Molecular evolution in different subtypes of multifocal hepatocellular carcinoma

Keyue Ding (DingKeyue@mayo.edu)
Mayo Clinic
Xia Tang
Fudan University
Lei Xiang
Chongqing Medical University
Qingshu Li
Chongqing Medical University
Yue Shao
Chongqing Medical University
Li Wan
Chongqing Medical University
Dachun Zhao
Peking Union Medical College Hospital
Xiaoyuan Li
Peking Union Medical College Hospital
Songfeng Wu
Beijing Qinglian Biotech Co., Ltd
Haijian Wang
Fudan University
Dewei Li
Chongqing University Cancer Hospital

Article

Keywords: Multifocal hepatocellular carcinoma (MF-HCC), molecular evolution, dynamic mutation signature, quantitative metastatic timing, and preneoplastic arising clone

Posted Date: February 2nd, 2023

DOI: https://doi.org/10.21203/rs.3.rs-2515298/v1

Additional Declarations: (Not answered)

Molecular evolution in different subtypes of multifocal hepatocellular carcinoma

Xia Tang, M.S.[1], Lei Xiang, M.M.[2], Qingshu Li, Ph.D.[3], Yue Shao, M.D.[4], Li Wan, M.M.[2], Dachun Zhao, M.M.[5], Xiaoyuan Li, M.M.[6], Songfeng Wu, Ph.D.[7], Haijian Wang, Ph.D.[1,*], Dewei Li, M.D.[8,2,*], and Keyue Ding, Ph.D.[9,*]

[1] Shanghai Pudong Hospital and Pudong Medical Center of Fudan University, State Key Laboratory of Genetic Engineering and Collaborative Innovation Center for Genetics and Development, School of Life Sciences, Fudan University, Shanghai, 200438, China

[2] Department of Hepatobiliary Surgery, The First Affiliated Hospital of Chongqing Medical University, Chongqing, 400016, China

[3] Department of Pathology, Chongqing Medical University, Chongqing, 400016, P.R. China

[4] Department of Cardiothoracic Surgery, The First Affiliated Hospital of Chongqing Medical University, Chongqing, 400016, China

[5] Department of Pathology, Peking Union Medical College Hospital, Beijing, 100730, P.R. China

[6] Department of Oncology, Peking Union Medical College Hospital, Beijing, 100730, China

[7] Beijing Qinglian Biotech Co., Ltd, Beijing, 102206, China

[8] Hepatobiliary and Pancreatic Cancer Center, Chongqing University Cancer Hospital, Chongqing, 400030, China

[9] Department of Medicine, Mayo Clinic, Rochester, MN 55905, United States

*Corresponding authors

Haijian Wang, Ph.D.

Address: #2005 Songhu Road, Yangpu District, Shanghai, 400030, P.R. China

Telephone: +86 137 6186 0169 Fax: +86 21 3124 6678 Email: haijiangwang@fudan.edu.cn;

Dewei Li, M.D.

Address: #181 Hanyu Rd, Shapingba District, Chongqing, 400044, P.R. China

Telephone: +86 136 1831 7772 Fax: +86 23 6389 3697 Email: deweili406@sina.com;

Keyue Ding, Ph.D.

Address: 200 1st Street Southwest Rochester, MN 55905, United States Telephone: +1 507 538-6420 Email: Ding.Keyue@mayo.edu

Abstract

Background: Multifocal hepatocellular carcinoma (MF-HCC) accounts for >40% of HCCs, demonstrating poor prognosis than single primary HCCs. Characterizing the dynamic changes of mutation signature along with clonal evolution, quantifying intrahepatic metastatic timing, and investigating the genetic structure in preneoplastic stage underlying different subtypes of MF-HCC is important for understanding their molecular evolution and developing precision management strategy.

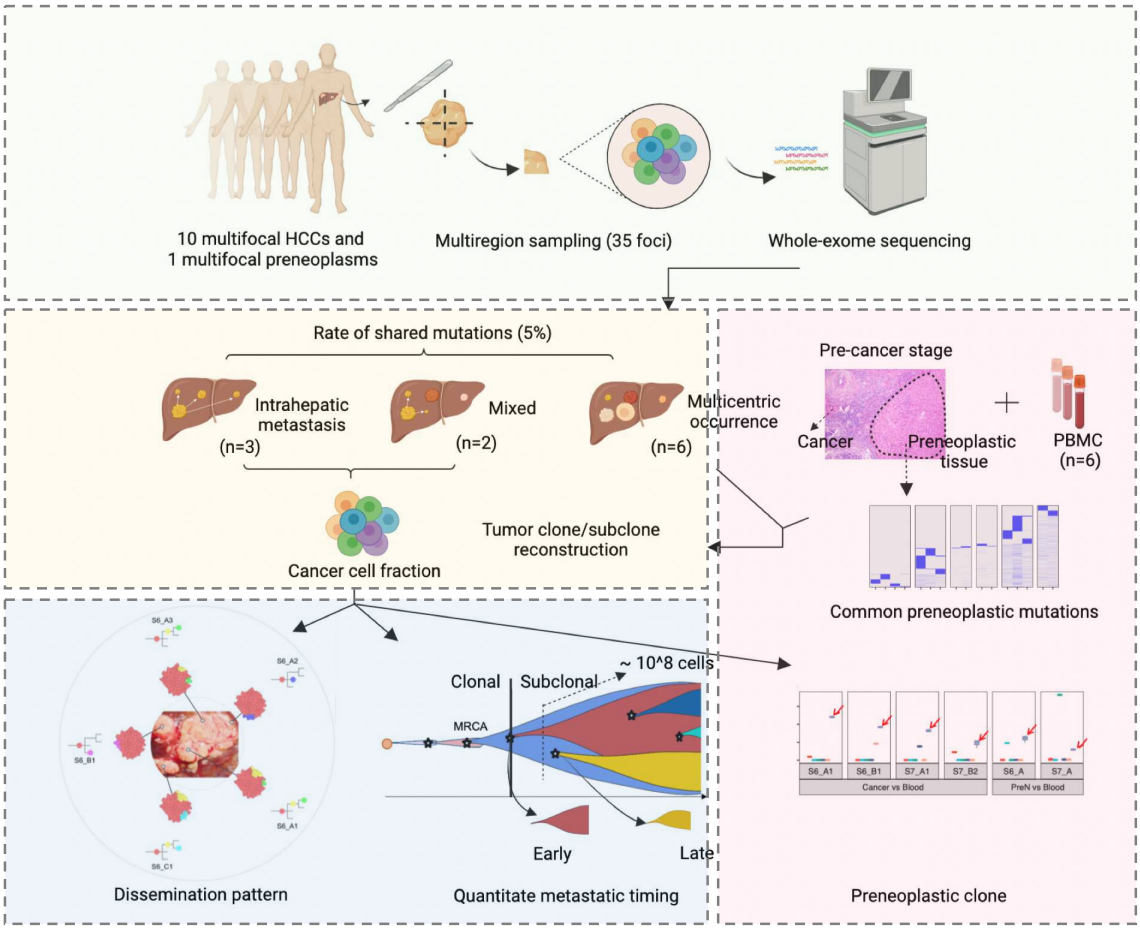
Methods: We conducted whole-exome sequencing in 74 tumor samples from spatially distinct regions in 35 resected tumor lesions and adjacent non-cancerous tissues in 11 patients with synchronous MF-HCCs, 15 histologically confirmed preneoplastic lesions, and six samples from peripheral blood mononuclear cell. We also used the previously published MF-HCC cohorts (*n*=9) as an independent validation dataset.

Results: We classified MF-HCCs into three subtypes, including intrahepatic metastasis, multicentric occurrence, and the mixed intrahepatic metastasis and multicentric occurrence. The dynamic changes in mutation signatures between tumor subclonal expansions demonstrated varied etiologies underlying the clonal progression in different MF-HCC subtypes, especially the contribution of aristolochic acid exposure decreased along with subclonal expansions. The clonal evolution in intrahepatic metastasis exhibited an early metastatic seeding that occurred at 10^{-4}–0.01 cm^3 in primary tumor volume, below the limitation in clinical detection. These findings were validated in another independent cohort. In addition, for multicentric occurrence MF-HCC,

mutational footprints in the preneoplastic lesions revealed common preneoplastic arising clones, evidently being ancestor of distinct tumor sites.

Conclusion: Our study comprehensively characterized the dynamic mutational signature, early metastatic timing, and common preneoplastic clones in multicentric occurrence that characterized the varied tumor clonal evolutionary history underlying different subtypes of MF-HCC, and provided important implications for optimizing personalized clinical management for MF-HCC.

Keywords: Multifocal hepatocellular carcinoma (MF-HCC), molecular evolution, dynamic mutation signature, quantitative metastatic timing, and preneoplastic arising clone



Introduction

Approximately 41-75% of HCC patients presented multifocal lesions at diagnosis with significant anatomical and biochemical heterogeneity (i.e., multifocal hepatocellular carcinoma, MF-HCC) (1), exhibiting an increased tumor burden with poor prognosis than single primary HCCs (2). MF-HCCs can occur synchronously or metachronously either from intrahepatic metastasis or multicentric occurrence (3,4), which different MF-HCC subtypes exhibit varied biological behaviors and demonstrate different aggressiveness. Comprehensively understanding of the molecular features of MF-HCCs could bringing new insights into the pathogenesis and management in clinical practice, including surgical prioritization and allocation, adjuvant therapy, and follow-up after resection (3).

Next-generation sequencing based investigation have benefits for increasing pathological discrimination of MF-HCCs (3), and revealed pervasive genomic and transcriptional heterogeneity across (interfocal) and within (intratumoral) lesions (5), which may result in different responsiveness to targeted drugs (6). Furthermore, MF-HCCs exhibited distinct genetic structure and evolutionary history (7–9), and constrained by heterogeneous immunogenomic footprints and escape mechanisms (10). However, molecular features during clonal evolution, i.e., dynamic change of mutation signature, quantitatively intrahepatic metastatic timing, and precancerous molecular carcinogenesis of multicentric occurrence has yet to be vigorously investigated.

Molecular features such as mutational signatures changed significantly during tumor evolution. For example, mutation process is associated with tumor subclone

emergence (11), and changes of mutational signatures coincided with the contribution of etiology change in shaping tumor clonal evolution. Previous studies have shown that mutational processes related to UVB exposure in melanomas were remarkably stable between tumor subclonal expansions, while the activity of mutational signatures changed ubiquitously in gastrointestinal solid tumors (12). Tracking the mutational signature change could beneficially inform possible timely intervene of extrinsic etiology.

Synchronously MF-HCCs with intrahepatic metastasis indicated the aggressiveness of metastatic capability. The traditional forward-time model posited that metastatic seeding occurred late in linear progression, while the recently proposed 'looking backward' model showed that metastasis might have happened before the primary site could be clinically detected (13,14). Clocking the metastatic timing with quantitative evidence could reflect the evolution features in an enhanced time resolution and lay the foundation in clinical decision-making. However, quantitative evidence for metastatic timing of synchronous MF-HCC has yet to be established. In addition, preneoplastic lesions arising from cirrhosis may be significantly predisposed to multiple *de novo* tumors, as posited in a 'field effect' hypothesis (9). The process that preneoplastic lesions hatched multicentric tumors (i.e., multicentric tumorigenic potential) and the mutation footprint before the malignant transformation was largely unknown.

Here, we performed whole-exome sequencing in 74 tumor samples from spatially distinct regions in 35 resected foci and adjacent non-cancerous tissues in 11 patients with synchronous MF-HCCs, 15 histologically confirmed preneoplastic lesions, and six peripheral blood mononuclear cell (PBMC) samples. We included a previously

published MF-HCC cohort (*n*=9) as an independent validation dataset. We aimed to delineate their molecular features during evolution, including the dynamic changes for the single base substitution (SBS) signatures between subclonal expansions, quantitative intrahepatic metastatic timing, and the genetic structure in the preneoplastic lesions of multicentric occurrence, to benefit the precision oncology in the clinical management of MF-HCCs.

Methods and materials

The study subjects

We recruited 11 patients diagnosed with synchronous MF-HCC who underwent surgical resection without chemotherapy. Two experienced pathologists confirmed the diagnoses of MF-HCC in 10 patients and one with multiple preneoplastic lesions (*Figure S1*). No evidence for 'clinically detected' extrahepatic metastases was shown. We performed multiregion sampling for 74 regions at different spatial locations from 35 resected foci (*Table S1*) if the tumor size allowed. In addition, the adjacent non-cancerous tissues and PBMC were collected. We also sampled tissues from the pathologically confirmed preneoplastic lesion adjacent to each lesion in six multicentric occurrence patients (*Table S2*). The ellipsoid tumor volume was calculated as $V = \frac{4}{3}\pi \times a \times b \times c$, where *a* and *b* are the half-length of two axes of the tumor measured based on the magnetic resonance imaging (MRI) or computed tomography (CT) scan, and *c* is the mean of *a* and *b* (*Table S3*). The study was approved by the Institute Review Board (IRB) of the First Affiliated Hospital of Chongqing Medical University, and all patients signed the written informed consent.

DNA extraction and whole-exome sequencing

Total DNA was extracted from the formalin-fixed paraffin-embedded (FFPE) tumor, adjacent non-tumor, and preneoplastic tissues, and PBMC using the QIAamp DNA Mini Kit (QIAGEN, Hilden, Germany). The libraries were constructed by the SureSelect Target Enrichment System (G7530-90000) (Agilent Technologies, Santa Clara, CA), and whole-exome sequencing was performed on the Illumina Hiseq-4000

platform, as described previously (15). The sequencing depth was averaged 142.6 (± 58.8) for tissues and 38.8 (± 5.2) for blood samples.

Identification of somatic mutations and copy number alterations

Following the Genome Analysis Toolkit (GATK) (v4.1.6) best practice guide (16), we mapped paired-end sequencing reads (150bp length) to the reference sequence (hg38) using BWA (v0.7.15). We removed PCR duplicates, realigned reads around indels, and recalibrated base quality. Somatic mutations (i.e., single nucleotide variants (SNVs) and small insertions and deletions (indels)) were called by Mutect2 (17) and were further filtered by 'FilterMutectCalls'. We finally annotated somatic mutations using 'Funcotator', and all nonsilent mutations (i.e., missense, nonsense, splice_site, frameshift indels, and inframe indels) remained for further analyses. We identified somatic copy number alterations (sCNAs) segments, together with tumor cellularity and ploidy using Sequenza (18), and then assessed significant sCNAs in all sequenced samples using GISTIC2 (19).

Assessment of MF-HCC origins

To assess MF-HCC origins, we calculated the proportion of shared mutations in pairwise tumor samples within and between lesions. A threshold of 5% for the shared mutations in any two samples from different lesions (i.e., interfocal) was suggested for discriminating intrahepatic metastasis MF-HCC, given rare shared somatic mutations among multicentric pairs and potential sequencing artifacts (3).

Quantification for tumor heterogeneity

We used the mutant-allele tumor heterogeneity (MATH)-score for measuring intratumor heterogeneity according to a vector of variant allelic fractions (VAFs) from a tumor sample (20). The MATH-score was calculated as the percentage ratio of the median absolute deviation (MAD) of its VAF values to the median of the distribution of VAFs, adjusted by 1.4826 (MATH_score = MAD $\times$ 1.4826/median(VAFs)). A more heterogeneous tumor with a higher MATH score tends to have a wider distribution of VAFs among all mutation loci and centers at a lower fraction. A cutoff value of 32 MATH units was used to distinguish high- from low-heterogeneity tumors (21).

Reconstruction of tumor subclones

A Bayesian clustering method - PyClone - was used for grouping nonsilent somatic mutations into putative clusters while estimating their cellular prevalence (or cancer cell fraction, CCF) and accounting for allelic imbalances (i.e., segmental copy-number changes and normal-cell contamination) (22). Using the estimated cellular frequencies of mutations, we inferred the clone tree by exhaustively exploring all possible trees, reporting those with the highest likelihood, and assigning mutations to the tree nodes using citup (clonality inference in multiple tumor samples using phylogeny) (23). Finally, the tumor clonal evolution was visualized by mapscape (24). To enable a high-confidence clonal analysis, we included somatic nonsilent mutations with VAF $>$ 0.01 and a minimum sequencing depth of 20$\times$. The clonal mutation was defined from the estimated CCF ($\geq$ 0.8), whereas the subclonal mutation with CCF $<$ 0.8. We finally calculated the Jaccard similarity index (JSI) - a measure for gauging the

similarity and diversity of tumor samples - to infer the metastatic pattern. JSI was defined as $JSI = W_s/(L_m + L_p + W_s)$, where W_s was the shared subclonal mutations between primary and metastasis, L_p was the number of the private clonal mutations in primary, and L_m was the private clonal mutations in metastasis (25). Polyclonal dissemination was considered when $JSI > 0.3$, and monoclonal dissemination otherwise (14).

Changes of mutational signature between subclonal expansions

A rigorous method has been developed for deciphering mutation signatures (26). First, we extracted the *de novo* mutational signatures using the negative matrix factorization (NMF) algorithm, where an NMF rank was estimated to be three. Signatures were considered similar to the known COSMIC signatures (27) if the cutoff > 0.85. Then, we identified the optimal contribution of COSMIC signatures to the mutational profile of the samples. Finally, a bootstrapped refitting was performed to verify the refitting stability, given the number of bootstrap iterations of 500 and the maximum difference in original versus reconstructed cosine similarity between two iterations of 0.002. All analyses were implemented in the *R* 'MutationalPatterns' package (v3.4) (28).

To characterize the changes of mutational signatures during subclonal expansions, we stratified mutations with different cancer cell fractions (CCF) into five stages, i.e., CCF in [1.0, 0.8), [0.8, 0.6), [0.6, 0.4), [0.4, 0.2), and [0.2, 0), reflecting pseudo-time in clonal evolution. The same approach was used to extract and map

mutation signatures from each CCF stage, and its dynamics along with clonal evolution was thereby tracked.

Inference of intrahepatic metastatic timing

Inferring the primary site from synchronous lesions in MF-HCC remained uncertain based on clinicopathological criteria, although a lesion with the most significant volume was usually considered. We inferred the primary site in a specific spatial location (i.e., a sequenced tumor sample) using the mean of cellular prevalence assigned to each cluster (22), given that all subclones were derived from the ancestral tumor clone and the primary site harbored the cluster with the most significant cellular prevalence. To quantify the evolutionary dynamics of metastasis, then, we first estimated the time from primary tumor initiation to metastasis (T_{pm}) and from metastasis to dissection (T_{md}) using the proposed framework (13). The H value, defined as $H=L_m/L_p+1$, was used to estimate the metastatic timing, where $H<20$ was considered early dissemination. Then, we estimated the primary tumor size at the time of dissemination (N_d) and the likelihood of tumor evolution mode (e.g., neutral (N) or subclonal selection (S)) in primary and metastasis. Early dissemination was defined as the upper bound of $N_d < 10^8$ cells (approximately 1 cm^3 in volume); otherwise, later dissemination (i.e., $N_d \geq 10^8$). We implemented these procedures using SCIMET (13).

Characterization of mutations in preneoplastic lesions in multicentric occurrence patients

The underlying mechanism of the ‘synchronous’ occurrence of multiple independent lesions was unclear. However, independent lesions from the same field

were genetically correlated, and a typical ‘preneoplastic clone’ may arise before malignant transformation (29). We expected to move backward to the precancer stage to investigate the molecular background for the multicentric occurrence, and try to find genetical ancestral for the synchronously distinct tumor sites. Therefore, for MF-HCCs patients with multicentric occurrence, we collected pathologically confirmed preneoplastic tissues adjacent to each tumor lesion for whole-exome sequencing (*Table S2*). In addition, we used their PBMCs to surrogate normal liver tissues to identify somatic mutations accumulated in the preneoplastic lesions and combine with mutations identified in cancer tissues for further analyses. We implemented the same somatic mutation calling and clonal/subclonal reconstruction protocol in these collected precancer lesions.

External validation

We used whole-exome sequencing data from nine MF-HCC patients (NCBI SRA accession number: SRP062373) (2) for independently validating the dynamic changes of SBS signatures and early intrahepatic metastatic seeding.

Results

Distinct somatic mutational landscapes in MF-HCCs subtypes

We summarized the clinical characteristics of the 11 patients in *Table 1*, and pathological reviews confirmed surgically-collected tumor or preneoplastic lesions (*Table S1*). All patients were male from 24 to 66 years (median 59) at diagnosis, and 10 had a history of chronic hepatitis B virus (HBV) infection. These patients harbored 2-6 synchronous lesions (*Figure S1*) with varied tumor sizes (*Table S3*) and most lesions showed moderate or poor differentiation. The Child-Pugh score for cirrhosis mortality was mild or moderate, and AJCC cancer staging showed stage III in three patients. Six patients showed signs of vascular invasion, and special pathological features were noted, including a highly aggressive multinodular mass along with the portal vein tumor thrombosis (p551), a primary tumor accompanied by satellite nodules (p424), and a nodule-in-nodule phenotype (p331). Nine patients exhibited local recurrence with an average of 24 months of follow-up.

Using the threshold (5%) of shared mutations in any two samples from different lesions (*Figure S2*), we classified these patients into three subtypes, including the intrahepatic metastasis (IM, *n*=3), multicentric occurrence (MO, *n*=6), and the mixed intrahepatic metastasis and multicentric occurrence (IM_MO, *n*=2). The mixed one suggested that different mechanisms may have operated on multiple lesions during the tumor development. However, the subtype was not associated with clinical variables partly due to the small sample size (*Table S4*).

Distinct somatic mutational landscapes showed different heterogeneity at different levels, as demonstrated by interpatient > interfocal > intratumor (*Figure 1A*). An average of 119 nonsilent somatic mutations per sample (16-548, 0.24-8.4/MB) was identified (*Figure 1B*). Intrahepatic metastasis showed a significantly increased tumor mutation burden than the remaining two subtypes (p=8e-5 and p=6e-5, respectively). A high intratumor heterogeneity (MATH-score>32) was noted in the majority of tumor samples (*Figure 1C*). It differed significantly among patients (analysis of variance (ANOVA), p=3e-4), interfocal (p=2e-3) but not intratumoral samples (p=0.1), although not associated with recurrence (*Table S5*). Of note, mutations in canonical HCC drivers (n=111, curated from 606 potential cancer driver genes (30–32), see *Table S6*) were also varied in three subtypes (e.g., *TP53*, *AXIN1*, *SMARCA4*, and *NTRK3*) (*Figure S3A*). The mutational burden in drivers in intrahepatic metastasis was more significant than other subtypes (p=3e-5 and p=1e-4, respectively). Clonal drivers were expected to have a relatively higher proportion, although clonal passengers might have potential roles in carcinogenesis (*Figure S3B*). The clonal and subclonal drivers differed significantly among subtypes (*Figure S3C*), e.g., a relatively higher proportion of clonal drivers in intrahepatic metastasis. Overall, driver mutations were enriched in known signalling pathways involved in cancer, and several exclusive pathways were exhibited in different subtypes (*Figure S3D*), e.g., the ‘p53 signaling pathway’ in intrahepatic metastasis and the ‘PI3K-Akt signaling pathway’ in the mixed subtype. Of note, ‘hepatitis B’ pathway in multicentric occurrence implicated the essential role of HBV infection. These findings showed that the vast genetic structures coupled with etiological backgrounds, contributed to the complex phenotype of MF-HCC.

Finally, we identified a total of 57 somatic copy number alterations (sCNAs) peaks (residual q<0.05 after removing segments shared with higher peaks) (*Figure S4*). Strikingly, some sCNAs were universal although varied significantly at different levels (*Figure 1D*). For example, amplification in 4p11, 5q11.1, or 10q11.21 was noted in nearly all lesions in p331 and p460, and a subset of sCNAs was identified nearly in all independent lesions in multicentric occurrence patients (e.g., 11p15.5 in p900 and p058). These patterns suggested that sCNAs leading to DNA damage during the MF-HCC development may have occurred earlier and remained relatively stable throughout tumor evolution (33).

Dynamic changes of mutational signatures in the development of MF-HCCs

Three *de novo* mutational signatures contributed moderately, but differed in these three subtypes. For example, the SBS22-like signature (aristolochic acid (AA) exposure) enriched in intrahepatic metastasis, the SBSA-like signature in multicentric occurrence, and the SBS5-like signature (tobacco smoking) in the combined subtype (*Figure 2A*). However, most multicentric occurrence MF-HCCs did not exhibit dominant signature, indicating that its etiology is mostly unknown (*Figure 2B*). Refitting the mutation type with the COSMIC signatures showed a mild or moderate contribution (<0.5), including SBS22, SBS5, SBS24 (aflatoxin exposure), SBS25 (unknown etiology), and SBS26 (defective DNA mismatch repair) (*Figure S5*). In patient p424, SBS22 and SBS24 were exclusively dominant among all tumor sites, coincided with their intrahepatic metastases with similar mutation signatures in each sites (*Figure S5*). In summary, the heterogeneity of mutation signature was less pervasive than genetic

aberrations between and within these subtypes, but one or two etiologies were dominant in each patient.

Dynamic changes in the mutation signature underlying molecular evolution could reflect the extrinsic or intrinsic etiology shift in shaping the cancer progression. Of the 74 samples with sufficient SNVs, 58 (78%) had an activity change of more than 6% in one or more signatures. We detected an average of 1.82 change of the activity of mutational signature per sample. Overall, mutation signature activity was unstable between subclonal expanisions, especially in intrahepatic metastasis patients (*Figure 2C*). However, in the combined subtype, we did not note that the activity of mutational signature changed significantly (*Figure 2C* middle), possibly due to the mixed phenotype exacerbating difficulty in tracking the etiology change. Specifically, SBS22, SBS24, SBS29 (tobacco chewing), and SBS89 (unknown) significantly changed during the tumor evolution (*Figure 2D*). SBS29 increasingly contributed to the subclonal expansions in intrahepatic metastasis patients (+18%), highlighted its role in driving cancer progression over time. However, the contribution of SBS22 (AA exposure) in intrahepatic metastasis decreased dramatically (-45%, *Figure 2D* left), suggesting a potential role of AA exposure in the early carcinogenesis instead of driving the cancer development. The increased SBS22 (+11%) in multicentric occurrence also supported its role in cancer initiation from the other way around (*Figure 2D* right). The same change was also observed in SBS24 (aflatoxin exposure), which decreased in intrahepatic metastasis (-6%) but increased in multicentric occurrence (+9%). In the independent validation cohort (2), similar changes of SBS signatures (SBS22 and

SBS24) also consistently highlighted the roles of AA and aflatoxin exposure in early carcinogenesis of MF-HCC (*Figure S6*).

Quantitative evidence for early intrahepatic metastasis timing

The reconstructed tumor clones/subclones and their phylogeny revealed either linear (p551) or branching evolution (p424 and p321) in MF-HCCs (*Figure 3A* and *Figure S7*), and highlighted that the metastases were disseminated from spatial-specific primary site. Similar to the mutational landscape, the tumor clonal components differed remarkably among interfocal samples than intratumoral samples. A complex metastatic pattern (i.e., with both linear and branching evolution) was evident in the mixed subtype (p331 and p460) (*Figure 3B*). Of note, two samples in p460 constituted 'pure' tumor clones (i.e., red in S3_A2 and yellow in S5_A1), indicating a sign of independent origins. The refined spatially clonal evolution pattern enabled identifying the specific lesion in primary site for initiating metastasis. For example, metastatic sites (S8_B{1,2} and S8_C{1-4}) in p460 were seeded from a specific region (S8_B3) within the inferred primary lesion (S8_B); whereas in p331 with a 'nodule-in-nodule' phenotype, one metastatic satellite tumor (S6_C) harbored multiple clones shared with S6_A{1-3}, seeded from S6_A1. The Jaccard similarity index (JSI) further supported the metastatic pattern, i.e., a low JSI indicating a linear (or monophyletic) evolutionary pattern (p551) and a high JSI indicating branching (or polyphyletic) dissemination (p424 and p331) (*Figure 3C*).

Metastatic seeding was initiated when the primary tumor was most likely at 10^{-4}-0.01 cm^3 in volume (corresponding to 10^4 and 10^6 cells) (*Figure 4A*), where a volume of

1 cm^3 (at a number of cells of 10^8) was usually required to be clinically detected. For example, the inferred primary site (S8_A) in p221 initiated metastatic seeding at 10^5 cells with a similar mutation rate (μ=0.6, per cell division in exonic regions). Two evolutionary scenarios were exhibited in primary/metastasis (P/M) pairs, i.e., selection/neutral (S/N) (e.g., all P/M pairs in p424) and selection/selection (S/S) (e.g., six out of eight P/M pairs in p221) with selective subclonal evolution. A relatively lower *H* value (<20) in all P/M pairs also supported an early metastatic seeding (*Figure 4B*). The estimated median time from primary tumor initiation to metastasis (T_{pm}) and from metastasis to dissection (T_{md}) was 253 (range: 202-304) and 175 (123-203) days, respectively (*Table S7*). The *H* value was positively correlated with N_d (*Figure 4B*) and T_{pm} (*Figure 4C*), suggesting that a larger size of the primary tumor was related to a long time from initiation to dissemination due to the required time for tumor cell proliferation (*Figure 4D*). One independent validation cohorts also confirmed our findings that early metastatic seeding in the majority of metastatic sites (in six out of nine intrahepatic metastatsis patients in the cohort (2))(*Figure S8*). Our findings provided direct quantitative evidence that early metastatic seeding before clinically detected was relatively common in intrahepatic metastasis MF-HCC.

Common mutations in preneoplastic lesions in multicentric occurrence MF-HCC

Using PBMC as a surrogate for normal liver, we noted that somatic mutations accumulated in preneoplastic lesions differed from tumors (*Figure S9A-B*). The mutational burden was significantly greater than in cirrhotic tissues in 12 solitary HBV-associated single primary HCC patients in our previous study (15) (p=0.01) (*Figure S9C*). Of common somatic mutations identified in the preneoplastic lesions (*Figure 5A*),

five patients showed a proportion of shared nonsilent somatic mutations >5% among preneoplastic lesions (*Figure 5B*), suggesting that the divergence of independent lesions may have occurred after malignant transformation. Furthermore, somatic mutations in several known HCC drivers (e.g., *APOB*, *ALB*, *BIRC6*, *AKAP9*, and *BRCA2*) were identified in preneoplastic stage (*Figure 5C*), which were enriched significantly in canonical cancer signaling (e.g., Rap1, PI3K-Akt, and MAPK) and ECM-receptor interaction pathways (*Figure S9D*). These results highlighted their carcinogenesis potential in the preneoplastic stage. The comparison of SBS signatures between preneoplastic lesions (*Figure S9E*) and tumor stage (*Figure S5*) showed that, with the progression from preneoplastic lesions to cancer, the dominant SBS6 (defective DNA mismatch repair) decreased, whereas SBS22 (AA exposure) - hardly detected in the preneoplastic lesions - emerged in cancer tissues (*Figure S5*). Therefore, an internal defect in DNA mismatch repair and exposure to exogenous AA and tobacco chewing may underline carcinogenesis in multicentric occurrence MF-HCC.

Given commonly accumulated mutations in preneoplastic lesions (*Figure S10*), we hypothesized that 'preneoplastic arising clones' could be identified in the preneoplastic lesion and its corresponding tumor sample. Therefore, we reconstructed the clonal structure by combining somatic mutations by comparing tumor samples and preneoplastic lesions with blood. In p090, clone #9 (i.e., a preneoplastic arising clone), as the red arrow indicated, was typically identified in preneoplastic and tumor tissues with a relatively higher CCF, indicative of its role in the preneoplastic stage. The observation was applied to clone #5 in p360 and clone #11 in p058. Although a

preneoplastic arising clone was not apparent in p900, preneoplastic lesions (i.e., S6_A_Pre and S8_A_Pre) demonstrated shared nonsilent mutations (*Figure 5B and 5C*). We could not detect a significant clone in p465 and p406, partly due to insufficient sampling for spatially heterogeneous tumors. HCC-related driver genes commonly mutated in preneoplastic lesions and their roles underlying hepatocarcinogenesis were summarized in *Table S8*. Taken together, we provided initial evidence that preneoplastic lesions in multicentric occurrence patients were genetically correlated before the malignant transformation.

Discussion

Whole-exome sequencing for a synchronous MF-HCC cohort allowed us to explore the molecular evolution underlying its progression. In the present study, we longitudinally characterized the distinct dynamic changes in SBS signatures among different MF-HCC subtypes, provided quantitative evidence for early intrahepatic metastatic seeding, and demonstrated that 'preneoplastic arising clones' carrying oncogenic drivers were predisposed to multicentric carcinogenesis. We validated our findings in an independent MF-HCC cohort, increasing the robustness of our conclusion to serve as the evidence for precision oncology in clinical practice.

The distinct genetic structure resulted in unpredictable cancer progression and conferred challenge in clinical management (11), known as pervasive tumor heterogeneity (*Figure 1*). In MF-HCC, the quantified significant intratumor heterogeneity (*Figure 1C*) (i.e., the average MATH-score of 42, range: 10.8-74.9) was similar to the previous estimation (39; range: 12.9-68.5) (34). Although multiple lesions in multicentric occurrence patients were expected to be genetically distinct, significantly greater interfocal heterogeneity in intrahepatic metastasis patients implicated an advanced divergence of metastasis from its primary site (*Figure 1A-B*). Quantitatively, the MATH-score varied significantly interfocal (p=0.002) rather than intratumor. A high interfocal molecular heterogeneity has been demonstrated in multifocal lung cancer (35) and primary prostate cancer (36,37), affecting how genomic-based management can be achieved.

The post-expansion mutations carry the signature of subclonal expansions that activate mutation process (11). Thus, dynamic changes in SBS signatures may enhance our understanding of the varied etiologies underlying the progression of MF-HCC. In addition to confirming the AA exposure (SBS22) (38,39), we demonstrated that, the contribution of AA exposure sharply declined between subclonal expansions in intrahepatic metastasis patients (*Figure 2*). Nevertheless, it inclined with tumor evolution in multicentric occurrence patients, indicating its role in the initial carcinogenesis of MF-HCC. These results suggested a prevention strategy for exogenous carcinogens (e.g., aflatoxin exposure and tobacco smoking) or drug selection for adjuvant chemotherapy. Additionally, banning the prescribed some AA-containing herbal remedies for HBV infection is required, especially in East Asia (40).

We demonstrated that metastases were seeded from spatial-specific primary sites in synchronous MF-HCCs (*Figure 3*). We also provided quantitative evidence for early intrahepatic metastasis, where metastasis were disseminated when the primary lesion was in 10^{-4}–0.01 cm^3 (*Figure 4*). Early metastatic seeding may occur in colorectal cancer (13), lung cancer (41), and pancreatic tumors (42,43). Intrahepatic metastases and tumor thrombi can occur early in HCC progression based on the distance between the intrahepatic metastatic site and primaries (2). Compared with 5-10 years from initiation to being clinically detected in other cancers (12), a much shorter time (<one year) from the most recent common ancestor to dissemination or from dissemination to surgery was inferred in intrahepatic metastasis patients, indicating their aggressive history. Due to a higher risk of systemic metastases, tumor staging for intrahepatic metastasis MF-HCC might be modified by taking into account early metastasis. Patients

with intraheaptic metastasis were staged as IIIa (p551 and p331) and IIb (p221, p424 and p460), advanced than patients with multicentric origins which three of them were staged as Ib. These findings showed that staging for intrahepatic metastasis MF-HCC might be modified by taking into account early intrahepatic metastasis. Lamarca *et al.* (44) recently showed that intrahepatic cholangiocarcinoma with early intrahepatic metastases has a worse outcome than other early stages and suggested modifying the current AJCC staging by considering early intrahepatic metastasis.

Although multicentric occurrence has been suggested to arise from different preneoplastic lesions of the cirrhotic liver (45), it may not be thoroughly independent during its evolutionary history (*Figure 5*). Common preneoplastic mutations enriched in the cancer-related pathways demonstrated their carcinogenesis potential before the malignant transformation. Preneoplastic arising clones were shown at least in three out of six patients, whereas the clones identified specifically in cancer tissues showed mutual independence. Thus, the multicentric occurrence was not thoroughly independent across all stages during carcinogenesis. Identifying such mutations in preneoplastic lesions can enable risk-stratifying these preneoplastic lesions that require intervention. In non-alcoholic fatty liver disease associated HCC patients, somatic mutations and epigenetic changes were identified in the background liver, and thus, rigorous surveillance for the emergence of HCC in such cases was required (46).

In summary, our study provided a comprehensive characterization for the molecular features including dynamic mutation signature, intrahepatic metastatic timing, and common preneoplastic clones in multicentric occurrence that delineated the varied tumor clonal evolutionary history underlying different subtypes of MF-HCC. This study

presents a novel insight into the characterization underlying the cancer progression and may accelerate personalized clinical management for MF-HCC.

Acknowledgments

None.

Conflict of interest statement

The authors have no conflicts of interest to declare.

Financial support statement

This study was supported by the National Natural Science Foundation of China, NSFC (No. 81470898 for D.L.).

Authors contributions

L.X., Y.S., L.W., and D.L. recruited patients and collected clinical information; Q.L. and D.Z. reviewed pathological reports; K.D. conducted computation; K.D. and X.T. performed data analysis; X.L., S.W., and H.W. participated in the discussion; K.D. and X.T. wrote the manuscript; K.D. and D.L. provided funding support, K.D. designed the project, and K.D. and D.L. conceived the project. All authors approved the manuscript.

Data avaliability

Supplementary materials including tables and figures. Whole-exome sequencing data for MF-HCC tumor and its paired adjacent non-cancer tissues, PBMC, and preneoplastic lesions have been submitted to the Sequence Read Archive (SRA) (https://submit.ncbi.nlm.nih.gov/) under the accession number SUB10411606. The source code is available upon request.

References

1. Xie D, Fan H, Ren Z, Fan J, Gao Q. Identifying clonal origin of multifocal hepatocellular carcinoma and its clinical implications. Clinical and Translational Gastroenterology. 2019;10:e00006.

2. Xue R, Li R, Guo H, Guo L, Su Z, Ni X, et al. Variable intra-tumor genomic heterogeneity of multiple lesions in patients with hepatocellular carcinoma. Gastroenterology. 2016;150:998–1008.

3. Furuta M, Ueno M, Fujimoto A, Hayami S, Yasukawa S, Kojima F, et al. Whole genome sequencing discriminates hepatocellular carcinoma with intrahepatic metastasis from multi-centric tumors. Journal of Hepatology. 2017;66:363–373.

4. Torrecilla S, Sia D, Harrington AN, Zhang Z, Cabellos L, Cornella H, et al. Trunk mutational events present minimal intra- and inter-tumoral heterogeneity in hepatocellular carcinoma. Journal of Hepatology. 2017;67:1222–1231.

5. Lu L-C, Hsu C-H, Hsu C, Cheng A-L. Tumor heterogeneity in hepatocellular carcinoma: facing the challenges. Liver Cancer. 2016;5:128–138.

6. Xu LX, He MH, Dai ZH, Yu J, Wang JG, Li XC, et al. Genomic and transcriptional heterogeneity of multifocal hepatocellular carcinoma. Annals of Oncology. 2019;30:990–997.

7. Ally A, Balasundaram M, Carlsen R, Chuah E, Clarke A, Dhalla N, et al. Comprehensive and integrative genomic characterization of hepatocellular carcinoma. Cell. 2017;169:1327–1341.e23.

8. Zhai W, Lim TK-H, Zhang T, Phang S-T, Tiang Z, Guan P, et al. The spatial organization of intra-tumour heterogeneity and evolutionary trajectories of metastases in hepatocellular carcinoma. Nature Communications. 2017;8:4565.

9. Ding X, He M, Chan AWH, Song QX, Sze SC, Chen H, et al. Genomic and epigenomic features of primary and recurrent hepatocellular carcinomas. Gastroenterology. 2019;157:1630–1645.e6.

10. Dong L, Peng L, Ma L, Liu D, Zhang S, Luo S, et al. Heterogeneous immunogenomic features and distinct escape mechanisms in multifocal hepatocellular carcinoma. Journal of Hepatology. 2020;72:896–908.

11. Dentro SC, Leshchiner I, Haase K, Tarabichi M, Wintersinger J, Deshwar AG, et al. Characterizing genetic intra-tumor heterogeneity across 2,658 human cancer genomes. Cell. 2021;184:2239–2254.e39.

12. Gerstung M, Jolly C, Leshchiner I, Dentro SC, Gonzalez S, Rosebrock D, et al. The evolutionary history of 2,658 cancers. Nature. 2020;578:122–128.

13. Hu Z, Ding J, Ma Z, Sun R, Seoane JA, Shaffer JS, et al. Quantitative evidence for early metastatic seeding in colorectal cancer. Nature genetics. 2019;51:1113–1122.

14. Hu Z, Curtis C. Looking backward in time to define the chronology of metastasis. Nature Communications. 2020;11:3213.

15. Zhang G, Tang X, Liang L, Zhang W, Li D, Li X, et al. DNA and RNA sequencing identified a novel oncogene VPS35 in liver hepatocellular carcinoma. Oncogene. 2020;39:3229–3244.

16. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nature Genetics. 2011;43:491–498.

17. 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. Nature Biotechnology. 2013;31:213–219.

18. Favero F, Joshi T, Marquard AM, Birkbak NJ, Krzystanek M, Li Q, et al. Sequenza: allele-specific copy number and mutation profiles from tumor sequencing data. Annals of Oncology. 2015;26:64–70.

19. 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 Biology. 2011;12:R41–R41.

20. Mroz EA, Rocco JW. MATH, a novel measure of intratumor genetic heterogeneity, is high in poor-outcome classes of head and neck squamous cell carcinoma. Oral Oncology. 2013;49:211–215.

21. Mroz EA, Tward AD, Tward AM, Hammon RJ, Ren Y, Rocco JW. Intra-tumor genetic heterogeneity and mortality in head and neck cancer: analysis of data from The Cancer Genome Atlas. PLoS Medicine. 2015;12:e1001786.

22. Roth A, Khattra J, Yap D, Wan A, Laks E, Biele J, et al. PyClone: statistical inference of clonal population structure in cancer. Nature Methods. 2014;11:396–398.

23. Malikic S, McPherson AW, Donmez N, Sahinalp CS. Clonality inference in multiple tumor samples using phylogeny. Bioinformatics. 2015;31:1349–1356.

24. Smith MA, Nielsen CB, Chan FC, McPherson A, Roth A, Farahani H, et al. E-scape: interactive visualization of single-cell phylogenetics and cancer evolution. Nature Methods. 2017;14:549–550.

25. Makohon-Moore AP, Zhang M, Reiter JG, Bozic I, Allen B, Kundu D, et al. Limited heterogeneity of known driver gene mutations among the metastases of individual patients with pancreatic cancer. Nature genetics. 2017;49:358–366.

26. Alexandrov LB, Kim J, Haradhvala NJ, Huang MN, Ng AWT, Wu Y, et al. The repertoire of mutational signatures in human cancer. Nature. 2020;578:94–101.

27. Tate JG, Bamford S, Jubb HC, Sondka Z, Beare DM, Bindal N, et al. COSMIC: the Catalogue Of Somatic Mutations In Cancer. Nucleic acids research. 2019;47:D941–D947.

28. Blokzijl F, Janssen R, Boxtel R van, Cuppen E. MutationalPatterns: comprehensive genome-wide analysis of mutational processes. Genome Medicine. 2018;10:33.

29. Brand KG, Buoen LC. Polymer Tumorigenesis : multiple preneoplastic clones in priority order with clonal inhibition. Experimental Biology and Medicine. 4:1154–1158.

30. Bailey MH, Tokheim C, Porta-Pardo E, Sengupta S, Bertrand D, Weerasinghe A, et al. Comprehensive characterization of cancer driver genes and mutations. Cell. 2018;173:371–385.e18.

31. Caruso S, Calatayud A-L, Pilet J, Bella TL, Rekik S, Imbeaud S, et al. Analysis of liver cancer cell lines identifies agents with likely efficacy against hepatocellular carcinoma and markers of response. Gastroenterology. 2019;157:760–776.

32. Martínez-Jiménez F, Muiños F, Sentís I, Deu-Pons J, Reyes-Salazar I, Arnedo-Pac C, et al. A compendium of mutational cancer driver genes. Nature Reviews Cancer. 2020;20:555–572.

33. Duan M, Hao J, Cui S, Worthley DL, Zhang S, Wang Z, et al. Diverse modes of clonal evolution in HBV-related hepatocellular carcinoma revealed by single-cell genome sequencing. Cell Research. 2018;28:359–373.

34. Gao Q, Wang Z-C, Duan M, Lin Y-H, Zhou X-Y, Worthley DL, et al. Cell culture system for analysis of genetic heterogeneity within hepatocellular carcinomas and response to pharmacologic agents. Gastroenterology. 2017;152:232–242.e4.

35. Ma P, Fu Y, Cai M-C, Yan Y, Jing Y, Zhang S, et al. Simultaneous evolutionary expansion and constraint of genomic heterogeneity in multifocal lung cancer. Nature Communications. 2017;8:823.

36. Løvf M, Zhao S, Axcrona U, Johannessen B, Bakken AC, Carm KT, et al. Multifocal primary prostate cancer exhibits high degree of genomic heterogeneity. European Urology. 2019;75:498–505.

37. Fontugne J, Cai PY, Alnajar H, Bhinder B, Park K, Ye H, et al. Collision tumors revealed by prospectively assessing subtype-defining molecular alterations in 904 individual prostate cancer foci. JCI Insight. 2022;7:e155309.

38. Lu Z, Luo Q, Zhao L, Shi Y, Wang N, Wang L, et al. The mutational features of aristolochic acid–induced mouse and human liver cancers. Hepatology. 2020;71:929–942.

39. Llovet JM, Kelley RK, Villanueva A, Singal AG, Pikarsky E, Roayaie S, et al. Hepatocellular carcinoma. Nature Reviews Disease Primers. 2021;7:6.

40. Alvin NG, 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. Science Translational Medicine. 2017;9.

41. Tang W-F, Wu M, Bao H, Xu Y, Lin J-S, Liang Y, et al. Timing and origins of local and distant metastases in lung cancer. Journal of Thoracic Oncology. 2021;16:1136–1148.

42. Tang X, Shao Y, Yi X, Newey PJ, Li D, Ding K. Metastatic timing and genetic heterogeneity in the evolution of a pancreatic neuroendocrine tumor. American Journal of Gastroenterology. 2020;116:844–847.

43. Tang X, Shao Y, Li D, Ding K. Quantitative evidence for early metastasis in a pancreatic acinar cell carcinoma patient. Pancreas. 2022;51:e27–e29.

44. Lamarca A, Santos-Laso A, Utpatel K, Casta AL, Stock S, Forner A, et al. Liver metastases of intrahepatic cholangiocarcinoma: implications for an updated staging system. Hepatology (Baltimore, Md.). 2021;73:2311–2325.

45. Morimoto O, Nagano H, Sakon M, Fujiwara Y, Yamada T, Nakagawa H, et al. Diagnosis of intrahepatic metastasis and multicentric carcinogenesis by microsatellite

loss of heterozygosity in patients with multiple and recurrent hepatocellular carcinomas. Journal of Hepatology. 2003;39:215–221.

46. Hagiwara S, Nishida N, Ueshima K, Minami Y, Komeda Y, Aoki T, et al. Accumulation of genetic and epigenetic alterations in the background liver and emergence of hepatocellular carcinoma in patients with non-alcoholic fatty liver disease. Cells. 2021;10:3257.

Figure legends

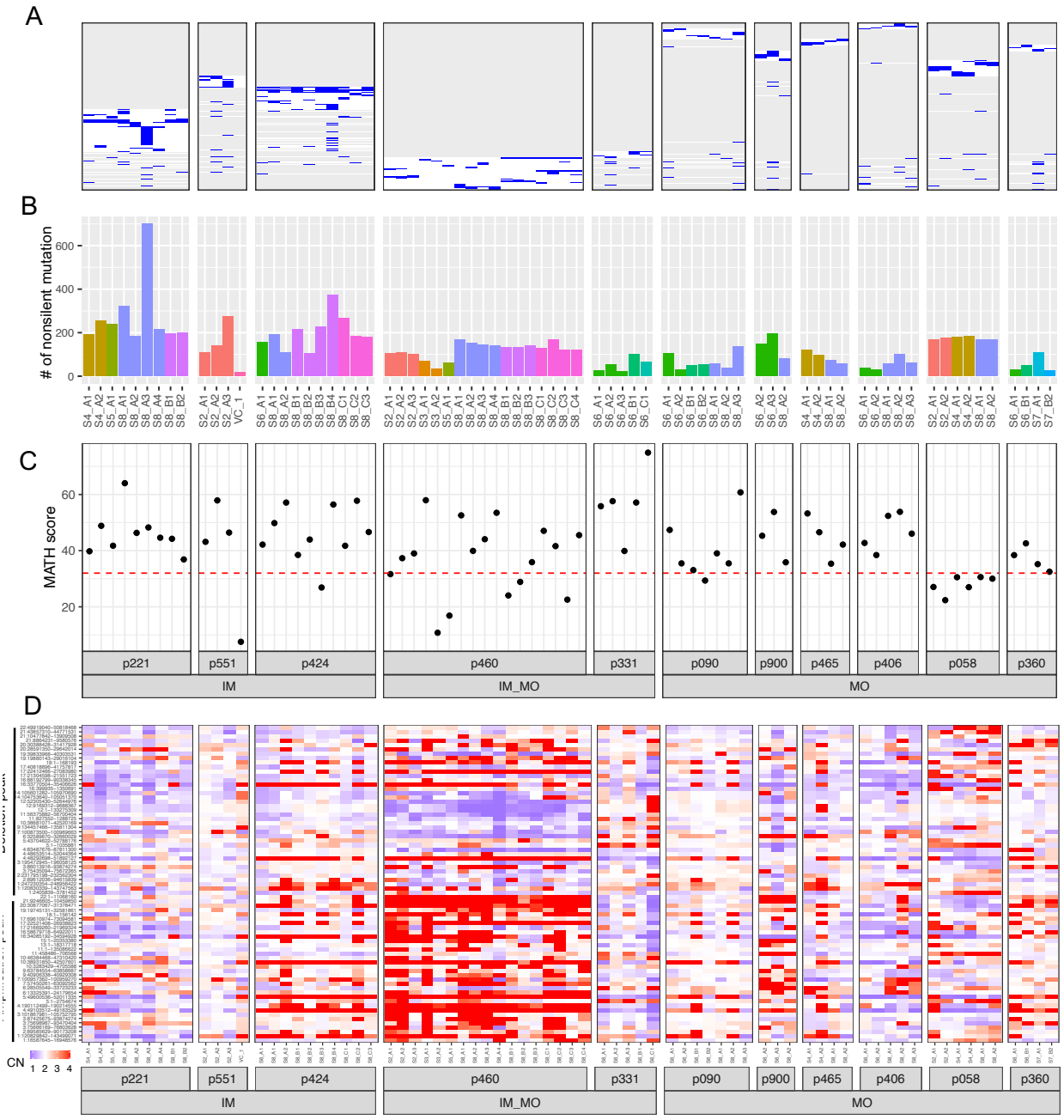


Figure 1. Tumor genomic heterogeneity in MF-HCC patients. (A). The nonsilent mutational landscape in sequenced tumor samples. The y-axis showed nonsilent mutations identified in all samples. (B). The nonsilent mutational burden in each

sequenced sample. *VC*, the portal vein tumor thrombus. Different colors represented anatomical locations (e.g., S4_A). (C). The mutant-allele tumor heterogeneity (MATH) score. (D). Distribution of somatic copy number amplification (red) and deletion (blue). IM, intrahepatic metastasis; MO, multicentric occurrence; and IM_MO, the mixed IM and MO.

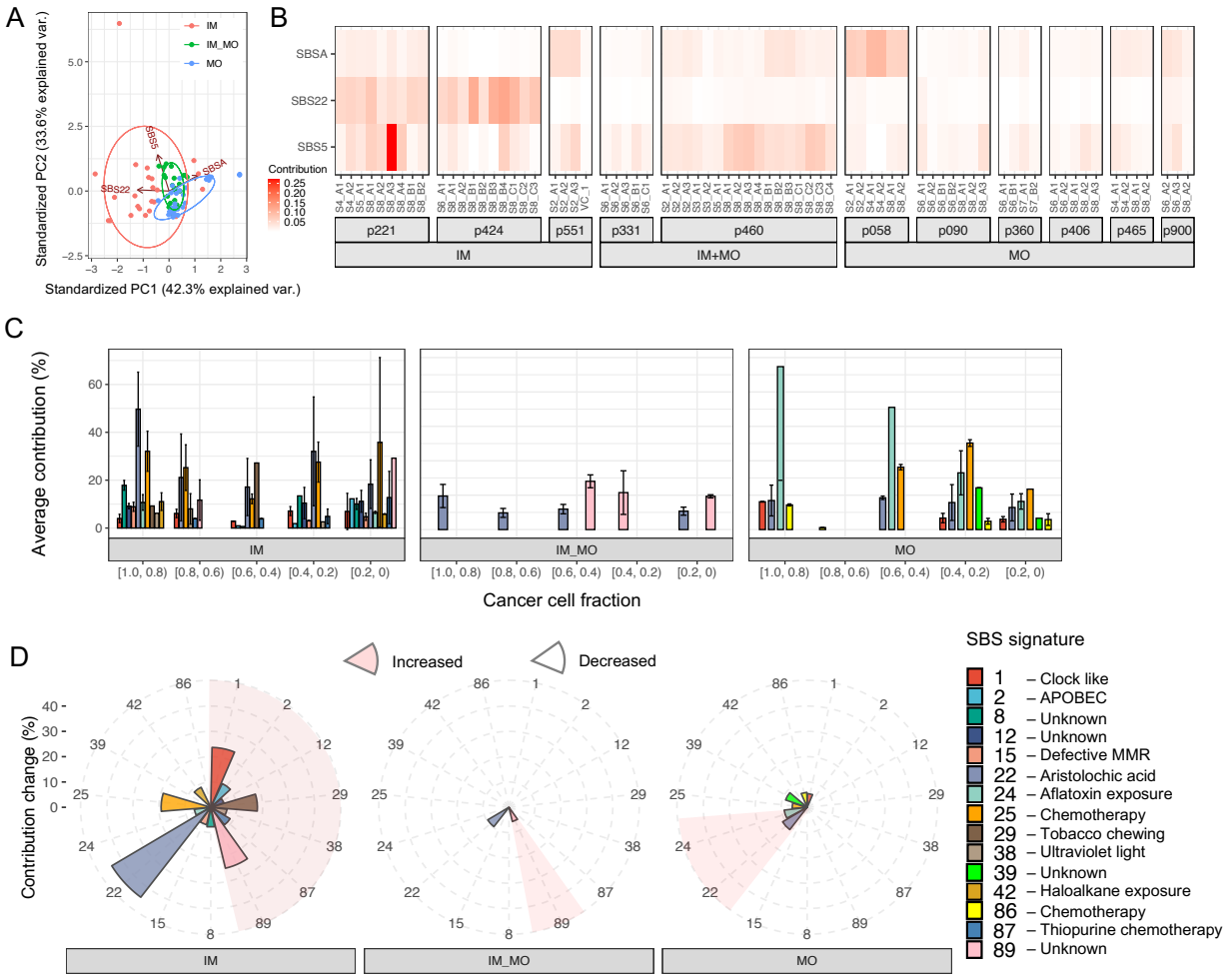


Figure 2. Mutation signatures exhibited dynamic changes along with MF-HCC evolution. (A). Principal component analysis for *de novo* mutational signatures. (B). Negative matrix factorization estimated three *de novo* mutational signatures. (C). Dynamic changes of the refitted SBS signatures with cancer cell fraction (CCF). (D) Changes of SBS signatures (contribution>5%) with a decreased CCF.

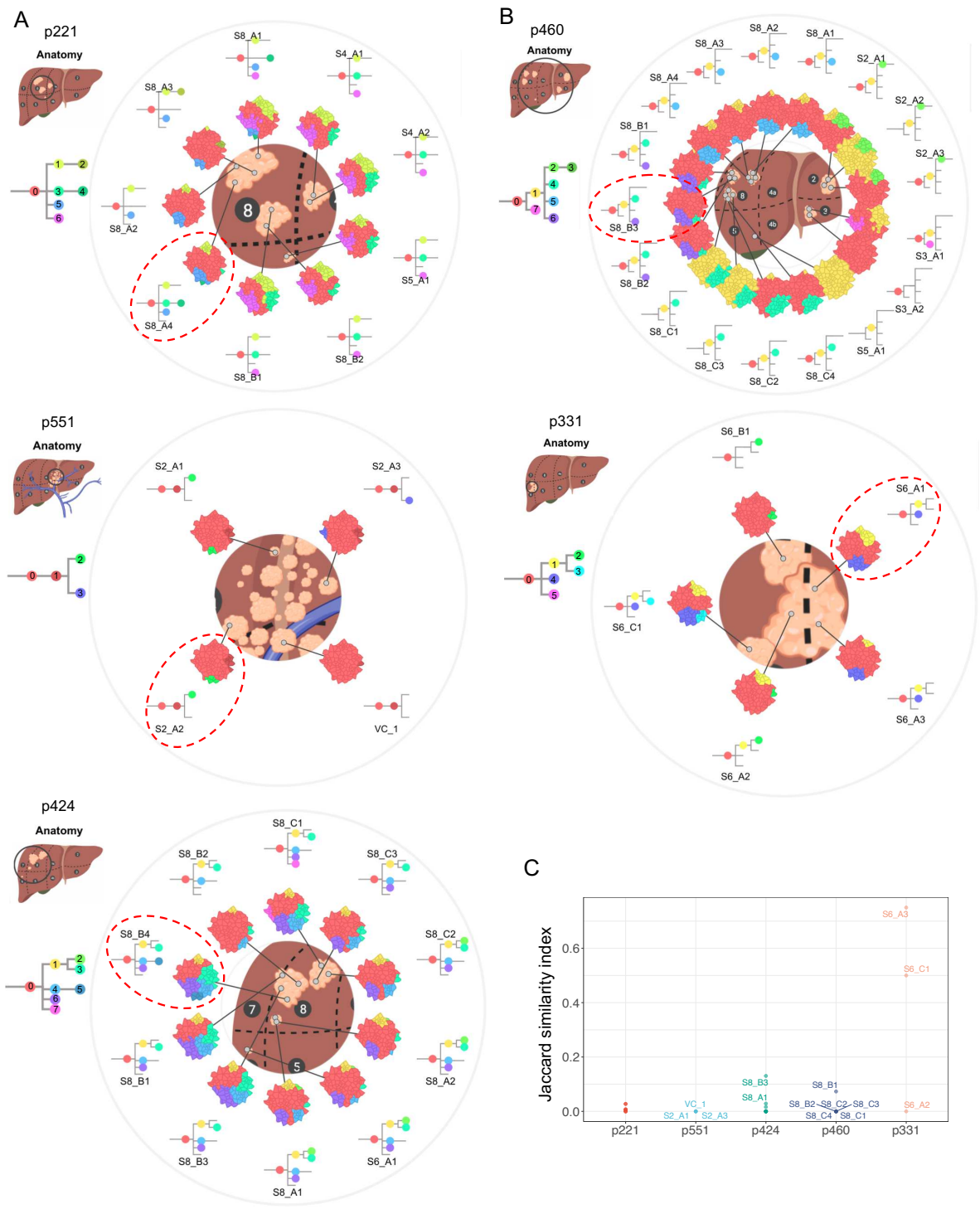


Figure 3. Intrahepatic metastasis occurred in spatial-specific primary site. Clonal inference from multi-regional samples clearly indicated the metastasis occurred in spatial-specific primary site in MF-HCCs patients with intrahepatic metastasis patients

(A) and the mixed intrahepatic metastasis and multicentric occurrence (B). The primary lesion was marked with dashed red circle. Distinct clones were present in different colors with numbers for each patient. The clonal composition was scaled to the fraction of each clone in each sample. ‘0’ represented ancestor clone. Samples were zoomed in the inner circle around the anatomical diagram, where the outer circle was constituent clones from the overall clonal tree. (C). The Jaccard similarity index (JSI) for each tumor sample.

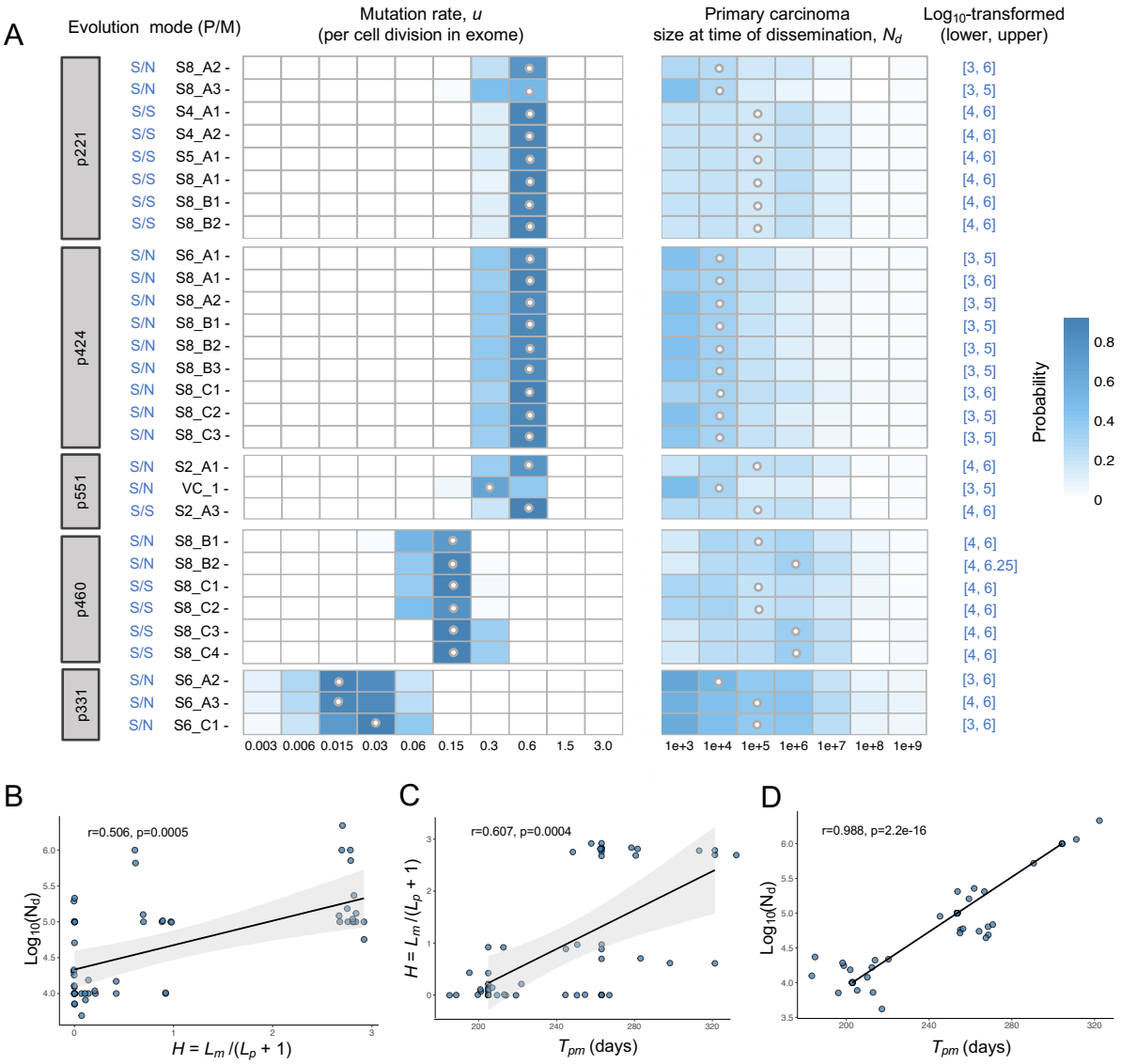


Figure 4. Quantitative evidence for early intrahepatic metastasizing timing. (A). The inferred metastatic timing. *N*, neutral evolution; and *S*, subclonal selection. P, primary tumor; M, metastasis tumor; and N_d, the primary tumor size (cell number) at metastatic seeding. (B). The correlation between *H* and N_d. (C). The correlation between the time from the primary tumor initiation to dissemination (T_{pm}, days) and N_d. (D). The correlation the time from the primary tumor initiation to dissemination (T_{pm}, days) and *H*.

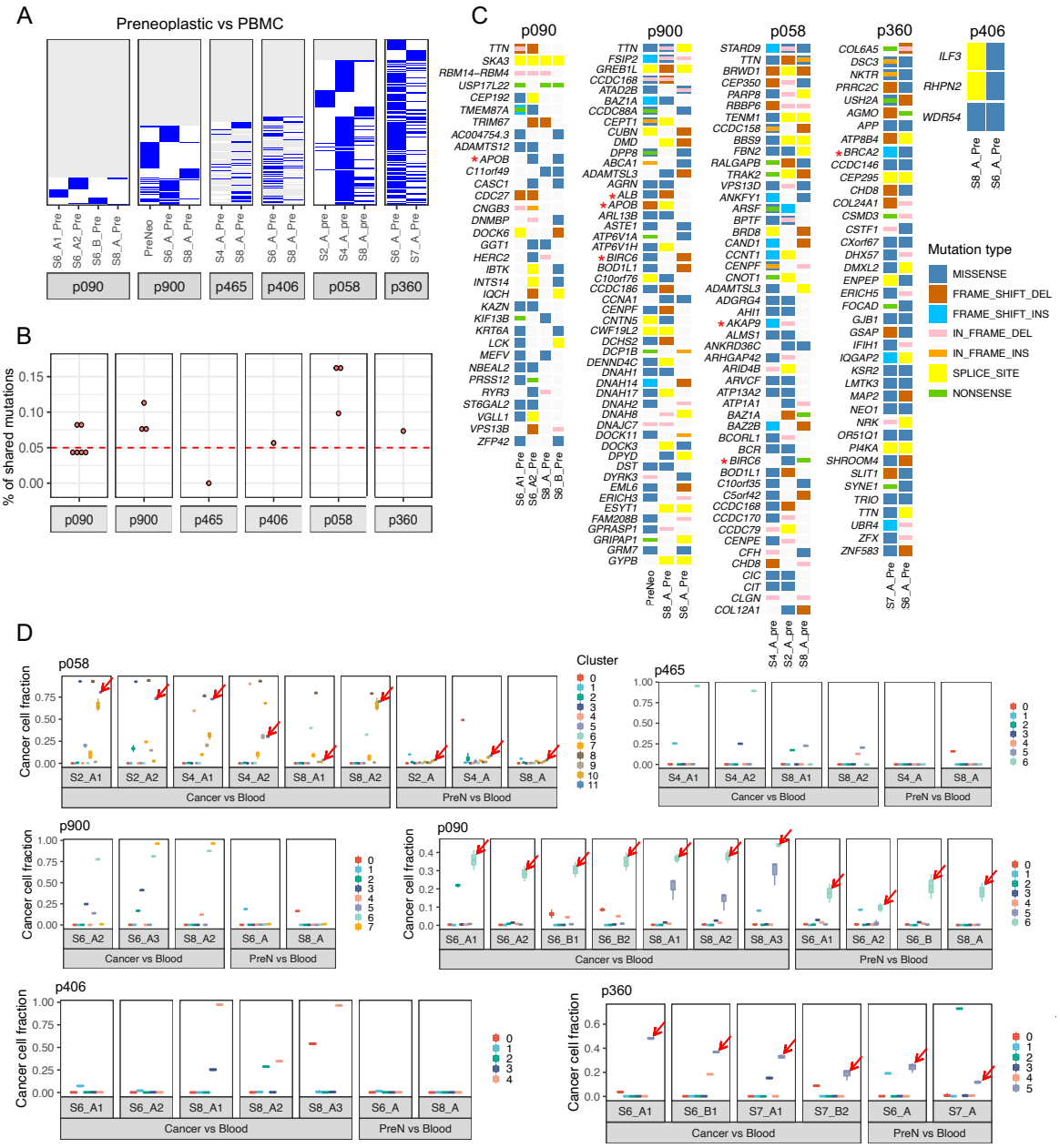


Figure 5. Genetic footprints in the preneoplastic lesions in multicentric occurrence patients. (A) A nonsilent mutations landscape in the preneoplastic lesions. (B). The rate of shared nonsilent mutation in any two preneoplastic lesions. (C). Common nonsilent mutations were identified in each patient. Mutations in HCC drivers were indicated as a red star. (D). The mean frequency of cluster was inferred for each sample in multicentric

occurrence patients, which common preneoplastic clone was indicated with red arrows. PreN, preneoplastic tissue.

Supplementary Files

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

- suppMFHCC20230109.pdf