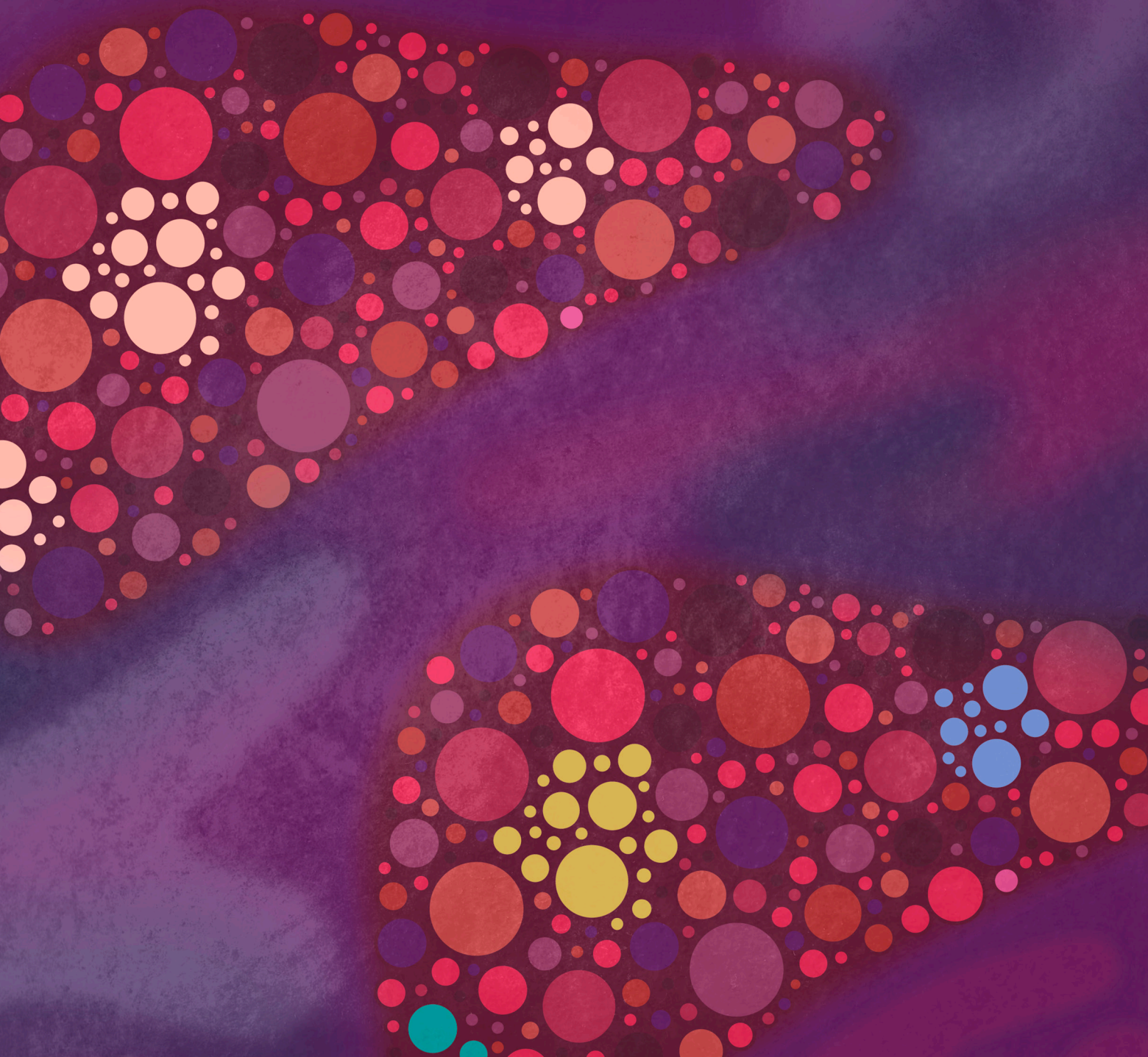
RESEARCH ARTICLE

Intrahepatic Microbial Heterogeneity in Multifocal Hepatocellular Carcinoma and Its Association with Host Genomic and Transcriptomic Alterations

Yinghong Lu[1], Lixia Xu[2], Weikang Chen[3], Weixin Liu[1], Ying Zhang[2], Qianying Zhou[3], Na Wang[3], Yongxin Zhang[3], Haojie Bai[3], Shu Xu[3], Pingmei Huang[1], Kaili Fu[1], Wenxuan Xie[4], Xin Liu[4], Xueliang Wang[1], Chi Chun Wong[1], Ming Kuang[4], and Jun Yu[1,3]



ABSTRACT The signature of the intrahepatic microbiome in multifocal hepatocellular carcinoma (HCC) and its association with genomic alterations remain elusive. In this study, we performed multiomics profiling of 242 HCC tumor nodules and 58 adjacent nontumor tissues from 58 patients with multifocal HCC, revealing heterogeneous microbial communities in multifocal HCC. The presence of bacteria in HCC nodules was confirmed by Gram stain, lipopolysaccharide, lipoteichoic acid staining, and transmission electron microscopy. Mutational profiling stratified patients into intrahepatic metastasis (IM)–HCC and multicentric occurrence (MO)–HCC. Bacterial communities differed between IM and MO nodules ($P = 0.01$). A nine-bacterium biomarker panel could distinguish IM nodules from MO nodules with an AUROC of 0.795. The epithelial–mesenchymal transition pathway was upregulated in IM nodules and correlated with IM-enriched bacteria. IM-enriched bacteria such as *Enterococcus faecalis* and *Streptococcus anginosus* promoted HCC cell migration, invasion, and tumor progression in orthotopic HCC mouse models by inducing an immunosuppressive microenvironment and epithelial–mesenchymal transition. Collectively, the intrahepatic microbiome contributes to the heterogeneity and pathogenesis of multifocal HCC.

SIGNIFICANCE: We reveal intraindividual heterogeneous microbial communities among nodules from patients with multifocal HCC. IM-HCC–enriched bacteria promote tumor growth and influence the tumor microenvironment of HCC. Our work highlights the necessity of considering bacterial heterogeneity as biomarkers and targets for multifocal HCC therapeutic intervention.

See related commentary by Dzutsev and Trinchieri, p. 1540

INTRODUCTION

Liver cancer is the third leading cause of cancer-associated deaths worldwide, with hepatocellular carcinoma (HCC) being the most common type of liver cancer, accounting for ~80% of the primary cases (1, 2). Approximately 41% to 75% of patients with HCC present with multifocal tumors at initial diagnosis (3). Multifocal HCC could arise from intrahepatic metastasis (IM) from a primary HCC or multicentric occurrence (MO) stemming from independent clones, as demonstrated by us and others (4, 5). Although surgical resection is a radical treatment option for HCC, HCC multifocality increases operation difficulty and the risk of postoperative recurrence (6, 7). Micrometastases are believed to be a significant trigger for recurrence and distant metastasis, particularly in IM-HCC (8, 9), which indicates a worse prognosis compared with MO-HCC (10). Nevertheless, the molecular characteristics and microenvironmental factors of multifocal HCC remain understudied, hampering therapeutic decision-making and patient prognostication.

Emerging evidence has revealed the intratumoral microbiota as a critical player in cancer initiation and progression by inducing DNA mutations, activating oncogenic and metastasis pathways, and promoting inflammation (11, 12). Each tumor type harbors its own unique bacterial signature (13). Recent studies have demonstrated that a dysbiotic intrahepatic microbiota in HCC, such as the enrichment of *Stenotrophomonas maltophilia*, could promote hepatocarcinogenesis (14). Moreover, intrahepatic microbial signatures could be utilized for the molecular subtyping of patients with HCC with prognostic value (15, 16). However, most HCC profiling studies thus far have only collected one biopsy per patient, which is far from representative given the significant heterogeneity of the intratumoral microbiome (17). Furthermore, the intertumoral heterogeneity of the microbiome in multifocal HCC and its potential impact on host signal transduction remain largely unclear.

In this study, we collected 242 HCC nodule tissues and matched adjacent nontumor liver tissues from 58 patients with multifocal HCC, 3 to 11 tumor nodules, and 1 paired adjacent nontumor liver tissue per patient. Microbiota profiling was performed by 16S rRNA gene sequencing. In parallel, host gene alterations and expression were determined by whole-exome sequencing (WES) and RNA sequencing (RNA-seq), respectively. We revealed significant heterogeneity in microbial communities in multifocal HCC. Based on mutation profiling by WES, we stratified the patients

[1]Institute of Digestive Disease and Department of Medicine and Therapeutics, State Key Laboratory of Digestive Disease, Li Ka Shing Institute of Health Sciences, The Chinese University of Hong Kong, Hong Kong SAR, China. [2]Department of Oncology, The First Affiliated Hospital, Sun Yat-sen University, Guangzhou, China. [3]Institute of Precision Medicine, The First Affiliated Hospital, Sun Yat-sen University, Guangzhou, China. [4]Center of Hepato-Pancreate-Biliary Surgery, The First Affiliated Hospital, Sun Yat-sen University, Guangzhou, China.

Y. Lu, L. Xu, and W. Chen contributed equally to this article.

Corresponding Authors: Jun Yu, Department of Medicine and Therapeutics, Prince of Wales Hospital, The Chinese University of Hong Kong, Hong Kong, China. E-mail: junyu@cuhk.edu.hk; and Ming Kuang, Center of Hepato-Pancreato-Biliary Surgery, The First Affiliated Hospital, Sun Yat-sen University, Guangzhou 510275, China. E-mail: kuangm@mail.sysu.edu.cn

Cancer Discov 2025;15:1630–48

doi: 10.1158/2159-8290.CD-24-1259

into IM-HCC with a high incidence of trunk mutations and strong genetic similarity, MO-HCC with no shared mutations and extremely low genetic similarity, and mixed IM/MO-HCC with a combination of branch and private mutations. We found that IM-HCC and MO-HCC harbor distinct bacterial signatures, and a microbial biomarker panel was devised that effectively distinguishes the two different types of multifocal HCC. Integrative analyses of host transcriptome and microbiome profiles showed that metastasis-related signatures were upregulated in IM and correlated with IM-enriched bacterial taxa. A biofunctional study confirmed that the IM-enriched bacteria promoted HCC cell migration/invasion *in vitro* and tumor growth and metastasis *in vivo*, suggesting that the intrahepatic microbiome drives IM-HCC. Our study illuminates the host and microbial heterogeneity in multifocal HCC and highlights its potential contribution to disease progression.

RESULTS

The Microbial Community Features of Multifocal HCC

A total of 300 liver tissue samples were collected from 58 patients with multifocal HCC (Supplementary Table S1), including 242 tumor nodule samples (3–11 nodules per patient) and 58 matched adjacent nontumor liver tissues. Samples were subjected in parallel to 16S rRNA gene sequencing, WES, and bulk RNA-seq to elucidate the microbiome, genome, and transcriptome profiles of multifocal HCC, followed by integrative analysis and functional validation (Fig. 1A).

Microbial profiling generated 7.74 ± 1.40 (10^4, mean ± SD) reads per sample. After removing potential contaminating DNA sequences (Supplementary Fig. S1A; Supplementary Tables S2 and S3), each sample was rarefied to 1,880 reads prior to diversity analysis (Supplementary Fig. S1B). The overall phenotypic patterns of the microbial community revealed no significant differences in microbial diversity (both α- and β-diversity) between tumor tissues and the corresponding adjacent nontumor tissues ($P = 0.94$ and $P = 0.72$, respectively; Supplementary Fig. S1C). This result aligns with a previous study showing that the microbial load of paratumor tissue is similar to that of the tumor (13). The differential bacteria between tumor tissues and corresponding adjacent nontumor tissues are presented in Supplementary Fig. S1D. Next, we analyzed the bacterial composition in each tumor nodule and adjacent normal sample from all 58 patients to provide an overview of individual microbiota profiles. At the phylum level (Fig. 1B), the microbiota in patients with multifocal HCC was dominated by *Proteobacteria* (mean 38.38%), followed by *Firmicutes* (mean 17.55%) and *Actinobacteria* (mean 11.45%). At the genus level (Supplementary Fig. S1E), the top three most abundant genera were *Acinetobacter* (mean 2.87%), *Streptococcus* (mean 2.56%), and *Lactobacillus* (mean 2.14%). Clustering analysis showed that only 3% of the patients ($n = 2$ out of 58) exhibited similar microbiome composition across tumor nodules, whereas the majority ($n = 56/58$) displayed significant heterogeneity in their intertumoral microbiome (Supplementary Fig. S2), suggesting intraindividual variations in microorganisms across tumor nodules in patients with multifocal HCC.

Validation of Intrahepatic Microbiota in Human Multifocal HCC

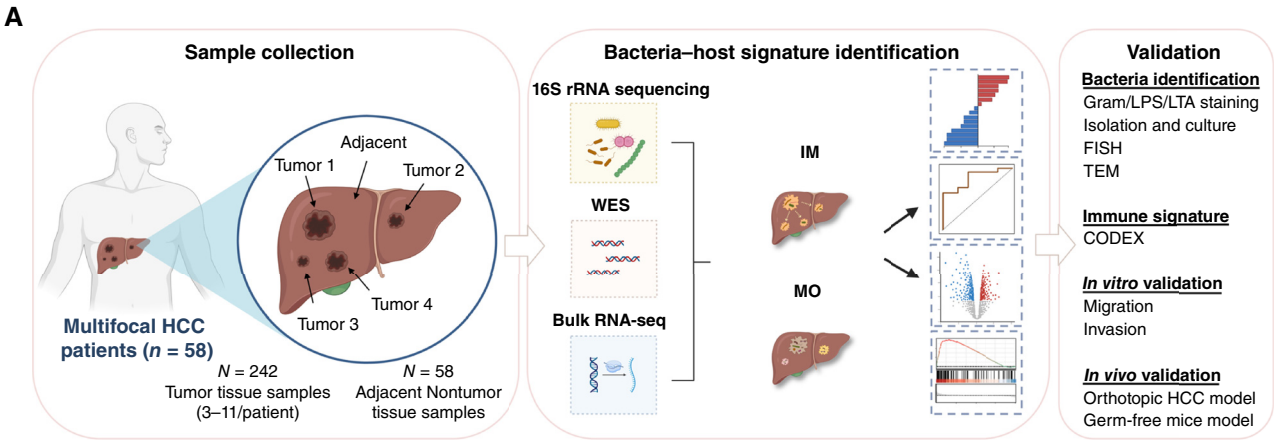
To confirm the presence of bacteria in multifocal HCC tissues, we performed Gram stain (Fig. 1C) and IHC staining of lipopolysaccharide (LPS, for Gram-negative bacteria) and lipoteichoic acid (LTA, for Gram-positive bacteria; Fig. 1C) in tumor nodules ($N = 40$). As shown in Fig. 1C and Supplementary Fig. S3A and S3B, intratumoral bacteria were localized in the perinuclear region as punctuated dots according to Gram, LPS, and LTA staining, respectively, containing approximately 108 bacteria per mm^2 and occupying approximately 10.1% ± 7.5% (mean ± SD, LPS ratio) and 6.7% ± 4.9% (mean ± SD, LTA ratio) of the tumor tissues by area (Fig. 1C). High-resolution electron microscopy indicated that bacteria were mainly distributed in the cytoplasm of HCC samples (Fig. 1D). These findings confirmed the presence of bacteria in multifocal HCC tissues.

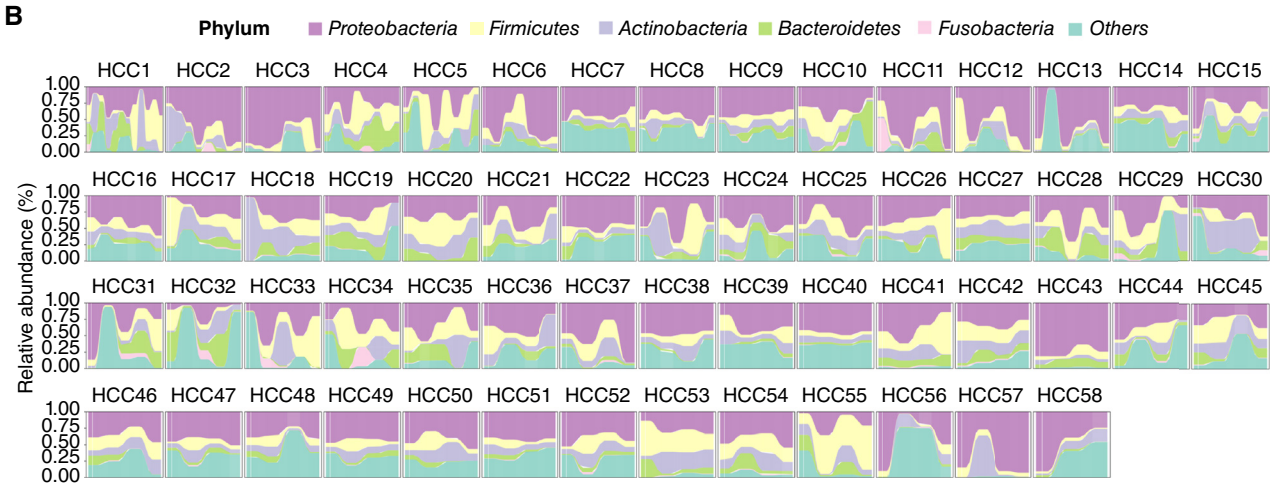
Heterogeneous Microbial Community in Multifocal HCC

Considering the high variation in microbial composition among HCC nodules within individuals, we determined a group of high-variable microbes (HVM) for each patient. Representative genera were chosen based on the following three criteria: normalized abundance of microbes at the genus level between HCC nodules and corresponding adjacent nontumor tissue ≥1, differences in abundance between HCC nodules and adjacent nontumor tissue ≥0.1%, and percent coefficient of variation among HCC nodules ≥50. The number of HVMs in each individual is listed in Supplementary Fig. S4A. Overall, we identified 347 internodule HVMs; each patient had 94 HVMs on average. Among these, *Streptococcus* (prevalence 91.4%), *Stenotrophomonas* (86.2%), and *Lactobacillus* (82.8%) were the most prevalent HVMs in patients with multifocal HCC (Supplementary Fig. S4B). Genera that met the HVM criteria were used to create a microbial mapping for each HCC tissue. Notably, we observed that the abundance of HVMs varied significantly across HCC nodules from the same patient, and network mapping analysis revealed that each nodule also had its own unique HVMs (Fig. 2). These findings collectively inferred that heterogeneous microbial communities reside in tumor nodules from the same patient with multifocal HCC.

Genomic Heterogeneity Correlates with Microbial Heterogeneity in Multifocal HCC

To explore the host genetic mutations that might arise during tumor evolution, we performed WES in 235 tumor nodules and 56 adjacent nontumor tissue samples from 56 patients with multifocal HCC. We identified 29,226 nonsynonymous and 11,014 synonymous somatic mutations, and validation by Sanger sequencing ($n = 1,560$) confirmed a true discovery rate of 95.3%. Focusing on HCC driver mutations (Supplementary Table S4), *TP53* (34%), *PCLO* (16%), and *CTNNB1* (16%) were among the most frequently mutated genes in our cohort (Fig. 3A). We next measured the intraindividual intertumor genomic heterogeneity by calculating the Jaccard index [intertumor heterogeneity (ITH) index] for each patient. Results revealed that more than half of the patients





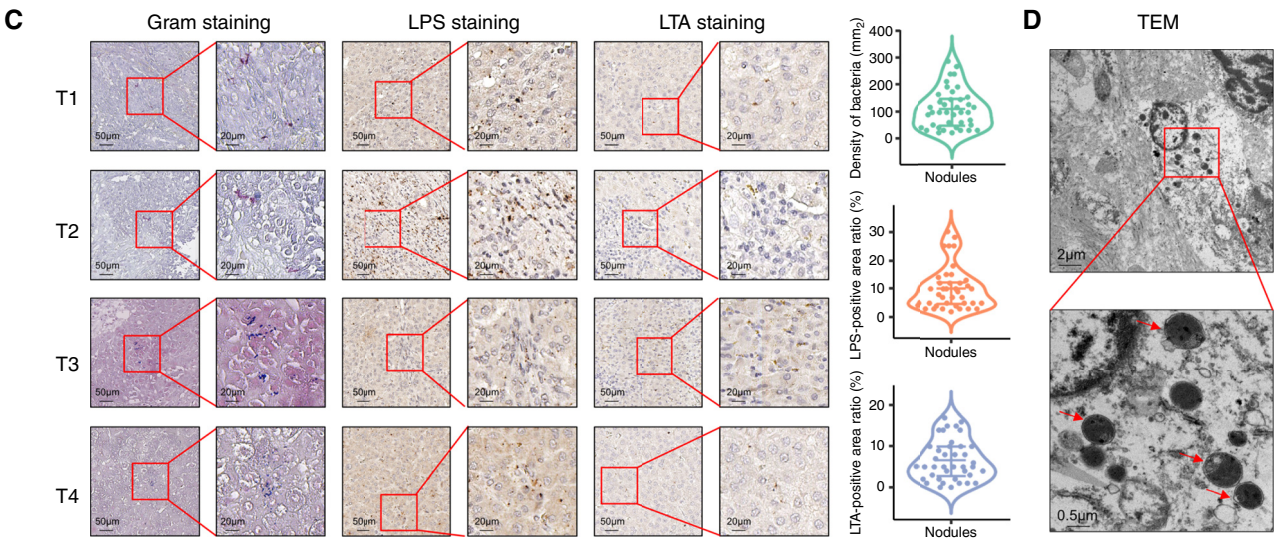


Figure 1. Microbial distribution of multifocal HCC. **A,** Flowchart of sample collection and study design. A total of 58 patients with multifocal HCC were enrolled in this study. Tumor nodule samples ($N = 242$, 3–11/patient) and adjacent nontumor liver tissue samples ($N = 58$) were collected to perform the 16S rRNA gene sequencing, WES, and bulk RNA-seq, followed by integrative analysis and functional validation. **B,** Alluvial plot depicting the relative abundance of microbial composition at the phylum level (top 5) within each patient. Multiple tumor nodules (right-sided) and adjacent nontumor tissues (left-sided) of each individual are separated by a white line. **C,** Representative image of Gram staining, bacterial LPS, and LTA staining in the multifocal HCC liver tissue. The quantification of positive signals is presented as mean ± SD on the right. Scale bars, 50 μm (left) and 20 μm (right). "T1–4" refers to tumor nodules. **D,** Representative images of transmission electron microscopy (TEM) indicating intracellular bacteria (red arrow) within multifocal HCC liver tissue. Scale bars, 2 μm (top) and 0.5 μm (bottom). (**A,** Created with BioRender.com.)

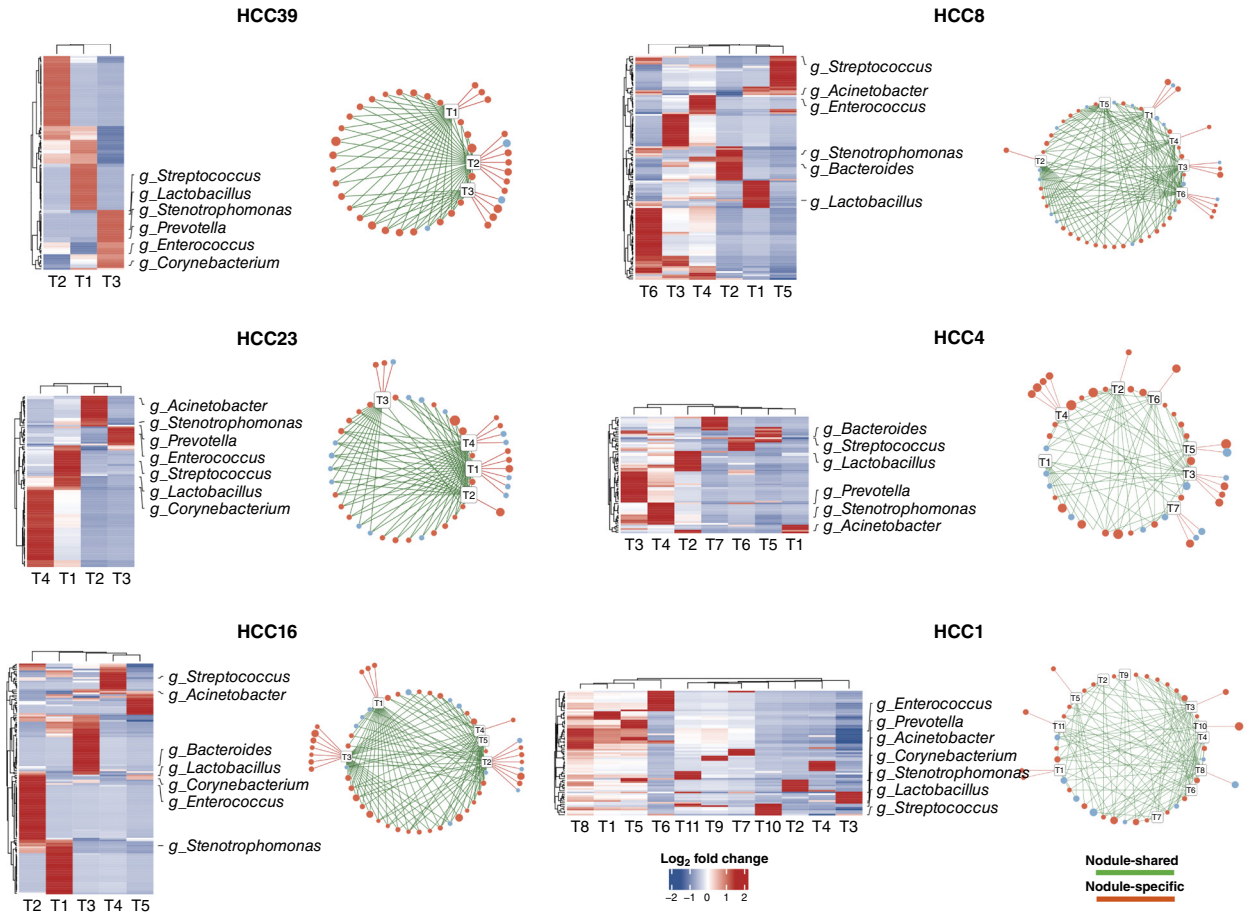


Figure 2. Heterogeneous microbial community in multifocal HCC. Microbial profile of HVMs (see "Methods") in six representative patients with multifocal HCC with different tumor nodule samples (3–11 nodules). Heatmaps show the abundance of microbiota in each tumor biopsy after normalizing with its corresponding adjacent normal tissue at the genus level in each patient (left). Selected microbes are labeled based on the rank of HVMs in multifocal HCC. Network maps illustrate the tumor nodule-shared (green lines) and nodule-specific (orange lines) microbes among HVMs in each patient (right). Each circle represents a microbial genus, with the circle size and color corresponding to the levels of normalized abundance and differential abundance relative to the adjacent normal tissue, respectively. "T1–11" refers to tumor nodules.

exhibited high ITH levels (median ITH index: 0.63; Fig. 3A). These results suggested that the intertumor genomic mutation presented significant heterogeneity in multifocal HCC.

To ask if the host genomic heterogeneity is associated with the intrahepatic microbiome, we applied the Mantel test. Indeed, the host genomic heterogeneity was positively correlated with microbial heterogeneity (Supplementary Fig. S5A), implying potential cross-talk between host genomic and intrahepatic microbiome in multifocal HCC.

Discrimination of IM-HCC and MO-HCC by Phylogenetic Analysis

To decipher whether the multifocal HCC arises from a primary origin (IM) or multicentric origins (MO), we first grouped nonsynonymous somatic mutations into trunk mutations (present in all nodules), branch mutations (shared by some nodules), or private mutations (present in only one nodule) for each patient (Fig. 3B). Meanwhile, we constructed the phylogenetic tree and calculated the genetic similarity (Euclidean distance) for each patient, and the results showed that patients with multifocal HCC can be stratified into MO-HCC ($n = 5$, with no shared mutations and extremely low genetic similarity), IM-HCC [$n = 40$, with a high incidence of trunk mutations (8%–90%) and strong genetic similarity], and mixed IM/MO-HCC ($n = 11$, a combination of branch and private mutations, Fig. 3B; Supplementary Fig. S5B and S5C). Clonal dynamics of multifocal HCC were further analyzed using the PyClone algorithm, confirming the presence of common clone clusters in IM-HCC but not in MO-HCC (Supplementary Fig. S5D). We also compared the tumor mutational burden (TMB) among IM nodules from IM-HCC (IM), partial IM nodules from mixed IM/MO-HCC (MIX_IM), partial MO nodules from mixed IM/MO-HCC (MIX_MO), and MO nodules from MO-HCC (MO). The MO nodule showed a higher TMB than other nodule types, especially for MIX_MO ($P = 0.03$; Supplementary Fig. S5E) and MIX_IM ($P = 0.02$; Supplementary Fig. S5E). We compared the mutation signatures among these four types of tumor nodules, but no significant differences were found (Supplementary Fig. S5F).

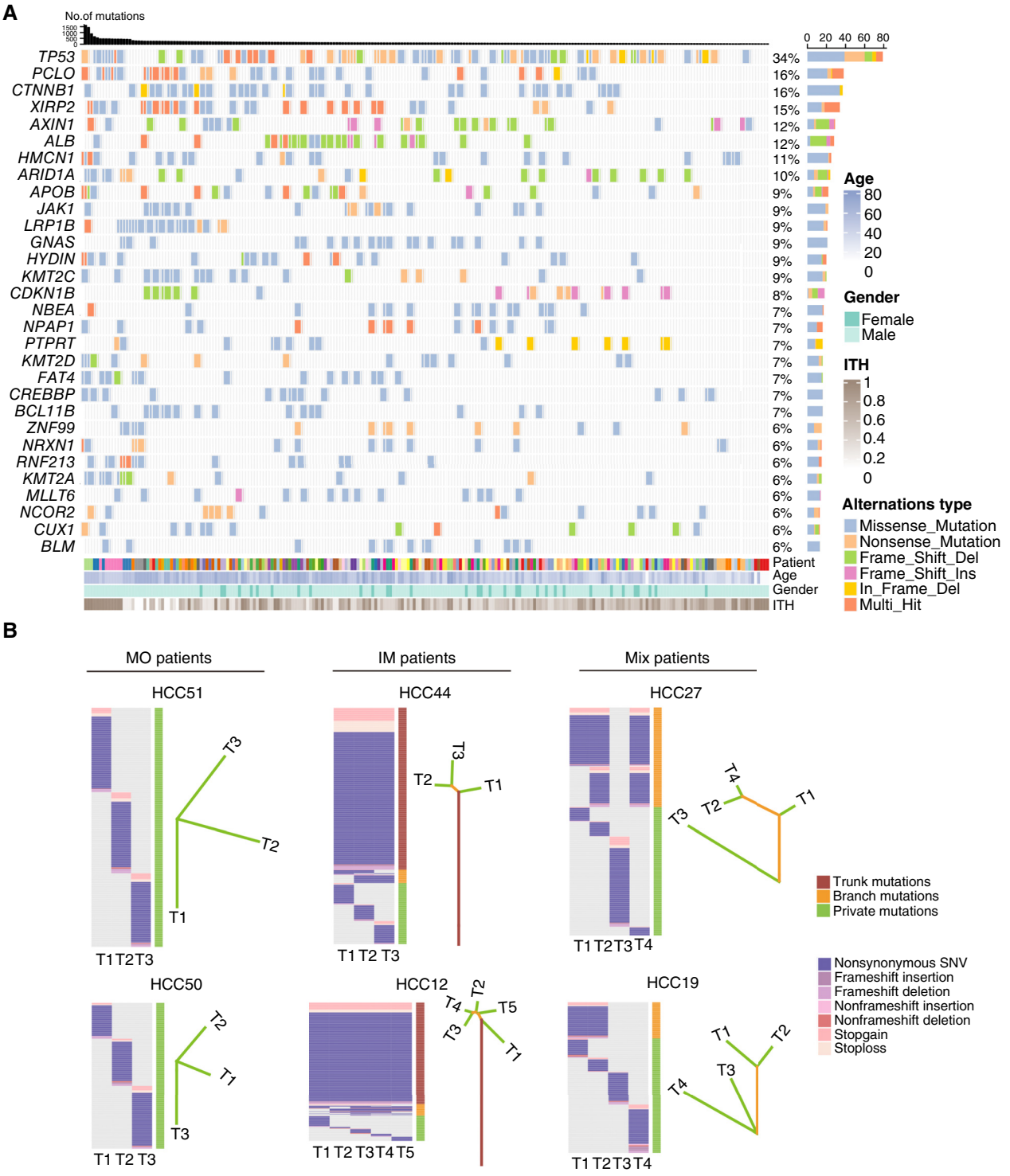


Figure 3. Intertumor genetic heterogeneity in multifocal HCC. **A,** Mutational landscape of the top 30 potential driver genes across all the tumor regions from 56 patients with multifocal HCC. (Top) The mean number of clonal and subclonal alterations in exonic regions of each tumor. (Bottom) The patients' age, gender, and ITH index information. The ITH index was measured by calculating the ratio of shared variants to all variants between tumor nodules (Jaccard similarity coefficient) within each patient (see "Methods"). **B,** Mutation and evolutionary profiles of six representative patients with multifocal HCC, including two MO patients, two IM patients, and two mix patients. Heatmaps show the regional distribution of all nonsynonymous mutations in multifocal HCC (left). The columns next to each heatmap show three categories of mutations: trunk mutations (mutations present in all tumors, red-brown), branch mutations (shared by some but not all nodules, orange), and private mutations (unique to one nodule, green). Mutation types are demonstrated by different colors. Phylogenetic trees are constructed using a maximum parsimony algorithm based on all nonsynonymous mutations identified in each patient (right). The trunk, shared branches, and private branches are represented with red-brown, orange, and green, respectively. The length of each line is proportional to the number of mutations.

There were no significant differences in clinical characteristics between patients with IM-HCC and those with MO-HCC, including age, sex, body mass index, HBV status, HCV status, alanine aminotransferase, aspartate aminotransferase, TBil, and DBil (Supplementary Table S5). These findings suggest that multiple tumors from the same patient with multifocal HCC can exhibit differential evolutionary patterns. Sequencing only the primary lesion (or any single lesion) cannot completely characterize the genomic landscape of multifocal HCC.

Distinct Intratumor Microbiota in IM-HCC and MO-HCC

Having shown that multifocal HCC could have distinct evolutionary paths, we next asked whether the intratumoral microbiota composition varies among different nodule types. We discovered that the microbial community (β-diversity) was significantly different among these four types of nodules [permutational multivariate analysis of variance (PERMANOVA); $P < 0.001$; Fig. 4A], especially between IM and MO nodules (PERMANOVA; $P = 0.01$). In addition, a lower α-diversity (Shannon and Chao1 index) was observed in IM nodules compared with MO nodules ($P = 0.003$ by Shannon and $P = 0.034$ by Chao1 index, respectively; Fig. 4B). There was no significant difference in the microbial community between the IM/MO nodules and their corresponding adjacent nontumor tissues (Supplementary Fig. S6A and S6B). The overall bacterial composition at the phylum and genus levels in the four types of nodules is illustrated in Fig. 4C and D. Compared with MO nodules, the genera *Enterococcus* and *Fusobacterium* were enriched in IM nodules, whereas *Stenotrophomonas*, *Corynebacterium*, *Burkholderia*, and *Clostridium* were depleted (Fig. 4E). At the species level, *Streptococcus minor* and *Streptococcus anginosus* were also enriched in IM nodules compared with MO nodules (Fig. 4E). To evaluate whether bacterial markers could differentiate between IM and MO nodules, we randomly divided the samples into training and validation cohorts at a 6:4 ratio. Using a panel of nine differentially abundant taxa (Supplementary Fig. S7A), we constructed a random forest model with repeated fivefold cross-validation. As shown in Fig. 4F, this model achieved an AUC of 0.795 (95% confidence interval, 0.658–0.932) in differentiating IM from MO nodules. Due to the imbalance in case numbers, we also performed an area under the precision–recall (PR) curve analysis to minimize false positive rates (18, 19). The PR analysis confirmed that the nine markers could effectively classify IM from MO with an AUC of 0.831 (Supplementary Fig. S7B). Collectively, our results suggest that distinct microbiota inhabit IM and MO nodules in patients with multifocal HCC.

IM Tumor Nodules Display Enhanced Metastatic Potential

We further conducted RNA-seq on the collected samples to elucidate the functional gene expression alteration between IM and MO nodules. We identified a total of 615 differentially expressed genes, of which 434 genes were upregulated and 181 genes were downregulated in IM nodules compared with MO nodules (Fig. 5A). Notably, a number of metastasis-associated genes, including *AGR2*, *HOXB9*, and *CEACAM6*, were significantly upregulated in IM nodules. Gene set enrichment analysis showed that the epithelial–mesenchymal transition (EMT) pathway emerged as the top enriched pathway in IM nodules, followed by cell cycle–related pathways such as G_2–M and E2F signaling, as well as angiogenesis (Fig. 5B). These findings are consistent with reports (20) suggesting a higher metastatic capacity in IM nodules.

IM-Enriched Bacteria Correlate with Metastasis-Related Gene Expression in IM Nodules

To see if the intratumoral microbes are associated with host gene expression, we utilized Spearman correlation to characterize bacteria–host gene associations. We identified 3,622 genes and 139 taxa showing significant associations in IM nodules (Supplementary Fig. S8A, FDR < 0.05). Interestingly, we found that host genes significantly associated with IM-enriched taxa (such as *Enterococcus* spp., *S. anginosus*, *Fusobacterium* spp., and *S. minor*) are EMT-related (Fig. 5C), which were distinct from association patterns in MO nodules (Supplementary Fig. S8B). Specifically, *S. minor* positively correlated with the metastasis-promoting transcription factor (*SNAI2*), whereas it showed an inverse correlation with the tight junction marker occludin (*OCLN*). *S. anginosus* demonstrated a positive correlation with SPARC/osteonectin, cwcv, and kazal-like domain proteoglycan 1 (*SPOCK1*) and cell adhesion molecules (*CEACAM*). *Enterococcus* spp., on the other hand, was associated with the glucose transporter protein (*SLC2A1*) and *VEGFA* (both of which are established drivers in EMT; Fig. 5D). Taken together, these analyses revealed that IM-enriched pathogenic bacteria may participate in IM by activating EMT-related signaling pathways in HCC.

IM-Enriched Bacteria Promote HCC Cell Metastasis Phenotypes

To investigate the potential function of IM-enriched bacteria in promoting tumor metastases, we chose IM-enriched bacterial species, such as *Enterococcus faecalis* and *S. anginosus*, to perform functional validation. *E. faecalis* was selected as the representative bacterium for *Enterococcus* spp. after species-level identification, showing its high prevalence in IM nodules (Supplementary Table S6). First, we performed FISH with specific probes (Supplementary Table S7) to target these pathobionts, confirming the presence of *E. faecalis* and *S. anginosus* in the multifocal HCC samples analyzed (Fig. 6A). We successfully cultivated multiple bacterial species from the fresh liver homogenates of patients with HCC by anaerobic and aerobic culture, including the pathobiont *E. faecalis* (Supplementary Fig. S9A; Supplementary Table S8). As the abundance of these bacteria was correlated with metastasis-related genes in IM-HCC, we further evaluated the effect of bacterial co-culture on the invasion and migration ability of HCC cells. Indeed, the wound-healing assay showed that *E. faecalis* (isolated from HCC) and *S. anginosus* (ATCC 33397) both significantly promoted wound closure compared with control cells (Fig. 6B). Consistently, transwell migration and invasion assays revealed that *E. faecalis* and *S. anginosus* significantly increased HCC cell migration and invasion, respectively (Fig. 6C and D). Together, these data demonstrated that the IM-enriched pathobionts *E. faecalis* and *S. anginosus* could enhance HCC metastatic capacity.

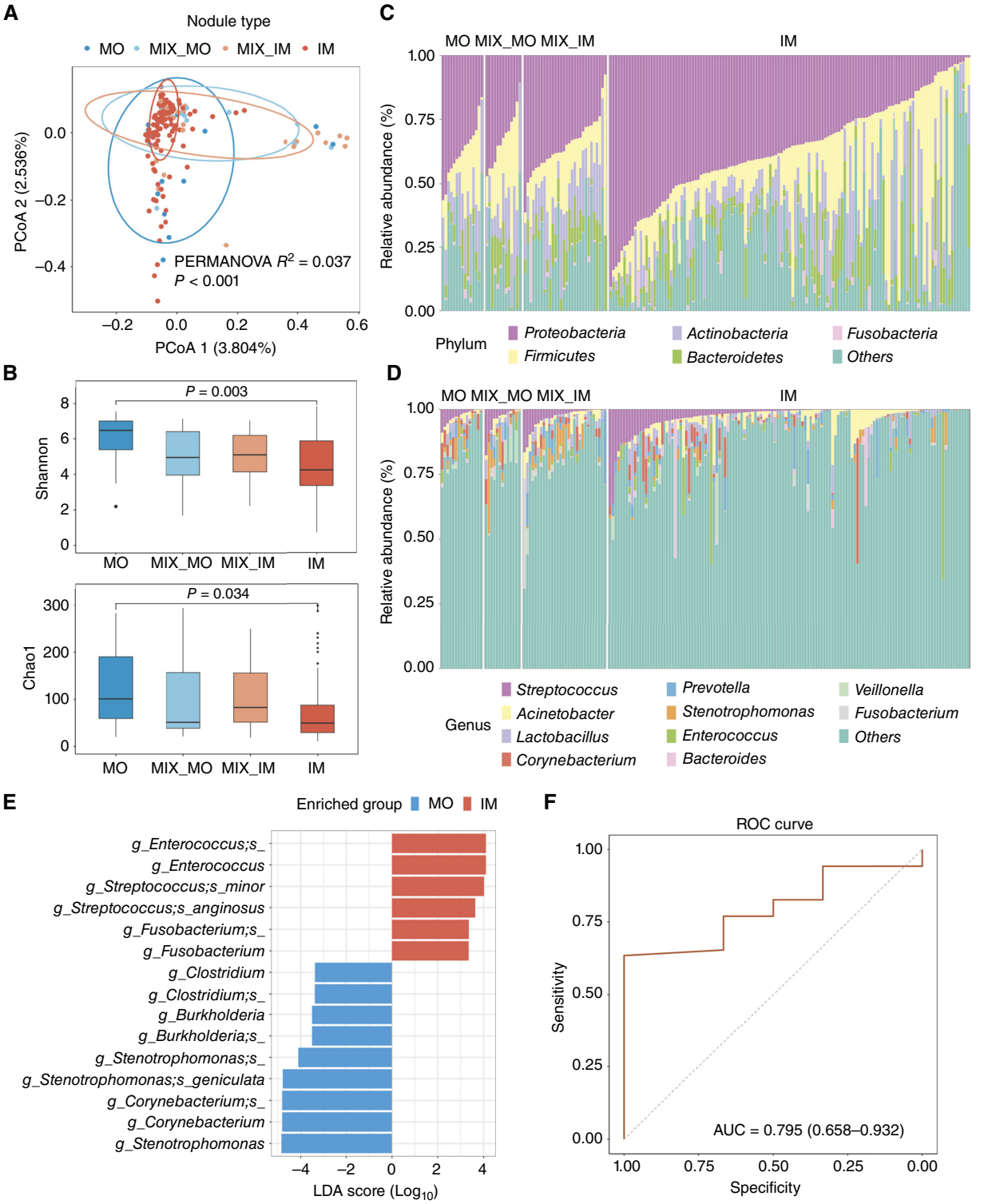


Figure 4. Characterization of the microbiome in IM and MO nodules of multifocal HCC. **A,** Principal coordinates analysis (PCoA) based on Bray–Curtis distances, showing the microbial community structure in MO, MIX_MO, MIX_IM, and IM nodule tissues. Community dissimilarities are tested by PERMANOVA analysis. **B,** Boxplot of α-diversity based on the Shannon and Chao1 indices comparing the MO, MIX_MO, MIX_IM, and IM nodule tissues. Data are presented as mean ± SD. *P* values are determined using the Mann–Whitney U test. **C** and **D,** Stacked column bar plot depicting the relative abundance of the phylum and genus composition in the MO, MIX_MO, MIX_IM, and IM nodule tissues. **E,** Linear discriminant analysis effective size (LEfSe) analysis showing the differentially abundant taxa at the genus and species levels between IM and MO nodules. Threshold value with linear discriminant analysis (LDA) score >3. **F,** ROC curve evaluating the predictive ability of nine differentially abundant taxa in discriminating IM from MO using a random forest model with repeated fivefold cross-validation. The AUC is 0.795 (95% confidence interval, 0.658–0.932).

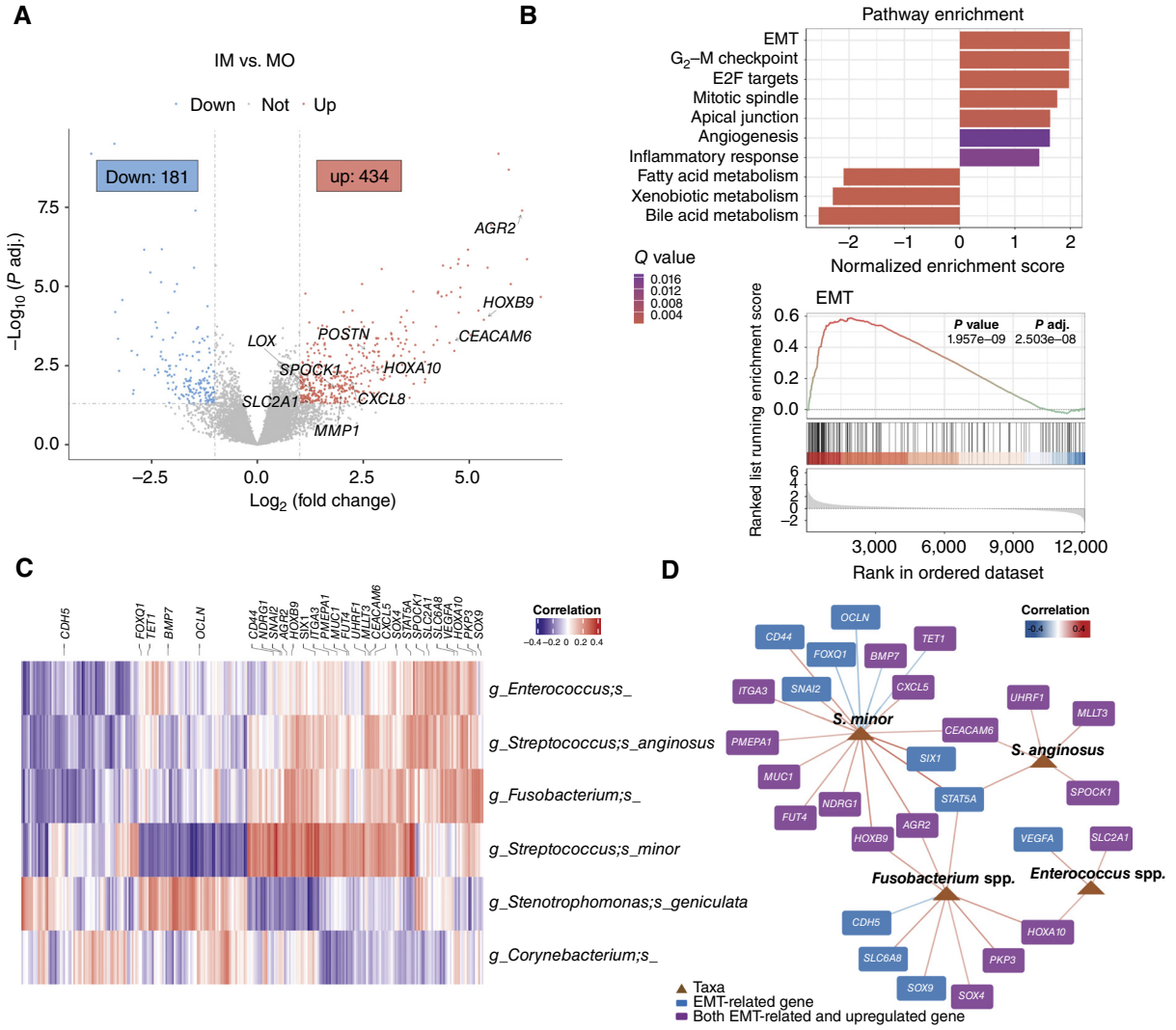


Figure 5. IM tumors exhibit enhanced metastatic activities and are correlated with intratumoral microbiota. **A,** Volcano plot showing differentially expressed genes between IM and MO tumor nodules identified using the DESeq2 program. Cutoffs for significant genes are set at an adjusted *P* value < 0.05 and $\log_2$ fold change >1. Highlighted genes associated with metastasis are labeled. **B,** Gene set enrichment analysis indicating hallmark signaling pathways that are differentially regulated between IM and MO nodules with an adjusted *P* value < 0.05. (Bottom) The enrichment plot for the EMT pathway. **C,** Heatmap illustrating the correlation pattern between differential microbial taxa (rows) and host genes (columns) identified by Spearman correlation in IM nodules (FDR < 0.05). **D,** Network analysis showing enriched intratumoral microbes in IM nodules and their correlations with host genes related to metastasis. Triangular nodes (brown) represent intrahepatic microbes; rectangular nodes (blue) represent host genes associated with EMT; and rectangular nodes (purple) represent EMT-related host genes that are significantly upregulated in IM nodules. Edge color represents positive (blue) or negative (red) associations, and edge width represents the strength of association (Spearman's rho).

IM-Enriched Bacteria Promote Tumor Progression and Shape Immunosuppressive Microenvironment

To recapitulate these findings *in vivo*, we established an orthotopic HCC model via the intrahepatic injection of RIL-175 liver cancer cells co-cultured with *E. faecalis*, *S. anginosus*, *Escherichia coli MG1655* (bacteria control), or PBS (vehicle) into specific pathogen–free (SPF) C57BL/6 mice (Fig. 6E). We found that *E. faecalis* and *S. anginosus* significantly accelerated tumor growth compared with *E. coli* or PBS (Fig. 6E). Consistently, Ki-67 staining demonstrated significantly increased tumor cell proliferation in the *E. faecalis* and *S. anginosus* groups compared with the control groups (Supplementary Fig. S9B). IHC analyses of EMT markers revealed downregulation of E-cadherin and upregulation of N-cadherin and Snail in *E. faecalis*– and *S. anginosus*–treated groups compared with control groups, indicative of elevated EMT (Fig. 6F), in support of our *in vitro* observations. Notably, the administration of antibiotics (ampicillin and gentamicin) abolished the protumorigenic (Fig. 6E), proliferative (Supplementary Fig. S9B), and EMT-inducing effects (Supplementary Fig. S9C) of *E. faecalis* and *S. anginosus*. Concordant results were obtained in a subcutaneous mouse model (Supplementary Fig. S9D), reinforcing the tumor-promoting effect of *E. faecalis* and *S. anginosus*.

We then explored the role of pathogenic bacteria in shaping the tumor microenvironment of HCC in this orthotopic mouse model. Intratumoral immune cells were profiled by

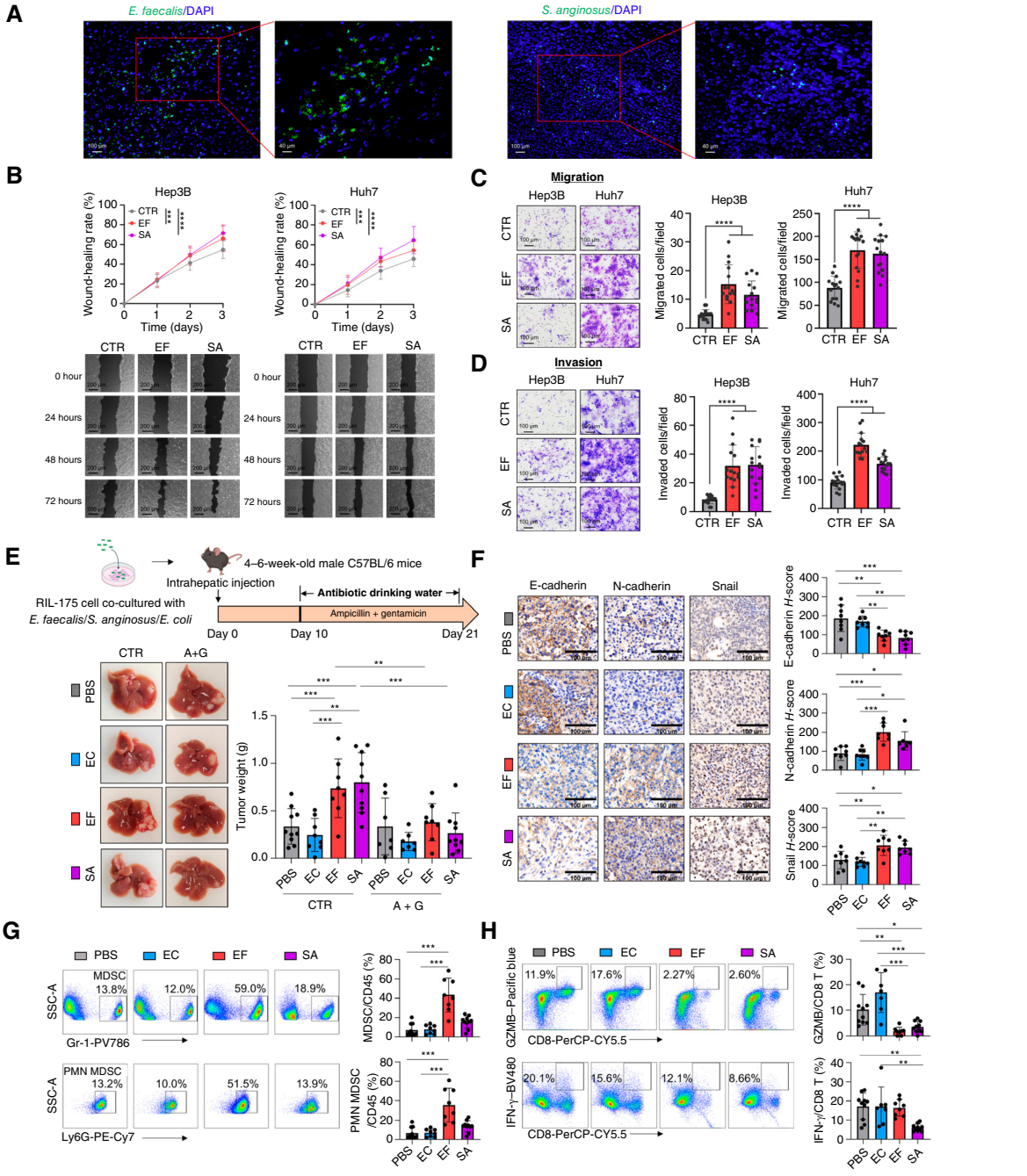


Figure 6. Intrahepatic bacteria promote HCC tumor growth and metastasis. **A,** Representative FISH images displaying the detection of *E. faecalis* (left) and *S. anginosus* (right) in liver tissues from patients with multifocal HCC. Bacterial probes (green); nuclei stained with DAPI (blue). Scale bars, 100 μm (left) and 40 μm (right). Probe sequences are provided in Supplementary Table S7. **B,** Wound-healing assay of human HCC cell lines (Hep3B and Huh7) co-cultured with *E. faecalis* (EF) and *S. anginosus* (SA) at a multiplicity of infection of 50 for 20 hours. *P* values were determined using one-way ANOVA with multiple comparisons. Representative images of the wound-healing assay at the indicated time points (0, 24, 48, and 72 hours) are shown below (Scale bar, 200 μm). **C** and **D,** Representative images and quantification of (**C**) migration and (**D**) invasion assays in human HCC cell lines (Hep3B and Huh7) co-cultured with EF and SA. Scale bars, 100 μm. *P* values are determined using the Mann–Whitney U test. **E,** Schematic diagram illustrating the construction of an SPF C57BL/6 mouse model with orthotopic HCC allografts. RIL-175 HCC cells are co-cultured with EF, SA, and *E. coli* (EC; $n = 7$–10/group) prior to intrahepatic injection. Mice are administered antibiotics [ampicillin (A) + gentamicin (G)] in drinking water starting from day 10 after injection. Representative images of liver tumors from different groups under control (CTR) and antibiotic treatment (A + G) conditions are shown. Statistical significance is assessed using one-way ANOVA with multiple comparisons. **F,** Representative images of IHC staining for EMT markers E-cadherin, N-cadherin, and Snail in tumor tissues from RIL-175 allografts ($n = 8$/group) in different groups (PBS, EC, EF, and SA). Scale bars, 100 μm. *P* values are determined using one-way ANOVA with multiple comparisons. **G** and **H,** Representative images of flow cytometry (left) and quantification (right) of (**G**) MDSCs and polymorphonuclear (PMN) MDSCs, (**H**) granzyme B-positive (GZMB$^+$) CD8$^+$ T cells, and IFN-γ$^+$ CD8$^+$ T cells in different groups (PBS, EC, EF, and SA; $n = 7$–10/group). Statistical significance is assessed using one-way ANOVA with multiple comparisons. Data are presented as mean ± SD. *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$; ****, $P < 0.0001$.

flow cytometry. *E. faecalis* increased the proportion of immunosuppressive myeloid-derived suppressor cells (MDSC), especially polymorphonuclear MDSCs (Fig. 6G). In both *E. faecalis*– and *S. anginosus*–treated groups, we observed reductions in $CD8^+$ T and $CD4^+$ T cells (Supplementary Fig. S9E). Moreover, the effector function of $CD8^+$ T cells was downregulated, including reduced granzyme B expression in *E. faecalis* and *S. anginosus* and reduced IFN-γ expression in the *S. anginosus*–treated group, indicating dampened antitumor immunity (Fig. 6H). Collectively, these findings suggest that IM-enriched pathobionts *E. faecalis* and *S. anginosus* could promote tumor growth and metastasis and induce an immunosuppressive microenvironment in HCC, highlighting the critical role of these bacteria in modulating HCC progression and immune evasion.

Spatial Analysis of Multifocal HCC Reveals Close Proximity of IM-Enriched Bacteria to Blood Vessels and Immunosuppressive MDSCs

The intrahepatic microbiota might originate from the gut through the gut–liver axis. By comparing the microbial profiles of stool samples and HCC tumor samples parallelly from a cohort of patients with single-lesion HCC, we identified identical bacterial species [at amplicon sequence variant (ASV) level], parallelly existing in both stool and tumor samples (Supplementary Fig. S10A–S10C). Notably, we observed a significantly elevated abundance of the identified bacterial taxa enriched in multilesion HCC compared with single-lesion HCC (Supplementary Fig. S10D–S10G), indicating that the bacteria are gained during the tumor progression process.

Next, we performed co-detection by indexing (CODEX) multiplexed staining of 16 human markers together with Eubacteria 16S rRNA FISH detection in the sequential slices of formalin-fixed, paraffin-embedded (FFPE) samples from IM-HCC and MO-HCC tumor nodules (Supplementary Table S9). We classified the human single cells into immune cells (B cells, $CD4^+$ T cells, $CD8^+$ T cells, dendritic cells, MDSCs, Treg, and tumor-associated macrophages), fibroblasts, and endothelial cells (Supplementary Fig. S11A–S11C) and observed the significant enrichment of MDSCs in IM-HCC compared with MO-HCC (Supplementary Fig. S11D). Moreover, MDSCs predominantly surrounded bacterial microniches in IM-HCC compared with MO-HCC (Supplementary Fig. S11E). Supporting the hematogenous origin of intrahepatic microbiota, the majority of intratumoral bacteria were found in close proximity to blood vessels in both IM-HCC and MO-HCC (Supplementary Fig. S11E). Hence, proximity to immunosuppressive cells and blood vessels may enable bacterial colonization in some cancer cells within HCC tumors.

IM-Enriched Bacteria Translocate from the Gut to the Liver and Promote HCC in Germ-Free Mice

We next sought to validate whether IM-enriched bacteria could translocate from the gut to the liver (Fig. 7). Germ-free mice were orally gavaged with two IM-enriched bacterial combinations, including *E. faecalis* plus *S. anginosus* (EF + SA) or *E. faecalis* and *S. anginosus* plus *Fusobacterium nucleatum* (EF + SA + FN) for 8 weeks, with *E. coli* MG1655 or PBS as controls (Fig. 7A). PCR (Fig. 7B), live bacterial culture (Fig. 7C), and FISH (Fig. 7D) all confirmed the presence of live bacteria (*E. faecalis*, *S. anginosus*, and *F. nucleatum*) in the livers of the corresponding bacteria-gavaged mice, whereas the livers of *E. coli*– and PBS-gavaged mice tested negative.

To determine if the translocated IM-enriched bacteria drive HCC tumorigenesis, we generated two orthotopic HCC models in germ-free mice via the intrahepatic injection of RIL-175 (Fig. 7E) and Hepa1-6 cells (Supplementary Fig. S12A) and gavaged the mice with two IM-enriched bacteria combinations (EF + SA or EF + SA + FN). Either EF + SA or EF + SA + FN accelerated tumor growth compared with the *E. coli* and PBS groups in the RIL-175 (Fig. 7E) and Hepa1-6 (Supplementary Fig. S12A) models. Bacterial culture from tumor homogenates and FISH confirmed the presence of live bacteria in tumors (Fig. 7F; Supplementary Fig. S12B–S12D). Furthermore, IM-enriched bacteria (EF + SA or EF + SA + FN) enhanced the tumoral expression of N-cadherin and Snail, whereas E-cadherin was downregulated (Fig. 7G; Supplementary Fig. S12E), corroborating the role of IM-enriched bacteria in promoting EMT.

DISCUSSION

In this study, we integrated microbiome, genome, and transcriptome analyses to study tumor nodules and matched adjacent nontumor tissues from patients with multifocal HCC. We demonstrated, for the first time, heterogeneous microbial communities among nodules from the same patients with multifocal HCC. WES-informed stratification of tumor nodules into IM- or MO-HCC also unraveled distinct microbial compositions among multifocal HCC subtypes. Furthermore, we identified that certain IM-enriched bacteria were correlated with metastasis-related gene signatures and promoted prometastatic phenotypes in HCC both *in vitro* and *in vivo*, implicating a role of intratumoral microbes in IM. Our study thus resolves the genetic and microbiome heterogeneity of multifocal HCC and offers biological insights into bacteria-oriented HCC intrahepatic progression and metastasis of multifocal HCC.

Over the past decade, the landscape of molecular alterations in multifocal HCC has been explored using genomics (4), transcriptomics (21), epigenomics (22), and proteomics (23) approaches. Nevertheless, little is known about the microbial profiles in this malignancy. In this study, we performed a comprehensive metagenomic evaluation of multifocal HCC by analyzing multiple nodules from each patient. We systematically identified a group of HVMs with large variations across distinct nodule tumors from the same patient. Identified HVMs were dominated by bacterial genera, such as *Streptococcus*, *Stenotrophomonas*, and *Lactobacillus*. *Streptococcus* genera, which are opportunistic pathogens in the human gut microbiome, and their species have been reported to promote gastrointestinal cancers, including gastric cancer and colorectal cancer (24, 25). Importantly, we noticed an association between the microbial variation and host gene mutation, implying intricate and dynamic cross-talk between tumor and its resident bacteria at individual nodules. These results indicate that individual tumor nodules in multifocal HCC are highly heterogeneous at genomic and metagenomic levels,

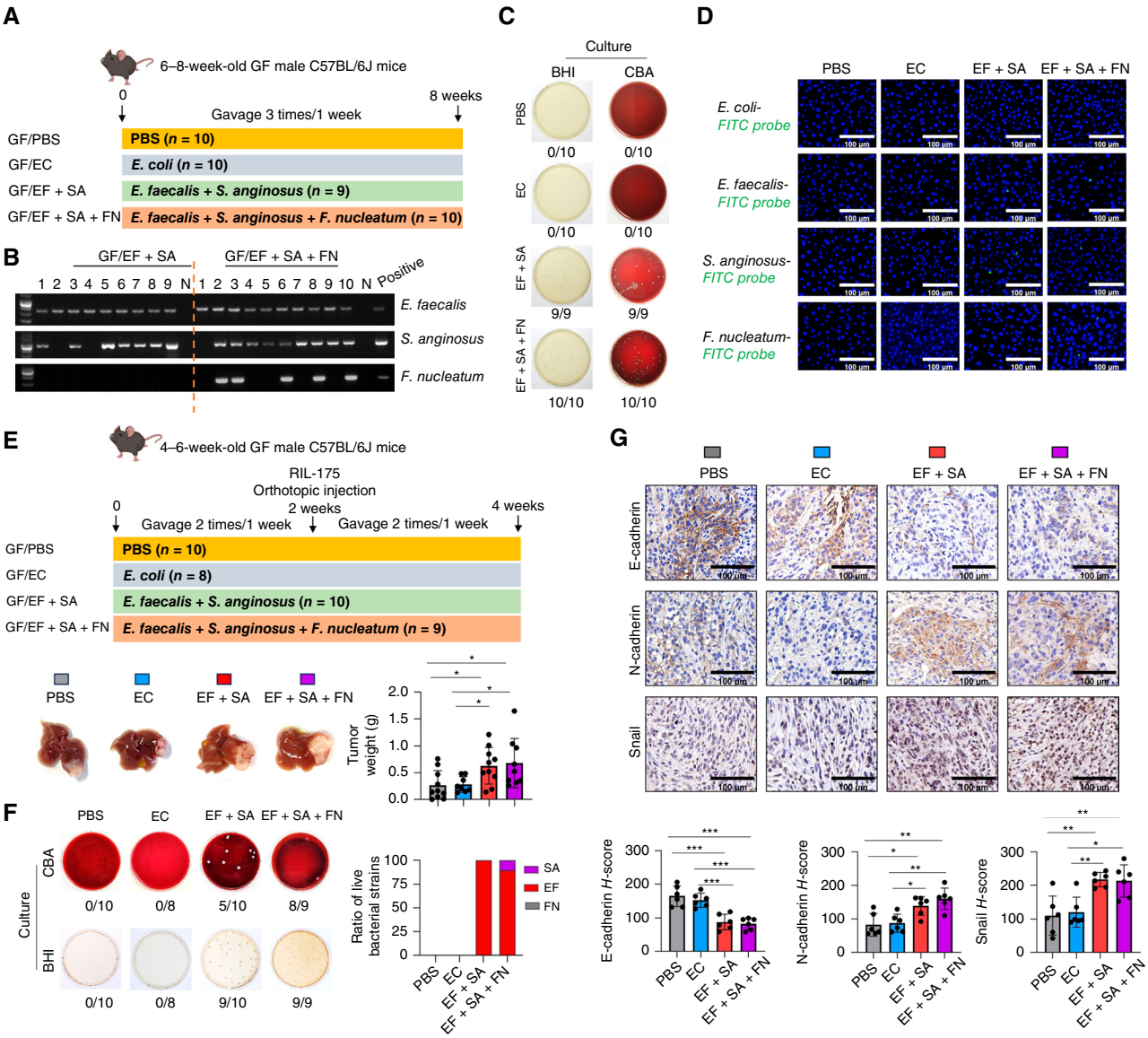


Figure 7. Bacteria translocate from the gut to the liver in the germ-free mouse model. **A,** Schematic of the experimental design for oral gavage of bacteria in germ-free (GF) C57BL/6J mice ($n = 9$–10/group). Mice are assigned to four groups: PBS (GF/PBS), *E. coli* (GF/EC), a combination of *E. faecalis* and *S. anginosus* (GF/EF + SA), and a combination of *E. faecalis*, *S. anginosus*, and *F. nucleatum* (GF/EF + SA + FN). **B,** PCR analysis of bacterial DNA in liver tissues from GF mouse models. Bacterial-specific PCR is conducted to detect the presence of *E. faecalis*, *S. anginosus*, and *F. nucleatum* in the GF/EF + SA and GF/EF + SA + FN groups. Lanes 1–10 show samples from individual mice. "N" represents the negative control, and "positive" denotes the positive control. Primer sequences are detailed in Supplementary Table S7. **C,** Photographs of bacterial cultures obtained from liver tissues of GF mice under anaerobic conditions. These cultures are obtained from Columbia blood agar (CBA) plates and BHI plates (the number of positive cultures per total samples is indicated at the bottom of each image). **D,** Representative FISH images displaying the detection of *E. faecalis*, *S. anginosus*, and *F. nucleatum* in liver tissues from different groups. Scale bars, 100 μm. **E,** Schematic showing the establishment of an orthotopic HCC model in GF C57BL/6J mice with RIL-175 cell lines ($n = 8$–10/group). Mice are assigned to four groups: GF/PBS, GF/EC, GF/EF + SA, and GF/EF + SA + FN. Representative tumor images from each group are shown, and statistical significance is assessed using one-way ANOVA with multiple comparisons. **F,** Photographs of bacterial cultures derived from GF RIL-175 allografts under anaerobic conditions from CBA plates and BHI plates (the number of positive cultures per total samples is indicated at the bottom of each image). A stacked bar chart displaying the relative abundance of culturable bacterial strains found in liver tumor tissues was obtained by randomly picking bacterial colonies through 16S rRNA full-length sequencing. **G,** Representative IHC staining images of EMT markers E-cadherin, N-cadherin, and Snail in tumor tissues from GF RIL-175 allografts ($n = 6$/group). Scale bars, 100 μm. Statistical significance is assessed using one-way ANOVA with multiple comparisons. Data are presented as mean ± SD. *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$; ****, $P < 0.0001$.

and thus, multiple samplings from different nodules in this study provide a comprehensive representation of the entire host–bacteria ecosystem. Corroborating our metagenomic identification, we confirmed the existence of intratumoral bacteria in HCC using Gram staining and transmission electron microscopy. Both gram-positive and gram-negative bacteria were present, as evidenced by positive LTA and LPS staining, respectively.

Multifocal HCC could arise either metachronously or synchronously from IM or MO, and currently, there are no established criteria for distinguishing between the two. We leveraged our next-generation sequencing dataset and phylogenetic analysis to differentiate our cohort into IM-HCC, MO-HCC, and mixed IM/MO-HCC, showing that IM-HCC is the prevailing type of multifocal HCC, consistent with previous findings reported by us (4) and others (26). Based on our classification, we compared the microbiota composition in IM and MO nodules and observed a lower α-diversity in IM nodules. In keeping with our finding, a previous study has suggested that intratumoral diversity can serve as a predictor of cancer prognosis (27). A high microbial diversity of the intratumoral microbiome is associated with increased immune cell infiltration and better survival in pancreatic cancer (28). We further identified a panel of nine bacterial taxa for differentiating between IM and MO with an AUC of 0.795, indicating that bacterial biomarkers might aid in the discrimination of multifocal HCC. Further validation in independent cohorts is warranted to assess the translational potential of microbial biomarkers in guiding personalized therapeutic strategies.

We sought to elucidate the metagenomic features of IM-HCC that might exacerbate malignant phenotypes. Enriched intratumor bacteria in IM nodules, such as *Streptococcus*, *Enterococcus*, and *Fusobacterium*, have been associated with liver inflammation and carcinogenesis (29–31). Intratumoral *Streptococcus* species have been shown to promote the tumor metastasis cascade by enabling the survival of circulating tumor cells (32). A protein produced by commensal strains of *Enterococcus* can disrupt adherence junctions and promote intestinal inflammation (33). *Fusobacterium*, another IM-enriched bacterium, is a causative pathogen in colorectal cancer, and it promotes prometastatic potential in colorectal cancer and breast cancer (34, 35). This indicates that IM-enriched bacterial pathogens might contribute to cancer progression. Indeed, accumulating evidence has highlighted the role of the intratumoral microbiota in tumor metastasis by modulating cancer cell–intrinsic properties such as cancer stemness, EMT, adhesion programs, inflammation, and mechanical stress response (11, 36). Thus, the microbial features of IM-HCC are largely consistent with its propensity to metastasize compared with MO-HCC.

EMT is considered a crucial initiating step in tumor invasion and metastasis. In our analysis, EMT emerged as the top enriched pathway in IM nodules. Moreover, EMT-related genes were correlated with IM-enriched bacterial signatures. *CEACAMs* were positively correlated with *Streptococcus* species, including *S. anginosus* and *S. minor*, whereas *SLC2A1* is positively correlated with *Enterococcus* spp. *CEACAMs* have diverse functions in cell adhesion and intra- and intercellular signaling and are involved in cancer progression, angiogenesis, and metastasis (37). However, *SLC2A1*, also known as glucose transporter type 1, is associated with tumor progression, metastasis, and poor prognosis in multiple human solid tumors (38, 39). Functionally, the co-culture of *E. faecalis* and *S. anginosus* significantly enhanced the cell migration and invasion capability of HCC cells. These observations were validated in HCC orthotopic mouse models, in which *E. faecalis* and *S. anginosus* increased tumor growth and induced EMT *in vivo*. Tumor-resident microbes modulate the tumor immune microenvironment to promote tumor progression (40). Corroborating this notion, *E. faecalis* colonization in HCC tumors enriched MDSCs while impairing cytotoxic T-cell function. Together, intratumoral pathogens in IM nodules may contribute to metastatic potential in IM-HCC through EMT activation and restraining antitumor immunity.

An important question arises about the origin of these intrahepatic bacteria, as the liver has long been considered sterile. We further provided metagenomic and experimental evidence that specific intrahepatic microbiota in HCC tumors originally translocate from the gut to the liver via the gut–liver axis (41). This echoes our recent work showing the causality of HCC-associated gut dysbiosis in contributing to hepatocarcinogenesis (42).

In summary, our study demonstrated that microbial communities are heterogeneous in individual tumor nodules of patients with multifocal HCC and deciphered the tumor-bacteria cross-talk associated with HCC progression and IM. Our discoveries imply that intratumoral microbiota act as a potential contributor to disease progression in multifocal HCC, highlighting the necessity of considering bacterial heterogeneity as biomarkers and targets for therapeutic intervention.

METHODS

Patients and Sample Collection

A total of 242 HCC nodule tissues and 58 matched adjacent nontumor liver tissues were collected from 58 patients with multifocal HCC, with 3 to 11 tumor nodules and 1 paired adjacent nontumor liver tissue per patient, from the Department of Oncology at the First Affiliated Hospital, Sun Yat-sen University, from January 2013 to January 2020. None of the patients had undergone systemic therapy for their disease prior to the curative resection surgery. Multiple tumor nodule samples and one adjacent nontumor liver tissue sample were collected per patient. Adjacent nontumor tissue was collected outside the marginal zone, more than 5 cm away from the tumor nodules. Diagnosis was confirmed by two experienced pathologists. Detailed clinical information for the included patients is provided in Supplementary Table S1. Additionally, we collected stool samples ($n = 20$) and HCC tumor tissue samples ($n = 35$) from a cohort of 35 patients with single-lesion HCC from the First Affiliated Hospital, Sun Yat-sen University (42). Samples were collected in sterile microtubes, snap-frozen in liquid nitrogen, and kept at −80°C until DNA extraction. The study was approved by the Ethics Committee of the First Affiliated Hospital, Sun Yat-sen University [Protocol Number (2020) 353], and conducted under the ethical guidelines of the Declaration of Helsinki. Written informed consent was obtained from all enrolled patients.

DNA Extraction and 16S Sequencing

Microbial DNA was extracted from the tissue samples using cetyltrimethylammonium bromide according to the protocol adapted from Sambrook and Russell (43). The V3–V4 region of the bacterial 16S rRNA gene was amplified through PCR using the universal primers (341F: 5′-CCTACGGGNGGCWGCAG-3′ and 805R: 5′-GACT ACHVGGGTATCTAATCC-3′). PCR conditions included an initial denaturation at 98°C for 30 seconds, followed by 32 cycles of denaturation at 98°C for 10 seconds, annealing at 54°C for 30 seconds, extension at 72°C for 45 seconds, and finally, a 10-minute extension at 72°C. PCR products were confirmed with 2% agarose gel electrophoresis. AMPure XP beads (Beckman Coulter Genomics) were used

to purify and recover PCR products, and the recovered products were subjected to fluorescence quantification by Qubit (Invitrogen). The size and quantity of the amplicon library were assessed using the Agilent 2100 Bioanalyzer (Agilent) and the Library Quantification Kit for Illumina (Kapa Biosciences), respectively. The libraries were sequenced on the NovaSeq PE250 platform.

Contamination Controls

To control for potential contamination during sample collection and DNA extraction procedures, environmental and negative controls were incorporated into the study. Two Eppendorf tubes, each containing 1 mL of sterile PBS, were left open in the operating room during tissue sample collection to serve as environmental controls. In addition, four independent negative controls were taken from the reagents used as controls in the DNA extraction, library construction, and sequencing steps. Both environmental controls and negative controls were subjected to 16S rRNA gene sequencing. The decontam R package (version 1.20.0) was then employed to remove any contamination in the tissue samples at the ASV level by the "prevalence" method with a threshold of 0.5 ($P < 0.05$; ref. 44). We further removed all ASVs identified in negative controls from the tissue samples.

Sequence Curation and Annotation

Raw paired-end reads of 16S rRNA gene sequence data were quality-filtered and analyzed using QIIME 2 (version 2022.08; RRID:SCR_021258) software. The DADA2 algorithm was applied to reduce sequencing errors and dereplicate sequences using default parameters. Sequence reads were preprocessed by trimming to 400 base pairs and retaining all reads with <2 expected errors (maxEE = 2 in DADA2) before dereplicating reads that encoded the ASVs. The curated sequences were classified taxonomically using the Greengenes 16S rRNA gene reference database (RRID:SCR_002830) at a 99% identity cutoff by NCBI local alignment. Further microbial community analysis, including α- and β-diversity, was calculated with the phyloseq R package (version 1.44.0; RRID:SCR_013080) after rarefying. Alpha diversity was evaluated using Chao1 and Shannon indices, and the differences between groups were tested using the Wilcoxon rank-sum test. For β-diversity, principal coordinate analysis on the Bray–Curtis distance matrix was performed using the vegan R package (version 2.6.4; RRID:SCR_011950). The community dissimilarities among groups were determined by PERMANOVA with 1,000 iterations. Next, we summarized the ASV table at the species and genus taxonomic levels and performed prevalence and abundance-based filtering to retain taxa found at a 0.001 relative abundance in at least 10% of the samples. Due to the inconclusive annotation from 16S sequencing at the species level, particular genera were mapped to the species level using BLAST to align representative sequences to the NCBI nucleotide collection (nt; https://blast.ncbi.nlm.nih.gov/; RRID:SCR_011824). Resulting hits were sorted by max score, query cover, e-value, and percent identity.

Determination of HVMs

We defined a group of HVMs, in which abundance was higher in tumor nodule samples compared with corresponding adjacent nontumor tissue, as well as exhibiting high variance within tumor nodules of each patient. This method is reported in our previous publication (17). The relative abundance of microbiota in each biopsy was normalized to 1,000,000 reads using scaling normalization. To ensure that the normalized values were nonzero, a pseudocount of 1 was added prior to mathematical division and logarithmic transformation. Briefly, microbes were designated as HVMs if they met the following three criteria: (i) the normalized abundance of microbes at the genus level among nodule samples to $\log_2$ fold difference relative to the corresponding adjacent nontumor tissues from the same individual ≥1, (ii) the difference in relative abundance between nodule samples and adjacent nontumor tissues ≥0.1%, and (iii) the percent coefficient of variation of microbes in nodule samples ≥50. Microbes showing significant differences in abundance between nodules and adjacent nontumor tissues were identified based on criteria (i) and (ii), whereas criterion (iii) was used to detect microbes with large internodule abundance variation. Microbes that passed these filters were used to generate a heatmap for each patient. A patient-specific bacterial network mapping was then created by Cytoscape (version 3.10.1; RRID:SCR_003032) to illustrate both shared and unique HVMs within each individual's tumor nodules.

Differential Abundance Bacteria and Classification Model

The differentially abundant functional compositions of the microbial communities were analyzed by both the linear discriminant analysis effective size algorithm with the linear discriminant analysis cutoff >3 and the Wilcoxon rank-sum test with $P < 0.05$. A random forest classifier based on the selected differential abundance bacteria was implemented with the caret R package (version 6.0.94; RRID:SCR_021138). Samples were randomly split into training and validation cohorts with a 6:4 ratio. ROC and PR curves were generated using the pROC R package (version 1.18.5; RRID:SCR_024286) and the PRROC R package (version 1.3.1).

DNA Extraction and WES

Total DNA was extracted from the above frozen tissues using the QIAGEN DNeasy Blood & Tissue Kit (QIAGEN) following the manufacturer's protocol. The qualified genomic DNA of tumor and matched adjacent nontumor tissue from patients with HCC was fragmented to 200 to 300 bp by Covaris technology with resultant library, and then adapters were ligated to both ends of the fragments. Next, the extracted DNA was amplified by ligation-mediated PCR and then purified and hybridized to the NimbleGen or Agilent human exome array for enrichment. Nonhybridized fragments were washed out. Both noncaptured and captured ligation-mediated PCR products were subjected to real-time PCR to estimate the magnitude of enrichment. Each captured library was then loaded onto a HiSeq X Ten platform, and sequences of each individual were generated as 150 bp paired-end reads. The average sequencing depth was 205× for tumors.

Reads Mapping and Variation Detection

Cutadapt (version 1.18; RRID:SCR_011841) and FastQC (version 0.11.9; RRID:SCR_014583) were used to remove adapter sequences to obtain raw sequencing reads and perform quality control, and then, high-quality reads were mapped to the NCBI human reference genome (hg19) using Burrows–Wheeler Aligner (RRID:SCR_010910; ref. 45) with default parameters. We performed local realignment of the original BAM alignment using GATK (RRID:SCR_001876; ref. 46), followed by Picard to mark duplicate reads. Somatic substitutions were detected by MuTect (version 1.1.4; RRID:SCR_000559; ref. 47) based on Burrows–Wheeler Aligner alignment, and high-confidence somatic single-nucleotide variants were called if the following criteria were met: (i) Both the tumor and nontumor samples should be covered sufficiently (≥10×) at the genomic position; (ii) the variants should be supported by at least 5% of the total reads in the tumor while being less than 1% in the normal; and (iii) the variants should be supported by at least five reads in the tumor. High-confidence somatic insertions and deletions (indels) were detected using Strelka (version 2.8.4; RRID:SCR_005109; ref. 48) and called using the following steps: (i) Candidate somatic indels were predicted with the GATK somatic indel detector with default parameters and (ii) high-confidence somatic indels were defined after filtering germline events. All high-confidence somatic mutations were filtered

out by the Single Nucleotide Polymorphism Database (version 135; RRID:SCR_002338) site, which is commonly polymorphic without known medical impact. The remaining mutations were annotated with ANNOVAR (RRID:SCR_012821; ref. 49) and subjected to subsequent analyses. The specificity of this method has been validated in a previous study (4).

Identification of Potential Driver Mutations and the Intertumor Heterogeneity Evaluation

To identify potential driver mutations in this study, we first nominated potential cancer genes by comparing all mutated genes against: (i) COSMIC Cancer Gene Census (December 2016; ref. 50), (ii) mutation data provided by Gao and colleagues (51), and (iii) The Cancer Genome Atlas HCC sequencing project (detailed information in Supplementary Table S2; ref. 52). Next, all nonsynonymous variants in these potential cancer genes were evaluated. The ITH index was measured by calculating the ratio of shared variants to all variants between tumor nodules (Jaccard similarity coefficient) within each patient, in which no shared mutations were recorded as 0, whereas completely identical mutations were recorded as 1 between two tumors (53).

Construction of Phylogenetic Tree and Clonal Evolution

Phylogenetic trees for each patient were constructed using MEGA 11 (RRID:SCR_000667; ref. 54). Sequences of 20 bp in length surrounding the somatic nonsynonymous mutations (single-nucleotide variants and indels) were extracted to construct the phylogenetic trees of each patient based on the maximum parsimony algorithm. All phylogenetic trees were optimized using Adobe Illustrator (RRID:SCR_010279). The genomic distance was calculated by taking the Euclidean distance of the mutations between tumor regions. Mutations present in any tumor from one patient were turned into a binary matrix, in which the rows were mutations and the columns were tumors. This matrix was clustered, and the pairwise distance between any two tumors was determined. We applied PyClone (version 0.13.1; RRID: SCR_016873; ref. 55) to perform subclonal analysis.

TMB and Mutational Signature Analysis

TMB of a tumor sample was calculated by the number of nonsynonymous somatic mutations per megabase in coding regions. We inferred mutational signatures (C>A, C>G, C>T, T>A, T>C, and T>G) through linear decomposition using the R package deconstructSigs (version 1.8.0) with default parameters.

RNA Extraction and RNA Sequencing

Beads (Invitrogen) with oligo (dT) were used to isolate poly(A) mRNA after total RNA was collected. Fragmentation buffer was added to interrupt mRNA into short fragments. Using these short fragments as templates, a random hexamer primer was used to synthesize the first-strand cDNA. The second-strand cDNA was synthesized using buffer, dNTPs, RNase H, and DNA polymerase I. Short fragments were purified with the QIAquick PCR extraction kit (QIAGEN), resolved with EB buffer for end reparation and adding poly(A), and connected with sequencing adaptors. For amplification with PCR, we selected suitable fragments as templates according to agarose gel electrophoresis analysis. Finally, the library was sequenced using the HiSeq X Ten platform, and 150 bp paired-end reads were generated.

RNA-seq Data Analysis

For host gene expression, the biomaRt R package (version 2.56.1; RRID:SCR_019214) was used to filter data for protein-coding genes. We filtered out lowly expressed genes to retain those that are expressed in at least half of the samples. Filtered gene expression read count data was performed variance stabilizing transformation using the R package DESeq2 (version 1.40.2; RRID:SCR_000154). We used the 25% quantile of variance across samples as a cutoff to filter out genes with low variance. Performing these steps for RNA-seq data resulted in a unique host gene expression matrix, including 12,126 genes for downstream analysis. Differentially expressed genes were calculated using DESeq2. An adjusted $P < 0.05$ and $|\log_2 FC| > 1$ (where FC = fold change) were considered to be statistically significant. The clusterProfiler R package (version 4.8.3; RRID:SCR_016884) was used for gene set enrichment analysis and visualization. Hallmark gene sets used in this article were downloaded from the Molecular Signatures Database (RRID:SCR_016863). The EMT-related gene was collected from the dbEMT database (RRID:SCR_024692; ref. 56).

Mice

SPF male C57BL/6J mice were purchased from GemPharmatech Co., Ltd. and raised at the First Affiliated Hospital of Sun Yat-sen University. Germ-free male C57BL/6J mice were bred and housed at the Germ-Free Mouse Research Facility, GemPharmatech Co., Ltd. Mice were raised in SPF or sterile facilities, following a cycle of 12 hours of light and 12 hours of dark conditions. Food (for SPF mice: 1010058, Jiangsu Xietong; for germ-free mice: 1019018, Jiangsu Xietong) and sterile water were provided *ad libitum*. Mice were randomly assigned to various experimental groups and experimental treatments. All mice were euthanized by pentobarbital sodium and cervical dislocation.

Bacteria Isolation, Culture, and Strain Identification

E. faecalis was isolated from tumor tissues of patients with HCC. *S. anginosus* and *F. nucleatum* were purchased from the ATCC (ATCC 33397 and ATCC 25586, respectively). *E. coli* MG1655 (ATCC 700926) was used as a control for *E. faecalis* and *S. anginosus*. *E. faecalis*, *S. anginosus*, and *E. coli* were cultured on brain heart infusion (BHI) agar (CM1136B; Thermo Fisher Scientific) plates and maintained in BHI broth (CM1135B; Thermo Fisher Scientific) medium in an anaerobic jar (Hardy Diagnostics) at 37°C. *F. nucleatum* was cultured in fluid thioglycollate medium (HB5191, Hope Bio-Technology) in an anaerobic jar (C-31, Mitsubishi Gas Chemical). Anaerobic conditions were created by AnaeroGen (Thermo Fisher Scientific).

Fresh liver tissues from mice were collected in sterile tubes with BHI broth under aerobic and anaerobic conditions. The tissues were washed six times with sterile PBS and chopped into pieces. Tissue pieces were homogenized using TissueLyser II (85300, QIAGEN), and 100 μL of the tissue homogenate was directly spread onto different agar plates, including blood agar plates, chocolate agar plates, MacConkey agar plates, and Columbia blood agar plates, in both anaerobic and aerobic conditions at 37°C. Colony-forming units per mg of tissue were counted. Full-length 16S rRNA sequencing was used for strain identification.

Cell Lines

Human HCC cell line Hep3B was obtained from the ATCC (RRID: CVCL_0326), and Huh7 (RRID: CVCL_0336) was purchased from the Japanese Collection of Research Bioresources Cell Bank. Both cell lines were grown in DMEM (Thermo Fisher Scientific), supplemented with 10% FBS and 1% penicillin/streptomycin in a humidified atmosphere containing 5% CO_2. The murine hepatoma cell line Hepa1-6 was obtained from the ATCC (RRID: CVCL_0327), and the RIL-175 cells were kindly provided by Dr. Greten and Prof. Zender. Hepa1-6 cells were cultured in DMEM with 10% FBS and 1% penicillin/streptomycin, whereas RIL-175 cells were maintained in RPMI 1640 medium containing 10% FBS and 1% penicillin/streptomycin. Cells were tested monthly for *Mycoplasma* using the MycoBlue Mycoplasma Detector (Vazyme, #D101).

SPF and Germ-Free Orthotopic and Subcutaneous Mouse Models

For the SPF orthotopic mouse model, mouse HCC cell lines (RIL-175) were orthotopically implanted into the left liver lobe of 4- to 6-week-old male C57BL/6J mice (5×10^5 cells in 10–20 μL PBS). For the SPF subcutaneous mouse models, RIL-175 cells were subcutaneously injected into the right flanks of 4- to 6-week-old male C57BL/6J mice (1×10^6 cells in 100 μL PBS). For germ-free mice, 6- to 8-week-old male C57BL/6J mice were selected for bacterial gavage. For the germ-free orthotopic mouse model, mouse HCC cell lines (RIL-175 or Hepa1-6) were orthotopically implanted into the left liver lobe of 4- to 6-week-old male C57BL/6J mice (1×10^5 cells in 10–20 μL PBS). The number of mice in each group is shown in the figure legends. At the endpoint, mice were anesthetized with sodium pentobarbital intraperitoneally. Tumor volume was measured by length (a) and width (b) using a digital caliper and calculated as tumor volume = $a \times b^2/2$. Tumor burden was the summation of tumor volumes of each liver. Liver tissues were collected in 4% paraformaldehyde and 2% glutaraldehyde or snap-frozen. For bacterial culture, fresh liver tissues were collected in the biosafety cabinet. Fresh liver tissues were collected for flow cytometry.

Antibiotic Treatments on SPF Orthotopic and Subcutaneous Mouse Models

To specifically eliminate intratumor microbiota, we selected a combination of ampicillin and gentamicin (MCE) to eliminate bacteria (32). For the orthotopic mouse model, these antibiotics were administered to mice through drinking water at a dose of 1 mg/mL. For the subcutaneous mouse model, ampicillin and gentamicin were administered via intratumoral injection at a dose of 2 mg/mL. All animal experiments were approved by the Animal Experimentation Ethics Committee of the First Affiliated Hospital of Sun Yat-sen University [No. (2024)179].

Flow Cytometry Analysis of Immune Cells

Tissue samples were dissociated into single cells using RPMI 1640 (Gibco, C11875500BT), and red blood cells were eliminated with a red blood cell lysis buffer. The cells were then filtered through a 70-μm cell strainer and resuspended in a staining buffer. Cells were incubated with Fixable Viability Stain 700 (BD Biosciences, #564997) for 15 minutes at room temperature, followed by incubation with Fc Receptor Blocking Solution (BD Biosciences, #553141) for 15 minutes at 4°C and stained with anti-CD45 (BD Biosciences, #564279, 1:400), anti-CD11b (BioLegend, #101206, 1:100), anti-Gr1 (BD Biosciences, #740850, 1:200), anti-Ly6C (BioLegend, #128016, 1:200), and anti-Ly6G (BD Biosciences, #560601, 1:200) for 30 minutes at 4°C in the dark. For T cells, after stimulation with the Leukocyte Activation Cocktail containing BD GolgiPlug (BD Biosciences, #550583) for 4 to 6 hours at 37°C, surface marker staining was performed at 4°C for 30 minutes with anti-CD45 (BD Biosciences, #564279, 1:400), anti-CD3e (BD Biosciences, #557596, 1:200), anti-CD4 (BD Biosciences, #563106, 1:200), and anti-CD8 (BioLegend, #100734, 1:200). For the assessment of intranuclear cytokines, cells were permeabilized with BD fixation and permeabilization solution (BD Biosciences, #554722) and stained intracellularly with anti-IFNγ (BD Biosciences, #564336, 1:100) and anti-granzyme B (Thermo Fisher Scientific, #25-8898-82, 1:100) antibodies at 4°C for 1 hour. After washing, the samples were analyzed using a Cytek Aurora (Cytek Biosciences), and the data were processed using FlowJo version 10.

IHC and Gram Stain Assays

FFPE HCC tissue samples were subjected to IHC and Gram staining. The FFPE slides were deparaffinized, rehydrated, and were stained with bacterial LPS antibody [mAb WN1 222-5; #HM6011; RRID: AB_2750644; 1:200 dilution; Hycult Biotech], bacterial LTA antibody (mAb 55; #HM2048; RRID: AB_57466; 1:400 dilution; Hycult Biotech), E-cadherin (Cell Signaling Technology, Cat# 3195S, 1:1,000, RRID:AB_2291471), N-cadherin (Cell Signaling Technology, Cat# 13116S, 1:1,000, RRID:AB_2687616), Snail (ServiceBio, GB11260, 1:200), and Ki-67 (Cell Signaling Technology, Cat# 12202S, 1:1,000, RRID:AB_2620142). The positive controls for LPS and LTA staining were performed on colon tissue according to the manufacturer's instructions. Bacterial Gram stain was performed using a Gram Stain Kit (Modified Brown & Brenn; #BTN160634; Baiao Laibo Technology). Slides were inspected under a fluorescence microscope at 50× and 100× (Olympus, BX63F #9072693). Quantitative analysis was performed using ImageJ (version 1.51, NIH).

Transmission Electron Microscopy

Fresh HCC tissues were fixed with 2.5% glutaraldehyde containing 0.1 mol/L sodium cacodylate. Samples were then fixed using 1% osmium tetroxide, followed by dehydration with an increasing concentration gradient of ethanol and propylene oxide. Samples were then embedded, cut into 50-nm sections, and stained with 3% uranyl acetate and lead citrate. Images were acquired using a JEM-1200 electron microscope (JEOL).

FISH

FITC-labeled probes for specific bacterial taxa (Supplementary Table S7) were used to detect bacteria in paraffin-embedded liver tissue slides. Hybridization was performed with 10 pmol of oligonucleotide probe in 10 μg/mL buffer (0.9 mol/L NaCl, 20 mmol/L Tris/HCl, 0.01% SDS, and 10% formamide) at 40°C for 24 hours in a dark, humidified chamber. Slides were then incubated in preheated washing buffer (0.9 mol/L NaCl, 20 mmol/L Tris/HCl, pH 7.2) at 55°C for 20 minutes, washed gently in PBS three times, and mounted with DAPI–antifade solution (ab104139, Abcam). FISH images were captured by a fluorescent microscope (BX63F, Olympus).

PCR

Bacterial genomic DNA from mouse liver tissue was extracted using the Quick-DNA Fecal/Soil Microbe Miniprep kit (D6010, Zymo Research). PCR was performed to determine the presence of specific bacterial taxa (Supplementary Table S7) in mouse liver using Premix Taq DNA Polymerase (R028A, Takara Bio). The PCR mix contained 100 ng of DNA in a final 10 μL reaction volume. The PCR was performed according to the manufacturer's instructions. After the reaction, PCR products were used for agarose gel electrophoresis analysis.

Wound-Healing Assay

For live bacteria co-culture, 5×10^5 HCC cells per well were seeded in a six-well plate overnight. Cells were then co-cultured with bacteria (multiplicity of infection = 50) under anaerobic conditions for 20 hours, after which medium supplemented with gentamicin (200 μg/mL) was added for 1 hour to kill the bacteria. After 24 hours, confluent cultures were scratched with sterile 1,000 μL pipette tips. At specified intervals, wounds were observed under a light microscope. Wound closure was determined by TScratch software.

Transwell Migration and Invasion Assays

For the migration assay, HCC cells co-cultured with bacteria under identical conditions as mentioned above were stained with 0.4% Trypan Blue solution (Thermo Fisher Scientific) and counted under a light microscope. Next, 2×10^5 HCC cells were resuspended in 200 μL serum-free medium and seeded in Falcon transwell inserts (8.0 μm pore size). Complete DMEM with 10% FBS and 1% penicillin/streptomycin was added to the lower chamber. After 48 to 96 hours,

transwell inserts were fixed with 4% paraformaldehyde for 15 minutes and stained with crystal violet for 15 minutes. For the invasion assay, 2×10^5 HCC cells were seeded into Matrigel-coated transwell inserts (Corning), and after 72 to 120 hours, the transwell inserts were fixed and stained. Migrated/invaded cells were observed under a light microscope. Five random fields were counted for each sample.

CODEX Staining, Image Processing, and Cell Segmentation

The FFPE tissue sections from IM-HCC ($N = 6$) and MO-HCC ($N = 4$) tumor nodules were stained using the CODEX (Akoya Biosciences) staining kit following the manufacturer's instructions. Briefly, tissue sections were deparaffinized, rehydrated, subjected to antigen unmasking using 1× citrate buffer (0.01 mol/L), and equilibrated in CODEX staining buffer. An antibody cocktail was prepared by mixing CODEX blocking buffer with predetermined conjugated antibodies to a total volume of 200 μL, which was added to the sections followed by incubation at room temperature for 3 hours (Supplementary Table S9). After poststaining fixing, the sections were transferred to a storage buffer and prepared for CODEX imaging. The CODEX Instrument Manager software was used to obtain raw images at a resolution of 0.377 μm per pixel, which were further preprocessed by the CODEX Processor to align images across cycles, align and stitch tiles across regions, and subtract autofluorescence before downstream analysis. The processed images were visualized using the Multiple Analysis Viewer in ImageJ (version 2.3.0). High-quality regions within the processed CODEX images were selected for single-cell recognition (cell segmentation), which was implemented in QuPath (version 0.5.1; RRID:SCR_018257; ref. 57) with the StarDist algorithm.

Quality Control and Definition of Cell Type

Cells with the sum of the intensity of cell nucleus dyes above the first percentile were included for further analysis. All mass channels per image were individually censored at the 99th percentile to remove outliers. Then, we scaled the data using Seurat (version 5.2.1; RRID:SCR_016341) and removed batch effects using ComBat in the sva package (version 3.48.0; RRID:SCR_012836). For cell type definition, we used 15 markers to cluster the retained cells using the Rphenograph R package (version 0.99.1; RRID:SCR_016919; k = 20) and defined nine cell types according to the expression of markers (Supplementary Fig. S9). The k value corresponding to the "elbow point" was selected with reference to the PG_elbow function.

Statistical Analysis

Statistical analyses and graphical visualization were performed using open-source R software (version 4.3.0) or GraphPad Prism 9 software. Quantitative data between groups were compared using the Student t test or the Wilcoxon rank-sum test. Categorical data were analyzed using χ^2 or Fisher's exact test. The Mantel test was performed to verify the overall correlation between dissimilarity matrices of host genomic alteration (Euclidean distance) and gut microbiome abundance (Bray–Curtis distance) using the vegan R package. Spearman correlation was used to identify microbial taxa and host gene correlation. One-way ANOVA was used to compare the differences among multiple groups. All statistical tests were two-sided, and a P value <0.05 was considered statistically significant. To account for multiple tests, P values were adjusted using the Benjamini–Hochberg FDR correction.

Data Availability

The raw 16S rRNA sequencing data of liver tissues from patients with multilesion HCC reported in this article have been deposited at the Genome Sequence Archive of the China National Center for Bioinformation (CNCB) under accession number CRA023894. The 16S rRNA sequencing data of liver tissues and stools from patients with single-lesion HCC have been deposited at the Genome Sequence Archive of CNCB under accession numbers CRA019234 and CRA019232, respectively. The processed expression matrix and annotation table of RNA-seq and WES data have been deposited into the Open Archive for Miscellaneous Data database of CNCB with accession IDs: OMIX009938 and OMIX009942, respectively. The code for analysis in this study is available on GitHub (https://github.com/YinghongLu/mHCC).

Authors' Disclosures

No disclosures were reported.

Authors' Contributions

Y. Lu: Data curation, formal analysis, investigation, methodology, writing–original draft, writing–review and editing. **L. Xu:** Resources, data curation, investigation. **W. Chen:** Data curation. **W. Liu:** Formal analysis. **Ying Zhang:** Data curation. **Q. Zhou:** Data curation. **N. Wang:** Data curation. **Yongxin Zhang:** Data curation. **H. Bai:** Data curation. **S. Xu:** Data curation. **P. Huang:** Data curation. **K. Fu:** Data curation. **W. Xie:** Data curation. **X. Liu:** Data curation. **X. Wang:** Data curation. **C.C. Wong:** Writing–review and editing. **M. Kuang:** Resources, supervision, funding acquisition. **J. Yu:** Conceptualization, resources, supervision, project administration.

Acknowledgments

This study was supported by the Noncommunicable Chronic Diseases-National Science and Technology Major Project (2023ZD0500200), ITF Guangdong-Hong Kong Technology Cooperation Funding Scheme (GHP/086/21GD), National Natural Science Foundation of China (82372914), key projects in Guangdong Province (2022B1515120031), RGC Research Impact Fund Hong Kong (R4032-21F), Strategic Seed Funding for Collaborative Research Scheme CUHK (3133344), Strategic Impact Enhancement Fund at the CUHK (3135509), and Impact Case for RAE CUHK (3134277).

Note

Supplementary data for this article are available at Cancer Discovery Online (http://cancerdiscovery.aacrjournals.org/).

Received August 30, 2024; revised April 1, 2025; accepted April 24, 2025; posted first April 27, 2025.

REFERENCES

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2021;71:209–49.
2. Rumgay H, Ferlay J, de Martel C, Georges D, Ibrahim AS, Zheng R, et al. Global, regional and national burden of primary liver cancer by subtype. Eur J Cancer 2022;161:108–18.
3. Xie D-Y, Fan H-K, Ren Z-G, Fan J, Gao Q. Identifying clonal origin of multifocal hepatocellular carcinoma and its clinical implications. Clin Transl Gastroenterol 2019;10:e00006.
4. Xu LX, He MH, Dai ZH, Yu J, Wang JG, Li XC, et al. Genomic and transcriptional heterogeneity of multifocal hepatocellular carcinoma. Ann Oncol 2019;30:990–7.
5. Dong L-Q, Peng L-H, Ma L-J, Liu D-B, Zhang S, Luo S-Z, et al. Heterogeneous immunogenomic features and distinct escape mechanisms in multifocal hepatocellular carcinoma. J Hepatol 2020;72:896–908.
6. Tabrizian P, Jibara G, Shrager B, Schwartz M, Roayaie S. Recurrence of hepatocellular cancer after resection: patterns, treatments, and prognosis. Ann Surg 2015;261:947–55.

7. Sherman M. Recurrence of hepatocellular carcinoma. N Engl J Med 2008;359:2045–7.
8. Shi J-Y, Xing Q, Duan M, Wang Z-C, Yang L-X, Zhao Y-J, et al. Inferring the progression of multifocal liver cancer from spatial and temporal genomic heterogeneity. Oncotarget 2016;7:2867–77.
9. Zhou S-L, Zhou Z-J, Song C-L, Xin H-Y, Hu Z-Q, Luo C-B, et al. Whole-genome sequencing reveals the evolutionary trajectory of HBV-related hepatocellular carcinoma early recurrence. Signal Transduct Target Ther 2022;7:24.
10. Sun Y, Wu P, Zhang Z, Wang Z, Zhou K, Song M, et al. Integrated multi-omics profiling to dissect the spatiotemporal evolution of metastatic hepatocellular carcinoma. Cancer Cell 2024;42:135–56.e17.
11. Fu A, Yao B, Dong T, Cai S. Emerging roles of intratumor microbiota in cancer metastasis. Trends Cell Biol 2023;33:583–93.
12. Yang L, Li A, Wang Y, Zhang Y. Intratumoral microbiota: roles in cancer initiation, development and therapeutic efficacy. Signal Transduct Target Ther 2023;8:35.
13. Nejman D, Livyatan I, Fuks G, Gavert N, Zwang Y, Geller LT, et al. The human tumor microbiome is composed of tumor type-specific intracellular bacteria. Science 2020;368:973–80.
14. Liu B, Zhou Z, Jin Y, Lu J, Feng D, Peng R, et al. Hepatic stellate cell activation and senescence induced by intrahepatic microbiota disturbances drive progression of liver cirrhosis toward hepatocellular carcinoma. J Immunother Cancer 2022;10:e003069.
15. Li S, Xia H, Wang Z, Zhang X, Song T, Li J, et al. Intratumoral microbial heterogeneity affected tumor immune microenvironment and determined clinical outcome of HBV-related HCC. Hepatology 2023;78:1079–91.
16. Sun L, Ke X, Guan A, Jin B, Qu J, Wang Y, et al. Intratumoural microbiome can predict the prognosis of hepatocellular carcinoma after surgery. Clin Transl Med 2023;13:e1331.
17. Liu W, Zhang X, Xu H, Li S, Lau HC-H, Chen Q, et al. Microbial community heterogeneity within colorectal neoplasia and its correlation with colorectal carcinogenesis. Gastroenterology 2021;160:2395–408.
18. Ozenne B, Subtil F, Maucort-Boulch D. The precision–recall curve overcame the optimism of the receiver operating characteristic curve in rare diseases. J Clin Epidemiol 2015;68:855–9.
19. Liu S, Roemer F, Ge Y, Bedrick EJ, Li ZM, Guermazi A, et al. Comparison of evaluation metrics of deep learning for imbalanced imaging data in osteoarthritis studies. Osteoarthritis Cartilage 2023;31:1242–8.
20. Torrecilla S, Sia D, Harrington AN, Zhang Z, Cabellos L, Cornella H, et al. Trunk mutational events present minimal intra- and inter-tumoral heterogeneity in hepatocellular carcinoma. J Hepatol 2017;67:1222–31.
21. Xue R, Chen L, Zhang C, Fujita M, Li R, Yan SM, et al. Genomic and transcriptomic profiling of combined hepatocellular and intrahepatic cholangiocarcinoma reveals distinct molecular subtypes. Cancer Cell 2019;35:932–47.e8.
22. Ding X, He M, Chan AWH, Song QX, Sze SC, Chen H, et al. Genomic and epigenomic features of primary and recurrent hepatocellular carcinomas. Gastroenterology 2019;157:1630–45.e6.
23. Ng CKY, Dazert E, Boldanova T, Coto-Llerena M, Nuciforo S, Ercan C, et al. Integrative proteogenomic characterization of hepatocellular carcinoma across etiologies and stages. Nat Commun 2022;13:2436.
24. Boleij A, Tjalsma H. The itinerary of Streptococcus gallolyticus infection in patients with colonic malignant disease. Lancet Infect Dis 2013;13:719–24.
25. Fu K, Cheung AHK, Wong CC, Liu W, Zhou Y, Wang F, et al. Streptococcus anginosus promotes gastric inflammation, atrophy, and tumorigenesis in mice. Cell 2024;187:882–96.e17.
26. Xue R, Li R, Guo H, Guo L, Su Z, Ni X, et al. Variable intra-tumor genomic heterogeneity of multiple lesions in patients with hepatocellular carcinoma. Gastroenterology 2016;150:998–1008.
27. Gopalakrishnan V, Spencer CN, Nezi L, Reuben A, Andrews MC, Karpinets TV, et al. Gut microbiome modulates response to anti-PD-1 immunotherapy in melanoma patients. Science 2018;359:97–103.
28. Riquelme E, Zhang Y, Zhang L, Montiel M, Zoltan M, Dong W, et al. Tumor microbiome diversity and composition influence pancreatic cancer outcomes. Cell 2019;178:795–806.e12.
29. Ponziani FR, Bhoori S, Castelli C, Putignani L, Rivoltini L, Del Chierico F, et al. Hepatocellular carcinoma is associated with gut microbiota profile and inflammation in nonalcoholic fatty liver disease. Hepatology 2019;69:107–20.
30. Zhong X, Cui P, Jiang J, Ning C, Liang B, Zhou J, et al. Streptococcus, the predominant bacterium to predict the severity of liver injury in alcoholic liver disease. Front Cell Infect Microbiol 2021;11:649060.
31. He Y, Zhang Q, Yu X, Zhang S, Guo W. Overview of microbial profiles in human hepatocellular carcinoma and adjacent nontumor tissues. J Transl Med 2023;21:68.
32. Fu A, Yao B, Dong T, Chen Y, Yao J, Liu Y, et al. Tumor-resident intracellular microbiota promotes metastatic colonization in breast cancer. Cell 2022;185:1356–72.e26.
33. Steck N, Hoffmann M, Sava IG, Kim SC, Hahne H, Tonkonogy SL, et al. Enterococcus faecalis metalloprotease compromises epithelial barrier and contributes to intestinal inflammation. Gastroenterology 2011;141:959–71.
34. Chen S, Zhang L, Li M, Zhang Y, Sun M, Wang L, et al. Fusobacterium nucleatum reduces METTL3-mediated m^{6}A modification and contributes to colorectal cancer metastasis. Nat Commun 2022;13:1248.
35. Parhi L, Alon-Maimon T, Sol A, Nejman D, Shhadeh A, Fainsod-Levi T, et al. Breast cancer colonization by Fusobacterium nucleatum accelerates tumor growth and metastatic progression. Nat Commun 2020;11:3259.
36. Battaglia TW, Mimpen IL, Traets JJH, van Hoeck A, Zeverijn LJ, Geurts BS, et al. A pan-cancer analysis of the microbiome in metastatic cancer. Cell 2024;187:2324–35.e19.
37. Beauchemin N, Arabzadeh A. Carcinoembryonic antigen-related cell adhesion molecules (CEACAMs) in cancer progression and metastasis. Cancer Metastasis Rev 2013;32:643–71.
38. Yan S, Wang Y, Chen M, Li G, Fan J. Deregulated SLC2A1 promotes tumor cell proliferation and metastasis in gastric cancer. Int J Mol Sci 2015;16:16144–57.
39. Yang S, Qian L, Li Z, Li Y, Bai J, Zheng B, et al. Integrated multi-omics landscape of liver metastases. Gastroenterology 2023;164:407–23.e17.
40. Liu Y, Wong CC, Ding Y, Gao M, Wen J, Lau HC, et al. Peptostreptococcus anaerobius mediates anti-PD1 therapy resistance and exacerbates colorectal cancer via myeloid-derived suppressor cells in mice. Nat Microbiol 2024;9:1467–82.
41. Albillos A, de Gottardi A, Rescigno M. The gut-liver axis in liver disease: pathophysiological basis for therapy. J Hepatol 2020;72:558–77.
42. Wang X, Fang Y, Liang W, Cai Y, Wong CC, Wang J, et al. Gut-liver translocation of pathogen Klebsiella pneumoniae promotes hepatocellular carcinoma in mice. Nat Microbiol 2025;10:169–84.
43. Sambrook J. Molecular cloning : a laboratory manual. In: Russell DW, editor. 3rd ed. ed. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory; 2001.
44. Davis NM, Proctor DM, Holmes SP, Relman DA, Callahan BJ. Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data. Microbiome 2018;6:226.
45. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics 2010;26:589–95.
46. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 2010;20:1297–303.
47. Cibulskis K, Lawrence MS, Carter SL, Sivachenko A, Jaffe D, Sougnez C, et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. Nat Biotechnol 2013;31:213–9.
48. Saunders CT, Wong WS, Swamy S, Becq J, Murray LJ, Cheetham RK. Strelka: accurate somatic small-variant calling from sequenced tumor-normal sample pairs. Bioinformatics 2012;28:1811–7.

49. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res 2010;38:e164.
50. Futreal PA, Coin L, Marshall M, Down T, Hubbard T, Wooster R, et al. A census of human cancer genes. Nat Rev Cancer 2004;4:177–83.
51. Gao Q, Wang ZC, Duan M, Lin YH, Zhou XY, Worthley DL, et al. Cell culture system for analysis of genetic heterogeneity within hepatocellular carcinomas and response to pharmacologic agents. Gastroenterology 2017;152:232–42.e4.
52. Cancer Genome Atlas Research Network. Comprehensive and integrative genomic characterization of hepatocellular carcinoma. Cell 2017;169:1327–41.e23.
53. Zhang Y, Chang L, Yang Y, Fang W, Guan Y, Wu A, et al. Intratumor heterogeneity comparison among different subtypes of non-small-cell lung cancer through multi-region tissue and matched ctDNA sequencing. Mol Cancer 2019;18:7.
54. Tamura K, Stecher G, Kumar S. MEGA11: molecular evolutionary genetics analysis version 11. Mol Biol Evol 2021;38:3022–7.
55. Roth A, Khattra J, Yap D, Wan A, Laks E, Biele J, et al. PyClone: statistical inference of clonal population structure in cancer. Nat Methods 2014;11:396–8.
56. Zhao M, Liu Y, Zheng C, Qu H. dbEMT 2.0: an updated database for epithelial-mesenchymal transition genes with experimentally verified information and precalculated regulation information for cancer metastasis. J Genet Genomics 2019;46:595–7.
57. Bankhead P, Loughrey MB, Fernández JA, Dombrowski Y, McArt DG, Dunne PD, et al. QuPath: open source software for digital pathology image analysis. Sci Rep 2017;7:16878.