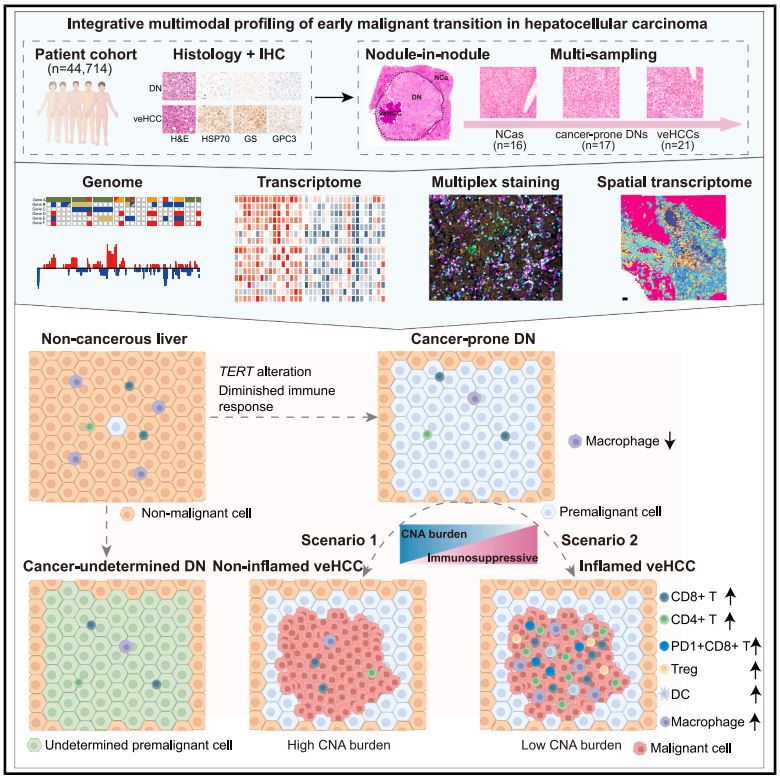
Cancer Cell

Article

Molecular insights into early malignant transition of hepatocellular carcinoma

Graphical abstract



Highlights

- *TERT* alterations may predispose dysplastic nodules (DNs) to malignant transition
- Copy number alterations account for malignant transition in HCC
- Arising from immune inactive DNs, veHCCs display immune surveillance and evasion
- Two potential major evolutionary scenarios are involved in HCC initiation

Authors

Zhengtao Zhang, Hong Li, Lingli Chen, ..., Hui Dong, Yuan Ji, Lijian Hui

Correspondence

lihong01@sibs.ac.cn (H.L.), ji.yuan@zs-hospital.sh.cn (Y.J.), ljhui@sibcb.ac.cn (L.H.)

In brief

Zhang et al. present a comprehensive genomic and immunological analysis of the premalignant-to-malignant transition in hepatocellular carcinoma, showing *TERT* alterations predispose to malignancy, while copy number alterations (CNAs) account for the transition. They propose two major evolutionary scenarios, CNA-dominant progression and inflamed progression with immune evasion.

Zhang et al., 2026, Cancer Cell *44*, 1–17
May 11, 2026
https://doi.org/10.1016/j.ccell.2026.03.005

CellPress

Cancer Cell

CellPress
OPEN ACCESS

Article

Molecular insights into early malignant transition of hepatocellular carcinoma

Zhengtao Zhang,[1,2,15] Hong Li,[3,15,*] Lingli Chen,[4,15] Tao Lu,[5] Xinyi Shentu,[6] Yuanhua Liu,[7] Zhixuan Tang,[3] Shuyi Ji,[8] Zhixin Qiu,[10] Yining Zou,[4] Hong Wen,[4] Jing Han,[4] Zhengzeng Jiang,[4] Jie Fan,[9] Xinbo Gao,[1,2] Dianfan Li,[1] Agavni Mesropian,[11,12] Roser Pinyol,[11,12] Josep M. Llovet,[11,12,13,14] Hui Dong,[5] Yuan Ji,[4,*] and Lijian Hui[1,6,16,*]

[1]The Research Laboratory of Liver Molecular Pathology, Key Laboratory of Multi-cell Systems, Shanghai Institute of Biochemistry and Cell Biology, Center for Excellence in Molecular Cell Science, Chinese Academy of Sciences, Shanghai 200031, P.R. China
[2]University of Chinese Academy of Sciences, Beijing 101408, P.R. China
[3]Shanghai Institute of Nutrition and Health, University of Chinese Academy of Sciences, Chinese Academy of Sciences, Shanghai 200031, P.R. China
[4]Department of Pathology, Zhongshan Hospital, Fudan University, Shanghai 200032, P.R. China
[5]Department of Pathology, Eastern Hepatobiliary Surgery Hospital, Naval Medical University, Shanghai 200438, P.R. China
[6]School of Life Science and Technology, ShanghaiTech University, Shanghai 201210, P.R. China
[7]State Key Laboratory of Cardiovascular Diseases and Medical Innovation Center, Shanghai East Hospital, School of Medicine, Tongji University, Shanghai 200120, China
[8]Institute for Regenerative Medicine, Shanghai East Hospital, Tongji University School of Medicine, Shanghai 200123, P.R. China
[9]Department of Pathology, Huashan Hospital, Fudan University, Shanghai 200040, P.R. China
[10]Department of Anesthesiology, Shanghai Key Laboratory of Perioperative Stress and Protection, Zhongshan Hospital, Institute for Translational Brain Research, Center for Clinical Neuro-AI, State Key Laboratory of Brain Function and Disorders, MOE Frontiers Center for Brain Science, Fudan University, Shanghai 200032, P.R. China
[11]Liver Cancer Translational Research Group, Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), Liver Unit, Hospital Clínic, Universitat de Barcelona, Barcelona, Catalonia, Spain
[12]Facultat de Medicina i Ciències de la Salut, Universitat de Barcelona (UB), Barcelona, Spain
[13]Mount Sinai Liver Cancer Program, Division of Liver Diseases, Tisch Cancer Institute, Icahn School of Medicine at Mount Sinai, New York, NY, USA
[14]Institució Catalana de Recerca i Estudis Avançats, Barcelona, Catalonia, Spain
[15]These authors contributed equally
[16]Lead contact
*Correspondence: lihong01@sibs.ac.cn (H.L.), ji.yuan@zs-hospital.sh.cn (Y.J.), ljhui@sibcb.ac.cn (L.H.)
https://doi.org/10.1016/j.ccell.2026.03.005

SUMMARY

Understanding the molecular drivers of the premalignant-to-malignant transition is essential for early cancer detection and intervention, yet this process remains poorly characterized due to limited access to evolutionarily related lesions. Here, we comprehensively profiled the genomic and immunological landscapes of 21 very early hepatocellular carcinomas (veHCCs) arising within 17 cancer-prone dysplastic nodules (DNs). Notably, 82% of cancer-prone DNs harbored *TERT* alterations, suggesting a predisposing rather than causative role of *TERT* alteration in malignant transition. Notably, the accumulation of copy number alterations (CNAs), rather than single-nucleotide variants, is strongly associated with malignant transition. Intriguingly, different from the paradigm that HCC arises in chronic inflammation, cancer-prone DNs displayed immune inactivity. Strikingly, 43% of veHCCs showed an inflamed yet immune-evasive phenotype. These findings propose two evolutionary scenarios of early hepatocarcinogenesis, CNA-dominant progression and inflamed progression with early immune evasion, underscoring the potential of immunotherapy for early intervention.

INTRODUCTION

The transition from premalignant lesions to early cancers represents a critical stage in carcinogenesis. While observation or surveillance is typically recommended for premalignant nodules, diagnosis of malignancy requires immediate medical intervention.[1–3] Understanding this transition is essential for early prevention and treatment.[4] Previous studies using evolutionarily unrelated premalignant samples and cancers have provided statistical inferences about genetic and microenvironmental changes during malignant transition.[5–11] However, only a subset of premalignant lesions transform into cancers, and most eventually regress.[12] Thus, many premalignant lesions previously characterized may not be informative. Moreover, because the

Please cite this article in press as: Zhang et al., Molecular insights into early malignant transition of hepatocellular carcinoma, Cancer Cell (2026), https://doi.org/10.1016/j.ccell.2026.03.005

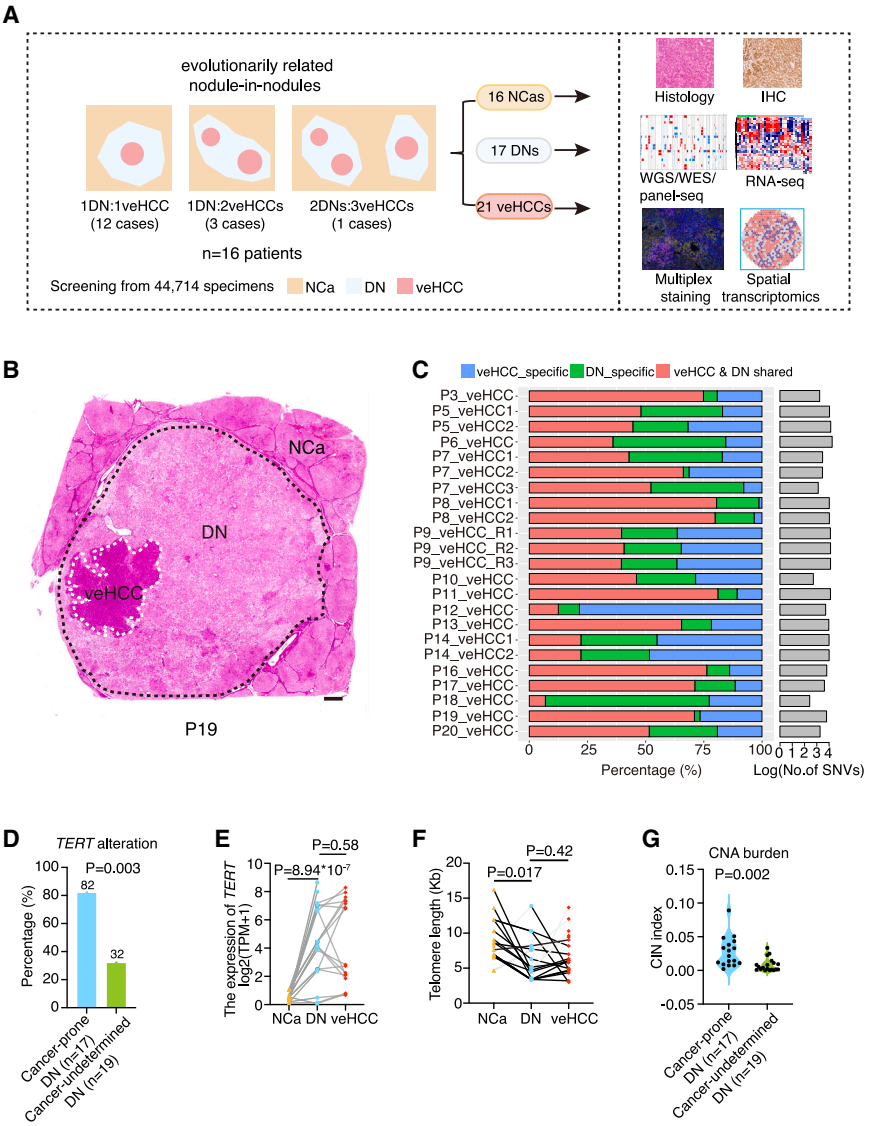


Figure 1. Sample collection and characterization

(A) Schematic illustration of the experimental design for characterizing DNs and veHCCs with a nodular-in-nodule pattern. The cartoon illustrating spatial transcriptomics was created with BioRender. NCa, non-cancerous tissue; DN, dysplastic nodule.

(B) Whole-slide H&E images of P19. White dashed lines represent the veHCC region, black dashed lines represent the DN region. Scale bars, 1 mm.

(C) The percentage of overlapped SNVs between DNs and veHCCs in each case (left) and the total number of SNVs detected in each case (right). Three distinct regions (R1, R2, and R3) of P9_veHCC were subjected to sequencing and were counted as one veHCC.

(D) The percentage of cancer-prone and cancer-undetermined DNs that acquired *TERT* alterations.

(E) Expression of *TERT* in NCas, DNs, and veHCCs based on RNA-seq data. The average *TERT* expression in P8_DN_R1 and P8_DN_R2 was calculated and used for analysis. Those DN and veHCC samples that acquired *TERT* alterations were selected for analysis.

(F) Changes in telomere length during the carcinogenesis from NCa to veHCC. Only DNs and veHCCs harboring *TERT* alterations were analyzed. Black lines denote instances where telomere length was shortened or remained stable. Gray lines indicate cases where telomere length was notably extended in DNs or veHCCs.

(G) Chromosome instability (CIN) index of cancer-prone and cancer-undetermined DNs. The violin plot shows all data points. Statistical significance is determined by Fisher's exact test (D), by a linear mixed model (E and F), and by two-tailed unpaired Student's t-tests (G). See also Figure S1 and S2; Tables S1 and S2.

tumor microenvironment changes dynamically, unrelated samples complicate the prediction of microenvironmental features during malignant transition.

Only evolutionarily related premalignant and malignant samples provide deterministic evidence. Premalignant lesions, specifically dysplastic nodules (DNs), containing early cancers that develop in a nodule-in-nodule pattern, can provide a snapshot of malignant transition. These nodule-in-nodule lesions capture the molecular events that predispose to malignancy and are critical for demarcating genetic and microenvironmental alterations driving this process. However, such lesions are extremely rare in clinical samples, as the coexistence of DNs and cancer is brief due to the high proliferative rate of malignant cells.

Hepatocellular carcinoma (HCC), a leading cause of cancer-related mortality,[13] exemplifies this complexity. HCC typically arises from DNs, which are premalignant lesions with an estimated 30% risk of progression into HCCs.[12,14,15] Previous studies have analyzed HCC initiation using evolutionarily unrelated DNs and HCCs.[5–7] In this study, we leverage nodule-in-nodule lesions[16] as a unique opportunity to examine the development of very early HCCs (veHCCs) within DNs.[17,18]

RESULTS

Evolutionary related DNs and veHCCs from nodule-in-nodule samples

We screened 44,714 FFPE human liver specimens from three liver cancer centers within a 9-year interval for nodule-in-nodule samples containing DNs and veHCCs (Figure 1A). 17 nodule-in-nodule samples from 16 HBV-infected patients were collected (Figures 1A and S1A–S1Q; Table S1). High-grade DNs[14,19] and veHCCs[20] were blindly and independently validated by three pathologists through the histological evaluation of H&E-stained tissues (Figure 1B) and immunohistochemical (IHC) staining of GPC3, HSP70, and GS (Figures S1R and S1S).

These samples were subjected to WGS (at approximately 35×) and RNA-seq (Table S2A). Seventeen DNs (referred to as cancer-prone DNs) shared 7–85% of single-nucleotide variants

Please cite this article in press as: Zhang et al., Molecular insights into early malignant transition of hepatocellular carcinoma, Cancer Cell (2026), https://doi.org/10.1016/j.ccell.2026.03.005

(SNVs) with their evolutionarily related veHCCs (21 veHCCs in total; Figures 1C and S2A). Notably, paired DNs and veHCCs exhibited similar mutational signature patterns (Figure S2B). Moreover, follow-up magnetic resonance imaging (MRI) data from patient 3 (P3) validated the evolution from DN to veHCC in this case (Figure S2C). These data collectively demonstrated their evolutionary relationships.

To confirm that SNVs could be captured from FFPE samples, we compared the SNV profiles of two HCC samples with those of freshly frozen and FFPE tissues. In line with a previous reported method,[21] 94.4% of SNVs detected in freshly frozen samples were confirmed in FFPE tissues, showing that FFPE tissues largely retained the original SNVs (Figure S2D).

Differences between cancer-prone and cancer-undetermined DNs

A total of 19 DNs, with no carcinoma-*in-situ* lesions, and their paired non-cancerous tissues (NCas) were additionally obtained from 19 HBV-infected patients. These DNs, which had an undetermined cancer potential (referred to as cancer-undetermined DNs), were used for comparative analyses to identify pivotal molecular events enriched in cancer-prone DNs (Table S1A and S1B). Key mutations between cancer-prone and cancer-undetermined DNs were analyzed from a list of 37 liver cancer functional genes (CFGs) cataloged from 8 published HCC cohorts.[7,22–28] Notably, *TERT* alterations were detected in 82% of cancer-prone DNs (Figure 1D), including HBV integration, promoter mutation, and copy number gain of the *TERT* gene. In contrast, *TERT* alterations were observed in 6 of the 19 cancer-undetermined DNs (32%) (Figure 1D), which was in line with previous studies.[5,6] These data indicate a high prevalence of *TERT* alterations in cancer-prone DNs ($p = 0.003$).

In line with previous findings,[7] *TERT* alterations upregulated the expression of *TERT* in cancer-prone DNs and matched veHCCs (Figure 1E). However, significantly shortened telomeres were observed in 79% of cancer-prone DNs carrying *TERT* alterations (Figure 1F). This trend remained consistent in cancer-prone DNs with increased *TERT* expression, with 73% (8/11 DNs) showing shortened telomeres (Figures S2E and S2F). Notably, telomere length was maintained or even shortened in 65% of veHCCs (Figure 1F). These data indicate that telomere attrition persists during malignant transition, even in the presence of *TERT* alterations.

Upon comparing copy number alterations (CNAs) between cancer-prone and cancer-undetermined DNs, significantly increased CNAs were found in cancer-prone DNs (Figure 1G, $p = 0.002$). Notably, Chr7p (*EGFR*), Chr7q (*MET*), and Chr8q (*MYC*) gains were enriched in 24%, 24%, and 12% of cancer-prone DNs, respectively (Figures S2G–S2I).

Copy number alteration is strongly associated with malignant transition

Consistent with previous findings,[5,7] an increased burden of both protein-altering SNVs and CNAs was observed in veHCCs when evolutionarily unrelated DNs and HCCs were compared (Figures S3A and S3B). However, no significant difference was found in the number of total SNVs (Figure S3C, $p = 0.50$) or protein-altering SNVs between paired cancer-prone DNs and veHCCs (Figure 2A, $p = 0.25$; Table S3), suggesting that the DN-to-veHCC transition was not associated with the SNV burden. Next, we analyzed SNV-induced CFG mutations relevant to malignant transition. Notably, the number of CFG mutations showed no significant increase in veHCCs compared to their paired DNs (Figure S3D, $p = 0.08$). Additionally, acquired CFG mutations were detected in veHCCs, such as *ARID1A* mutations in P5_veHCC2 and P7_veHCC2, whereas 71% of veHCCs showed no acquisition of additional SNVs in these CFGs (Figure S3E). To determine these CFG mutations in later-stage HCCs, we compared veHCCs with later-stage HCCs from the Chinese Liver Cancer Atlas (CLCA, $n = 494$) and The Cancer Genome Atlas dataset (TCGA, $n = 363$). *TP53* and *CTNNB1* exhibited lower mutation frequencies in veHCCs than in later-stage HCCs (*TP53*, 19% vs. 52% in CLCA and 30% in TCGA; *CTNNB1*, 14% vs. 21% in CLCA and 26% in TCGA; Table S4A). Eight other CFGs showed comparable mutation frequencies, including *TERT*, *ARID1A,* and *PTEN*.

In contrast to SNVs, veHCCs exhibited a significantly higher CNA burden (>1 Mb) compared to paired DNs (Figures 2B, S3F, and S3G). Markedly, 89% of the arm-level CNAs exhibited an increased frequency in veHCCs, such as losses of Chr10q and Chr17p and gains of Chr1q and Chr8q (Figure 2C). Importantly, 10 CFG-harboring CNAs were unique to veHCCs, including Chr9q (*TSC1*), Chr13q (*RB1*) losses, and Chr12p (*KRAS*) gain (Figure 2D). Arm-level CNAs enriched in veHCCs were also common in later-stage HCCs, with 6 out of 11 arm-level CNAs occurring at comparable frequencies in both groups (Table S4B), including gains of Chr7p/Chr7q, Chr5p, Chr12p, and losses of Chr10q, Chr17p.

Notably, a veHCC-specific extended loss in Chr1p (15–28 Mb) was detected in 19% of veHCCs (4/21) but was absent in DNs (Figure 2D). In matched DNs, P6_DN and P9_DN showed shorter adjacent Chr1p losses (0–15 Mb) (Figures S3H and S3I), suggesting a progressive loss of Chr1p during malignant transition. Two tumor suppressor genes (TSGs), *ARID1A and NR0B2*, were identified in this veHCC-specific loss region, both showing reduced expression in veHCCs (Figure S3J).

We further expanded the analysis to 803 pan-cancer oncogenes[29] and 1,218 pan-cancer TSGs.[30] Consistently, veHCCs harbored more oncogene gains and TSG losses (Figure S3K) compared to paired DNs. Furthermore, veHCCs exhibited enriched expression of chromosomal instability related genes[31] (Figure S3L) and elevated DNA repair and chromosomal breakage signatures (Figures S3M–S3P). Notably, *TP53* alterations were significantly enriched in veHCCs with high chromosomal instability (CIN, see STAR Methods) (0.072 ± 0.035) compared with those with low CIN (0.002 ± 0.016) (Figure S3Q, $p = 0.02$; Table S5). Together, these findings further highlight the importance of CNAs during malignant transition.

Key genetic alterations from DNs to malignant transition

Next, integrative analysis of SNVs, CNAs, and HBV insertions was performed in 37 CFGs during malignant transition (Table S6). In total, 33 CFGs were altered in cancer-prone DNs and matched veHCCs (Figure 2E; Table S6C). Among them, 22 were altered in both DNs and veHCCs, including *TERT* (82% vs. 86%), *TP53* (29% vs. 57%), *CTNNB1* (12% vs. 24%), and *MYC* (12% vs. 24%) (Figure 2F).

Please cite this article in press as: Zhang et al., Molecular insights into early malignant transition of hepatocellular carcinoma, Cancer Cell (2026), https://doi.org/10.1016/j.ccell.2026.03.005

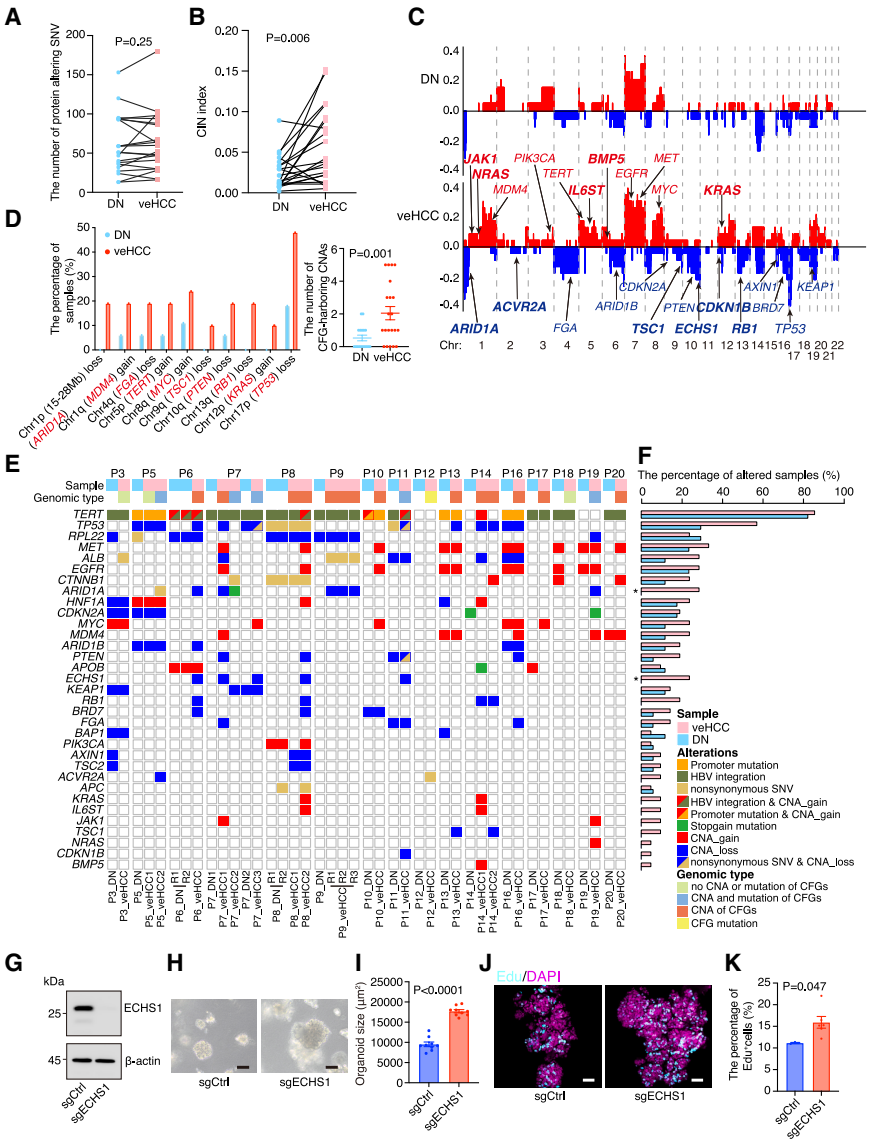


Figure 2. Key genomic alterations for the malignant transition of DNs

(A) The number of protein-altering SNVs in the DNs and veHCCs of each case.

(B) The CIN index of evolutionarily related DNs and their paired veHCCs.

(C) The frequency of CNAs along the genome. The upper panel indicates the frequency of CNAs in DNs, while the lower panel indicates the frequency of CNAs in veHCCs. CFGs are labeled, with those highlighted in bold representing CFGs that are either exclusively found in veHCCs or occur with greater frequency compared to those in DNs.

(D) The frequencies of CFG-harboring CNAs (left) and the number of CFG-harboring CNAs in DNs and veHCCs (right). Data are presented as mean ± s.e.m. CFGs located within the altered genomic regions are indicated in red.

(E) Heatmap shows the altered landscape of CFGs in DNs and veHCCs.

(F) Frequency of CFGs altered in DNs and veHCCs. *$p < 0.05$.

(G) Immunoblot analysis of ECHS1 in organoids with or without *ECHS1* knock out. Ctrl, control.

(H) Representative brightfield images of organoids with or without *ECHS1* knock out, respectively. Scale bars, 100 μm.

(I) The sizes of organoids with or without *ECHS1* knock out after 10 days of culture. Each dot represents the average size of 200–500 organoids ($n = 9$ for each).

(J) Representative EdU staining images of organoids with or without *ECHS1* knock out, respectively. Scale bars, 50 μm.

(K) The percentage of EdU-positive cells in organoids with or without *ECHS1* knock out. Each dot represents the average size of 100–300 organoids. Statistical significance is determined by a linear mixed model (A, B, D), by Fisher's exact test (F), and by two-tailed unpaired Student's t-tests (I, K). Independent experiments were biologically repeated in duplicate for (G), in triplicate for (H, I), and one representative result is shown. $n = 3$ for sgCtrl organoids and $n = 5$ for sgECHS1 organoids for (J, K). Data are presented as mean ± s.e.m (I, K). See also Figure S3 and S4; Tables S3, S4, S5, and S6.

Upon analyzing SNVs and CNAs in malignant transition-related CFGs, we found that 61% of veHCCs acquired additional CNAs in CFGs without additional mutations, while 22% of veHCCs accumulated both additional CNAs and mutations in CFGs. Only one veHCC acquired a single additional CFG mutation, but no CNAs in the CFGs (*ACVR2A*^{H196Q} in P12_veHCC). Consistently, the number of CNA-induced CFG alterations was significantly higher than that of SNV-induced CFG alterations in veHCCs (Figure S3R). These findings further suggest that, while CFG mutations contribute to malignant transition, the acquisition of CNAs in CFGs is a major event during malignant transition (83% of veHCCs in total).

Alterations in 11 CFGs were detected exclusively in veHCCs (Figure 2F). Among these, *ARID1A* alterations occurred in 6 veHCCs, including copy number loss ($n = 4$) and loss-of-function mutations ($n = 2$) (Figure 2F). Although *ECHS1* loss of function has been reported to be associated with HBV-related HCCs,[23] its role in liver carcinogenesis remains unclear. Notably, copy number loss of *ECHS1* was frequent in veHCCs ($n = 5$) (Figures 2E and 2F).

Given the limited sample size, we decided to validate the functions of these genes by an organoid-based *in vitro* model of liver cancer initiation using human induced hepatocytes (hiHeps).[32] We first assessed the roles of the frequently altered oncogenes *c-MYC* and *CTNNB1* by activating them in hiHep organoids. Overexpression of either gene significantly enhanced organoid growth (Figures S4A–S4C). Furthermore, we evaluated the functional roles of 7 candidate genes, 2 exhibited upregulation, and 5 underwent loss-of-function (see STAR Methods). Among these genes, 2 significantly increased organoid growth (Figures S4D–S4H). Specifically, ECHS1 depletion induced a robust pro-proliferative phenotype, with organoids exhibiting markedly increased

Please cite this article in press as: Zhang et al., Molecular insights into early malignant transition of hepatocellular carcinoma, Cancer Cell (2026), https://doi.org/10.1016/j.ccell.2026.03.005

size (Figures 4G–4I) and an elevated percentage of EdU-positive cells (Figures 4J and 4K). Similarly, the downregulation of another candidate gene, FGA, by either CRISPR-mediated knockout (Figures S4I–S4K) or shRNA-mediated knockdown (Figures S4L–S4N), significantly increased organoid size. Collectively, these results provide experimental evidence linking *ECHS1* and *FGA* alterations to the early malignant transition in liver carcinogenesis.

Immune-desert phenotype in premalignant DNs

Gene expression was examined during malignant transition. Upon analyzing the differentially expressed genes (DEGs) between cancer-prone and cancer-undetermined DNs, only 52 DEGs were identified between the two cohorts of DNs (Figure S5A, $p < 0.05$). Of these 52 genes, 20 were significantly upregulated in cancer-prone DNs, which were enriched for mitochondrial ATP synthesis and small molecule metabolic processes (Figure S5B). The remaining 32 genes were upregulated in cancer-undetermined DNs. Notably, 20 of these were snoRNAs, whose roles in early tumorigenesis remain largely elusive (Figure S5C).

Subsequently, we focused on the gene expression dynamics in evolutionarily related DNs and veHCCs. Using NCas as a control, cancer-prone DNs exhibited different expression profiles following principal component analysis (PCA). Interestingly, although cancer-prone DNs showed remarkable individual specificity in genomic variation, the expression profiles of these DNs clustered closely with each other (Figure 3A). Furthermore, cancer-prone DNs were separated from their evolutionarily related veHCCs, supporting the idea that, despite accumulated genetic alterations, DNs had not yet developed a cancerous phenotype. Notably, the expression profiles of veHCCs were dispersed from one another and showed increased expression heterogeneity (Figure 3A).

DEG analysis of NCas, cancer-prone DNs, and veHCCs further revealed three main modules that characterized the expression dynamics during early malignancy (Figure S5D). The largest module (module 1, 777 genes), which exhibited an increased expression during tumorigenesis, was associated with proliferation-related pathways (Figure S5D). Although cell cycle-related genes were induced in DNs, the expression of the key cell-cycle arrest gene, *CDKN2A*, was also upregulated (Figure S5E). Accordingly, the number of proliferative cells was elevated in veHCCs but not in DNs, as determined by Ki67 staining (Figure S5F). The second module (module 2, 215 genes) showed a gradual decline in expression during tumorigenesis and was enriched for liver metabolism pathways (Figure S5D). Among the 80 metabolism-related gene sets, 5 were reduced in DNs, including aspartate and asparagine metabolism (Figure S5G), and 40 were further diminished in veHCCs, including the fatty acid and hepatic vitamin metabolism, suggesting partial dedifferentiation and reprogramming during malignant transition. Enhanced glycolysis and glucose metabolism were detected (Figure S5H), which have previously been reported to provide favorable energy sources for cancer cells.[33]

Interestingly, a third module (module 3, 703 genes) presented a down-up expression pattern and was enriched in pathways related to epithelial-mesenchymal transition (EMT), angiogenesis, and immune-related pathways, such as the inflammatory response and IL2-STAT5 signaling (Figure S5D), suggesting dynamic microenvironmental changes during malignant transition. Immune response activation during HCC initiation was specifically evaluated. Previous studies have demonstrated that liver cancer frequently develops in an inflammatory microenvironment.[13,34–36] However, the expression of inflammation-related genes indicated that pro-tumorigenic inflammatory signaling was reduced in cancer-prone DNs, including TNFα/NF-κB, IL6/JAK/STAT3, and NLRP3-inflammasome signaling (Figure 3B). Specifically, when compared with NCas, cancer-prone DNs showed the downregulation of genes encoding pro-inflammatory type I interferons (e.g., *IFI44L* and *IFI16*), cytokines (e.g., *IL33* and *IL1B*), and toll-like receptors (e.g., *TLR1 TLR2*, and *TLR4*) (Figure 3C, $p < 0.05$), which were reported as important players in inflammation-related HCC development using mouse models.[35,37] Furthermore, cancer-prone DNs displayed the downregulation of genes involved in inflammation-associated fibrosis (e.g., *COL4A2* and *COL6A3*) and angiogenesis (e.g., *FGF2* and *MMP2*) (Figure 3C, $p < 0.05$). Interestingly, these inflammation-related signatures were also at low levels in 8 of 11 (73%) cancer-undetermined DNs (Figure S5I).

The abundance of different microenvironmental cell types was estimated by deconvoluting the expression data using MCP-counter.[38] The NCas exhibited high abundance scores for fibroblasts, endothelial cells, and monocytic lineage cells, consistent with the cirrhotic and chronic inflammatory states of these samples.[35,39] However, abundance scores for all these cell types were significantly reduced in cancer-prone DNs (Figure 3D). Multiplex staining of the macrophage marker, CD163, further confirmed the significant reduction in macrophages in cancer-prone DNs (Figures 3E and 3F). Furthermore, T cell infiltration was also low in cancer-prone DNs, as determined by immunostaining for CD4 and CD8 (Figures 3G and 3H). Additionally, analysis of dynamic changes in microenvironmental cell abundance from NCas to later-stage HCCs[22,23] revealed that cancer-prone DNs exhibited markedly lower abundance scores of fibroblasts, endothelial cells, monocytic lineage cells, and T cells compared with other groups (Figure S5J). Together, these data indicate that cancer-prone DNs exhibit an immune-desert phenotype, suggesting that HCCs originate from an immune-inactive, rather than immune-active[35,37,40] microenvironment.

Immune evasion phenotype of inflamed veHCCs

To define the immune landscape of veHCC, we estimated immune cell abundances and calculated the activation scores for an inflamed signature comprising 20 core genes.[40] 43% of the veHCCs exhibited both high abundances of T cells and dendritic cells (DCs) (Figure 4A) and enriched scores for the inflamed signature (Figure 4B), which we termed inflamed veHCCs. Remarkably, veHCCs with lower CIN index (0.002 ± 0.016) exhibited significant enrichment of inflamed phenotype compared with those with higher CIN index (0.072 ± 0.035) (Figure S6A, $p = 0.001$). Reciprocally, inflamed veHCCs displayed a significantly reduced CNA burden compared with non-inflamed veHCCs (Figure 4C). In contrast, SNV burden (Figure S6B) and predicted neoantigen levels (Figure S6C) were comparable between the two classes of veHCCs. Together, the integration of transcriptional and genomic characteristics revealed two major veHCC subtypes: inflamed veHCCs with a low CNA burden and non-inflamed veHCCs with a high CNA burden.

Please cite this article in press as: Zhang et al., Molecular insights into early malignant transition of hepatocellular carcinoma, Cancer Cell (2026), https://doi.org/10.1016/j.ccell.2026.03.005

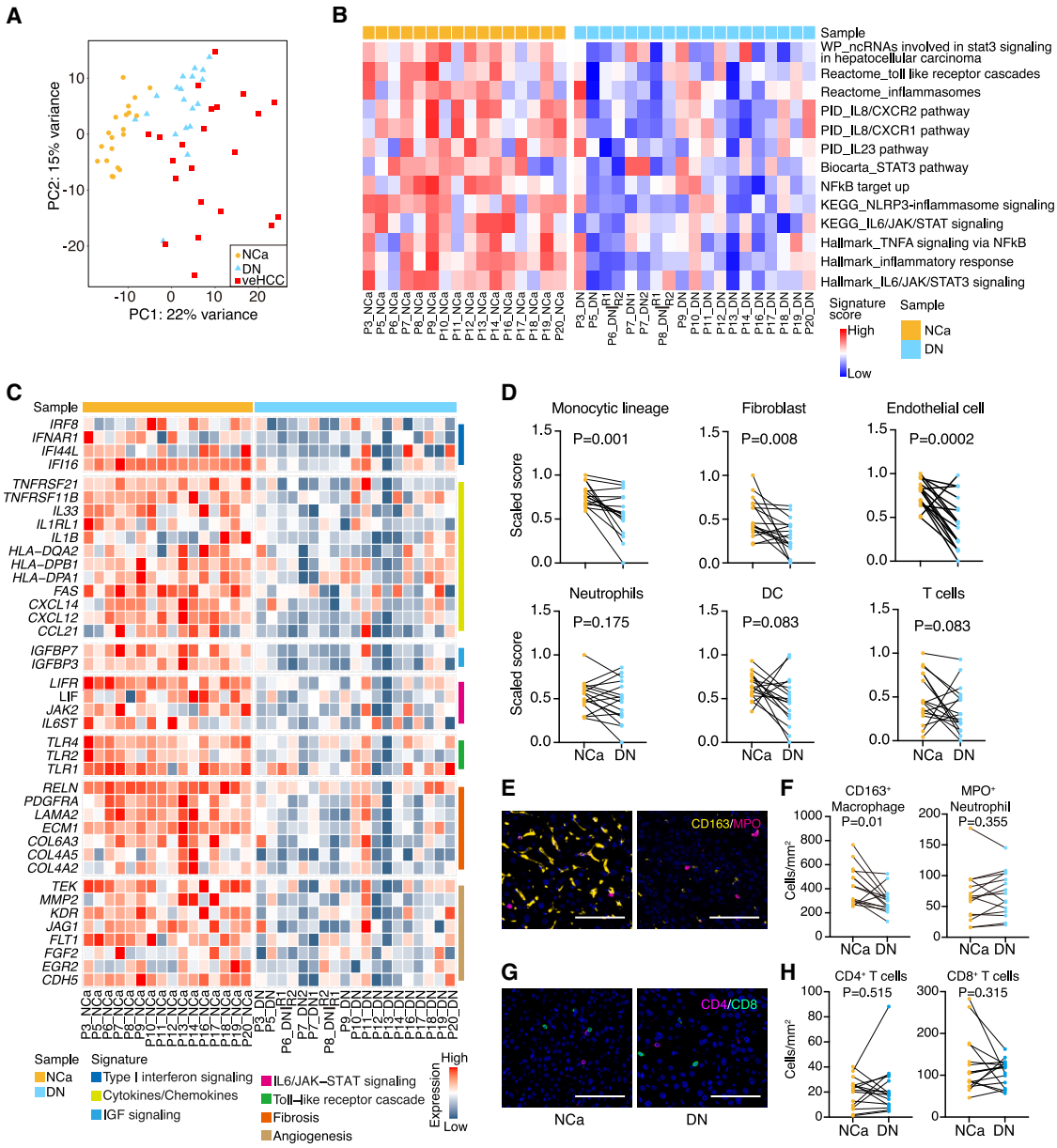


Figure 3. Immune desert phenotype in premalignant DNs

(A) Principal component analysis (PCA) was conducted using the top 500 variable genes among NCas, DNs, and veHCCs in evolutionarily related cases.

(B) Inflammation-associated pathway activity in NCas and DNs in evolutionarily related cases. The activity of each inflammation-associated pathway in each sample was determined using gene set variation analysis (GSVA).

(C) Reduced expression of inflammation-associated genes in DNs compared to their matched NCas.

(D) Estimated microenvironmental cell abundance from gene-expression profiles showing the changes in immune contexture in NCas and DNs. The abundance of six cell types was estimated from gene expression data. DC, dendritic cell.

(E) Representative immunofluorescence staining of CD163 and MPO of P11_NCa and DN. Scale bars, 100 μm.

(F) The densities of $CD163^+$ macrophages and MPO^+ neutrophils, quantified as the number of cells per tissue surface area by analyzing multispectral images.

(G) Representative immunofluorescence staining of CD4 and CD8 of P16_NCa and DN. Scale bars, 100 μm.

(H) The densities of $CD4^+$ and $CD8^+$ T cells. Statistical significance is determined by a two-tailed paired Student's *t* test (D, F, and H). See also Figure S5.

Compared to non-inflamed veHCCs, inflamed veHCCs showed increased expression of immune response-related genes, including chemokines (e.g., *CXCL12*, *CCL2*, and *CCL5*), HLA class II molecules (e.g., *HLA-DRA* and *HLA-DOA*), and immune cell markers (e.g., *CD3D*, *CD4*, and *CD8A*) (Figure S6D, $p < 0.05$). Additionally, multiplex staining of CD4 and CD8 confirmed that inflamed veHCCs exhibited high infiltration of $CD4^+$ and $CD8^+$ T cells (Figures 4D–4F).

Next, we determined the immune characteristics of the inflamed veHCCs with their related DNs. Unlike non-inflamed

Please cite this article in press as: Zhang et al., Molecular insights into early malignant transition of hepatocellular carcinoma, Cancer Cell (2026), https://doi.org/10.1016/j.ccell.2026.03.005

CellPress
OPEN ACCESS

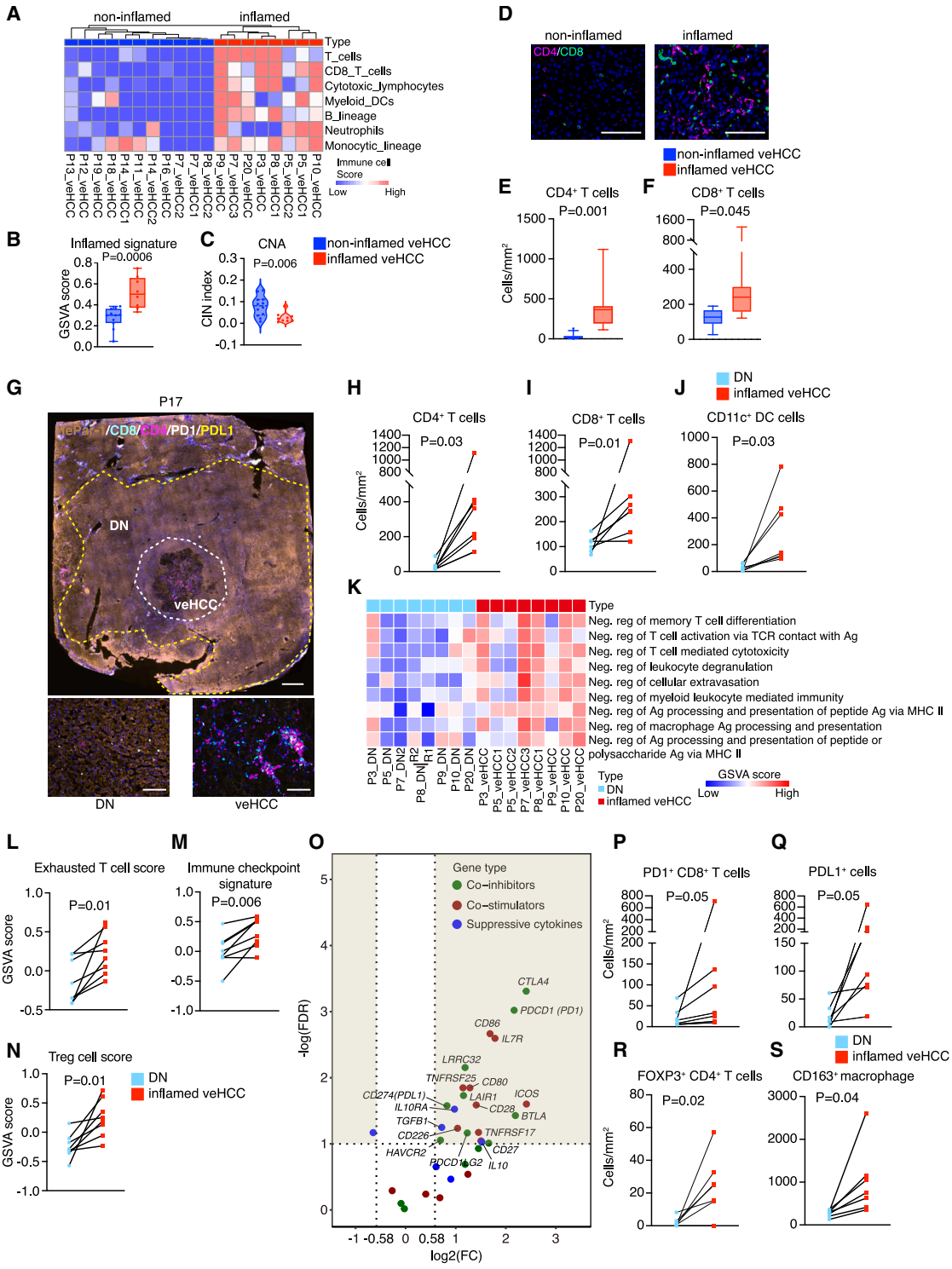


Figure 4. Adaptive immune activation and evasion in a subset of veHCCs

(A) Unsupervised clustering of veHCCs based on the estimated immune cell scores using the MCP-counter method. RNA-seq was not conducted for P6_veHCC, P9_veHCC_R2, and P17_veHCC due to insufficient material. To note, P6_veHCC was classified as non-inflamed veHCC, whereas P17_veHCC was classified as inflamed veHCC based on multiplex staining data. DC, dendritic cell.

(B) The activation scores of an inflamed signature comprising 20 genes, evaluated using GSVA.

(C) CNA burden in inflamed and non-inflamed veHCCs. The violin plot shows all data points.

(D) Representative immunofluorescence staining of CD4 and CD8 in an inflamed veHCC (P20_veHCC) and non-inflamed veHCC (P16_veHCC). Scale bars, 100 μm.

(legend continued on next page)

Please cite this article in press as: Zhang et al., Molecular insights into early malignant transition of hepatocellular carcinoma, Cancer Cell (2026), https://doi.org/10.1016/j.ccell.2026.03.005

veHCCs, which exhibited no enhanced immune response (Figure S6E), inflamed veHCCs showed increased expression of adaptive immune cell markers (e.g., *CD8A*, *CD4*, and *CD11c*) (Figure S6F) and a high abundance of T cells and DCs compared to their paired DNs (Figure S6G). Multiplex staining for CD4, CD8, and CD11c confirmed the increased infiltration of $CD4^+$ T cells, $CD8^+$ T cells, and DCs into inflamed veHCCs (Figures 4G–4J). Interestingly, genes involved in the bile acid pathway were downregulated in inflamed veHCCs (Figures S6H and S6I). Since bile acids have been shown to impede T cell responses in mouse liver cancer models,[41] it would be worthwhile to further investigate whether the downregulation of this pathway contributes to the development of the inflamed immune phenotype.

Notably, these inflamed veHCCs were different from their inflammatory and fibrotic NCa counterparts; they showed enhanced adaptive immune responses that were missing in their paired NCas, as assessed by MCP-counter analysis (Figure S6J) and multiplex staining (Figure S6K). Collectively, these data indicate that, in addition to inflammatory signaling activation, adaptive immune responses were enhanced in inflamed veHCCs.

Despite this pronounced immune activation, inflamed veHCCs simultaneously exhibited features of immune evasion. Pathways related to the negative regulation of T cell responses and antigen processing and presentation were already upregulated in inflamed veHCCs compared with paired DNs (Figure 4K). Consistently, inflamed veHCCs showed significantly higher scores for immune suppression-related signatures (Table S7), including signatures for exhausted T cells (Figure 4L), immune checkpoints (Figure 4M), and regulatory T cells (Treg) (Figure 4N). Signatures that recapitulated M2 macrophages and angiogenesis-related tumor-associated macrophages (TAMs) were also significantly enriched in these inflamed veHCCs (Figure S6L). Both types of macrophages have been shown to play key roles in tumor immune suppression.[42] Moreover, immune co-inhibitory (e.g., *CTLA4*, *PDL1*, and *PD1*), co-stimulatory (e.g., *CD80*, *CD86*, and *ICOS*), and suppressive molecules (e.g., *IL10* and *TGFB1*) were upregulated in these inflamed veHCCs (Figure 4O). Consistent with the upregulation of *TGFB1*, the TGF-β signaling showed increased activation scores in these inflamed veHCCs (Figure S6M). TGF-β has been previously associated with immune suppression[43] and the immune-exhausted subtype of HCCs.[40] Notably, immunostaining showed that the infiltration of $PD1^+CD8^+$ T cells, $PDL1^+$ cells, $FOXP3^+CD4^+$ T cells, and $CD163^+$ macrophages (Figures 4P–4S) increased in inflamed veHCCs. In addition to immune evasion, inflamed veHCCs exhibited significant enrichment of oncogenic driver pathways, such as IL6,[44] TNF-α,[45] EMT,[46] and KRAS signaling[47] (Figure S6N), consistent with early malignant progression to inflamed veHCCs.

Collectively, these findings indicate that inflamed veHCCs acquired features of immune evasion during the early malignant transition, characterized by robust adaptive immune activation coupled with the early establishment of immunosuppressive programs.

Spatial mapping of immune phenotypes in inflamed veHCCs

To define the spatial organization of the tumor microenvironment at the earliest stages of HCC development, we used the high-resolution spatial transcriptomics platform Visium HD to profile 3 inflamed veHCCs and 1 non-inflamed veHCC together with their paired DNs (Figure 5A). Using a single-cell reference, we deconvolved the Visium HD data and identified 9 major cell types (Figure 5B), which were validated by canonical marker expression, such as *ALB* (hepatocytes), *MAS4A1* (B cells), and *COL1A1* (cancer-associated fibroblasts, CAFs) (Figures S7A–S7D). Inflamed veHCCs showed increased abundances of $CD4^+$ and $CD8^+$ T cells, macrophages, and B cells compared with paired DNs (Figure 5C). CAFs were notably increased in 2 inflamed veHCCs (P7_veHCC3 and P20_veHCC) (Figure 5C). In contrast, the non-inflamed veHCC and its paired DN exhibited persistently low immune and stromal cell infiltration (Figure 5C).

We next analyzed the immune evasion features in inflamed veHCCs using spatial transcriptomics data. Spatial cell-cell interactions identified by SOAPy[48] showed that $CD8^+$ T cells engaged in extensive interactions with multiple cell types, including $CD4^+$ T cells, CAFs, and macrophages (Figures 5D–5F). Notably, $CD4^+$ T cells exhibited the highest number of ligand-receptor interactions with $CD8^+$ T cells in all 3 inflamed veHCCs (Figures 5D–5F). Spatially, $CD4^+$ and $CD8^+$ T cells were closely adjacent and formed distinct cell clusters (Figures 5G, 5H, and S7E). Analysis of Treg cell and exhausted T cell signature scores suggested pronounced activation of these cells in regions populated by $CD4^+$ and $CD8^+$ T cells (Figures 5I–5L, S7F, and S7G).

Notably, CAFs ranked among the top cell types interacting with $CD8^+$ T cells, particularly in the 2 inflamed veHCCs with high CAF abundance (Figures 5D–5F). Consistently, $CD8^+$ T cells were spatially proximal to CAFs (Figures 5M and 5N). Regions enriched for CAFs and $CD8^+$ T cells showed strong activation of myofibroblastic CAF (myoCAF) and TGF-β signatures (Figures 5O–5R). These findings are in line with bulk RNA-seq results and support a role for TGF-β-mediated immunosuppression. Given prior evidence implicating TAM in T cell suppression,[49,50] we specifically characterized TAM in inflamed

(E, F) The density of $CD4^+$ T cells (E) and $CD8^+$ T cells (F) in inflamed and non-inflamed veHCCs. Data are presented as mean ± SD.
(G) Representative multiplex staining of HepPar-1, PD1, PDL1, CD4, and CD8 of P17. Scale bars, 1 mm (upper) and 100 μm (lower).
(H–J) The densities of $CD4^+$ T (H), $CD8^+$ T (I), and $CD11c^+$ DC (J) cells in inflamed veHCCs and their paired DNs.
(K) The GSVA scores of gene sets associated with the negative regulation of immune responses in DNs and their matched inflamed veHCCs. Gene sets related to negative immune regulation were sourced from the Gene Ontology (GO) database.
(L–N) GSVA scores of exhausted T cell (L), immune checkpoint (M), and regulatory T cell (Treg) (N) signatures in DNs and paired inflamed veHCCs.
(O) Upregulated expression of co-inhibitors, co-stimulators, and suppressive cytokines in inflamed veHCCs compared with paired DNs. FDR <0.1 indicates significance.
(P–S) The densities of $PD1^+CD8^+$ T cells (P), $PDL1^+$ cells (Q), $FOXP3^+$ $CD4^+$ T cells (R), and $CD163^+$ macrophages (S) in DNs and paired inflamed veHCCs. Statistical significance is determined by a two-tailed unpaired Student's *t* test (B, C, E, and F), by two-tailed paired Student's *t* test (H–J, L–N, and P–S). For the box-and-whisker plot, the box indicates IQR, the line in the box indicates the median, the whiskers indicate points within Q3 + 1.5 × IQR and Q1 − 1.5 × IQR, and the points beyond whiskers indicate outliers (B, E, and F). See also Figure S6 and Table S7.

Please cite this article in press as: Zhang et al., Molecular insights into early malignant transition of hepatocellular carcinoma, Cancer Cell (2026), https://doi.org/10.1016/j.ccell.2026.03.005

CellPress
OPEN ACCESS

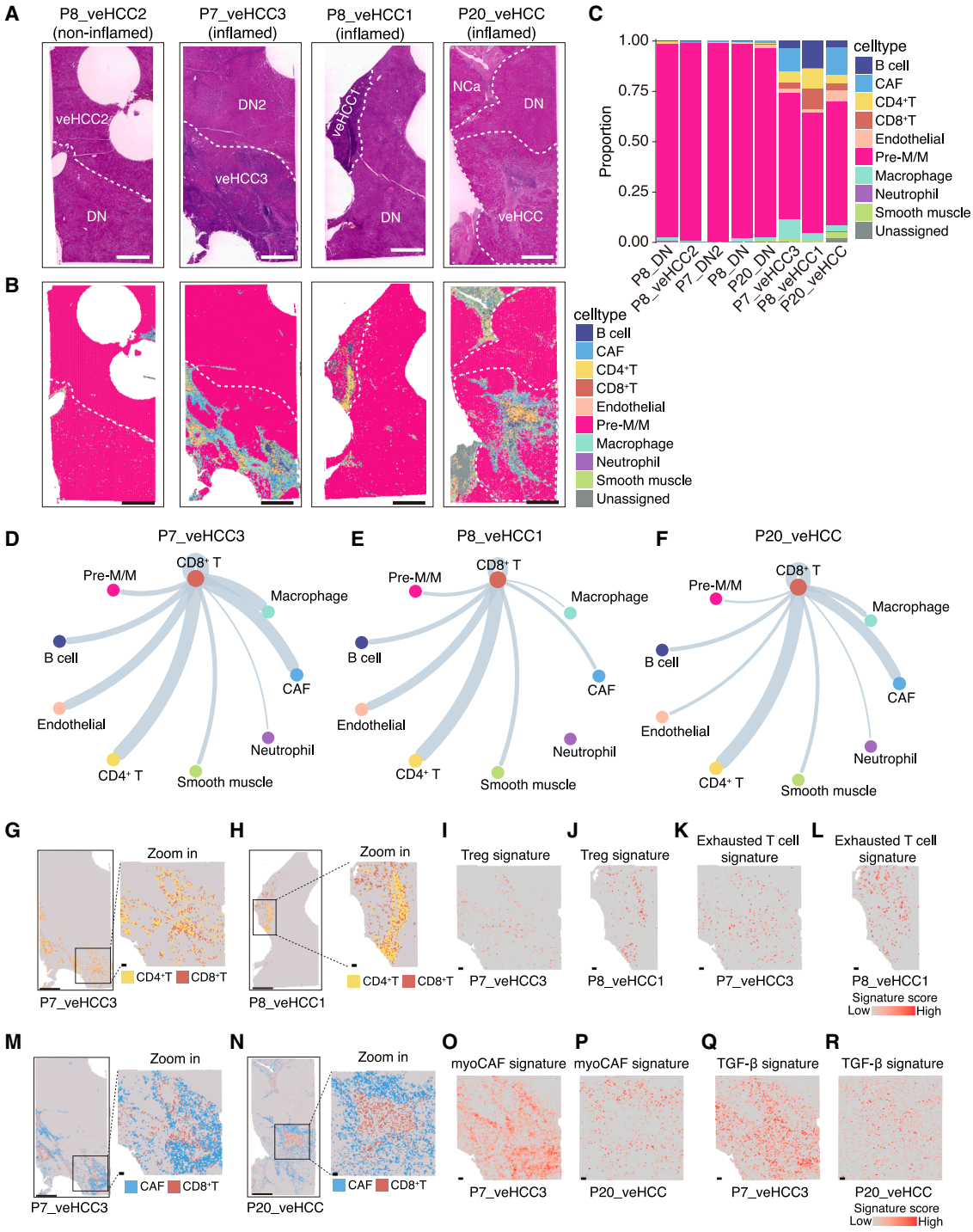


Figure 5. Spatial transcriptomic analysis of immune/stromal cell infiltration, cell-cell communication, and signature activation

(A) Whole-slide H&E images showing the regions of DN and veHCC in each case.

(B) Spatial distribution of the indicated cell types in the global view of Visium HD sections. Distributions were determined using RCTD deconvolution. "Pre-M" denotes premalignant hepatocytes, and "M" denotes malignant hepatocytes. Premalignant hepatocytes in DN regions and malignant hepatocytes in veHCC regions were grouped into a single cell cluster for subsequent analyses.

(C) The proportions of different immune and stromal cell types.

(D–F) Cell-cell communication networks between $CD8^+$ T cells and other cell types in P7_veHCC3 (D), P8_veHCC1 (E), and P20_veHCC (F). Edge width corresponds to the number of significant ligand-receptor pairs (affinity p value <0.05, strength >2).

(G and H) Spatial distribution of $CD4^+$ T cells and $CD8^+$ T cells in the global and local views of Visium HD sections from P7_veHCC3 and its paired DN (G), and P8_veHCC1 and its paired DN (H).

(legend continued on next page)

Please cite this article in press as: Zhang et al., Molecular insights into early malignant transition of hepatocellular carcinoma, Cancer Cell (2026), https://doi.org/10.1016/j.ccell.2026.03.005

veHCCs. Interactions were detected between $CD8^+$ T cells and macrophages (Figures 5D–5F), consistent with multiplex staining data showing a significant increase in infiltrating TAM in inflamed veHCCs (Figure 4S). In addition, inflamed veHCCs exhibited activation of IL6, TNF-α, EMT, and KRAS pathways (Figures S7H–S7K), further supporting the enrichment of oncogenic pathways in the bulk RNA-seq analyses (Figure S6N). Together, these data indicate that inflamed veHCCs activate inflammatory signaling while concurrently establishing an immunosuppressive microenvironment.

Integrative proposal of evolutionary scenarios of malignant transition

Collectively, our data suggest that both genomic alterations and immune microenvironmental changes occur during the early stages of HCC development. *TERT* alterations and reduced immune responses were key events in the development of cancer-prone DNs (Figure 6A). Based on integrated genomic and immune profiling from our dataset, we propose two potential evolutionary scenarios underlying DN-to-veHCC malignant transition: (1) CNA-dominant progression (Scenario 1) and (2) inflamed progression with immune evasion (Scenario 2) (Figure 6A). We also detected two outlier cases, P12_veHCC and P18_veHCC, which did not fall under either of these scenarios; neither exhibited additional CNAs in the CFGs nor increased immune responses compared to their paired DNs. P12_veHCC harbored a single point mutation in *ACVR2A*, whereas no additional CFG mutations were detected in P18_veHCC (Figure 2E).

To further delineate the evolutionary relationship of paired DNs and veHCCs, phylogenetic trees were constructed incorporating SNVs and CNAs affecting CFGs, enabling the annotation of trunk and branch events. CNA-induced CFG alterations were detected in the branches of all 6 non-inflamed veHCCs from Scenario 1 (Figures 6B, S8A, and S8C, Scenario 1). Representative cases included P6 and P14. In P6, *TERT* alterations were identified in the trunk, while many CFG-harboring CNAs accumulated in P6_veHCC, including Chr6q (*ARID1B*), Chr10q (*ECHS1*), Chr17p (*TP53*), and Chr13q (*RB1*) loss and the extended loss of Chr1p (*ARID1A*) (Figure 6B). In P14, losses of Chr13q (*RB1*) and Chr17p (*TP53*) emerged prior to the divergence of 2 veHCC lesions. P14_veHCC1 acquired lesion-specific CNAs, including gains in Chr5q (*IL6ST*), Chr6p (*BMP5*), and Chr12p (*KRAS*), and P14_veHCC2 additionally acquired loss of Chr9q (*TSC1*), suggesting CNA-mediated branch evolution (Figure S8A).

Phylogenetic trees of 9 inflamed veHCCs were analyzed in Scenario 2. 6 inflamed veHCCs (P5_veHCC2, P7_veHCC3, P9_veHCC, P10_veHCC, P17_veHCC, and P20_veHCC) acquired CNA-induced CFG alterations compared to their paired DNs (Figures S8B and S8C, Scenario 2), but the magnitude of CNA accumulation in these veHCCs was substantially lower compared with those in Scenario 1. Remarkably, the remaining 3 inflamed veHCCs (P3_veHCC, P5_veHCC1, and P8_veHCC1) exhibited nearly identical CNA profiles to their paired DNs (Figures 6C, S8B, and S8C). Moreover, P5_veHCC1 and P8_veHCC1 did not acquire additional mutations in any of the 37 CFGs, while P3_veHCC acquired only a point mutation in *ALB* (Figure 2E). These findings suggest that major cancer genes are absent in these veHCCs, indicating a potential role for immune evasion during the malignant transition of the 3 veHCCs.

We observed two cases (P7 and P8), in which spatially separated veHCCs developed from different evolutionary scenarios in the same patient (Figure S8C). P7 represented a typical clinical case of multifocal HCCs, with independent initiation from spatially separated DNs (Figure S9). In P8, however, two spatially separated veHCCs represented different evolutionary scenarios developed from the same DN (Figure 7A). P8_veHCC2 showed a large number of CNAs covering 14 chromosomes, which resulted in gains of oncogenes (*EGFR*, *MET,* and *KRAS*) and losses of tumor suppressors (*PTEN* and *RB1*) (Figures 7A and 7B), consistent with CNA-dominant malignant transition (Scenario 1). In contrast, P8_veHCC1 presented almost identical CNAs to P8_DN (Figure 7B) and lacked additional mutations in the CFGs. However, unlike P8_veHCC2, P8_veHCC1 showed an inflamed phenotype compared to P8_DN (Figures 7C–7E). Histological assessment, multiplex staining (CD3-T cell, CD20-B cell, CD21-follicular dendritic, and FOXP3-Treg cell), and signature expression analysis indicated the presence of tertiary lymphoid structures (TLSs) in the P8_veHCC1 (Figures 7F and 7G). TLS are key players in the response to checkpoint inhibitor-based immunotherapies[51–53] and have been shown to promote HCC initiation.[54] Furthermore, the increased infiltration of $PD1^+CD8^+$ T, $FOXP3^+CD4^+$ T cells, and $PDL1^+$ cells (Figure 7E) may mediate immune evasion and facilitate malignant transition from P8_DN to P8_ veHCC1 (Scenario 2).

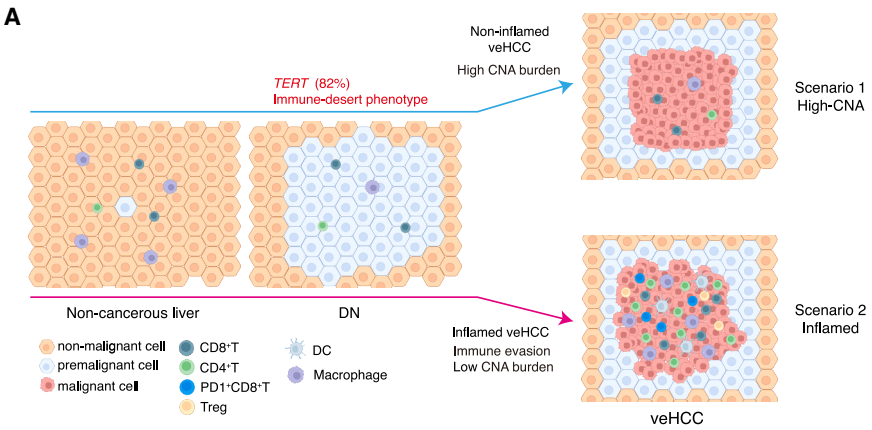
DISCUSSION

Unlike previous publications,[5,6] this study focused on evolutionarily related DNs and veHCCs, offering unique insights into the spatiotemporal genomic, transcriptomic, and immune microenvironmental changes during malignant transition to veHCCs. Our findings provide direct evidence of the molecular drivers and evolutionary scenarios underlying the malignant transition of HCC, challenging the existing paradigms of cancer development in several aspects. Our analysis of the CFGs and CNAs indicated that our samples covered most recurrent CFGs (e.g., *TERT*, *TP53*, and *CTNNB1*) and CNAs (e.g., Chr1q/Chr8q gain and Chr10q/Chr17p loss) observed in malignant HCCs, supporting their representativeness. Several striking biological findings emerged. Notably, *TERT* alterations and diminished immune responses were detected in over 80% of DNs, suggesting these lesions acquire oncogenic "priming" before histological malignancy.

(I and J) Regulatory T (Treg) cell signature scores in local views of Visium HD sections from P7_veHCC3 (I) and P8_veHCC1 (J).
(K and L) Exhausted T cell signature scores in local views of Visium HD sections from P7_veHCC3 (K) and P8_veHCC1 (L).
(M and N) Spatial distribution of CAF and $CD8^+$ T cells in the global and local views of Visium HD sections from P7_veHCC3 and its paired DN (M), and P20_veHCC and its paired DN (N).
(O and P) myoCAF signature scores in local views of Visium HD sections from P7_veHCC3 (O) and P20_veHCC (P).
(Q and R) TGF-β signature scores in local views of Visium HD sections from P7_veHCC3 (Q) and P20_veHCC (R). Scale bars for (A and B), 1 mm. Scale bars for (G, H, M, N), 1 mm for global views and 80 μm for local views. Scale bars for (I–L and O–R), 80 μm. See also Figure S7.

Please cite this article in press as: Zhang et al., Molecular insights into early malignant transition of hepatocellular carcinoma, Cancer Cell (2026), https://doi.org/10.1016/j.ccell.2026.03.005

CellPress
OPEN ACCESS



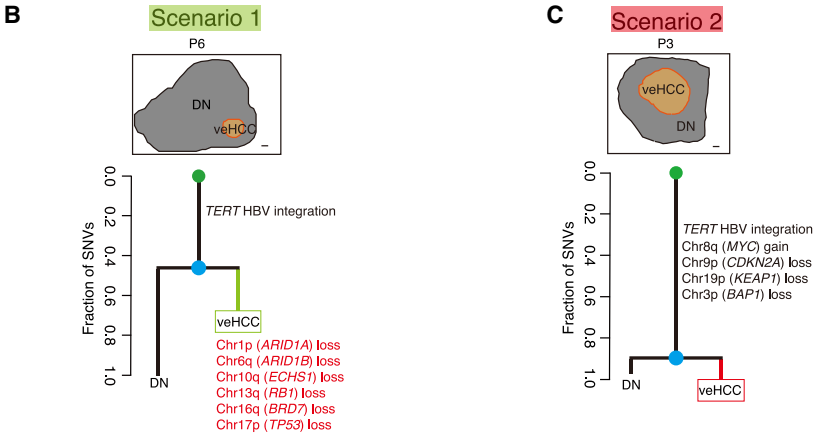


Figure 6. Evolutionary scenarios of malignant transition of DNs

(A) Schematic illustration of the progression of carcinogenesis from NCas to veHCCs. Two major evolutionary scenarios of the malignant transition from DNs to veHCCs are illustrated.

(B) Spatial relationship of the DNs and veHCCs in P6 (upper). Phylogenetic trees of P6 (lower) (Scenario 1). The trunk and branch lengths are proportional to the number of SNVs.

(C) Spatial relationship of the DNs and veHCCs in P3 (upper). Phylogenetic trees of P3 (lower) (Scenario 2). Scale bars for (B and C), 1 mm. See also Figure S8.

Furthermore, genomic and immune profiling revealed distinct evolutionary scenarios during early hepatocarcinogenesis. CIN, reflected by an increased CNA burden, was observed in over 80% of veHCCs, whereas approximately 50% of veHCCs concurrently exhibited an immunosuppressive tumor microenvironment. These parallel findings imply that CIN and immune microenvironment remodeling are strongly associated with the premalignant-to-malignant transition.

Telomerase activation caused by *TERT* alterations is a hallmark of HCC and is present in around 90% of cases.[15,55] Interestingly, *TERT* alterations were previously observed in 20% of undetermined DNs, including those either progressing to HCCs or eventually regressing.[5,6] However, *TERT* alterations were detected in 82% of cancer-prone DNs, indicating that *TERT* alterations may not be direct drivers of malignant transition but instead potentially function as gatekeepers. In this study, all nodules-in-nodules lesions occurred within the context of cirrhosis. Recent studies have identified *TP53* and *ARID1A* mutations, but not *TERT* alterations, in cirrhotic liver tissues.[56,57] Therefore, the results from the DN analysis[56] suggest that mutations in other cancer-related genes may not be as essential as *TERT* for the genesis of cancer-prone DNs. Additionally, our findings suggest the possibility of reevaluating the functions of *TERT* alterations in the malignant transition of other cancer types, such as melanoma, glioma, and bladder cancer.[58]

Human liver cancers are characterized by numerous genomic alterations, including SNVs and CNAs.[22,23] However, the sequential order of these alterations and their contributions to malignant transition remains largely unknown. We found that the SNV burden did not increase in veHCCs compared to their paired DNs. Furthermore, the acquisition of additional mutations in cancer-related genes was less common than the accumulation of new CNAs during malignant transition from DN to veHCC. This shifts the focus away from the conventional view that driver gene mutations are the major instigators of cancer development.[5,6,15,18] Instead, we showed that CNAs frequently accumulated during malignant transition, with copy number gains in oncogenes and losses in TSGs playing pivotal roles in the transition from DN to veHCC. Pathways related to chromosomal instability, chromosome breakage, and DNA repair were remarkably enhanced in veHCCs, suggesting that genomic instability may underlie these CNA-driven changes. Although *TERT* expression persisted throughout the transition from DN to veHCC, telomere attrition was not completely mitigated, which may also be correlated with chromosomal instability. Notably, after veHCCs develop from matched DNs, both veHCCs and the parental DN continue to accumulate unique SNVs and CNAs. It is possible that malignant transition to veHCC is a CNA-associated phase, whereas the progression from veHCC to HCC is a CNA- and SNV-associated phase.

It was previously unclear how the immune microenvironment evolves and participates in the transition from pre-malignancy to malignancy. Surprisingly, we found that DNs were dominated by an immune desert phenotype. Inflammation has been shown to enhance liver carcinogenesis by releasing excess cytokines and growth factors to the microenvironment.[34–36] Since veHCCs evolve from immune-inactive DNs, this suggests that there is a need for reevaluating the role of inflammation in liver carcinogenesis. Notably, when the immune states of veHCCs were analyzed, a subset of veHCCs showed an inflamed phenotype. Interestingly, these inflamed veHCCs had already acquired immune evasion-related features, suggesting that the inflamed

Please cite this article in press as: Zhang et al., Molecular insights into early malignant transition of hepatocellular carcinoma, Cancer Cell (2026), https://doi.org/10.1016/j.ccell.2026.03.005

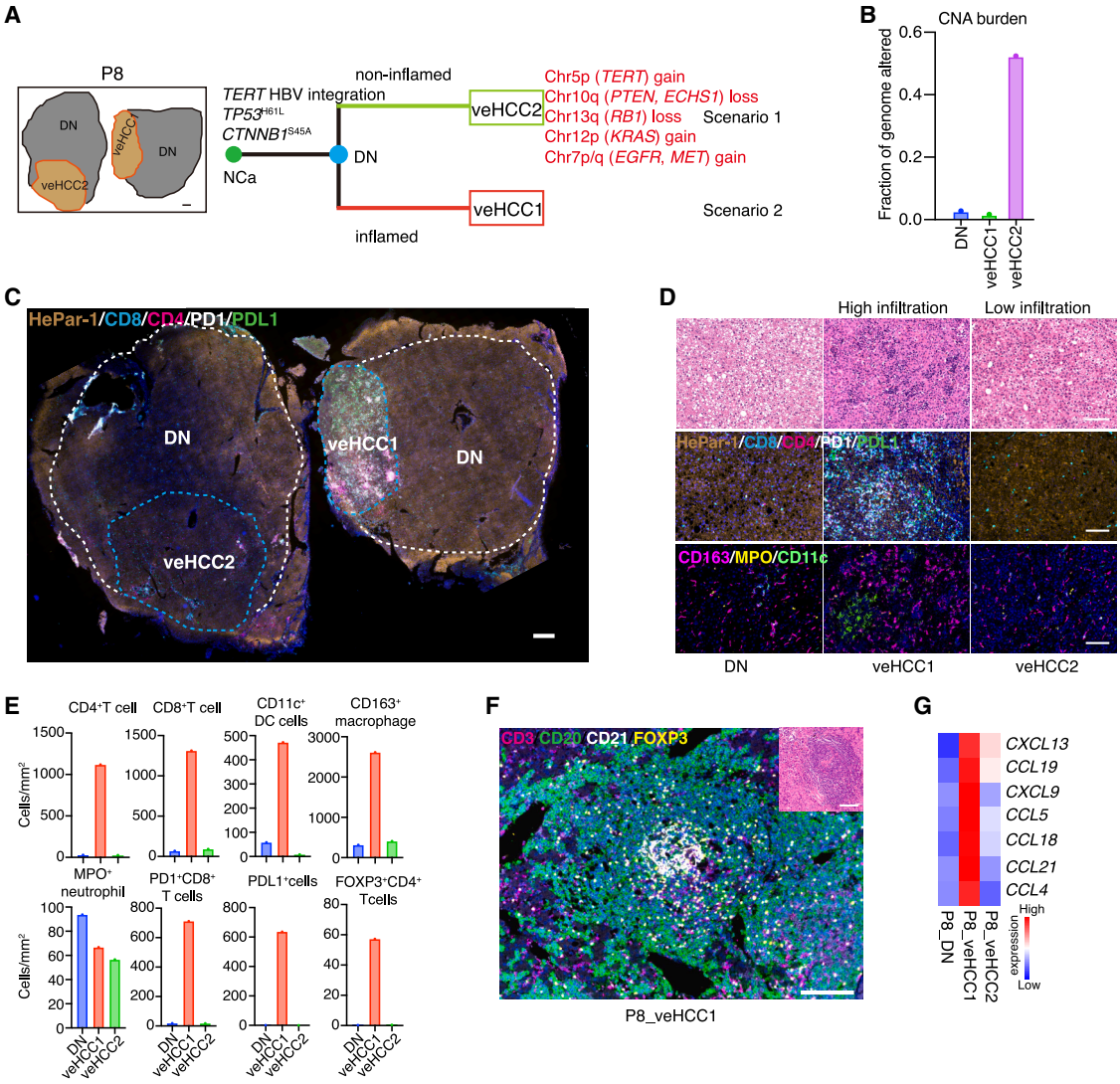


Figure 7. Evolutionary scenarios of P8

(A) Schematic illustration of the two evolutionary scenarios of the malignant transition from P8_DN to P8_veHCC1 and P8_veHCC2. The trunk and branch lengths are proportional to the number of SNVs.

(B) CNA burden in the DNs and paired veHCCs of P8.

(C) Multiplex staining of HepPar-1, PD1, PDL1, CD4, and CD8 of P8.

(D) Representative H&E staining (upper), multiplex staining of HepPar-1, PD1, PDL1, CD4, and CD8 (middle), and multiplex staining of CD163, MPO, and CD11c (bottom) of P8.

(E) The densities of $CD4^+$ and $CD8^+$ T cells, $CD11c^+$ DC cells, $CD163^+$ macrophages, MPO^+ neutrophils, $PD1^+$ $CD8^+$ T cells, $PDL1^+$ cells, and $FOXP3^+$ $CD4^+$ T cells in the DN and veHCCs of P8.

(F) Multiplex staining of CD3, CD20, CD21, and FOXP3 of tertiary lymphoid structures (TLSs) in P8_veHCC1. H&E staining shows the histology of TLS.

(G) Heatmap showing the upregulated expression of TLS-related signature genes in P8_veHCC1. Scale bars for (A and C), 1 mm. Scale bars for (D and F), 100 μm. See also Figure S9.

phenotype and immune evasion may coexist at a very early stage of HCC initiation and that immune evasion may contribute to malignant transition.

Two major evolutionary scenarios from DN to veHCC were observed: one with a high CNA burden (Scenario 1) and another with an inflamed phenotype displaying features of immune evasion (Scenario 2). Unlike previous studies that statically describe the correlation between aneuploidy and immune responses in later-stage cancers,[59,60] our findings emphasize the dynamic changes in CNAs and the immune microenvironment during malignant transition. Although the causative link between CNA burden and immune status remains unknown, it is interesting that veHCCs with high CNA burdens exhibited significant enrichment of *TP53*-related genomic abnormalities.[61,62] Furthermore, non-inflamed veHCCs were characterized with significantly low expression of *STING1*, a key player enabling immune surveillance evasion.[63] Nevertheless, it is worth investigating whether immune cells select against cells harboring CNAs, or

Please cite this article in press as: Zhang et al., Molecular insights into early malignant transition of hepatocellular carcinoma, Cancer Cell (2026), https://doi.org/10.1016/j.ccell.2026.03.005

non-inflamed veHCCs possess a permissive environment for CNA accumulation.

Among the cases in Scenario 2, P3_veHCC, although exhibiting a relatively high CIN index among inflamed veHCCs, displayed nearly identical CNAs to its matched DN. Similarly, P5_veHCC1 and P8_veHCC1 also shared nearly identical CNAs with their paired DNs. Furthermore, no SNVs in known HCC-driving or pan-cancer oncogenes[29] and TSGs[30] were acquired in P8_veHCC1. P3_veHCC presented with additional SNVs in the *ALB* and *UFL1* gene. Mutations in the *ALB* gene may be associated with its high expression, which renders the gene prone to mutagenesis.[57] In P5_veHCC1, three new SNVs were identified (*EEF1A1*, *NDST4*, and *CTNNA2*). *NDST4* and *CTNNA2* are expressed at very low levels and may not be involved in HCC formation. The roles of *EEF1A1* and *UFL1* in liver carcinogenesis remain unclear. Although it is premature to draw definitive conclusions, these findings collectively suggest a testable hypothesis: In some cases, a malignant phenotype may arise because of factors other than genetic alterations.[64] Notably, hepatocyte reprogramming to a progenitor-like state can occur during the HCC initiation, potentially contributing to early malignant transition. In this regard, it is worth investigating whether some cancer cells, even after acquiring mutations, can be reverted to non-malignant states.

Our findings provide several insights that may be valuable for clinical practice. First, the detection of *TERT* alterations as potential markers for cancer-prone DNs offers an opportunity for early stratification and non-invasive monitoring through circulating DNA analysis. Moreover, the high prevalence of *TERT* overexpression may serve as a possible target for immunotherapy to eliminate cancer-prone DNs and veHCCs.[65] Second, exploring CNAs and their associated genomic instability pathways may open new avenues for therapeutic intervention during the transition to malignancy. Synthetic lethal targets that exploit the vulnerabilities of cancers with specific chromosomal alterations have been identified, including the loss of chromosomes 1p,[66] 18q, or 16q,[67] and deletions at chromosome 9p21.3.[68,69] Third, immune checkpoint inhibitors could be beneficial for patients with inflamed veHCCs.

Limitation of this study

Due to the rarity of nodule-in-nodule samples, the final cohort size was limited despite an extensive screening of approximately 45,000 samples over a 9-year period at 3 large liver cancer centers. As a result, certain statistical analyses may be underpowered, and additional findings on CFGs and evolutionary scenarios might emerge in larger cohorts. Moreover, by focusing specifically on HCCs arising through the DN-HCC sequence, our study may not capture *de novo* hepatocarcinogenic pathways that lack identifiable dysplastic precursors. Future investigations integrating *in vivo* animal models and human organoid systems will be important to further delineate the mechanistic roles of the newly identified molecular and cellular alterations in early hepatocarcinogenesis.

RESOURCE AVAILABILITY

Lead contact

Requests for further information and resources should be directed to and will be fulfilled by the Lead Contact, Lijian Hui (ljhui@sibcb.ac.cn).

Materials availability

This study did not generate new unique reagents.

Data and code availability

The raw sequence data reported in this paper have been deposited in the Genome Sequence Archive (Genomics, Proteomics & Bioinformatics 2021) at the National Genomics Data Center (Nucleic Acids Res 2022), China National Center for Bioinformation/Beijing Institute of Genomics, and Chinese Academy of Sciences (GSA-Human: HRA008950), and are publicly accessible at https://ngdc.cncb.ac.cn/gsa-human. The detailed codes have been deposited at https://github.com/LiHongCSBLab/DN_eHCC_bulkOmics.

ACKNOWLEDGMENTS

We thank CQX, JQZ (Center for Excellence in Molecular Cell Science, CAS), BL (Shanghai Institute of Immunology, Shanghai Jiaotong University), PYY (Institute of Biophysics, CAS), LL (Nankai University), and DGL (Suzhou Institute of Systems Medicine) for their helpful discussions and suggestions. We thank all patients for their consent and participation in this study.

This study was supported by the National Natural Science Foundation of China (92368301, 82504075, 92168202), the Noncommunicable chronic disease-national science and technology major project (2024ZD525404), the NSFC (T2122018 and 32470707), the Shanghai Municipal Science and Technology Major Project, and the "Strategic Priority Research Program" of the Chinese Academy of Sciences (XDB1150000). JML was supported by grants from the European Commission (Horizon European Mission Cancer, THRIVE; 101136622); the NIH (RO1-CA273932-01, RO1DK56621, and RO1DK128289); the Samuel Waxman Cancer Research Foundation; the Spanish National Health Institute (MICINN, PID2022-139365OB-I00, funded by MICIU/AEI/10.13039/501100011033 and FEDER); (1) Cancer Research UK, United Kingdom; (2) Fondazione AIRC per la Ricerca sul Cancro, Italy and Fundación Científica de la Asociación Española Contra el Cáncer (Accelerator Award, HUNTER, C9380/A26813); the "la Caixa" Banking Foundation under agreement LCF/PR/SP23/52950009; Fundación Científica de la Asociación Española Contra el Cáncer (FAECC; Proyectos Generales, PRYGN223117LLOV; Reto AECC 70%; Supervivencia: RETOS245779LLOV; AECC-IDIBAPS Excellence Program Ref. EPAEC246711CLIN and the Generalitat de Catalunya/AGAUR (2021 SGR 01347). RP was supported by the Fundació de Recerca Clínic Barcelona - IDIBAPS and by a grant from the Spanish National Health Institute (MICINN, PID2022-139365OB-I00, funded by MICIU/AEI/10.13039/501100011033 and FEDER). AM was supported by Generalitat de Catalunya with an FISDUR fellowship (2021 FISDU 00338) from AGAUR and by mobility grants from the University of Barcelona, Montcelimar Foundation and Acadèmia de Ciències Mèdiques i de la Salut de Catalunya i de Balears Foundation.

AUTHOR CONTRIBUTIONS

Conceptualization: Z. Z., L. H., H. L., and Y. J.; methodology: Z. Z., H. L., Z. T., L. C., S. J., and D. L.; software, H. L., X. S., and Y. L.; experiments: Z. Z., X. G., and Z. J.; sample and histology: Y. J., L. C., H. D., T. L., H. W., J. F., J. H., and Y. Z.; writing: Z. Z., L. H., H. L., A. M., R. P., J. M L., Y. J., and Z. Q.; supervision: L. H., Y. J., and H. L.

DECLARATION OF INTERESTS

JML has received research support from Genentech and Roche, consulting and sponsored lecture fees from Eisai, Merck, Roche, Genentech, AstraZeneca, Bayer Pharmaceuticals, AbbVie, Sanofi, Moderna, Glycotest, Exelixis, and Boehringer Ingelheim, and is on the Data Safety Monitoring Board for Bristol Myers Squibb. All other authors declare no competing financial interests.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.ccell.2026.03.005.

Received: June 15, 2025
Revised: January 2, 2026
Accepted: March 4, 2026

REFERENCES

1. Fitzgerald, R.C., Antoniou, A.C., Fruk, L., and Rosenfeld, N. (2022). The future of early cancer detection. Nat. Med. *28*, 666–677. https://doi.org/10.1038/s41591-022-01746-x.
2. Crosby, D., Bhatia, S., Brindle, K.M., Coussens, L.M., Dive, C., Emberton, M., Esener, S., Fitzgerald, R.C., Gambhir, S.S., Kuhn, P., et al. (2022). Early detection of cancer. Science *375*, eaay9040. https://doi.org/10.1126/science.aay9040.
3. Vogel, A., Chan, S.L., Dawson, L.A., Kelley, R.K., Llovet, J.M., Meyer, T., Ricke, J., Rimassa, L., Sapisochin, G., Vilgrain, V., et al. (2025). Hepatocellular carcinoma: ESMO Clinical Practice Guideline for diagnosis, treatment and follow-up. Ann. Oncol. *36*, 491–506. https://doi.org/10.1016/j.annonc.2025.02.006.
4. Srivastava, S., Ghosh, S., Kagan, J., and Mazurchuk, R.; National Cancer Institute's HTAN Implementation (2018). The Making of a PreCancer Atlas: Promises, Challenges, and Opportunities. Trends Cancer *4*, 523–536. https://doi.org/10.1016/j.trecan.2018.06.007.
5. Torrecilla, S., Sia, D., Harrington, A.N., Zhang, Z., Cabellos, L., Cornella, H., Moeini, A., Camprecios, G., Leow, W.Q., Fiel, M.I., et al. (2017). Trunk mutational events present minimal intra- and inter-tumoral heterogeneity in hepatocellular carcinoma. J. Hepatol. *67*, 1222–1231. https://doi.org/10.1016/j.jhep.2017.08.013.
6. Nault, J.C., Calderaro, J., Di Tommaso, L., Balabaud, C., Zafrani, E.S., Bioulac-Sage, P., Roncalli, M., and Zucman-Rossi, J. (2014). Telomerase reverse transcriptase promoter mutation is an early somatic genetic alteration in the transformation of premalignant nodules in hepatocellular carcinoma on cirrhosis. Hepatology *60*, 1983–1992. https://doi.org/10.1002/hep.27372.
7. Schulze, K., Imbeaud, S., Letouzé, E., Alexandrov, L.B., Calderaro, J., Rebouissou, S., Couchy, G., Meiller, C., Shinde, J., Soysouvanh, F., et al. (2015). Exome sequencing of hepatocellular carcinomas identifies new mutational signatures and potential therapeutic targets. Nat. Genet. *47*, 505–511. https://doi.org/10.1038/ng.3252.
8. Cross, W., Kovac, M., Mustonen, V., Temko, D., Davis, H., Baker, A.M., Biswas, S., Arnold, R., Chegwidden, L., Gatenbee, C., et al. (2018). The evolutionary landscape of colorectal tumorigenesis. Nat. Ecol. Evol. *2*, 1661–1672. https://doi.org/10.1038/s41559-018-0642-z.
9. Mascaux, C., Angelova, M., Vasaturo, A., Beane, J., Hijazi, K., Anthoine, G., Buttard, B., Rothe, F., Willard-Gallo, K., Haller, A., et al. (2019). Immune evasion before tumour invasion in early lung squamous carcinogenesis. Nature *571*, 570–575. https://doi.org/10.1038/s41586-019-1330-0.
10. Risom, T., Glass, D.R., Averbukh, I., Liu, C.C., Baranski, A., Kagel, A., McCaffrey, E.F., Greenwald, N.F., Rivero-Gutiérrez, B., Strand, S.H., et al. (2022). Transition to invasive breast cancer is associated with progressive changes in the structure and composition of tumor stroma. Cell *185*, 299–310.e18. https://doi.org/10.1016/j.cell.2021.12.023.
11. Chen, B., Scurrah, C.R., McKinley, E.T., Simmons, A.J., Ramirez-Solano, M.A., Zhu, X., Markham, N.O., Heiser, C.N., Vega, P.N., Rolong, A., et al. (2021). Differential pre-malignant programs and microenvironment chart distinct paths to malignancy in human colorectal polyps. Cell *184*, 6262–6280.e26. https://doi.org/10.1016/j.cell.2021.11.031.
12. Craig, A.J., von Felden, J., Garcia-Lezana, T., Sarcognato, S., and Villanueva, A. (2020). Tumour evolution in hepatocellular carcinoma. Nat. Rev. Gastroenterol. Hepatol. *17*, 139–152. https://doi.org/10.1038/s41575-019-0229-4.
13. Llovet, J.M., Kelley, R.K., Villanueva, A., Singal, A.G., Pikarsky, E., Roayaie, S., Lencioni, R., Koike, K., Zucman-Rossi, J., and Finn, R.S. (2021). Hepatocellular carcinoma. Nat. Rev. Dis. Primers *7*, 6. https://doi.org/10.1038/s41572-020-00240-3.
14. International Consensus Group for Hepatocellular NeoplasiaThe International Consensus Group for Hepatocellular Neoplasia (2009). Pathologic diagnosis of early hepatocellular carcinoma: a report of the international consensus group for hepatocellular neoplasia. Hepatology *49*, 658–664. https://doi.org/10.1002/hep.22709.
15. Zucman-Rossi, J., Villanueva, A., Nault, J.C., and Llovet, J.M. (2015). Genetic Landscape and Biomarkers of Hepatocellular Carcinoma. Gastroenterology *149*, 1226–1239.e4. https://doi.org/10.1053/j.gastro.2015.05.061.
16. Choi, B.I., Lee, J.M., Kim, T.K., Dioguardi Burgio, M., and Vilgrain, V. (2015). Diagnosing Borderline Hepatic Nodules in Hepatocarcinogenesis: Imaging Performance. AJR Am. J. Roentgenol. *205*, 10–21. https://doi.org/10.2214/AJR.14.12655.
17. Nault, J.C., Paradis, V., Ronot, M., and Zucman-Rossi, J. (2022). Benign liver tumours: understanding molecular physiology to adapt clinical management. Nat. Rev. Gastroenterol. Hepatol. *19*, 703–716. https://doi.org/10.1038/s41575-022-00643-5.
18. Müller, M., Bird, T.G., and Nault, J.C. (2020). The landscape of gene mutations in cirrhosis and hepatocellular carcinoma. J. Hepatol. *72*, 990–1002. https://doi.org/10.1016/j.jhep.2020.01.019.
19. Roncalli, M., Terracciano, L., Di Tommaso, L., David, E., and Colombo, M.; Gruppo Italiano Patologi Apparato Digerente GIPAD; Società Italiana di Anatomia Patologica e Citopatologia Diagnostica/International Academy of Pathology Italian division SIAPEC/IAP (2011). Liver precancerous lesions and hepatocellular carcinoma: the histology report. Dig. Liver Dis. *43*, 361–372. https://doi.org/10.1016/S1590-8658(11)60592-6.
20. Llovet, J.M., Burroughs, A., and Bruix, J. (2003). Hepatocellular carcinoma. Lancet *362*, 1907–1917. https://doi.org/10.1016/S0140-6736(03)14964-1.
21. Van Allen, E.M., Wagle, N., Stojanov, P., Perrin, D.L., Cibulskis, K., Marlow, S., Jane-Valbuena, J., Friedrich, D.C., Kryukov, G., Carter, S.L., et al. (2014). Whole-exome sequencing and clinical interpretation of formalin-fixed, paraffin-embedded tumor samples to guide precision cancer medicine. Nat. Med. *20*, 682–688. https://doi.org/10.1038/nm.3559.
22. Cancer Genome Atlas Research Network Electronic address wheeler@bcm edu; Electronic address, w.b.e.; Cancer Genome Atlas Research Network (2017). Comprehensive and Integrative Genomic Characterization of Hepatocellular Carcinoma. Cell *169*, 1327–1341.e23. https://doi.org/10.1016/j.cell.2017.05.046.

23. Chen, L., Zhang, C., Xue, R., Liu, M., Bai, J., Bao, J., Wang, Y., Jiang, N., Li, Z., Wang, W., et al. (2024). Deep whole-genome analysis of 494 hepatocellular carcinomas. Nature *627*, 586–593. https://doi.org/10.1038/s41586-024-07054-3.

24. Guichard, C., Amaddeo, G., Imbeaud, S., Ladeiro, Y., Pelletier, L., Maad, I.B., Calderaro, J., Bioulac-Sage, P., Letexier, M., Degos, F., et al. (2012). Integrated analysis of somatic mutations and focal copy-number changes identifies key genes and pathways in hepatocellular carcinoma. Nat. Genet. *44*, 694–698. https://doi.org/10.1038/ng.2256.

25. Fujimoto, A., Totoki, Y., Abe, T., Boroevich, K.A., Hosoda, F., Nguyen, H.H., Aoki, M., Hosono, N., Kubo, M., Miya, F., et al. (2012). Whole-genome sequencing of liver cancers identifies etiological influences on mutation patterns and recurrent mutations in chromatin regulators. Nat. Genet. *44*, 760–764. https://doi.org/10.1038/ng.2291.

26. Kan, Z., Zheng, H., Liu, X., Li, S., Barber, T.D., Gong, Z., Gao, H., Hao, K., Willard, M.D., Xu, J., et al. (2013). Whole-genome sequencing identifies recurrent mutations in hepatocellular carcinoma. Genome Res. *23*, 1422–1433. https://doi.org/10.1101/gr.154492.113.

27. Fujimoto, A., Furuta, M., Totoki, Y., Tsunoda, T., Kato, M., Shiraishi, Y., Tanaka, H., Taniguchi, H., Kawakami, Y., Ueno, M., et al. (2016). Whole-genome mutational landscape and characterization of noncoding and structural mutations in liver cancer. Nat. Genet. *48*, 500–509. https://doi.org/10.1038/ng.3547.

28. Totoki, Y., Tatsuno, K., Covington, K.R., Ueda, H., Creighton, C.J., Kato, M., Tsuji, S., Donehower, L.A., Slagle, B.L., Nakamura, H., et al. (2014). Trans-ancestry mutational landscape of hepatocellular carcinoma genomes. Nat. Genet. *46*, 1267–1273. https://doi.org/10.1038/ng.3126.

29. Liu, Y., Sun, J., and Zhao, M. (2017). ONGene: A literature-based database for human oncogenes. J Genet Genomics *44*, 119–121. https://doi.org/10.1016/j.jgg.2016.12.004.

30. Zhao, M., Sun, J., and Zhao, Z. (2013). TSGene: a web resource for tumor suppressor genes. Nucleic Acids Res. *41*, D970–D976. https://doi.org/10.1093/nar/gks937.

31. Carter, S.L., Eklund, A.C., Kohane, I.S., Harris, L.N., and Szallasi, Z. (2006). A signature of chromosomal instability inferred from gene expression profiles predicts clinical outcome in multiple human cancers. Nat. Genet. *38*, 1043–1048. https://doi.org/10.1038/ng1861.

32. Sun, L., Wang, Y., Cen, J., Ma, X., Cui, L., Qiu, Z., Zhang, Z., Li, H., Yang, R.Z., Wang, C., et al. (2019). Modelling liver cancer initiation with organoids derived from directly reprogrammed human hepatocytes. Nat. Cell Biol. *21*, 1015–1026. https://doi.org/10.1038/s41556-019-0359-5.

33. Hanahan, D., and Weinberg, R.A. (2011). Hallmarks of cancer: the next generation. Cell *144*, 646–674. https://doi.org/10.1016/j.cell.2011.02.013.

34. Grivennikov, S.I., Greten, F.R., and Karin, M. (2010). Immunity, inflammation, and cancer. Cell *140*, 883–899. https://doi.org/10.1016/j.cell.2010.01.025.

35. Ringelhan, M., Pfister, D., O'Connor, T., Pikarsky, E., and Heikenwalder, M. (2018). The immunology of hepatocellular carcinoma. Nat. Immunol. *19*, 222–232. https://doi.org/10.1038/s41590-018-0044-z.

36. Taniguchi, K., and Karin, M. (2018). NF-kappaB, inflammation, immunity and cancer: coming of age. Nat. Rev. Immunol. *18*, 309–324. https://doi.org/10.1038/nri.2017.142.

37. Leone, V., Ali, A., Weber, A., Tschaharganeh, D.F., and Heikenwalder, M. (2021). Liver Inflammation and Hepatobiliary Cancers. Trends Cancer *7*, 606–623. https://doi.org/10.1016/j.trecan.2021.01.012.

38. Becht, E., Giraldo, N.A., Lacroix, L., Buttard, B., Elarouci, N., Petitprez, F., Selves, J., Laurent-Puig, P., Sautès-Fridman, C., Fridman, W.H., and de Reyniès, A. (2016). Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression. Genome Biol. *17*, 218. https://doi.org/10.1186/s13059-016-1070-5.

39. Hammerich, L., and Tacke, F. (2023). Hepatic inflammatory responses in liver fibrosis. Nat. Rev. Gastroenterol. Hepatol. *20*, 633–646. https://doi.org/10.1038/s41575-023-00807-x.

40. Montironi, C., Castet, F., Haber, P.K., Pinyol, R., Torres-Martin, M., Torrens, L., Mesropian, A., Wang, H., Puigvehi, M., Maeda, M., et al. (2023). Inflamed and non-inflamed classes of HCC: a revised immunogenomic classification. Gut *72*, 129–140. https://doi.org/10.1136/gutjnl-2021-325918.

41. Varanasi, S.K., Chen, D., Liu, Y., Johnson, M.A., Miller, C.M., Ganguly, S., Lande, K., LaPorta, M.A., Hoffmann, F.A., Mann, T.H., et al. (2025). Bile acid synthesis impedes tumor-specific T cell responses during liver cancer. Science *387*, 192–201. https://doi.org/10.1126/science.adl4100.

42. Cheng, S., Li, Z., Gao, R., Xing, B., Gao, Y., Yang, Y., Qin, S., Zhang, L., Ouyang, H., Du, P., et al. (2021). A pan-cancer single-cell transcriptional atlas of tumor infiltrating myeloid cells. Cell *184*, 792–809.e23. https://doi.org/10.1016/j.cell.2021.01.010.

43. Llovet, J.M., Castet, F., Heikenwalder, M., Maini, M.K., Mazzaferro, V., Pinato, D.J., Pikarsky, E., Zhu, A.X., and Finn, R.S. (2022). Immunotherapies for hepatocellular carcinoma. Nat. Rev. Clin. Oncol. *19*, 151–172. https://doi.org/10.1038/s41571-021-00573-2.

44. Johnson, D.E., O'Keefe, R.A., and Grandis, J.R. (2018). Targeting the IL-6/JAK/STAT3 signalling axis in cancer. Nat. Rev. Clin. Oncol. *15*, 234–248. https://doi.org/10.1038/nrclinonc.2018.8.

45. Balkwill, F. (2009). Tumour necrosis factor and cancer. Nat. Rev. Cancer *9*, 361–371. https://doi.org/10.1038/nrc2628.

46. Ramesh, V., Brabletz, T., and Ceppi, P. (2020). Targeting EMT in Cancer with Repurposed Metabolic Inhibitors. Trends Cancer *6*, 942–950. https://doi.org/10.1016/j.trecan.2020.06.005.

47. Singhal, A., Li, B.T., and O'Reilly, E.M. (2024). Targeting KRAS in cancer. Nat. Med. *30*, 969–983. https://doi.org/10.1038/s41591-024-02903-0.

48. Wang, H., Li, J., Jing, S., Lin, P., Qiu, Y., Yan, X., Yuan, J., Tang, Z., Li, Y., Zhang, H., et al. (2025). SOAPy: a Python package to dissect spatial architecture, dynamics, and communication. Genome Biol. *26*, 80. https://doi.org/10.1186/s13059-025-03550-5.

49. Cassetta, L., and Pollard, J.W. (2023). A timeline of tumour-associated macrophage biology. Nat. Rev. Cancer *23*, 238–257. https://doi.org/10.1038/s41568-022-00547-1.

50. Kloosterman, D.J., and Akkari, L. (2023). Macrophages at the interface of the co-evolving cancer ecosystem. Cell *186*, 1627–1651. https://doi.org/10.1016/j.cell.2023.02.020.

51. Helmink, B.A., Reddy, S.M., Gao, J., Zhang, S., Basar, R., Thakur, R., Yizhak, K., Sade-Feldman, M., Blando, J., Han, G., et al. (2020). B cells and tertiary lymphoid structures promote immunotherapy response. Nature *577*, 549–555. https://doi.org/10.1038/s41586-019-1922-8.

52. Petitprez, F., de Reyniès, A., Keung, E.Z., Chen, T.W.W., Sun, C.M., Calderaro, J., Jeng, Y.M., Hsiao, L.P., Lacroix, L., Bougoüin, A., et al. (2020). B cells are associated with survival and immunotherapy response in sarcoma. Nature *577*, 556–560. https://doi.org/10.1038/s41586-019-1906-8.

53. Cabrita, R., Lauss, M., Sanna, A., Donia, M., Skaarup Larsen, M., Mitra, S., Johansson, I., Phung, B., Harbst, K., Vallon-Christersson, J., et al. (2020). Tertiary lymphoid structures improve immunotherapy and survival in melanoma. Nature *577*, 561–565. https://doi.org/10.1038/s41586-019-1914-8.

54. Finkin, S., Yuan, D., Stein, I., Taniguchi, K., Weber, A., Unger, K., Browning, J.L., Goossens, N., Nakagawa, S., Gunasekaran, G., et al. (2015). Ectopic lymphoid structures function as microniches for tumor progenitor cells in hepatocellular carcinoma. Nat. Immunol. *16*, 1235–1244. https://doi.org/10.1038/ni.3290.

55. Nault, J.C., Ningarhari, M., Rebouissou, S., and Zucman-Rossi, J. (2019). The role of telomeres and telomerase in cirrhosis and liver cancer. Nat. Rev. Gastroenterol. Hepatol. *16*, 544–558. https://doi.org/10.1038/s41575-019-0165-3.

56. Zhu, M., Lu, T., Jia, Y., Luo, X., Gopal, P., Li, L., Odewole, M., Renteria, V., Singal, A.G., Jang, Y., et al. (2019). Somatic Mutations Increase Hepatic Clonal Fitness and Regeneration in Chronic Liver Disease. Cell *177*, 608–621.e12. https://doi.org/10.1016/j.cell.2019.03.026.

57. Brunner, S.F., Roberts, N.D., Wylie, L.A., Moore, L., Aitken, S.J., Davies, S.E., Sanders, M.A., Ellis, P., Alder, C., Hooks, Y., et al. (2019). Somatic mutations and clonal dynamics in healthy and cirrhotic human liver. Nature *574*, 538–542. https://doi.org/10.1038/s41586-019-1670-9.

58. Lorbeer, F.K., and Hockemeyer, D. (2020). TERT promoter mutations and telomeres during tumorigenesis. Curr. Opin. Genet. Dev. *60*, 56–62. https://doi.org/10.1016/j.gde.2020.02.001.

59. Davoli, T., Uno, H., Wooten, E.C., and Elledge, S.J. (2017). Tumor aneuploidy correlates with markers of immune evasion and with reduced response to immunotherapy. Science *355*, eaaf8399. https://doi.org/10.1126/science.aaf8399.

60. Taylor, A.M., Shih, J., Ha, G., Gao, G.F., Zhang, X., Berger, A.C., Schumacher, S.E., Wang, C., Hu, H., Liu, J., et al. (2018). Genomic and Functional Approaches to Understanding Cancer Aneuploidy. Cancer Cell *33*, 676–689.e3. https://doi.org/10.1016/j.ccell.2018.03.007.

61. Baslan, T., Morris, J.P., 4th, Zhao, Z., Reyes, J., Ho, Y.J., Tsanov, K.M., Bermeo, J., Tian, S., Zhang, S., Askan, G., et al. (2022). Ordered and deterministic cancer genome evolution after p53 loss. Nature *608*, 795–802. https://doi.org/10.1038/s41586-022-05082-5.

62. Karlsson, K., Przybilla, M.J., Kotler, E., Khan, A., Xu, H., Karagyozova, K., Sockell, A., Wong, W.H., Liu, K., Mah, A., et al. (2023). Deterministic evolution and stringent selection during preneoplasia. Nature *618*, 383–393. https://doi.org/10.1038/s41586-023-06102-8.

63. Burdett, N.L., and Christie, E.L. (2025). Genome doubling fuels ovarian cancer evolution and immune dysregulation. Nature *644*, 880–881. https://doi.org/10.1038/d41586-025-02059-y.

64. Jassim, A., Rahrmann, E.P., Simons, B.D., and Gilbertson, R.J. (2023). Cancers make their own luck: theories of cancer origins. Nat. Rev. Cancer *23*, 710–724. https://doi.org/10.1038/s41568-023-00602-5.

65. Gao, J., and Pickett, H.A. (2022). Targeting telomeres: advances in telomere maintenance mechanism-specific cancer therapies. Nat. Rev. Cancer *22*, 515–532. https://doi.org/10.1038/s41568-022-00490-1.

66. Viswanathan, S.R., Nogueira, M.F., Buss, C.G., Krill-Burger, J.M., Wawer, M.J., Malolepsza, E., Berger, A.C., Choi, P.S., Shih, J., Taylor, A.M., et al. (2018). Genome-scale analysis identifies paralog lethality as a vulnerability of chromosome 1p loss in cancer. Nat. Genet. *50*, 937–943. https://doi.org/10.1038/s41588-018-0155-3.

67. Neggers, J.E., Paolella, B.R., Asfaw, A., Rothberg, M.V., Skipper, T.A., Yang, A., Kalekar, R.L., Krill-Burger, J.M., Dharia, N.V., Kugener, G., et al. (2020). Synthetic Lethal Interaction between the ESCRT Paralog Enzymes VPS4A and VPS4B in Cancers Harboring Loss of Chromosome 18q or 16q. Cell Rep. *33*, 108493. https://doi.org/10.1016/j.celrep.2020.108493.

68. Borck, P.C., Boyle, I., Jankovic, K., Bick, N., Foster, K., Lau, A.C., Parker-Burns, L.I., Lubicki, D.A., Li, T., Borah, A.A., et al. (2025). SKI complex loss renders 9p21.3-deleted or MSI-H cancers dependent on PELO. Nature *638*, 1104–1111. https://doi.org/10.1038/s41586-024-08509-3.

69. Prindle, V., Richardson, A.E., Sher, K.R., Kongpachith, S., Kentala, K., Petiwala, S., Cheng, D., Widomski, D., Le, P., Torrent, M., et al. (2025). Synthetic lethality of mRNA quality control complexes in cancer. Nature *638*, 1095–1103. https://doi.org/10.1038/s41586-024-08398-6.

70. Huang, P., Zhang, L., Gao, Y., He, Z., Yao, D., Wu, Z., Cen, J., Chen, X., Liu, C., Hu, Y., et al. (2014). Direct reprogramming of human fibroblasts to functional and expandable hepatocytes. Cell Stem Cell *14*, 370–384. https://doi.org/10.1016/j.stem.2014.01.003.

71. Li, D., Fu, J., Du, M., Zhang, H., Li, L., Cen, J., Li, W., Chen, X., Lin, Y., Conway, E.M., et al. (2016). Hepatocellular carcinoma repression by TNFalpha-mediated synergistic lethal effect of mitosis defect-induced senescence and cell death sensitization. Hepatology *64*, 1105–1120. https://doi.org/10.1002/hep.28637.

72. Cable, D.M., Murray, E., Zou, L.S., Goeva, A., Macosko, E.Z., Chen, F., and Irizarry, R.A. (2022). Robust decomposition of cell type mixtures in spatial transcriptomics. Nat. Biotechnol. *40*, 517–526. https://doi.org/10.1038/s41587-021-00830-w.

73. Wolf, F.A., Angerer, P., and Theis, F.J. (2018). SCANPY: large-scale single-cell gene expression data analysis. Genome Biol. *19*, 15. https://doi.org/10.1186/s13059-017-1382-0.

74. Reynisson, B., Alvarez, B., Paul, S., Peters, B., and Nielsen, M. (2020). NetMHCpan-4.1 and NetMHCIIpan-4.0: improved predictions of MHC antigen presentation by concurrent motif deconvolution and integration of MS MHC eluted ligand data. Nucleic Acids Res. *48*, W449–W454. https://doi.org/10.1093/nar/gkaa379.

75. Wintersinger, J.A., Dobson, S.M., Kulman, E., Stein, L.D., Dick, J.E., and Morris, Q. (2022). Reconstructing Complex Cancer Evolutionary Histories from Multiple Bulk DNA Samples Using Pairtree. Blood Cancer Discov. *3*, 208–219. https://doi.org/10.1158/2643-3230.BCD-21-0092.

76. Hänzelmann, S., Castelo, R., and Guinney, J. (2013). GSVA: gene set variation analysis for microarray and RNA-seq data. BMC Bioinf. *14*, 7. https://doi.org/10.1186/1471-2105-14-7.

77. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 550. https://doi.org/10.1186/s13059-014-0550-8.

78. Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S.L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol. *14*, R36. https://doi.org/10.1186/gb-2013-14-4-r36.

79. Cibulskis, K., Lawrence, M.S., Carter, S.L., Sivachenko, A., Jaffe, D., Sougnez, C., Gabriel, S., Meyerson, M., Lander, E.S., and Getz, G. (2013). Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. Nat. Biotechnol. *31*, 213–219. https://doi.org/10.1038/nbt.2514.

80. Reble, E., Castellani, C.A., Melka, M.G., O'Reilly, R., and Singh, S.M. (2017). VarScan2 analysis of de novo variants in monozygotic twins discordant for schizophrenia. Psychiatr. Genet. *27*, 62–70. https://doi.org/10.1097/YPG.0000000000000162.

81. Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics *25*, 1754–1760. https://doi.org/10.1093/bioinformatics/btp324.

82. Ding, Z., Mangino, M., Aviv, A., Spector, T., and Durbin, R.; UK10K Consortium (2014). Estimating telomere length from whole genome sequence data. Nucleic Acids Res. *42*, e75. https://doi.org/10.1093/nar/gku181.

83. Ho, D.W.H., Sze, K.M.F., and Ng, I.O.L. (2015). Virus-Clip: a fast and memory-efficient viral integration site detection tool at single-base resolution with annotation capability. Oncotarget *6*, 20959–20963. https://doi.org/10.18632/oncotarget.4187.

84. Xi, R., Hadjipanayis, A.G., Luquette, L.J., Kim, T.M., Lee, E., Zhang, J., Johnson, M.D., Muzny, D.M., Wheeler, D.A., Gibbs, R.A., et al. (2011). Copy number variation detection in whole-genome sequencing data using the Bayesian information criterion. Proc. Natl. Acad. Sci. USA *108*, E1128–E1136. https://doi.org/10.1073/pnas.1110574108.

85. Blokzijl, F., Janssen, R., van Boxtel, R., and Cuppen, E. (2018). MutationalPatterns: comprehensive genome-wide analysis of mutational processes. Genome Med. *10*, 33. https://doi.org/10.1186/s13073-018-0539-0.

86. Wang, K., Li, M., and Hakonarson, H. (2010). ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. *38*, e164. https://doi.org/10.1093/nar/gkq603.

87. DePristo, M.A., Banks, E., Poplin, R., Garimella, K.V., Maguire, J.R., Hartl, C., Philippakis, A.A., del Angel, G., Rivas, M.A., Hanna, M., et al. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat. Genet. *43*, 491–498. https://doi.org/10.1038/ng.806.

88. Dai, M., Thompson, R.C., Maher, C., Contreras-Galindo, R., Kaplan, M.H., Markovitz, D.M., Omenn, G., and Meng, F. (2010). NGSQC: cross-platform quality analysis pipeline for deep sequencing data. BMC Genom. *11*, S7. https://doi.org/10.1186/1471-2164-11-S4-S7.

89. Letouzé, E., Shinde, J., Renault, V., Couchy, G., Blanc, J.F., Tubacher, E., Bayard, Q., Bacq, D., Meyer, V., Semhoun, J., et al. (2017). Mutational signatures reveal the dynamic interplay of risk factors and cellular processes during liver tumorigenesis. Nat. Commun. *8*, 1315. https://doi.org/10.1038/s41467-017-01358-x.

90. Vasaikar, S., Huang, C., Wang, X., Petyuk, V.A., Savage, S.R., Wen, B., Dou, Y., Zhang, Y., Shi, Z., Arshad, O.A., et al. (2019). Proteogenomic Analysis of Human Colon Cancer Reveals New Therapeutic Opportunities. Cell *177*, 1035–1049.e19. https://doi.org/10.1016/j.cell.2019.03.030.

91. Talevich, E., Shain, A.H., Botton, T., and Bastian, B.C. (2016). CNVkit: Genome-Wide Copy Number Detection and Visualization from Targeted DNA Sequencing. PLoS Comput. Biol. *12*, e1004873. https://doi.org/10.1371/journal.pcbi.1004873.

92. Kulman, E., Wintersinger, J., and Morris, Q. (2022). Reconstructing cancer phylogenies using Pairtree, a clone tree reconstruction algorithm. STAR Protoc. *3*, 101706. https://doi.org/10.1016/j.xpro.2022.101706.

93. Szolek, A., Schubert, B., Mohr, C., Sturm, M., Feldhahn, M., and Kohlbacher, O. (2014). OptiType: precision HLA typing from next-generation sequencing data. Bioinformatics *30*, 3310–3316. https://doi.org/10.1093/bioinformatics/btu548.

Please cite this article in press as: Zhang et al., Molecular insights into early malignant transition of hepatocellular carcinoma, Cancer Cell (2026), https://doi.org/10.1016/j.ccell.2026.03.005

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Mouse monoclonal anti-HePar-1	Agilent	Cat#IR624; RRID: N/A
Mouse monoclonal anti-CD4	ZSGB-BIO	Cat#ZM0418; RRID: N/A
Mouse monoclonal anti-CD8α	Cell Signaling Technology	Cat#70306; RRID: AB_2799781
Mouse monoclonal anti-PD-1	Cell Signaling Technology	Cat#43248; RRID: AB_2728836
Rabbit monoclonal anti-PD-L1	Cell Signaling Technology	Cat#13684; RRID: AB_2687655
Rabbit monoclonal anti-FOXP3	Cell Signaling Technology	Cat#98377; RRID: AB_2747370
Rabbit monoclonal anti-MPO	Abcam	Cat#ab208670; RRID: AB_2864724
Rabbit monoclonal anti-CD163	Cell Signaling Technology	Cat#93498; RRID: AB_2800204
Rabbit monoclonal anti-CD11c	Abcam	Cat#ab52632; RRID: AB_2129793
Rabbit monoclonal anti-ECHS1	Cell Signaling Technology	Cat#18420; RRID: N/A
Rabbit monoclonal anti-FGA	Abcam	Cat#ab92572; RRID: AB_10561758
Rabbit monoclonal anti-c-Myc	Abcam	Cat#ab32072; RRID: AB_731658
Mouse monoclonal anti-β-catenin	BD Biosciences	Cat#610154; RRID: AB_397555
Rabbit monoclonal anti-CD3	Abcam	Cat#ab16669; RRID: AB_443425
Mouse monoclonal anti-CD20	Agilent	Cat#M0755; RRID: AB_2282030
Mouse monoclonal anti-CD21	Leica Biosystems	Cat#NCL-CD21-2G9; RRID: AB_442056
Mouse monoclonal anti-HSP70	Santa Cruz Biotechnology	Cat#sc-24; RRID: AB_627760
Rabbit polyclonal anti-GS	Abcam	Cat#ab49873; RRID: AB_880241
Mouse monoclonal anti-GPC3	GeneTex	Cat#GTX633410; RRID: AB_2888342
Biological samples		
Nodule-in-nodule samples	Department of Pathology, Zhongshan Hospital, Fudan University; Department of Pathology, Eastern Hepatobiliary Surgery Hospital, Naval Medical University; Department of Pathology, Huashan Hospital, Fudan University	N/A
Evolutionarily unrelated DN and veHCC samples	Department of Pathology, Zhongshan Hospital, Fudan University; Department of Pathology, Eastern Hepatobiliary Surgery Hospital, Naval Medical University; Department of Pathology, Huashan Hospital, Fudan University	N/A
Chemicals, peptides, and recombinant proteins		
Recombinant human TGFα	PeproTech	Cat#100-16A
Recombinant human EGF	PeproTech	Cat#AF-100-15
Dexamethasone	Sigma-Aldrich	Cat#D4902
ITS-A	Gibco	Cat#51300-044
NeuroCult™ SM1	Stem cell	Cat#05711
Critical commercial assays		
Click-iT EdU Cell Proliferation Kit	Biyuntian Biotechnology	Cat#C0081S
Opal 4-color Manual IHC Kit	Akoya Biosciences	Cat#NEL810001KT
Pano 7-plex IHC Kit	Panovue	Cat#004100100

(*Continued on next page*)

Please cite this article in press as: Zhang et al., Molecular insights into early malignant transition of hepatocellular carcinoma, Cancer Cell (2026), https://doi.org/10.1016/j.ccell.2026.03.005

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
DNA/RNA sequencing data	This paper	GSA-Human: HRA008950
Spatial Transcriptomic data	This paper	GSA-Human: HRA008950
WES and RNA-seq data from TCGA	TCGA	https://www.cbioportal.org/
WGS and RNA-seq data from the Chinese Liver Cancer Atlas (CLCA)	Chen et al.[23]	https://ngdc.cncb.ac.cn/databasecommons/database/id/7406
Experimental models: Cell lines		
Human induced hepatocytes (hiHeps)	Huang et al.[70]	N/A
Oligonucleotides		
sgRNA target sequence: *ECHS1*#1: GCGCACTCACCCGAGGCGAA	This paper	N/A
sgRNA target sequence: *ECHS1*#2: TATCATCTATGCCGGTGAGA	This paper	N/A
sgRNA target sequence: *ECHS1*#3: ACCCGAGGCGAAGGGACGCC	This paper	N/A
sgRNA target sequence: *FGA*#1: AGGAGGCGTGCGTGGCCCAA	This paper	N/A
sgRNA target sequence: *FGA*#2: GGGCTGTTCGCATGAAAATT	This paper	N/A
sgRNA target sequence: *FGA*#3: TTTCGGCAGCCTGAATGACG	This paper	N/A
sgRNA target sequence: *NR0B2*#1: CGGGCCGGTGCTGCCTACAT	This paper	N/A
sgRNA target sequence: *NR0B2*#2: CCCGTAGCCGCTGCCTATGT	This paper	N/A
sgRNA target sequence: *NR0B2*#3: CCGTGAGGAGGACACGGGTC	This paper	N/A
sgRNA target sequence: *SDHB*#1: GTCTGGGTCCCATCGATAGA	This paper	N/A
sgRNA target sequence: *SDHB*#2: TCGCCCTCTCCTTGAGGCGC	This paper	N/A
sgRNA target sequence: *SDHB*#3: GGCCGGCAACCGGCGCCTCA	This paper	N/A
sgRNA target sequence: *FGB*#1: GTGATTCAGAACCGTCAAGA	This paper	N/A
sgRNA target sequence: *FGB*#2: AGTCAACACTACCGTCTTGA	This paper	N/A
sgRNA target sequence: *FGB*#3: TAGGGGGTGTTGTGTCCTAC	This paper	N/A
shRNA target sequence: *FGA*: CCTCAGCCAATAACCGTGATA	This paper	N/A
Non-targeting sequence: CAACAAGATGAAGAGCACCAA	Sigma-Aldrich	Cat#SHC002
Primer: *FGA* Forward: GGACAATGGCACTCTGAATCT	This paper	N/A
Primer: *FGA* Reverse: GTGACCATCAGGACCAATAACA	This paper	N/A
Primer: *SMO* Forward: GTCGGGCCTCCGGAATG	This paper	N/A
Primer: *SMO* Reverse: TTGTCTGTCCGAACCAAGGG	This paper	N/A
Primer: *REPIN1* Forward: TTCTCTAACACCTGGGGGCT	This paper	N/A

(*Continued on next page*)

Please cite this article in press as: Zhang et al., Molecular insights into early malignant transition of hepatocellular carcinoma, Cancer Cell (2026), https://doi.org/10.1016/j.ccell.2026.03.005

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Primer: *REPIN1* Reverse: ACGTTCCAGCATCGGTTCTT	This paper	N/A
Recombinant DNA		
pWPI-mouse-ΔN90-β-catenin	Li et al.[71]	N/A
pWPI-human-*c-Myc*	Li et al.[71]	N/A
pCDH-human-*SMO*	This paper	N/A
pCDH-human-*REPIN1*	This paper	N/A
Software and algorithms		
lme4 R package	N/A	https://CRAN.R-project.org/package=lme4
InForm image analysis software	Akoya Biosciences	https://investors.akoyabio.com/
Akoya Vectra 3	Akoya Biosciences	https://investors.akoyabio.com/
SOAPy (v1.0.1)	Wang et al.[48]	https://soapy-st.readthedocs.io/en/latest/
Spacexr (v2.2.1)	Cable et al.[72]	https://github.com/dmcable/spacexr
Space Ranger software (v4.0.1)	10x Genomics	https://www.10xgenomics.com/support/software/space-ranger/latest
Scanpy (v1.9.1) Python package	Wolf et al.[73]	https://github.com/scverse/scanpy
MCP-counter	Becht et al.[38]	https://github.com/ebecht/MCPcounter
NetMHCpan-4.1	Reynisson et al.[74]	https://services.healthtech.dtu.dk/services/NetMHCpan-4.1/
Pairtree	Wintersinger et al.[75]	https://github.com/morrislab/pairtree
GSVA R package (v1.38.2)	Hanzelmann et al.[76]	https://bioconductor.org/packages//release/bioc/html/GSVA.html
DESeq2 R package	Love et al.[77]	https://bioconductor.org/packages//release/bioc/html/DESeq2.html
TopHat2	Kim et at.[78]	https://ccb.jhu.edu/software/tophat/index.shtml
MuTect	Cibulskis et al.[79]	https://github.com/broadinstitute/mutect
VarScan 2	Reble et al.[80]	https://github.com/Jeltje/varscan2
Burrows-Wheeler Aligner (BWA) algorithm	Li et al.[81]	https://github.com/lh3/bwa
TelSeq	Ding et al.[82]	https://github.com/zd1/telseq
Virus-Clip	Ho et al.[83]	https://github.com/dwhho/Virus-Clip
BIC-Seq	Xi et al.[84]	https://dspace.mit.edu/handle/1721.1/69934
MutationalPatterns R package	Blokzijl et al.[85]	https://bioconductor.org/packages//release/bioc/html/MutationalPatterns.html
ANNOVAR	Wang et al.[86]	https://annovar.openbioinformatics.org/en/latest/
Genome Analysis Toolkit (GATK)	DePristo et al.[87]	https://gatk.broadinstitute.org/hc/en-us

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Patient cohort

Samples were collected from the Department of Pathology, Zhongshan Hospital, Fudan University; the Department of Pathology, Eastern Hepatobiliary Surgery Hospital, Naval Medical University; and the Department of Pathology, Huashan Hospital, Fudan University (China). This study was approved by the Ethics Committees of all three institutions.

To identify coexistent high-grade dysplastic nodules (DNs) and very early hepatocellular carcinomas (veHCCs) with a nodule-in-nodule pattern, more than 44,714 formalin-fixed paraffin-embedded (FFPE) liver specimens derived from the surgical resection of liver cancers performed within a 9-year interval (2014–2022) were screened. To diagnose high-grade DNs and veHCCs, pathological reviews were independently performed by three expert pathologists for each case. The following histological features were quoted: mean nodule size, steatosis, unpaired arteries, portal tract, stromal invasion, hepatic plate thickness, inflammatory infiltrate, ductular reaction, cytological atypia, pseudoglandular formation, cholestasis, clear-cell changes, ballooning, hyaline inclusions, Mallory bodies, and large- and small-cell changes. Diagnosis was based on WHO-defined histological criteria for hepatocellular neoplasia and followed a published method.[6,14] In cases of diagnostic difficulty, pathologists performed immunostaining for GPC3, HSP70,

Please cite this article in press as: Zhang et al., Molecular insights into early malignant transition of hepatocellular carcinoma, Cancer Cell (2026), https://doi.org/10.1016/j.ccell.2026.03.005

and GS, with positivity for at least two markers considered sufficient for early HCC diagnosis. Owing to the very early stage of these HCCs, we classified these nodule-in-nodule HCCs as veHCCs. In total, 17 evolutionarily related nodule-in-nodule lesions from 16 patients (one lesion from P3, P5, P6, P8–P14, and P16-–P20, and two spatially separated lesions from P7) were collected.

These nodule-in-nodule NCas, high-grade DNs and their paired veHCCs were obtained by microdissection and subjected to whole-genome sequencing (WGS, ~35 ×) and RNA-seq. For tissue sampling, we implemented a rigorous multistep sampling protocol to obtain DN and veHCC samples. Briefly, we followed three steps: (1) expert guidance: all microdissections were confirmed by pathologists with extensive experience in liver neoplasia; (2) spatial separation: the DN regions were selected at the maximum feasible distance from the nearest veHCC focus. In all cases, DN sampling was performed at least 1 mm from the veHCC; and (3) morphological exclusion: we carefully avoided areas with morphological evidence of tumor invasion or desmoplastic stromal reactions adjacent to the tumor.

Four cases (P1, P2, P4, and P15), although spatially overlapping, exhibited a <1% overlap in SNVs and no shared copy number alterations (CNAs) between DNs and veHCCs. Furthermore, we confirmed that no shared CFG mutations were identified between the veHCC and its paired DN in 3 cases. P15_DN shared only the hotspot mutation in the TERT promoter (chr5:1295228:G>A) with its paired veHCC, with no other shared CFG mutations. Collectively, these data indicate that the DNs and veHCCs from these four cases arose independently. We therefore classified these four DNs as cancer-undetermined DNs; together with an additional 15 high-grade DNs collected, we obtained a final cohort of 19 cancer-undetermined DNs. Fifteen individual veHCC samples were collected from 15 patients. These evolutionarily unrelated DNs and veHCCs were subjected to whole-exome sequencing (WES, ~230×), target sequencing (e.g., *TERT* promoter, *CTNNB1*, *TP53,* and *ARID1A*) (~1880×), and RNA-seq. CNAs and protein-altering SNVs were identified by analyzing WES data, while *TERT* promoter alterations (promoter mutations and HBV integration) were further identified by analyzing panel sequencing data.

All patients had hepatitis B virus (HBV) infection and developed cirrhosis, with ages ranging from 32 to 72 years. Detailed information on tissue coding and clinicopathological data is summarized in Tables S1A and S1B.

METHOD DETAILS

Functional studies using *in vitro* organoid models

Cell culture

Human induced hepatocytes (hiHeps) were generated from immortalized umbilical cord fibroblasts via overexpression of three transcription factors, FOXA3, HNF1A, and HNF4A.[70] Immortalization of the umbilical cord fibroblasts was achieved by overexpressing the SV40 large T (SV40LT) antigen. The resulting hiHep cells were cultured in hepatocyte-maintaining medium (HMM) as previously described.[70]

Lentiviral transfection in hiHep cells

In this study, we functionally characterized two distinct sets of genes identified in our study. The first set comprised genes with copy number gains, whose elevated expression was further validated using both our RNA-seq data and RNA-seq datasets retrieved from The Cancer Genome Atlas (TCGA). For these upregulated genes, we constructed overexpression plasmids by cloning the full-length cDNA sequences of target genes into the pCDH vector, which harbors a CMV promoter.

The second set included genes with copy number losses, whose reduced expression was validated via the same RNA-seq datasets (this study and TCGA). To achieve loss-of-function manipulation, we designed three distinct single-guide RNAs (sgRNAs) for each target gene and subcloned each sgRNA into the lentiCRISPR v2 vector (Addgene, #52961). To enhance knockout efficiency, the three sgRNAs were co-packaged into a single lentivirus.

Notably, for the functional validation of upregulated genes, we directly overexpressed these genes in hiHep cells and evaluated their capacity to induce organoid overgrowth. Conversely, for the loss-of-function gene validation, we performed targeted knockout of these genes in β-catenin$^{\Delta ex3}$-overexpressing organoids, an oncogene-driven organoid model, to investigate whether loss of these genes could facilitate the enlargement of β-catenin$^{\Delta ex3}$-overexpressing organoids. As demonstrated in our present study and previously published work, activating mutations in *CTNNB1* occur at the early stage of liver tumorigenesis.[5] Overexpression of β-catenin$^{\Delta ex3}$ was performed to mimic such activating mutations in *CTNNB1*.

Lentiviral plasmids carrying the target gene sequences or sgRNAs were transfected into 293FT cells together with the packaging plasmid psPAX2 (Addgene) and envelope plasmid pMD2.G (Addgene) for virus production. HiHep cells were infected with lentivirus for 16-24 hours, followed by a 24-48-hour recovery period prior to selection. Infected cells were selected using 10 μg/mL puromycin until all uninfected control cells had died. For the knockout of genes, hiHep cells were first transduced with lentiviruses to overexpress β-catenin$^{\Delta ex3}$,[71] subsequently, the corresponding genes were knocked out in these β-catenin$^{\Delta ex3}$-overexpressing cells via lentiviral transduction.

Culture of organoids and quantification of organoid size

HiHep cells successfully infected with lentivirus were trypsinized into single-cell suspensions. After cell counting, the cells were resuspended to a density of 3.75-5 × 10^4 cells/mL, then mixed with an equal volume of Matrigel. The Matrigel-cell mixture was plated at 40 μL per well in 24-well plates. Following solidification of the Matrigel for 20-30 minutes, 1 mL of HMM medium was added to each well. HiHep organoids were cultured for 6-10 days, and organoid size was automatically imaged and quantitatively analyzed using the CBM system (BioTek, Cytation 5).

Please cite this article in press as: Zhang et al., Molecular insights into early malignant transition of hepatocellular carcinoma, Cancer Cell (2026), https://doi.org/10.1016/j.ccell.2026.03.005

Click-iT EdU incorporation assay

The Click-iT EdU assay was performed in accordance with the manufacturer's instructions for the Click-iT EdU Cell Proliferation Kit (Biyuntian Biotechnology). Organoids maintained in HMM medium were incubated with 10 μM EdU for 30 minutes. The percentage of EdU-positive cells was quantified by normalizing the number of EdU-positive cells to the number of DAPI-positive cells, using the Opera Phenix Plus high-content imaging system and Harmony software (v5.3).

Preparation of tissues for sequencing

After acquiring the FFPE blocks of the lesions from the tissue bank, serial sections were prepared at a thickness of 5 μm from each lesion for hematoxylin and eosin (H&E) staining and immunostaining procedures. For nodule-in-nodule lesions, a thorough pathological review was conducted on each lesion section to delineate the areas corresponding to DNs and veHCCs. For individual DN or veHCC lesions, a similar pathological review was performed to accurately identify the respective areas of DNs or veHCCs. Subsequently, DN and veHCC tissues were meticulously harvested through microdissection. Additionally, NCas in each case were obtained via microdissection to enrich the parenchymal cells instead of removing the fibrous bands.

The harvested tissues were transferred to a 1.5 mL microcentrifuge tube, and deparaffinization solution was used to remove the paraffin. The Maxwell® 16 LEV RNA FFPE Purification Kit (Promega) was used to extract FFPE RNA, according to the manufacturer's instructions. RNA integrity was determined using a 2100/2200 Bioanalyzer (Agilent) with DV_{200} (percentage of RNA fragments >200 nt fragment distribution value) and quantified using a NanoDrop (Thermo Fisher Scientific). For FFPE DNA extraction, the Maxwell® 16 FFPE Plus LEV DNA Purification Kit (Promega) was used according to the manufacturer's instructions. The integrity and concentration of the total DNA were determined using agarose electrophoresis and a Qubit 3.0 fluorometer dsDNA HS Assay (Thermo Fisher Scientific).

Whole genome sequencing and analysis

Experiment

The DNA was sheared to an average target size of 350 bp using Covaris S220 Sonicator (Covaris). Fragmented DNA was purified using Sample Purification Beads (Illumina). Adapter-ligated libraries were prepared with the TruSeq Nano DNA Sample Prep Kits (Illumina) according to protocol instructions (Illumina).

DNA concentration of the enriched sequencing libraries was measured by the Qubit 3.0 fluorometer dsDNA HS Assay (Thermo Fisher Scientific). Size distribution of the resulting sequencing libraries was analyzed using Agilent Bioanalyzer 4200 (Agilent). Paired-end sequencing is performed on the Illumina HiSeq X TEN or NovaSeq 6000 system with 2x150 bp paired-end reads following protocol instructions (Illumina).

Preprocessing

Raw WGS reads were processed using an in-house script to remove low-quality reads and trim adapters.[88] The remaining reads were mapped to the human genome (hg19) using the BWA algorithm.[81] Duplicate reads were removed using Picard tools. SNVs and indels were identified and recalibrated using the Genome Analysis Toolkit (GATK).[87] Variants were hard-filtered using the parameters recommended by the GATK Best Practices. Similarity between paired samples was calculated based on the Jaccard similarity coefficient between the germline variants. Hierarchical clustering analysis was performed on all samples to determine their sources.

Somatic mutations

Somatic mutations were called using MuTect[79] or VarScan2[80] by comparing DNs or veHCCs with their matched NCas. The initial mutation calls were filtered using the recommended criteria for each algorithm. To obtain more confident somatic mutations, single nucleotide variations (SNVs) identified by the two algorithms were first retained and designated as "core mutations". ANNOVAR[86] was used to map mutations to proteins in order to estimate their potential functional effects. Nonsynonymous, stop-gain, stop-loss and splicing-related SNVs were regarded as protein-altering SNVs. Since multiple samples from the same patient are more likely to share mutations, the "core mutations" detected in one sample may also be genuine in another sample from the same patient even if they did not pass the filtering criteria of the two algorithms. To avoid miss true protein-altering SNVs, if a certain protein-altering SNV passed the filtering criteria in one sample of a patient, it was also be retained in other samples if it had mutated reads but not passed the criteria.

Mutation signature analysis

The mutation signatures were analyzed using the R package MutationalPatterns.[85] Somatic SNVs were classified into 96 possible mutation types based on the sequence context of mutated base (6 types of substitution × 4 types of 5′ base × 4 types of 3′ base). The mutation matrix was reconstructed using an optimal linear combination of known mutational signatures that occur frequently in liver cancers.[89] Only signatures with contributions >10% were plotted, and the remaining signatures were categorized as "Other." Associations between the mutation signatures and exogenous and endogenous risk factors were obtained from the COSMIC database.

Copy number alterations

Somatic copy number alterations (CNAs) were identified using BIC-Seq[84] software by comparing DNs or veHCCs with their matched NCas. BIC-seq returned the log2 ratios between the observed and expected number of reads in each segment. Segments with log2 ratio >0.2 or <−0.2 were regarded as somatic CNAs. For arm-level CNAs, we defined them as those in which the loss or gain fragments account for more than 60% of the total length of the corresponding chromosome arm.[7] Whole-chromosome gains and losses were included in this category. We specified such alterations as "arm-level CNAs". When describing CNAs that affect the CFG, we

Please cite this article in press as: Zhang et al., Molecular insights into early malignant transition of hepatocellular carcinoma, Cancer Cell (2026), https://doi.org/10.1016/j.ccell.2026.03.005

denoted them as "CNA (CFG)", for instance, Chr17p (*TP53*) loss. If a CNA occurs in the same arm as a CFG but does not span the gene, it is not counted as a CFG-harboring CNA.

Chromosome instability (CIN) index for each chromosome in each sample was inferred from the segmentation data using a weighted-sum approach in which the absolute values of the log2 ratios of all segments within a chromosome were weighted by the segment length and summed up.[90] Oncogenes and tumor suppressor genes were downloaded from https://bioinfo-minzhao.org/ongene/index.html and https://bioinfo.uth.edu/TSGene1.0, respectively.

HBV integration

The genome sequence of HBV was downloaded from the National Center for Biotechnology Information (NC_003977.2). HBV integration breakpoints were identified and annotated using Virus-Clip.[83] The sequencing reads were aligned with the HBV genome. Soft-clipped reads were then extracted to identify human and HBV integration breakpoints.

Telomere length

Telomere repeat sequences (TTAGGG repeats) were obtained from the WGS alignment files and converted to telomere length in kilobases by TelSeq.[82]

Genomic alterations of CFGs

In this study, 37 CFGs were collected based on the significantly mutated HCC genes identified in eight genome sequencing studies.[7,22–28] This included the following 19 genes reported in more than two studies: *TERT*, *TP53*, *CTNNB1*, *AXIN1*, *ALB*, *BAP1*, *KEAP1*, *NFE2L2*, *RB1*, *PIK3CA*, *RPS6KA3*, *KRAS*, *IL6ST*, *CDKN2A*, *ARID1A*, *ARID2*, *ACVR2A*, *APOB,* and *NRAS*. Additionally, 12 genes that were frequently mutated in HBV-related HCCs[23] were identified: *TSC1*, *TSC2*, *JAK1*, *BRD7*, *FGA*, *PTEN*, *HNF1A*, *PRDM11*, *CDKN1B*, *BMP5*, *RPL22,* and *ECHS1*. We also included six additional genes that play important roles in HCC development: *MET*, *EGFR*, *APC*, *MDM4*, *MYC,* and *ARID1B*. Table S6A summarizes the list of CFGs. CFG mutations identified by either MuTect[79] or VarScan2[80] were manually checked to avoid missing mutations with low scores, and single nucleotide polymorphisms present in the dbSNP (build 138) database or frequently occurring in the population were excluded. In addition to somatic mutations, CNAs and HBV integration were also analyzed for these CFGs. These different types of genomic alterations were visualized using an OncoPrint plot.

Whole exome sequencing and panel sequencing

WES libraries were prepared and captured using the SureSelectXT Human All Exon V8 kit (35.1 Mb, Agilent Technologies), following the manufacturer's instructions. The DNA library was sequenced using the Illumina NovaSeq 6000 system. After removing low-quality reads and trim adapters,[88] the remaining reads were mapped to the human genome (hg19) using the BWA algorithm.[81] The WES files were preprocessed using the same pipeline as WGS and mapped to the exon kit. The protein-altering SNVs identified by either MuTect[79] or VarScan2[80] were manually checked and a downstream filter comprised of the following criteria was used to obtain high quality SNVs: (1) coverage >8X; (2) variant allele frequency (VAF) $\geq$ 5%, with at least four variant supporting reads in the DN or veHCC samples, and VAF <1%, with no more than 2 variant supporting reads in NCas; (3) single nucleotide polymorphisms identified in the dbSNP (build 138) were removed. CNAs were identified by comparing veHCC or DN samples with their paired NCas using CNVkit.[91]

A gene panel was designed to encompass important liver cancer functional genes (*TP53*, *CTNNB1*, *ARID1A*, and the *TERT* promoter). Probes targeting these genes were designed using Agilent Technologies. Probes designed to target the HBV genome were used to detect HBV integration into *TERT* gene (Table S2B). DNA samples of 10–200 ng was fragmented using Agilent's SureSelect Enzymatic Fragmentation Kit and subsequently used for library construction following the manufacturer's protocol. Targeted libraries were submitted for sequencing, achieving a mean depth of 1880X. Raw data processing involved the removal of low-quality reads and trimming of adapters.[88] The remaining reads were mapped to the human genome (hg19) using the BWA algorithm.[81] SNVs were identified by MuTect[79] or VarScan2.[80] Protein-altering SNVs identified using either of the two tools were checked manually.

Notably, 19 cancer-undetermined DNs in this study were subjected to both WES and panel sequencing. CNAs and protein-altering SNVs were identified by analyzing WES data, while *TERT* promoter alterations (promoter mutations and HBV integration) were further identified by analyzing panel sequencing data.

RNA sequencing and analysis

Experiment

RNA purification, reverse transcription, library construction, and sequencing were performed according to the manufacturer's instructions (Illumina). The captured coding regions of the transcriptome from total RNA were prepared using the TruSeq® RNA Exome Library preparation Kit. For FFPE samples, RNA input for library construction was determined based on RNA quality. Generally, 20 ng of RNA is recommended for high-quality RNA and 20–40 ng RNA for medium-quality RNA. cDNA was then generated from the input RNA fragments using random priming during first- and second-strand synthesis, and sequencing adapters were ligated to the resulting double-stranded cDNA fragments. To create the final library, coding regions of the transcriptome were captured using sequence-specific probes. After library construction, a Qubit 3.0 fluorometer dsDNA HS Assay (Thermo Fisher Scientific) was used to quantify the concentration of the resulting sequencing libraries, and the size distribution was analyzed using an Agilent Bioanalyzer 2100 (Agilent). Paired-end 150 bp sequencing was performed using a NovaSeq 6000 S4 sequencer following the manufacturer's instructions (Illumina) in Mingma Technologies, Shanghai, China. RNA-seq was not performed on P6_veHCC and P17_veHCC because of insufficient tissue availability.

Please cite this article in press as: Zhang et al., Molecular insights into early malignant transition of hepatocellular carcinoma, Cancer Cell (2026), https://doi.org/10.1016/j.ccell.2026.03.005

Preprocessing

The raw reads were first trimmed to remove low-quality bases and adapter sequences. High-quality reads were mapped to the human GRCh38 genome using the TopHat2 algorithm.[78] Counts of individual transcripts were quantified using FeatureCounts. For the PCA analysis of all evolutionarily related samples, batch effects of the counts were removed using the function ComBat_seq in the R package 'sva' and then transformed using 'vst' from DESeq2.[77] PlotPCA in DESeq2 was used to extract the principal components with the top 500 variable genes.

Differential expression analysis

Differential analysis across groups was performed using DESeq2 with count data as input, and the batch and individual information were included in the design to compensate for the batch effect and individual differences. Genes with the absolute change >2 and adjust *P*-value <0.05 were chosen as significant differentially expressed genes (DEGs). The DEGs were clustered into groups with similar expression patterns using the degPatterns function from the DEGreport package.

Pathway activity and enrichment analysis

Pathway activity was estimated using the gene set variation analysis (GSVA).[76] Hallmark cancer pathways (Figure S5D) and DNA repair (Figures S3N–S3P), reactome metabolism (Figures S5G and S5H), and inflammation-related gene sets (Figure 3B) and Gene Ontology (GO)-negative immune regulation (Figure 4K) were exported from the MSigDB website. Immune evasion-related signatures, including signatures of exhausted T cells (Figure 4L), immune checkpoints (Figure 4M), Tregs (Figure 4N), and TGF-β (Figure S6M) were obtained from previously published studies (Table S7). A list of 70 CIN-associated signature genes[31] (Figure S3L) and a chromosome breakage-related gene set (Figure S3M) from the Human Phenotype Ontology Database were analyzed using pre-ranked gene set enrichment analysis.

Estimation abundance of different cell populations

Relative proportion (abundance) of immune cells and stromal cells was estimated by Microenvironment Cell Populations-counter (MCP-counter)[38] from the gene expression matrix (transcripts per million, TPM). MCP-counter produces for each sample an abundance score for $CD3^+$ T cells, $CD8^+$ T cells, cytotoxic lymphocytes, NK cells, B lymphocytes, cells originating from monocytes (monocytic lineage), myeloid dendritic cells, neutrophils, as well as endothelial cells and fibroblasts.

Comparability of genomic and RNA-sequencing data analysis

For annotation, we used RefSeq gene models harmonized by gene symbols across builds, ensuring comparability when mapping mutations via ANNOVAR and when summarizing expression data. For germline filtering, we applied dbSNP databases corresponding to hg19 for WGS/WES/panel sequencing. RNA-seq analysis did not involve germline filtering. Finally, any cross-reference between mutation and expression data was performed at the gene symbol level, eliminating potential biases from different genome builds.

Evolutionary scenarios

The estimated immune cell abundance and activation score of an inflamed signature stratified veHCCs into two distinct subgroups: an inflamed subgroup and a non-inflamed subgroup. We next compared CNAs (CIN index) between these two subgroups and found that non-inflamed veHCCs harbored a significantly higher CIN index than their inflamed counterparts. Furthermore, non-inflamed veHCCs exhibited a marked increase in CNAs during the transition from DNs to veHCCs. To further delineate the evolutionary trajectories of multiple samples derived from individual patients, genome-wide somatic mutations were subjected to Pairtree[75,92] for subclonal clustering and phylogenetic tree construction. By comprehensively integrating CNA profiles of tumors, immune cell infiltration patterns in the tumor microenvironment, and computationally inferred phylogenetic trees, we summarized the evolutionary process from DNs to veHCCs into two major scenarios: CNA-dominant malignant transition (Scenario 1) and inflamed phenotype-associated progression coupled with immune evasion (Scenario 2).

Neoantigen predictions

To examine the neo-epitopes, 8–11 mer peptides around the mutated sites were extracted. HLA genotyping was estimated based on RNA-seq data for all major and minor HLA Class I alleles using OptiType.[93] The binding affinities of these neo-epitopes to the corresponding HLA-A, HLA-B, and HLA-C alleles were estimated using NetMHCpan-4.1.[74] Neoantigens with a relative percentile rank $\leq$0.5% were classified as strong binders, while neoantigens with a percentile $\leq$2% were classified as weak binders.

Comparative analysis of genomic and histological features across veHCC subgroups

The differences in CIN index (veHCC−DN) were fitted using a two-component normal mixture distribution. The first component indicates minimal differences in CIN index between DNs and veHCCs, while the second component indicates significantly higher CIN indexes in veHCCs compared with DNs. veHCCs were stratified into two groups based on the mean plus one standard deviation of the first component (0.0266). Group 1 comprised 12 veHCCs, which exhibited comparable or slightly increased CIN indexes relative to their matched DNs (mean CIN index (veHCC−DN) = 0.002±0.016). Group 2 consisted of 9 veHCCs, the CIN indexes of which were significantly higher than that of their matched DNs (mean CIN index (veHCC−DN) = 0.072 ± 0.035). We compared the genomic, histological, and clinical characteristics between these two groups, and detailed information is provided in Table S5.

Please cite this article in press as: Zhang et al., Molecular insights into early malignant transition of hepatocellular carcinoma, Cancer Cell (2026), https://doi.org/10.1016/j.ccell.2026.03.005

Comparative analysis of somatic SNVs between FFPE and fresh-frozen samples

Somatic SNVs were identified using muTect and VarScan2, and results that passed the threshold criteria of both algorithms were regarded as true somatic SNVs. These SNVs were then used to calculate the number of overlapping SNVs between FFPE and fresh-frozen samples. For SNVs unique to either FFPE or fresh-frozen samples, we further examined the corresponding BAM files for the presence of mutant reads; SNVs with no mutant reads detected at all were defined as truly unique SNVs.

Comparison with later-stage HCC

To investigate whether the genetic variants identified in veHCC differ from those in later-stage HCC, and how their frequencies vary across these two groups, we utilized two additional datasets of later-stage HCC: (1) WES and RNA-seq data from 363 tumor samples retrieved from The Cancer Genome Atlas (TCGA)[22]; and (2) WGS and RNA-seq data from 494 tumor samples obtained from the Chinese Liver Cancer Atlas (CLCA), a cohort of Chinese patients whose genetic background and viral status are more consistent with those of our study cohort.[23] TCGA data were downloaded from the cBioPortal database, while CLCA data were acquired from its official website. CNA segments of each sample were subjected to GISTIC analysis to identify genomic regions with significant amplification or deletion across the cohort. Immune cell abundance was estimated using MCP-counter based on the normalized TPM expression matrices of all datasets. Given that WES does not adequately capture the promoter regions of genes, we specifically compared *TERT* alterations between our dataset and the CLCA cohort.

Spatial transcriptomic sequencing and analysis

Experiment

The FFPE samples were sent to Shanghai OE Biotech Co., Ltd. (Shanghai, China) for spatial transcriptomic library preparation and sequencing. Briefly, FFPE sections of 5 μm thickness were prepared using a Leica RM2235 Manual Rotary Microtome (Leica Microsystems, Germany). The sections were attached to the slides (Sigma-Aldrich, cat.no. P0425) and then deparaffinization, HE Staining, Imaging and decrosslinking of the sections were performed according to 10x Genomics protocol (CG000684). Probe hybridization, probe release and library construction were performed using the 10x Genomics Visium HD Slides (10x Genomics, cat.no. PN-2000970) and Visium HD Spatial Gene Expression Reagent kit (10x Genomics, PN-1000675 for Human) according to the User Guide (CG000685). The libraries were sequenced on the BGI DNBSEQ-T7 sequencing platform using PE100 mode.

Data processing

The FASTQ files were processed and aligned to GRCh38 human reference genome using Space Ranger software (v4.0.1) from 10x Genomics, with unique molecular identifier (UMI) counts summarized for each barcode. In addition to the native 2 μm feature size, Space Ranger outputs Visium HD data binned at 8 μm and 16 μm resolutions. Unless otherwise specified, downstream analyses were performed on the 16 μm resolution data, which provides a 64-fold increase in the average number of UMI reads per bin in the gene-barcode matrix compared to a 2 μm square. The filtered UMI count matrix was then analyzed using scanpy (v1.9.1) Python package.[73] To obtain the normalized gene expression data, library size normalization was processed using the normalize_total and log1p function.

Deconvolution

Deconvolution methods aim to identify the cell types and their relative proportions that contribute to the signal captured in a given area (spot, square, or bin). Here, we applied spot deconvolution to classify and label each bin based on cell types from our in-house Single Cell Atlas. Using Spacexr (2.2.1)[72] in "doublet" mode, recommended for high-resolution platforms like Visium HD, we assigned up to two cell types per bin. This output includes the top two cell types per bin, a weight matrix for all reference cell types across bins, and a categorical variable indicating bin class. A minimum UMI threshold of 0 was used for bin inclusion. For each bin, the top-ranked cell type was taken as the final annotation, while bins lacking confident assignment were labeled "Unassigned."

Identification of cell-cell communications between cell types using SOAPy

First, we performed standard preprocessing of the data for each sample using SOAPy (1.0.1)[48] and annotated the respective regions of 'DN' and 'veHCC'. Subsequently, we applied SOAPy to compute cell-cell communication separately for each region, using https://github.com/LewisLabUCSD/Ligand-Receptor-Pairs/blob/master/Human/Human-2020-Jin-LR-pairs.csv as the ligand-receptor database. SOAPy is a multifunction toolkit designed for spatial omics analysis; its communication module models contact and secreted signaling ligand-receptor pairs using distinct communication modes, followed by permutation tests to identify significant ligand-receptor pairs between cell types. After matching samples with the database, a total of 654 contact ligand-receptor pairs and 1,140 secreted ligand-receptor pairs were screened. For contact-dependent interactions, the 8 spatially adjacent neighbors were considered, whereas for secreted interactions, neighbors within a radius of 400 μm were included. The final communication results were filtered using thresholds of 'Strength' > 2 and 'Affinity' < 0.05.

Multiplex immunohistochemistry and multispectral image analysis

Slides (3–5 μm thick) were cut from the FFPE blocks, deparaffinized, rehydrated through an ethanol gradient, and unmasked by heating in citrate sodium solution (pH 6). The slides were stained using a PANO 7-plex IHC Kit (Cat# 004100100, Panovue) according to the manufacturer's instructions. HerPar (DAKO, IR624, 60 min incubation), CD4 (ZSGB-BIO, ZM0418, 30 min incubation), CD8 (CST, CST70306, 30 min incubation), PD1 (CST, CST43248, overnight incubation), and PDL1 (CST, CST13684, 30 min incubation) antibodies were applied, followed by incubation with horseradish peroxidase-conjugated secondary antibody for 15 min and tyramide signal amplification for 10 min. Slides were heat-treated using a microwave after tyramide signal amplification. Nuclei were stained

Please cite this article in press as: Zhang et al., Molecular insights into early malignant transition of hepatocellular carcinoma, Cancer Cell (2026), https://doi.org/10.1016/j.ccell.2026.03.005

with DAPI for 5 min after antigen labelling. For CD4 and FOXP3 (CST, CST98377), MPO (abcam, ab208670), CD163 (CST, CST93498), and CD11c (abcam, ab52632) co-staining, we used a TSA fluorescein evaluation kit (Akoya Biosciences) following the manufacturer's instruction. The slides were scanned using the Akoya Vectra 3 system with identical exposure times. Using spectral libraries that were previously built from images stained for each fluorophore (monoplex), the multispectral images obtained were unmixed using the inForm advanced image analysis software.

The entire area on each slide was stained and scanned. High-resolution images of the NCa, DN, and veHCC regions were captured in each case. For NCas, we captured images only from the parenchymal area, which did not contain fibrotic bands. The scoring method consisted of several automated steps, including tissue categorization, cell segmentation, and cell phenotyping. Using the integrated inForm image analysis software, multispectral images representative of different samples was selected and used to train the inForm software for tissue categorization, cell segmentation, and cell phenotyping. The settings learned from training on the representative images from different samples were saved within an algorithm that enabled batch analysis of all tissue slides. The abundance of infiltrating cells of different types was quantified by calculating cell counts per unit area (cells/mm^2).

QUANTIFICATION AND STATISTICAL ANALYSIS

No statistical methods were used to determine the sample size. The experiments were not randomized. The investigators were not blinded to the allocation during the experiments or outcome assessment. Since some samples were derived from the same patient, linear mixed effects models (LMM) were used to compare features among samples, such as telomere length, CIN index et al. LMM is a robust statistical approach that explicitly accounts for non-independent observations by incorporating random effects. Specifically, we included patient identifiers as a random effect in the model. The LMM was implemented using the R package lme4. Two-tailed Student's t-test, Fisher's exact test and Wicoxon rank sum test were employed for statistical analyses in this study. $P < 0.05$ indicates statistical significance.