Y.Q. provided patient tissue samples and analyzed clinical data. Q.Z., Z.H., Y.Y., P.L., and J.S. performed experiments and analyzed data. Y.S., Q.D., and L.C. provided scientific input and critical constructs. X.Z. and N.R. supervised the entire project, designed the experiments, wrote the manuscript, and revised the manuscript. X.Z. and N.R. have verified the underlying data.

Funding This study was supported by grants from the National Key Research and Development Program of China (2021YFC2500403), the National Natural Science Foundation of China (82103521, 82073208), the Sino-German Mobility Program (M-0603), the Special Foundation

for Science and Technology Basic Research Program (2019FY101103), the Shanghai Sailing Program (21YF1407500, 23YF1431400), China Postdoctoral Science Foundation (2021M690674), Program of Shanghai Academic Research Leader (20XD1400900), and the Shanghai Shen Kang Hospital Development Center New Frontier Technology Joint Project (SHDC12021109).

Data availability Data are available from The Cancer Genome Atlas (<http://cancergenome.nih.gov/>) and series GSE14520 in GEO (<https://www.ncbi.nlm.nih.gov/geo/>). Other data analyzed in this study are included in this article and its supplementary materials and will be made available to other researchers on reasonable request.

Declarations

Competing interests The Chenhao Zhou, Jialei Weng, Shaoqing Liu, Qiang Zhou, Zhiqiu Hu, Yirui Yin, Peng Lv, Jialei Sun, Hui Li, Yong Yi, Yinghao Shen, Qinghai Ye, Yi Shi, Qiong Zhu Dong, Chunxiao Liu, Xiaoqiang Zhu, Ning Ren declare that they have no competing interests.

Ethical statement The study design was conducted in accordance with both the Declarations of Helsinki and Istanbul and approved by the Ethics Review Board of Zhongshan Hospital, and all participants provided written consent.

References

1. Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer statistics, 2022. *CA Cancer J Clin.* 2022;72(1):7–33
2. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin.* 2018;68(6):394–424
3. Ahn SM, Jang SJ, Shim JH, Kim D, Hong SM, Sung CO, et al. Genomic portrait of resectable hepatocellular carcinomas: implications of RB1 and FGF19 aberrations for patient stratification. *Hepatology.* 2014;60(6):1972–1982
4. Schulze K, Imbeaud S, Letouzé E, Alexandrov LB, Calderaro J, Rebouissou S, et al. Exome sequencing of hepatocellular carcinomas identifies new mutational signatures and potential therapeutic targets. *Nat Genet.* 2015;47(5):505–511
5. Totoki Y, Tatsuno K, Covington KR, Ueda H, Creighton CJ, Kato M, et al. Trans-ancestry mutational landscape of hepatocellular carcinoma genomes. *Nat Genet.* 2014;46(12):1267–1273
6. Tang ZY, Ye SL, Liu YK, Qin LX, Sun HC, Ye QH, et al. A decade's studies on metastasis of hepatocellular carcinoma. *J Cancer Res Clin Oncol.* 2004;130(4):187–196
7. Yang Y, Nagano H, Ota H, Morimoto O, Nakamura M, Wada H, et al. Patterns and clinicopathologic features of extrahepatic recurrence of hepatocellular carcinoma after curative resection. *Surgery.* 2007;141(2):196–202
8. Li J, Han X, Yu X, Xu Z, Yang G, Liu B, et al. Clinical applications of liquid biopsy as prognostic and predictive biomarkers in hepatocellular carcinoma: circulating tumor cells and circulating tumor DNA. *J Exp Clin Cancer Res.* 2018;37(1):213
9. Erstad DJ, Tanabe KK. Prognostic and Therapeutic Implications of Microvascular Invasion in Hepatocellular Carcinoma. *Ann Surg Oncol.* 2019;26(5):1474–1493
10. Subbotin VM. Privileged portal metastasis of hepatocellular carcinoma in light of the coevolution of a visceral portal system and liver in the chordate lineage: a search for therapeutic targets. *Drug Discov Today.* 2018;23(3):548–564

11. Chan SL, Wong AM, Lee K, Wong N, Chan AK. Personalized therapy for hepatocellular carcinoma: Where are we now? *Cancer Treat Rev*. 2016;45:77–86
12. Heitzer E, Haque IS, Roberts CES, Speicher MR. Current and future perspectives of liquid biopsies in genomics-driven oncology. *Nat Rev Genet*. 2019;20(2):71–88
13. Weng J, Atyah M, Zhou C, Ren N. Prospects and challenges of circulating tumor DNA in precision medicine of hepatocellular carcinoma. *Clin Exp Med*. 2020;20(3):329–337
14. Zhang Z, Chen P, Xie H, Cao P. Using circulating tumor DNA as a novel biomarker to screen and diagnose hepatocellular carcinoma: A systematic review and meta-analysis. *Cancer Med*. 2020;9(4):1349–1364
15. Ng CKY, Di Costanzo GG, Terracciano LM, Piscuoglio S. Circulating Cell-Free DNA in Hepatocellular Carcinoma: Current Insights and Outlook. *Front Med (Lausanne)*. 2018;5:78
16. Chun YS, Pawlik TM, Vauthey JN. 8th Edition of the AJCC Cancer Staging Manual: Pancreas and Hepatobiliary Cancers. *Ann Surg Oncol*. 2018;25(4):845–847
17. Zhou C, Chen W, Sun J, Atyah M, Yin Y, Zhang W, et al. Low expression of WW domain-containing oxidoreductase associates with hepatocellular carcinoma aggressiveness and recurrence after curative resection. *Cancer Med*. 2018;7(7):3031–3043
18. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2010;26(5):589–595
19. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernysky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*. 2010;20(9):1297–1303
20. Cibulskis K, Lawrence MS, Carter SL, Sivachenko A, Jaffe D, Sougnez C, et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat Biotechnol*. 2013;31(3):213–219
21. Rosenthal R, McGranahan N, Herrero J, Taylor BS, Swanton C. DeconstructSigs: delineating mutational processes in single tumors distinguishes DNA repair deficiencies and patterns of carcinoma evolution. *Genome Biol*. 2016;17:31
22. Mermel CH, Schumacher SE, Hill B, Meyerson ML, Beroukhim R, Getz G. GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome Biol*. 2011;12(4):R41
23. Yin X, Jing Y, Cai MC, Ma P, Zhang Y, Xu C, et al. Clonality, heterogeneity, and evolution of synchronous bilateral ovarian cancer. *Cancer Res*. 2017;77(23):6551–6561
24. Iijima Y, Hirotsu Y, Mochizuki H, Amemiya K, Oyama T, Uchida Y, et al. Dynamic changes and drug-induced selection of resistant clones in a patient with egfr-mutated adenocarcinoma that acquired t790m mutation and transformed to small-cell lung cancer. *Clin Lung Cancer*. 2018;19(6):e843–e847
25. Carter SL, Cibulskis K, Helman E, McKenna A, Shen H, Zack T, et al. Absolute quantification of somatic DNA alterations in human cancer. *Nat Biotechnol*. 2012;30(5):413–421
26. Jiang P, Gu S, Pan D, Fu J, Sahu A, Hu X, et al. Signatures of T cell dysfunction and exclusion predict cancer immunotherapy response. *Nat Med*. 2018;24(10):1550–1558
27. Fujimoto A, Furuta M, Totoki Y, Tsunoda T, Kato M, Shiraishi Y, et al. Whole-genome mutational landscape and characterization of noncoding and structural mutations in liver cancer. *Nat Genet*. 2016;48(5):500–509
28. Zhou SL, Zhou ZJ, Hu ZQ, Song CL, Luo YJ, Luo CB, et al. Genomic sequencing identifies WNK2 as a driver in hepatocellular carcinoma and a risk factor for early recurrence. *J Hepatol*. 2019;71(6):1152–1163
29. Letouzé E, Shinde J, Renault V, Couchy G, Blanc JF, Tubacher E, et al. Mutational signatures reveal the dynamic interplay of risk factors and cellular processes during liver tumorigenesis. *Nat Commun*. 2017;8(1):1315
30. Ng AWT, Poon SL, Huang MN, Lim JQ, Boot A, Yu W, et al. Aristolochic acids and their derivatives are widely implicated in liver cancers in Taiwan and throughout Asia. *Sci Transl Med*. 2017. <https://doi.org/10.1126/scitranslmed.aan6446>
31. Zhang J, Zhang S. Discovery of cancer common and specific driver gene sets. *Nucleic Acids Res*. 2017;45(10):e86
32. Roth A, Khattri J, Yap D, Wan A, Laks E, Biele J, et al. PyClone: statistical inference of clonal population structure in cancer. *Nat Methods*. 2014;11(4):396–398
33. Genda T, Sakamoto M, Ichida T, Asakura H, Kojiro M, Narumiya S, et al. Cell motility mediated by rho and Rho-associated protein kinase plays a critical role in intrahepatic metastasis of human hepatocellular carcinoma. *Hepatology*. 1999;30(4):1027–1036
34. Ye QH, Qin LX, Forgues M, He P, Kim JW, Peng AC, et al. Predicting hepatitis B virus-positive metastatic hepatocellular carcinomas using gene expression profiling and supervised machine learning. *Nat Med*. 2003;9(4):416–423
35. Nowell PC. The clonal evolution of tumor cell populations. *Science*. 1976;194(4260):23–28
36. Llovet JM, Montal R, Sia D, Finn RS. Molecular therapies and precision medicine for hepatocellular carcinoma. *Nat Rev Clin Oncol*. 2018;15(10):599–616
37. Ikeda S, Tsigelny IF, Skjervek ÅA, Kono Y, Mendler M, Kuo A, et al. Next-Generation sequencing of circulating tumor DNA reveals frequent alterations in advanced hepatocellular carcinoma. *Oncologist*. 2018;23(5):586–593
38. Hu P, Liang Y, Hu Q, Wang H, Cai Z, He J, et al. SNX6 predicts poor prognosis and contributes to the metastasis of pancreatic cancer cells via activating epithelial-mesenchymal transition. *Acta Biochim Biophys Sin (Shanghai)*. 2018;50(11):1075–1084
39. Luo W, Liao M, Liao Y, Chen X, Huang C, Fan J, et al. The role of kinesin KIF18A in the invasion and metastasis of hepatocellular carcinoma. *World J Surg Oncol*. 2018;16(1):36
40. Vijaya Kumar A, Salem Gassar E, Spillmann D, Stock C, Sen YP, Zhang T, et al. HS3ST2 modulates breast cancer cell invasiveness via MAP kinase- and Tcf4 (Tcf712)-dependent regulation of protease and cadherin expression. *Int J Cancer*. 2014;135(11):2579–2592

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

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

Authors and Affiliations

Chen hao Zhou^{1,2,3} · Jialei Weng^{1,3} · Shaoqing Liu^{1,3} · Qiang Zhou^{1,3} · Zhiqiu Hu^{3,4} · Yirui Yin⁵ · Peng Lv⁵ · Jialei Sun¹ · Hui Li¹ · Yong Yi¹ · Yinghao Shen¹ · Qinghai Ye¹ · Yi Shi⁶ · Qiongzhu Dong^{3,4} · Chunxiao Liu² · Xiaoqiang Zhu^{7,8} · Ning Ren^{1,3,4} 

✉ Xiaoqiang Zhu
zhuxq@connect.hku.hk

✉ Ning Ren
ren.ning@zs-hospital.sh.cn

¹ Department of Liver Surgery and Transplantation, Key Laboratory of Carcinogenesis and Cancer Invasion, Ministry of Education, Liver Cancer Institute, Zhongshan Hospital, Fudan University, 180 Fenglin Road, Shanghai 200032, People's Republic of China

² Department of Molecular and Cellular Oncology, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA

³ Key Laboratory of Whole-Period Monitoring and Precise Intervention of Digestive Cancer of Shanghai Municipal Health Commission, Shanghai 201199, People's Republic of China

⁴ Institute of Fudan-Minhang Academic Health System, Minhang Hospital, Fudan University, Shanghai 201199, People's Republic of China

⁵ Department of Liver Surgery, Xiamen Branch, Zhongshan Hospital, Fudan University, Xiamen 361015, People's Republic of China

⁶ Biomedical Research Centre, Zhongshan Hospital, Fudan University, Shanghai 200032, People's Republic of China

⁷ State Key Laboratory for Oncogenes and Related Genes, Division of Gastroenterology and Hepatology, Key Laboratory of Gastroenterology and Hepatology, School of Medicine, Ministry of Health, Shanghai Institute of Digestive Disease, Renji Hospital, Shanghai Jiao Tong University, Shanghai 200001, People's Republic of China

⁸ School of Biomedical Sciences, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong SAR 999077, People's Republic of China