Multimodule characterization of immune subgroups in intrahepatic cholangiocarcinoma reveals distinct therapeutic vulnerabilities

Jian Lin, Yuting Dai, Chen Sang, Guohe Song, Bin Xiang, Mao Zhang, Liangqing Dong, Xiaoli Xia, Jiaqiang Ma, Xia Shen, Shuyi Ji, Shu Zhang, Mingjie Wang, Hai Fang, Xiaoming Zhang, Xiangdong Wang, Bing Zhang, Jian Zhou, Jia Fan, Hu Zhou, Daming Gao, Qiang Gao

ABSTRACT

Background Immune microenvironment is well recognized as a critical regulator across cancer types, despite its complex roles in different disease conditions. Intrahepatic cholangiocarcinoma (iCCA) is characterized by a tumor-reactive milieu, emphasizing a deep insight into its immunogenomic profile to provide prognostic and therapeutic implications.

Methods We performed genomic, transcriptomic, and proteomic characterization of 255 paired iCCA and adjacent liver tissues. We validated our findings through H&E staining (n=177), multiplex immunostaining (n=188), single-cell RNA sequencing (scRNA-seq) (n=10), in vitro functional studies, and in vivo transposon-based mouse models.

Results Integrated multimodule data identified three immune subgroups with distinct clinical, genetic, and molecular features, designated as GI1 (immunosuppressive, 25.1%), GI2 (immunoevasion, 42.7%), and GI3 (immunostimulatory, 32.2%). GI1 was characterized by excessive infiltration of neutrophils and immature dendritic cells (DCs). The hallmark of GI2 was the relatively higher tumor-proliferative activity and tumor purity. GI3 exhibited an enrichment of adaptive immune cells, natural killer cells, and activated DCs. These immune subgroups were significantly associated with prognosis and validated in two independent cohorts. Tumors with KRAS mutations were enriched in GI1 and associated with myeloid inflammation-dominated immunosuppression. Although tumor mutation burden was relatively higher in GI2, loss of heterozygosity in human leucocyte antigen and defects in antigen presentation undermined the recognition of neoantigens, contributing to immune-exclusion behavior. Pathological analysis confirmed that tumor-infiltrating lymphocytes and tertiary lymphoid structures were both predominant in GI3. Hepatitis B virus (HBV)-related samples tended to be under-represented in GI1, and scRNA-seq analyses implied that HBV infection indeed alleviated myeloid inflammation and reinvigorated antitumor immunity.

Conclusions Our study elucidates that the immunogenomic traits of iCCA are intrinsically heterogeneous among patients, posing great challenge and opportunity for the application of personalized immunotherapy.

WHAT IS ALREADY KNOWN ON THIS TOPIC

- Intrahepatic cholangiocarcinoma (iCCA) represents the second most common type of primary liver cancer, which is generally featured by a highly desmoplastic tumor microenvironment (TME). Sequencing efforts have improved the understanding of the molecular pathogenesis of iCCA; however, the comprehensive landscape of iCCA microenvironment phenotype remains largely unexplored.

WHAT THIS STUDY ADDS

- Our multimodule study reveals three distinct immune subgroups (GI1, GI2, and GI3) with specific immunological and non-immune profiles associated with different patient outcomes. Myeloid inflammation/KRAS mutation, antigen presentation/proliferation, and immune checkpoint could serve as potential therapeutic targets for the treatments of GI1, GI2, and GI3-related iCCA, respectively. Hepatitis B virus (HBV) infection could probably alleviate myeloid inflammation in iCCA, thus activates antitumor immunity and improves the efficacy of immunotherapies.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

- These findings extend our understanding of the immunological profile of iCCA. Defined immune subgroups provide novel insights into the heterogeneous TME and underlying therapeutic strategies.

INTRODUCTION

Intrahepatic cholangiocarcinoma (iCCA) accounts for 10%–15% of all primary liver cancer, with a gradual increase over the past decades. Despite the high postoperative recurrence rate, surgery is prior for patients diagnosed at an early stage. Unfortunately,
most patients are diagnosed at advanced stages, when only palliative chemotherapy is available, with a median survival shorter than 12 months. Emerging molecular targeted therapies against FGFR2 fusion and IDH1/2 mutations show modest survival benefits in a limited subgroup of patients with iCCA. Hence, identification of novel and effective therapeutics remains an urgent need.

In general, iCCA is featured by a highly desmoplastic tumor microenvironment (TME) with excessively infiltrating immune and stromal cells. Available evidence indicates that the inflammatory microenvironment is associated with iCCA progression and poor prognosis, highlighting TME as a potential target for the treatment of iCCA. Indeed, recent phase II/III clinical trials confirm that immune checkpoint inhibitors (ICIs) plus chemotherapy is promising in patients with iCCA. However, the responses and survival benefits are only limited to a small subset of patients, probably owing to the heterogeneous and complex TME. Therefore, a better understanding of the immunogenomic profile is urgently needed for designing novel immunotherapies to improve responses and patient outcomes.

Patient stratification based on immune profile can help delineate immunologically heterogeneous tumors and identify suitable patients for immunotherapy. We and others have applied bulk-tumor transcriptomic or proteomic profiles to define clinically relevant iCCA subtypes, collectively converging on intertumor immune heterogeneity. Nevertheless, the underlying cancer-cell-intrinsic properties that dictate the diversity in immune landscape remain less characterized. Integrated multidimensional and multimodule analyses of preclinical models and patient samples are needed and expected to deeply characterize the immunological portraits and identify novel personalized immunotherapeutic opportunities in each molecular subtype of iCCA.

In this study, we analyzed the transcriptomic profiles of TME using the largest single-centered iCCA cohort and classified 255 patients into three immune subgroups (IG1, IG2, IG3). By integrating multiple complementary approaches, including whole-exome sequencing (WES), proteomics, single-cell RNA sequencing (scRNA-seq), immunostaining, and mouse models, we delineated the distinct genetic and molecular features among these subgroups. IG1 was an immune-suppressed subgroup with the worst prognosis, harboring the highest myeloid infiltration and KRAS mutations; IG2 was characterized by tumor proliferation and antigen presentation defects; IG3 was featured by the activation of antitumor immunity and enrichment of tertiary lymphoid structures (TLSs), with a good prognosis. The substantial differences in immunological profile suggested that each immune subgroup need specific therapeutic strategies.

METHODS
The iCCA samples from 255 patients used for this study were collected from Zhongshan Hospital of Fudan University. Patients treated between December 2014 to October 2018 were randomly selected from our prospectively collected tissue bank. Detailed clinicopathological features are summarized in online supplemental table S1.

Full details of multi-omics analysis, multiplexed immunostaining of tissue microarray (TMA), pathological examination, mouse model construction, scRNA-seq analysis, functional experiments, and statistical analysis are provided in the online supplemental materials and methods.

RESULTS

Immun-centered classification of iCCA
To characterize the immune profile of iCCA with prognostic significance, multi-omics data derived from the Fudan University (FU-iCCA) cohort (n=255) were obtained (online supplemental figure S1A and online supplemental table S1). Consistent with iCCA’s inflammatory and desmoplastic features, our cohort had a relatively higher immune infiltration among The Cancer Genome Atlas (TCGA) Pan-Cancer datasets (online supplemental figure S1B). Unsupervised hierarchical clustering based on the mRNA levels of 170 prognostic genes from TME signature revealed three Immune subGroups, namely IG1, IG2, and IG3 (figure 1A; online supplemental figure S1C–E and online supplemental table S1). Pairing the transcriptomic and proteomic data from this signature created 103 mRNA–protein pairs, which showed an overall positive correlation (median r=0.6, multiple-test adjusted p<0.01; online supplemental figure S1F). Of note, the abundance of corresponding 103 proteins also clustered these patients into similar immune subgroups with a comparable overlap (online supplemental figure S1G).

As expected, the immune subgroups significantly differed in overall survival (p=1e-09, log-rank test) and recurrence-free survival (p=4e-07, log-rank test), where IG1 had the worst and IG3 showed the best prognosis (figure 1B). Multivariable Cox analysis further authenticated the immune subgroups as a significantly independent prognosticator (HR, 2.4; 95% CI, 1.3 to 4.5; p=4e-03; figure 1C). Carbohydrate antigen 19-9 (CA19-9, p=6.8e-11) and carcinoembryonic antigen (CEA, p=1.2e-06) were significantly higher in IG1, whereas albumin (ALB, p=9.9e-03) levels increased gradually from IG1 to IG3 (online supplemental figure S1H). IG1 had more tumors with necrosis (p=0.012) and advanced tumor node metastases (TNM) stage (p=2.7e-04) (figure 1D and online supplemental figure S1H). Of note, lymph node metastasis (p=0.019) and vascular invasion (p=0.038) were predominated in IG1, whereas endothelial cells were enriched in IG3. Increased lymphangiogenesis and reduced angiogenesis could make iCCA more invasive and impair treatment efficacy. Thus, immunostaining confirmed that compared with IG3, where extensive CD31+ blood vessels were observed, IG1 showed a marked
increase in podoplanin positive (PDPN+) lymphatic vessels (online supplemental figure S1I).

Within this prognostic TME signature, IG1 was enriched with pathways of neutrophil degranulation and epithelial–mesenchymal transition (EMT), suggesting an inflammatory and invasive phenotype. In contrast, leucocyte-mediated cytotoxicity and lymphocyte-mediated immunity pathways were predominant in IG3, indicating an immune-active milieu (figure 1A). As expected, xCell\textsuperscript{14} deconvolution of the RNA-seq data showed that IG1 was dominant by myeloid cells such as neutrophils and immature DCs (iDC), whereas IG3 was abundant in adaptive
immune cells, natural killer (NK) cells, and activated DCs (aDC). Immune infiltration in IG2 was relatively low and heterogeneous, accompanied by the scarcity of stromal cells, which might account for the highest tumor purity \((p=2.2e-04; \text{figure } 1D)\). Generally, the abundance of neutrophils \((p=0.011)\) and iDC \((p=0.045)\) predicted poor survival, whereas the accumulations of lymphocytes, such as CD4+ T cells, CD8+ T effector/memory (Tem) cells, and B cells, were associated with good prognosis (log-rank test; \text{figure } 1E).

Further pathological examination confirmed that tumors in IG3 contained more tumor-infiltrating lymphocyte (TIL) estimates (\text{figure } 1D; online supplemental figure S1J,K and online supplemental table S1). Indeed, TIL estimates correlated negatively with tumor purity \((p=5.3e-03)\) and positively with immune deconvolution \((\text{CD8}^+ \text{ T}, p=1.6e-06; \text{CD4}^+ \text{ T}, p=5.6e-05; \text{B cell}, p=3.8e-04; \text{figure } 1F)\). Similarly, multiplex immunostaining on TMAs confirmed that the intensities of CD15, CD66b, and CD1a were the strongest in IG1, whereas CD3, CD8, and CD20 were primarily expressed in IG3 (\text{figure } 1G and H; online supplemental figure 1I and online supplemental table S1), remarkably consistent with the xCell estimation.

We then performed unsupervised clustering based on the overall mRNAs with the top 25% median absolute deviation and identified three subgroups that were somehow concordant with the immune subgroups (\text{figure } 1A). Specifically, pairing immune subgroups and overall mRNA subgroups (RGs) showed that IG1 and IG3 highly overlapped with RG1 and RG2, respectively. Leveraging our TME gene signature, we also classified Jusakul et al.’s and Job et al.’s iCCA cohorts and observed similar patient subgrouping and survival stratification (online supplemental figure 2A–D). Likewise, clustering our RNA-seq data with the transcriptomic signatures from four previous studies8–10 12 resulted in a significant concordant patient allocation (online supplemental figure 2E–H).

**Distinct molecular features among the three immune subgroups**

To further illustrate the key biological processes in patient subgroups, we performed single sample Gene Set Enrichment Analysis (GSEA) using the 50 hallmark gene sets16 (online supplemental figure S3A and online supplemental table S2). Analysis of principal-component feature showed that PC1 could be explained by tumor purity with immune and stromal gene set vectors opposite to oncogenic and other vectors, whereas PC2 discriminated immune from stromal gene sets (online supplemental figure S3A–C). Unsupervised hierarchical clustering further verified the association between signaling pathways and immune subgroups, as well as separated clustering of immune and other genes (online supplemental figure SSD).

Pathway enrichment analysis incorporating both transcriptomics and proteomics data demonstrated distinct molecular features among the immune subgroups (\text{figure } 2A). IG1 was high in inflammatory response, interleukin 6 (IL-6)/janus kinase (JAK)/signal transducer and activator of transcription 3 (STAT3) signaling, and tumor necrosis factor-\(\alpha\) (TNF-\(\alpha\))/nuclear factor kappa-B (NF-\(\kappa\)B) signaling, indicating that these tumors may be driven by smoldering inflammation. Hypoxia and glycolysis pathways, which have been implicated in myeloid cell-mediated tumor immune evasions,17 18 were also predominant in IG1. IG2 were characterized by enrichment of cell-cycle transcriptional programs (G2M checkpoint, E2F targets, Myc targets) and DNA repair pathway, demonstrating the high proliferative activity. Considering the highest tumor burden in IG2, we adjusted the tumor purity19 and still found enriched proliferation-associated pathways in IG2 (online supplemental figure S3E and online supplemental table S2). IG3 was associated with downregulation of glycolysis and hypoxia pathways, as well as upregulation of the oxidative phosphorylation (OXPHOS) pathway, which has been highlighted as the critical aspects of immune activation20 (\text{figure } 2A).

We further focused on cancer-promoting and cancer-inhibitory inflammation. IG1 was enriched in proinflammatory cytokines and chemokines, including IL-1A, IL-1B, IL-6, CXCL3, CXCL5, and CCL18 (\text{figure } 2B). The mRNA and protein levels of most S100A family members, including S100A8/A9, were upregulated in IG1, which may be involved in cancer-promoting inflammation21 (\text{figure } 2C). Concurrently, increased mRNA and protein levels of CD8A, GZMA, and PRF1 confirmed the upregulation of cancer-inhibitory inflammation in IG3 (\text{figure } 2B). Moreover, a panel of immune signatures18 22–24 collectively showed that IG3 displayed elevated T-cell survival and Teffector signatures relative to other subgroups, whereas IG1 exhibited elevated elevated lymphoid-derived suppressor cells (MDSC), M2/M1 macrophage, myeloid inflammation, and proinflammation signatures (\text{figure } 2D and online supplemental table S2). Overall, our stratification identified two immune subgroups with antagonistic inflammatory phenotypes.

Recently, the cyclooxygenase 2 (COX-2)/prostaglandin E2 (PGE2) axis has been reported to be involved in regulating the antagonistic inflammatory phenotypes in a Pan-cancer study.25 Indeed, the transcript levels of PTGS2, encoding for COX-2, were significantly upregulated in IG1 (\text{figure } 2B,E), consistent with the excessive infiltration of PGE2-associated neutrophils and iDCs.26–27 As expected, patients with higher COX-2 mRNA levels had a worse prognosis \((p=1e-04, \text{log-rank test; } \text{figure } 2F)\). Strong positive correlations between COX-2 mRNA levels and cancer-promoting factors28 including IL-1A, IL-1B, and IL-6 were detected (\text{figure } 2G and online supplemental table S2). Infiltration of neutrophils and iDCs positively correlated with COX-2 mRNA level, whereas plasma cells, CD4+ T cells, and CD8+ Tem cells showed negative correlations (\text{figure } 2H and online supplemental table S2). The COX-2-associated inflammatory signature (COX-IS) was a predictor for response to ICIs in multiple cancers,29 and COX-IS decreased sequentially from IG1 to IG3 (online
supplemental figure S3F). Additionally, IG3 had the highest cytolytic score\(^29\) and immunophenoscore\(^30\) (online supplemental figure S3G,H), further supporting the hypothesis that IG3 could be more likely to respond to ICIs.

**KRAS mutation-associated myeloid inflammation was dominant in IG1**

We further investigated genomic alterations that could be associated with immunological features. IG2 had more frequent somatic copy number variations (SCNAs) (online supplemental figure S4A), in agreement with previous reports that SCNA burden negatively correlated with immune cell infiltration\(^31\) (online supplemental figure S4B). When selecting subtype-specific mutated genes, we found that nearly half of tumors in IG1 were KRAS mutants (figure 3A,B and online supplemental table S3). KRAS mutation was negatively associated with
most xCell-derived adaptive immune and NK cell signatures, but positively correlated with iDC and neutrophil signatures (figure 3C and online supplemental table S3). GSEA at mRNA and protein levels both identified neutrophil degranulation to be the pathway most strongly associated with KRAS mutation (figure 3D and online supplemental table S3). Coss et al, previously reported, consistent with previous report, KRAS mutation significantly and positively correlated with several inflammatory molecules, such as COX-2, IL-1A, IL-1B, IL-6, and S100A8/A9, resembling ...
the immunological feature of IG1 (figure 3E and online supplemental figure S4C). Tumors with KRAS mutation also displayed elevated myeloid inflammatory, COX-2, and MDSC signatures, as well as decreased cytolytic signature (figure 3F, online supplemental figure 4D–F and online supplemental table S3). In Jusakul’s cohort, KRAS mutation also led to elevated neutrophil infiltration, COX-2/S100A8 expression, and neutrophil degranulation pathway (online supplemental figure S4G–J).

Considering that PTGS2 was the most significantly upregulated gene associated with myeloid inflammation both in IG1 and KRAS mutant tumors, we hypothesized that KRAS mutation may upregulate COX-2 to produce PGE2, which promoted myeloid inflammation. Indeed, ectopic expression of KRASG12D in iCCA cells significantly upregulated the mRNA and protein levels of COX-2 and COX-1, as well as increased PGE2 levels in the culture supernatant (online supplemental figure 4K–O). Transwell assay revealed that, compared with KRASWT, the culture supernatants from KRASG12D significantly enhanced neutrophil migration (online supplemental figure S4P), which could be hampered by COX-2 inhibitor, celecoxib (online supplemental figure S4Q). KRASG12D culture supernatant-treated neutrophils showed stronger immunosuppressive activity, as assessed by IL-2 production and apoptosis of preactivated Jurkat T cells (online supplemental figure S4R,S).

We next generated iCCA mouse models using onco-genic drivers myr-AKT and YAP5127A in combination with KRASWT (AY) or KRASG12D (AYK) via the sleeping beauty system and hydrodynamic tail vein injection34 (online supplemental figure S5A). Immunohistochemistry confirmed that AYK, but not AY tumors, were conspicuously infiltrated with neutrophils (positive for Ly6G, S100A8, and S100A9), while devoid of CD8+ T cells and B cells (figure 3G,H and online supplemental figure S5B,C). scRNA-seq analyses revealed that the proportions of neutrophil and macrophage were significantly higher in AYK tumors, whereas T/NK cells, B cell, and stromal cells were more abundant in AY tumors, consistent with multi-omics’ results (figure 3I and online supplemental figure 5D–F). Neutrophils in AYK tumors showed higher myeloid inflammation and elevated expression of inflammatory markers like S100a8, S100a9, Pigs2, and Illh, accompanied by stronger exhausted phenotypes of T/NK cells (figure 3J and online supplemental figure S5G). Of note, KRAS mutation upregulated the expression of COX-2 both in tumor cells and the microenvironment (figure 3G,H and online supplemental figure S5G). Thus, excessive infiltration of neutrophils may account for KRAS mutation-associated myeloid inflammation. Overall, KRAS mutation had a critical role in subverting anti-tumor immunity to myeloid inflammation-associated immunosuppression (online supplemental figure S5H).

**Antigen presentation defects are associated with immune exclusion in IG2**

Defects in antigen presentation could interrupt neoantigen recognition and lead to tumor immune evasion, irrespective of tumor mutation burden (TMB). We thus explored the underlying roles of these avenues of immune evasion among the three immune subgroups (figure 4A). Despite relatively higher TMB and tumor neoantigen burden (TNB) in IG2, defects in antigen presentation through loss of heterozygosity in human leucocyte antigen (HLA LOH, p=2e-03) or alterations of the other components of antigen presentation machinery (APM, p=0.065) were frequently observed in IG2 (figure 4A,B; online supplemental figure S6A and online supplemental table S4). Such defects may hamper the cross-priming of naive T cells and subsequent tumor recognition by primed T cells, which may explain the reduction in T cell receptor/B cell receptor (TCR/BCR) diversity and RNA-seq reads mapping to VDJ loci in IG2 (online supplemental figure 6B–D).

We further calculated the immunoeediting score for each tumor by quantifying the ratio of observed to expected numbers of predicted HLA class-I binding neoantigens, with a score of <1 indicating the presence of immunoeediting.35 Immunoeediting scores decreased continuously from IG1 to IG3 (figure 4C and online supplemental table S4). Within each immune subgroup, only IG2 tumors with HLA LOH had a significantly higher immunoeediting score, suggesting that tumor subclones in IG2 may be prone to HLA LOH-associated immune evasion (figure 4D). Overall, 15% (n=38) of patients exhibited HLA LOH, which had decreasing trends of overall survival (p=0.009, log-rank test) and recurrence-free survival (p=0.024, log-rank test) compared with the HLA intact patients (figure 4E,F).

The expression of immune checkpoints may reflect a cancer-adaptive immune response to an active immune system. We found a subgroup-specific pattern of these molecules, where IG2 had relatively few exclusively upregulated immune checkpoints (online supplemental figure S6E). HLA LOH correlated with some immune checkpoint molecules, such as programmed death ligand 1, and refined TMB as a biomarker of ICIs.36 Comparing the multi-omics profiles between tumors with and without HLA LOH revealed differentially expressed mRNA and proteins that regulated metabolic reprogramming, cell-cycle transcriptional programs, and immune processes. HLA LOH was enriched with higher TMB/TNB, higher tumor purity, decreased interferon gamma pathway, and decreased major histocompatibility complex class I signatures. Likewise, T-cell survival and T-effector signatures were significantly reduced in tumors with HLA LOH (online supplemental figure S6F). Therefore, APM defects, especially HLA LOH, might involve in immune anergy and afterwards immune-exclusion despite the relatively high TMB in IG2 (figure 4G).
Tertiary lymphoid structures are related to antitumor immunity in IG3

TLSs provide a pivotal microenvironment for generating antitumor immune responses. We used a previous TLS signature to estimate TLS distribution and found that the mRNA levels of those genes, together with estimated TLS scores, were relatively upregulated in IG3 (Figure 5A, B and online supplemental table S5). Pathological examination also confirmed that intratumoral TLSs were significantly accumulated in IG3 (Figure 5C, D). Patients with both primary follicles (FL-I) and secondary follicles (FL-II) had higher estimated TLS scores compared with those with only Aggregates (Agg) and no TLSs (TLSs-), confirming the accuracy of TLS score (Figure 5E and online supplemental figure S7A). Consistent with our latest report, intratumoral TLSs predicted a better prognosis (Figure 5F-H), and TLSs+ (Agg, FL-I, and FL-II) patients were characterized by early TNM stage (p=0.018), lack of microvascular invasion (p=0.033), absence of intrahepatic metastasis (p=0.011), and small tumor size (p=8.1e-04; online supplemental figure S7A and online supplemental table S5).

Figure 4  Antigen presentation defects occur frequently in IG2. (A) Antigen presentation machinery (APM) defects in each immune subgroup (Pearson’s χ² test or Fisher’s exact test), (B, C) Comparison of tumor mutation burden (TMB) (B) and immunoediting score (C) among the three immune subgroups (Wilcoxon rank-sum test), (D) Immunoediting score were plotted by HLA LOH status and immune classification (Wilcoxon rank-sum test), (E, F) Kaplan-Meier curves for overall survival (E) and recurrence-free survival (F) based on loss of heterozygosity in human leukocyte antigen (HLA LOH) status (log-rank test). (G) Model illustrating how HLA LOH may lead to immune escape in IG2.
Multi-omics data indicated that TLS+ tumors showed specific downregulation of pathways of G2M checkpoint, inflammatory response, and EMT, whereas upregulated ones in metabolism and peroxisome (online supplemental figure S7A). As expected, T cells and B cells, as well as coinhibitors, were enriched in TLS+ tumors, suggesting an immune-activated phenotype (online supplemental figure S7B,C). Moreover, both cytolytic scores and COX-IS predicted a better response to immunotherapy for TLS+ iCCA patients (online supplemental figure S7D,E). Thus, immunosurveillance in IG3 might be partially attributed to intratumoral TLSs (figure 5I), which predicted a better response to immunotherapy in this subgroup.
HBV infection is negatively associated with myeloid inflammation in iCCA

Hepatitis B virus (HBV) infection is the well-known risk factor for iCCA; however, little is known about the immunogenomic profile of HBV-related iCCA. We found that patients with HBV infection (n=68) were under-represented in IG1 (p=0.022), mutually exclusive with KRAS mutation (p=3.3e-03) and coexisted with TP53 mutation (p=0.036) (figure 6A; online supplemental figure S8A and online supplemental table S6). HBV-positive tumors displayed an elevated cytolytic score, decreased myeloid inflammation signature and decreased COX-2 mRNA levels (figure 6B). Recent study proposed that ICIs showed efficacy for patients with HBV-positive hepatocellular carcinoma (HCC), rather than non-viral HCC.\(^{39}\) Considering the negative impact of myeloid inflammation on immunotherapy,\(^{40}\) the COX-IS score predicted worse responses to ICIs in HBV-negative subgroup (figure 6B). The transcriptome of patients with HBV-positive iCCA resembled the expression profile of other solid tumors that responded to ICIs\(^{41}\) (figure 6C). Thus, the decreased myeloid inflammation observed in HBV-positive iCCA may define a population that could benefit from ICIs.

We additionally collected tumors and adjacent nontumor tissues from five HBV-positive and five HBV-negative iCCAs for scRNA-seq analyses (online supplemental figures S1A and S8B). Increased proportions of T/B/NK cells and reduced proportions of myeloid cells were observed in HBV-positive tumors (figure 6D), but HBV infection did not impact the proportions of each cluster in adjacent nontumor tissues (data not shown). We then classified the 10 patients into the closest immune subgroups according to the enriched gene sets (online supplemental figure S8C). At the CI of 0.7, four patients were excluded, with one (P06T), three (P03T/P09T/P10T), and two (P13T/P15T) samples grouping into IG1, IG2, and IG3, respectively. P06T, the only one remaining in IG1, was also the only patient with KRAS mutation (p=0.036) (figure 6H). The proportions of inflammatory CD4 T cells and myeloid cells. A total of 13 clusters emerged (figure S8D).

We then classified the 10 patients into the closest immune subgroups according to the enriched gene sets (online supplemental figure S8J). In HBV-positive patients, the top downregulated genes in macrophages were characterized by downregulated inflammatory responses, whereas NK cells upregulated cytotoxicity and allograft rejection responses (figure 6K and online supplemental figure S8H–J). Collectively, HBV infection in iCCA may alleviate myeloid inflammation and promote antitumor immunity. Hence, HBV infection was slightly associated with longer overall survival in our cohort (p=0.091, log-rank test; figure 6L), consistent with a previous study.\(^{45}\)

**DISCUSSION**

Tremendous sequencing studies have improved our understanding of the molecular pathogenesis of iCCA.\(^{8-11,33}\) However, little is known about the immunogenomic profile of iCCA and how to leverage this information to maximize responses to immune therapies. Herein, we provided an immunogenomic characterization of iCCAs, involving WES, RNA-seq, proteomics, scRNA-seq and mouse model analyses that could enable the discovery of novel biological insights and identification of promising targets for precise oncological practice.

We identified three distinct immune subgroups with diverse clinical, immune, genetic, and molecular features. IG1 was defined by the enrichment of neutrophils and iDCs, as well as KRAS mutations. KRAS mutation is associated with the recruitment of myeloid cells and MDSCs into the TME, which then drives immune suppression.\(^{46}\) Consistently, KRAS mutation may probably be the primary cause of myeloid inflammatory and immunosuppression...
in IG1. IG2 was an immune-exclusion subgroup with high tumor purity, strong proliferative activity, and defects in APM. Some tumors in IG2 could potentially have an ancestral immune-active microenvironment, and then gradually became “exclusion” during the enrichment of HLA LOH (figure 4G). IG3 was dominated by adaptive immune cells, aDCs, and NK cells, demonstrating an immune-activated phenotype with the enrichment of...

Figure 6  Influences of hepatitis B virus (HBV) infection on the immune microenvironment of intrahepatic cholangiocarcinoma (iCCA). (A) Associations of HBV infection with indicated features (analysis of variance (ANOVA), Pearson’s χ² test, or Fisher’s exact test). (B) Comparisons of representative signatures in patients with and without HBV infection (Wilcoxon rank-sum test). (C) Subclass mapping with other solid tumors treated with anti-PD-1 mAb (anti-programmed cell death protein 1 monoclonal antibody). (D) The proportion of indicated immune cells in tumor samples. (E) Uniform Manifold Approximation and Projection (UMAP) plot showing the subtypes of T/natural killer (NK) cells. (F) Boxplot showing the fractions of T/NK cells (ANOVA). (G) UMAP plot showing the subtypes of myeloid derived cells. (H) Boxplot showing the fractions of myeloid-derived cells (ANOVA). (I) Violin plot indicating the mRNA levels of COX-2 (upper panel) and proinflammatory scores (bottom panel) in myeloid subgroups. (J,K) Pathway enrichment analysis based on differentially expressed genes that associated with HBV infections in myeloid cells (J), CD4+ T cells, NK cells, and macrophages (K). (L) Kaplan-Meier curves of overall survival based on HBV infection status (log-rank test).
TLSs. The immunogenomic features of the three immune subgroups were reproducible in multiple independent cohorts, supporting that our classification was suitable and reliable.

Chronic inflammation induced by HBV infection increases the risk of HCC and iCCA. Using high-dimensional CyTOF analysis, Lim et al, revealed that HBV-related HCC microenvironment is more immunosuppressive and exhausted than that of non-viral-related HCC. In contrast, we found that HBV infections alleviated myeloid inflammation and enhanced the cytolytic activity of NK cells in iCCA, indicating that this common etiology may lead to distinct microenvironments between HCC and iCCA. Recently, three randomized phase III clinical trials in more than 1600 patients with advanced HCC revealed that immunotherapy was superior in patients with viral-related HCC. However, it is still unknown whether underlying viral hepatitis impacts on ICIs’ response in iCCA. We here proposed that HBV infection in iCCA could counteract the excessive myeloid inflammation and potentially reinvigorate antitumor immunity, probably favoring the efficacy of ICIs. The reduced myeloid inflammation in patients with virus-related iCCA might be due to the selection for KRAS wild type tumors during persistent HBV infection. Another rational explanation is the negative correlation between chronic
HBV infection and innate immune response, especially the function of neutrophils.49

Therapeutically, this study aimed to characterize iCCA immune subgroups amenable to specific individualized therapies (figure 7 and online supplemental table S7). IG1 might be responsive to myeloid targeted therapies, such as CX-C motif chemokine receptor 2 (CXCR2) and colony stimulating factor receptor (CSFR) inhibitors to deplete or reprogram tumor-associated neutrophils and M2 macrophages, respectively. Recent success in selective inhibitors of KRAS G12C may spark a renewed enthusiasm in targeting KRAS mutations in this immune-suppressive subgroup. IG2 may acquire better outcomes from specific CDK inhibitors, as well as strategies that regulate HLA and APM transcription. The enrichment of TLSs in IG3 and its potential role in the response to ICIs treatment may help guide clinical application of ICIs. Notably, the three immune subgroups showed distinct expression patterns of immune checkpoints, underscoring the need for personalized ICIs.

Our study had some limitations. The use of a retrospective cohort with varied clinicopathological features and treatments may require further prospective validation. In addition, immunostaining on TMAs for estimating immune cell infiltration was likely to over-represent or under-represent the density or distribution of each immune subset. In conclusion, our integrative multi-module analyses provide a comprehensive understanding of iCCA’s immune landscape. The three immune subgroups with specific immune and non-immune profiles may have implications for the design of clinical trials using novel therapeutic strategies in iCCA.

Contributors Conceptualization: QG, DG, and HZ; Methodology: JL, YD, BX, and QG; Formal analysis: JL, YD, BX, LD, XS, SJ, MW, and HF; Investigation: JL, GS, CS, MZ, XX, JM, and SZ; Resources: QG, DG, HZ, JF, JZ, BX, WX, and XZ; Data curation: QG, YD, and JL; Writing: JL and QG; Visualization: YD and JL; Supervision, QG; Funding acquisition: QG and JL. QG is responsible for the overall content as guarantor.

Funding This work was supported by the National Natural Science Foundation of China (Grants 82130077, 8196128025, and 82103314), Research Projects from the Science and Technology Commission of Shanghai Municipal (Grants 20JC1418900 and 19XJ1420700), Jinshan Hospital Flexible Mobile Talent Research Startup Fund (RXXC-2020-1), the Science and Technology Commission of Shanghai Municipal (20JC1418900), and China Postdoctoral Science Foundation (Grant 2020M681181).

Competing interests None declared.

Patient consent for publication Not required.

Ethics approval This study involves human participants and was approved by the Research Ethics Committee of Zhongshan Hospital, and written informed consent was obtained from each patient. Participants gave informed consent to participate in the study before taking part.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available in a public, open access repository. Data are available on reasonable request.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See http://creativecommons.org/licenses/by-nc/4.0/.

CORID ids

Jian Lin http://orcid.org/0000-0001-6457-7026
Hu Zhou http://orcid.org/0000-0001-7006-4737

REFERENCES


Supplementary Materials and methods

Multimodule analysis

Bulk-tumor RNA-seq (n = 255) was utilized to stratify the FU-iCCA cohort. Then, WES (n = 249) and proteomics (n = 208) data elucidated the trans-omics connections and underlying molecular mechanisms. Meanwhile, H&E staining (n = 177), multiplex immunostaining (n = 188), and scRNA-seq (n = 10) were used to validate the immune features. The details of WES, RNA-seq, and proteomics analyses were conducted as described in our recent publication\(^1\) and also briefly summarized as follows:

DNA Extraction and WES analysis

Genomic DNA was extracted from tumor and non-tumor liver tissues using QIAamp Fast DNA tissue kit (QIAGEN) according to manufacturer’s protocol. DNA was quantified by the Qubit 3.0 (Invitrogen) and NanoDrop 2000 (Thermo Scientific). WES libraries were prepared and captured using the QuarPrep EZ DNA Library Kit and QuarHyb Reagent kit (DynastyGene Biotechnologies) following the manufacturer’s instructions. The DNA library with 150 bp paired-end reads was sequenced with Illumina NovaSeq 6000 System. WES was conducted with a mean coverage depth of 205X (range: 90-300X) for tumor samples and 97X (range: 47-162X) for adjacent non-tumor liver samples. The exome sequencing data was first aligned to GRCh37/hg19 sequence reference using BWA MEM v 0.7.10 (http://bio-bwa.sourceforge.net/). After alignment, the GATK best practice workflow was applied on mapping results, including marking duplicates with picard v2.22.8

**RNA Extraction and RNAseq analysis**

Total RNA was extracted and purified from fresh frozen tissues using the Whole RNA extraction kit (TIAN GEN). RNA integrity was measured on an Agilent 2100 Bioanalyzer (Agilent Technologies). RNA with enough amount and good quality were used to prepare the transcriptome library. The cDNA synthesis, end-repair, A-base addition, and ligation of the Illumina index adapters were performed according to QuarPrep RNA Library Kit (DynastyGene Biotechnologies). Library quality was measured on an Agilent 2100 Bioanalyzer for product size and concentration. Paired-end libraries were sequenced by an Illumina NovaSeq 6000 System, with a sequence coverage of 139 million paired reads. RNA-seq data analysis identified 20,173 protein-coding genes with an average of 14,255 genes per sample, covering
the majority of the genes in proteomics. Sequence data were removed adaptors and low-quality reads using the Trimmomaticv 0.36. The cleaned sequence data were aligned to human reference genome (UCSC hg19 assembly) by STAR2(2.7.3a) in two-pass mode (with parameters: --chimSegmentMin 30 --chimJunctionOverhangMin 10 --alignSJDBoverhangMin 10 --alignIntronMax 200000 --alignSJstitchMismatchNmax 5 -1 5 5). The cleaned sequence reads were used for further qualification of gene expression. transcripts per million (TPM) values of each gene and transcript were calculated by Salmon with parameters of –seqBias --gcBias –posBias.

Proteomic Sample Preparation and Proteomic Analysis

Samples for MS were prospectively collected and 500 μL SDS lysis buffer was added into the powdered tissues for protein extraction, and sonicated at 20% amplitude for the total working time of 2 min with 5 s on and 5 s off (JY92-IIDN, Ningblio Scientz Biotechnology Co., LTD, China). The proteins were denatured at 95 °C for 5 min. Lysates were centrifuged at 12,000 g for 10 min to remove the insoluble debris and retain the supernatant for proteomic experiment. The tryptophan-based fluorescence quantification method was used to determine the protein concentration (Thakur et al., 2011). The equivalent of protein was digested by filter-aided sample preparation (FASP) procedure with 10 kDa centrifugal filter tubes (Millipore) and centrifuged at 12,000 g at 22 °C. Concentration for eluted peptides was determined by BCA protein quantification kit. 400 μg peptides were dried by vacuum freeze-drying for the following experiment. The mixed samples proteins for the ‘internal reference’ were
also digested by FASP as well as the other protein samples. 41 μL x 2 anhydrous acetonitrile was added into two sets TMT reagents (0.8 mg) and mixed with 400 μg peptides (dissolved in 200 μL 100 mMTEAB) per channel. The samples were incubated for 1 hr at room temperature, and then quenched the labeling reaction by adding 16 μL 5% hydroxylamine into the samples and was incubated for 15 min at room temperature. The pooled samples of the labeled peptides in one set were dried by vacuum freeze-drying for the following desalting experiment through the C18 solid-phase extraction (3M Empore). 0.1% formic acid was used to resolve the peptides and 1 μg resolved peptides for proteomic analysis were separated on an Easy nLC 1000 UHPLC system (Thermo Fisher Scientific) with a 120 min LC gradient at 300 nL/min (Buffer A: 0.1% formic acid in water; Buffer B: 0.1% formic acid in acetonitrile) of a home-made 75 mm x200 mm diameter C18 column (ReproSil-Pur C18-AQ, 3.0 μm resin (Dr. Maisch GmbH, Germany)). The column was heated to 50°C using an in-house column heater and the gradient was set as 2%–5% B in 1 min; 5%–25% B in 94 min; 25%–40% B in 15 min; 40%–60% B in 3 min; 60%–100% B in 1 min; 100% B in 6 min. The spray voltage of the Q Exactive HF-X mass spectrometer was set at 1,800 V in positive ion mode and the ion transfer tube temperature was set at 320 °C. Xcalibur software was used for data-dependent acquisition. The 24 benchmark fractions were analyzed with a 90 min LC gradient and the gradient was set as 2%–5% B in 1 min; 5%–25% B in 67 min; 25%–40% B in 13 min; 40%–60% B in 3 min; 60%–100% in 1 min; 100% B in 5 min. The parameters of Q Exactive HF-X mass spectrometer were set the same as iCCA proteomic samples.
Download of TCGA data and calculation of immune scores

Updated FPKM gene expression data, clinical data and sample information in TCGA database were obtained from the Genomic Data Commons (https://portal.gdc.cancer.gov/) using the R package TCGAbiolinks. Only primary solid tumor samples were enrolled in the analysis of immune score, which was calculated use R package ESTIMATE.

Microenvironment-related gene expression analysis and clustering

Gene expression data was log-transformed using log2 (FPKM+1). We first selected immune-cell-related markers of 21 immune cells and 2 stromal cells from the xCell database. Then correlations between expression level of all genes and prognosis were calculated. For each gene, patients were separated into two subgroups by the mean value of the gene expression level, and the Cox proportional hazards regression model was used to calculated P value between subgroups and prognosis using R package survival. 4,782 prognosis-related genes were filtered with P < 0.05. After intersected with immune-cell-related markers, 170 genes were selected as pivotal microenvironment genes. Unsupervised hierarchical clustering method was used to classify 255 iCCA patients into different subgroups with clustering method ‘ward.D2’ and distance ‘manhattan’. To identify the optimal cluster number, the ConsensusClusterPlus package and silhouette analysis in cluster package was used to assess clustering stability. Finally, three subgroups were identified (IG1, IG2, IG3).
**Calculation of differentially expressed genes/proteins and functional enrichment analysis**

The 170 pivotal microenvironment genes could be divided into two gene subgroups, IG1 enrichment and IG3 enrichment. We performed gene ontology (GO) enrichment analysis to identify the enrichment of functional pathways of two gene subgroups using the R package clusterProfiler. Differentially expressed genes and proteins between different conditions were calculated using R package limma. Genes with adjusted $P$ value < 0.05 and fold change > 1.5 ($\log_2\text{FC} > 0.58$) were set as significantly up-regulated, whereas with adjusted $P$ value < 0.05 and fold change < 0.67 ($\log_2\text{FC} < -0.58$) were set as significantly down-regulated. Proteins with adjusted $P$ value < 0.05 and fold change > 1.25 ($\log_2\text{FC} > 0.32$) were set as significantly up-regulated, whereas with adjusted $P$ value < 0.05 and fold change < 0.8 ($\log_2\text{FC} < -0.32$) were set as significantly down-regulated. Another functional enrichment analysis method, Gene Set Enrichment Analysis (GSEA), was also used to analysis differentially regulated pathways via R package clusterProfiler. Gene set enrolled in the GSEA analysis were download from the Molecular Signatures Database (MSigDB, v7.1) of the Broad Institute. HALLMARK gene sets (H) and MSigDB curated gene sets of KEGG (C2), GO (C5), REACTOME (C2) were used to perform GSEA in a 1,000-gene-set with a two-sided permutation.

**Calculation of immune-related and tumor-related scores**
For the calculation enrichment score of different types of immune cells in each sample, we used single sample GSEA (ssGSEA) via GSVA package, with gene-sets of these immune cells downloaded from xCell database\(^4\). For the calculation of immune score (represents the infiltration of immune cells in tumor tissue) and stromal score (captures the presence of stroma in tumor tissue), R package ESTIMATE was used to predict the presence of infiltrating stromal/immune cells in tumor tissues using gene expression matrix\(^3\). The immunophenoscore of each patient was calculated using the algorithm and R script provided by Pornpimol Charoentong et al\(^7\). The mutation structure of KRAS were plotted using ProteinPaint.

**Neoantigen predictions**

Neoantigens were predicted by NetMHC (v4.0) and NetMHCpan (v4.1). The candidate neoantigens were filtered follow the criteria: 1) predicting as binders (IC50 \textless 500 nM in NetMHC or Score\_EL > 0.1 in NetMHCpan). 2) neoantigen-derived genes are expressed in bulk RNA-seq data.

**HLA genotyping and HLA-LOH prediction**

POLYSOLVER was applied to investigate HLA gene genotypes. LOHHLA\(^8\) was used to predict the LOH events for HLA genes based on the POLYSOLVER results. HLA LOH events were identified using the parameter \(P < 0.001\). The coverage depth on HLA gene locations of tumor and normal samples of each patient were manually checked to verify the HLA LOH events.
**Immunoeediting analysis**

We followed the method developed by Rooney et al.\(^9\,10\) to analyze immunoeediting in each patient. TCGA-HCC data was used as the driven null model. Referencing the published algorithm\(^8\), 1,827 synonymous SNVs and 4,927 nonsynonymous SNVs were enrolled in the analysis. To estimate immunoeediting, 3,390 mutational spectra were considered.

**Somatic Copy Number Alteration (SCNA) analysis**

For each patient, SCNAs were inferred by CNVkit\(^11\) with default parameters and identified by Genomic Identification of Significant Targets in Cancer (GISTIC, version 2.0)\(^12\). The significantly gained or lost SCNA regions were determined by the default parameters in GISTIC and log\(_2\) ratio cut-off of ±0.8 was used to define SCNA amplification and deletion. The value of SCNA burden was calculated as the percentage of genes with amplification, gain, loss or deletion in each patient.

**TCR and BCR analyses**

Mapped RNA-sequencing reads were used to allelotype each patient by profiling TCR and BCR sequences with MiXCR as previously described\(^13\).

**Multiplexed immunostaining of Tissue MicroArray (TMA).** The TMA of the FU-iCCA cohort was used to validate our immune subgroups. Multiplex staining of
CD1a (ab108309, Abcam), CD3 (ab135372, Abcam), CD8 (ab93278, Abcam), CD15 (ab135377, Abcam), CD66b (ab197678, Abcam), CD20 (ab78237, Abcam), CK19 (ab52625, Abcam), PDPN (ab236529, Abcam), and CD31 (ab76533, Abcam) was performed by the Vectra Automated Quantitative Pathology Imaging and Analysis platform through multispectral imaging system and inFormTM image analysis software (PerkinElmer). Following the manufacturer’s instruction (PerkinElmer, Opal® Kit), we scanned the whole field of each sample by using the PerkinElmer Vectra3® platform and quantified the results by using PerkinElmer Vectra3® platform as is described before	extsuperscript{14}. All quantifications were evaluated blinded to patient clinical outcomes.

**Pathological examination.** The presence of intra-tumoral TLSs was assessed morphologically on H&E staining slides, using a previously published scale	extsuperscript{15}. Briefly, TLSs were classified as: i) Aggregates (Agg): vague, ill-defined clusters of lymphocytes; ii) Primary follicles (FL-I): round-shaped clusters of lymphocytes without germinal center formation and iii) Secondary follicles (FL-II): follicles with germinal center formation. Cases were also further scored according to TLS maturation stages: i) Agg iCCA: tumors with only Agg and no FL-I or FL-II; ii) FL-I iCCA: tumors with at least FL-I, with or without Agg and without FL-II and iii) FL-II iCCA: tumors with at least 1 FLII regardless of the presence of Agg and FL-I.

Pathology TILs were estimated from H&E staining slides using international established guidelines, using a previously published scale	extsuperscript{16}. Briefly, the relative
proportion of stromal area to tumor area was determined from the pathology slide of a given tumor region. TILs were reported for the stromal compartment (= per cent stromal TILs).

**Mouse model construction**

Six weeks old female FVB/N mice were ordered from Shanghai Branch of Beijing Vital River Laboratory Animal Technologies Co. Ltd. maintained under SPF housing with a maximum of five mice per cage. The experiments were performed following the institutional guidelines strictly and were approved by the Institutional Animal Care and Use Committee (IACUC) of the Shanghai Branch of Beijing Vital River Laboratory Animal Technologies Co. Ltd. (2017-0014). A sterile 0.9% NaCl solution/plasmid mix was prepared containing DNA. We prepared 20 μg of pT3-EF1α-HA-myr-Akt (Addgene, 31789), 30 μg of pT3-EF1a-YAPS127A (Addgene, 86497), 10 μg of pT3-EF1a-KRASWT or 10 μg of pT3-EF1a-KRASG12D, and a 10:1 ratio of transposon to SB-luc transposase–encoding plasmid dissolved in 2 mL of 0.9% NaCl solution. Mice were injected with the solution into the lateral tail vein with a total volume corresponding to 10% of body weight in 6 to 8 seconds. Vectors for hydrodynamic delivery were produced using the QIAGEN plasmid PlusMega kit. Equivalent DNA concentration between different batches of DNA was confirmed to ensure reproducibility among experiments.

**Single cell RNA sequencing of mouse iCCA models**
Tumors from AY (n = 6) and AYK (n = 6) iCCA mouse models were dissected using the Miltenyi Mouse Tumor Dissociation Kit and gentleMACS Octo-Dissociator (Miltenyi) following manufacturer’s instructions. After filtering through a 70-mm filter, tumors were selected and droplet-based isolation of single cells was performed. Then, the cells were resuspended in 500 μL sample buffer and placed on ice and cell viability was tested using a BD Rhapsody Scanner instrument. Calcein AM (Invitrogen) and Drap7 (BD Biosciences) were added to the cell suspension to label the living and dead cells separately. Whole transcriptome libraries were prepared using the BD Rhapsody single-cell whole-transcriptome amplification workflow. Sequencing libraries were prepared using random priming PCR of the whole-transcriptome amplification products to enrich the 3′ end of the transcripts linked with the cell label and UMI. Sequencing libraries were quantified using a High Sensitivity DNA chip (Agilent) on a Bioanalyzer 2200 and the Qubit High Sensitivity DNA assay (Thermo Fisher Scientific). The library for each sample was sequenced by HiSeq X (Illumina, San Diego, CA) on a 150 bp paired-end run.

**scRNA-seq data analyses for iCCA samples.** For 10× Genomics scRNA-seq data analyses, raw reads were aligned to human reference genome (hg38, 2020-A) using Cell Ranger software (v6.0.0). The alignment reference and software were both provided by 10x Genomics (https://support.10xgenomics.com). The same software was used for unique molecular identifier (UMI) counting and filtration using default parameters. Raw count data were then analyzed with the R package Seurat (v4.0.0).
Cells that expressed less than 500 genes or over 20% mitochondrial RNA were filtered out. The expression matrix was merged and normalized using a global-scaling normalization method using Seurat. 2,000 variable genes were selected for dimensionality reduction. Principal component analysis (PCA) was performed on the variable genes and the resolution parameter to identify clusters was set to 0.8. Batch effects were corrected using ComBat. For visualization purposes, uniform manifold approximation and projection (UMAP) was performed. Clusters were identified using the top markers found in each cluster with an adjusted \( P \) value \( \leq 0.05 \) and an average \( \log_2 \) fold change \( \geq 0.1375 \). Markers to identify cell populations were selected from xCell database\(^4\). We also used another software SingleR to verify cell populations. In the analysis of gene expression similarity between bulk RNA-seq data and scRNA-seq data, we calculated the mean expression value of scRNA-seq samples. The ssGSEA algorithm was used to calculate the enrichment score of scRNA-seq samples and bulk RNA-seq samples using IG1 and IG3 enrich genes. Z-scores were calculated from the enrichment scores.

**Functional experiments**

**Cell lines**

HuCCT1, HCCC9810, HL60, and Jurkat cells are kept by Dr. Daming Gao’s lab. These cell lines were cultured in RPMI.1640 medium supplemented with 10% FBS, 100 units of penicillin and 100 mg/mL streptomycin.

**Plasmids and lentivirus-mediated construction of stable cell lines**
The Myc-tagged coding sequence of human \textit{KRAS} and \textit{KRAS G12D} was cloned into the lentiviral vector pLEX-MCS-CMV-puro (Thermo Scientific Open Biosystems) to generate corresponding expression plasmids. pLEX-MCS-CMV-puro lentiviral packaging and generation of HuCCT1 (\textit{KRAS}_G12D \ iCCA cell line) and HCCC9810 (\textit{KRAS}_\text{WT} \ iCCA cell line) stable cell lines by infection were performed according to the protocol before\textsuperscript{17}.

**Real-time quantitative PCR (qRT-PCR)**

Total RNA was extracted from cells using TRIzol Reagent (Invitrogen, Thermo Fisher Scientific) according to the manufacturer’s instructions. Total RNA was reverse transcribed into first-strand cDNA using the ABScript II RT Master Mix for qPCR Kit (ABclonal). The cDNAs were then used for real-time PCR (qPCR) on a CFX96 Touch Real-Time quantitative PCR System (Bio-Rad) using TB Green Premix® Ex Taq II (Tli RNaseH Plus; Takara). Human \textit{GAPDH} was served as the internal control. The relative quantification of gene expression was analyzed by using the $2^{-\Delta\Delta C_t}$ method.

**Western blot analysis**

Cells were lysed in EBC lysis buffer (50 mM Tris HCl, pH 8.0, 120 mM NaCl, 0.5% Nonidet P-40) supplemented with protease inhibitors (Selleck Chemicals) and phosphatase inhibitors (Selleck Chemicals). 30 μg total proteins were separated by 10% SDS-PAGE gel and blotted with indicated primary antibodies. Primary antibodies used for western blot analysis were as follows: Myc (2276s, Cell Signaling Technology), COX-1 (9896s, Cell Signaling Technology), COX-2 (12282s, Cell
Signaling Technology), β-actin (4970s, Cell Signaling Technology). All western blot gel images were obtained with an Minichemi 610 chemiluminescent imager (Sagecreation, Beijing, China).

**Elisa**

The protein levels of PGE2 and IL-2 in culture medium were quantified by enzyme-linked immunosorbent assay (ELISA) kits (KGE004B, R&D Systems and 70-EK102-48, MultiSciences). Cell-free supernatants were collected and tested according to manufacturer’s protocols. Absorbance at 450 nm was measured using a microplate reader.

**Transwell assay**

Neutrophil model derived from the differentiation of HL60 (dHL60) cells after 120 h culture in RPIM1640 medium supplemented with 10% FBS and 1.25% dimethyl sulfoxide (DMSO). The supernatant of HuCCT1/HCCC9810 and the complete culture medium were used to prepare the conditional culture medium at the ratio of 3:1. For the neutrophil migration assay, $1 \times 10^5/200 \mu l$ neutrophils (from induced HL60 cells) were added to each upper chamber (5 μm, LABSELECT, 14331). The lower chambers contained conditioned medium from HuCCT1/HCCC9810 cells (incubated for 12 h). The number of neutrophils migrating from the upper chamber was counted microscopically.

**Apoptosis detection**

Jurkat T cells were pre-activated with human CD3/CD28 T cell Activator (STEMCELL, 10971) for 24 h. At the same time, dHL60 cells were treated with
indicated culture supernatants from HuCCT1 and HCCC9810. Then, the activated Jurkat T cells were cocultured with culture supernatants-treated dHL60 cells at the ratio of 1:1 for 24 hours. Cells were suspended in FACS buffer containing 1% FBS and 0.1% NaN3 and CD3+ Jurkat T cells were gated for further apoptosis analysis. The data were collected on an LSR-Fortessa X20 Flow Cytometer (BD Biosciences). The annexin V/PI apoptosis flow kit was purchased from BD Biosciences (559763).

**Immunohistochemistry (IHC)**

Tumor issues from mouse models were collected and fixed in 10% formalin overnight and embedded in paraffin. FFPE sections were prepared for staining using standard protocols for xylene and alcohol gradient for deparaffination. Antigen retrieval was performed in the pressure cooker (95℃ for 30 min) to remove aldehyde links formed during initial fixation of tissues. Slides were incubated with primary antibodies, including anti-COX-2 (Cell Signaling Technology, 12282s), anti-Ly6G (Abcam, ab238132), anti-S100A8 (Proteintech, 15792-1-AP), anti-S100A9 (Proteintech, 14226-1-AP), anti-CD8 (Abcam, ab217344) and anti-CD19 (Abcam, ab245235). Sections were incubated and then developed with Dako REAL™ EnVision™ Detection System (DAKO, K5007). The whole IHC slide was scanned and quantified with automated acquisition system (TissueFAXS Plus, TissueGnostics GmbH, Austria).

**Statistical analysis and visualization.** All of the statistical analyses were performed with R software (version 4.0.2, [http://www.R-project.org](http://www.R-project.org)). Student’s t-Test and Wilcoxon rank sum test were utilized to compare continuous and categorical variables.
between two subgroups or conditions, such as immune-related score, functional-related score, clinical features and SCNA levels in each subgroup. For multi-group comparison, we used ANOVA to estimate the $P$ value. Correlation matrices of immune cells were calculated using Pearson correlation and visualized using R package corrplot. Correlation plots between two features such as immune score and stromal score were also calculated using Pearson correlation. Survival curves were estimated with the Kaplan-Meier method and compared using a log-rank test. The Kaplan-Meier survival curves were plotted by R ggsurvplot package. Variables associated with overall survival were identified using univariate Cox proportional hazards regression models. Significant factors in univariate analysis were further subjected to a multivariate Cox regression analysis in a forward LR manner. The FDR correction was utilized in multiple tests to decrease false positive rates. R package ggplot2 and heatmap were used for visualization. For functional experiments, each was repeated at least three times independently, and results were expressed as mean ± standard error of the mean (SEM). Statistical analysis was performed using GraphPad Prism (version 8).

References


3. Yoshihara K, Shahmoradgoli M, Martinez E, et al. Inferring tumour purity and stromal...


12. Mermel CH, Schumacher SE, Hill B, et al. GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human


Supplementary Figure legends

Figure S1 Workflow and consensus clustering of multi-omics data of the FU-iCCA cohort. (A) Workflow of the iCCA multi-omics study. iCCA tumor samples from a consecutive cohort of patients were obtained for WES (n = 249), RNA-seq (n = 255), proteomics (n = 208), H&E (n = 177), multiplex immunostaining (n = 188), and scRNA-seq (n = 10) analyses. (B) ESTIMATE immune scores from our cohort and other TCGA cancer types. (C) The procedure for constructing compendium of microenvironmental cells and clustering. (D,E) Consensus cumulative distribution function (CDF) plot (D) and delta area (change in CDF area) plot (E) of 170 prognostic gene-based classification were shown. (F) Correlations between mRNA and protein abundance in 103 mRNA-protein pairs from 170 prognostic microenvironment genes. (G) Unsupervised hierarchical clustering of the FU-iCCA cohort based on the 103 proteins identified three subgroups. (H) Relative abundance and proportion of samples harboring the indicated clinical covariate in the given immune subgroup (Wilcoxon rank sum test, Pearson’s chi-square test, or Fisher’s exact test). (I) Representative multiplex immunostaining images to show the distribution of PDPN+ lymphatic vessels and CD31+ blood vessels among the three immune subgroups (left panel). Quantification of staining intensities for the indicated markers are shown (ANOVA, right panel). (J) Representative H&E images from the FU-iCCA cohort were shown to display the indicated pathological TIL estimates. (K) Box plot showing the pathological TILs among three immune subgroups (ANOVA). (L) Representative multiplex immunostaining images to show the distribution of
CD66b+ neutrophils among the three immune subgroups (left panel). Quantification of staining intensities for the indicated markers are shown (right panel). Green arrows: CK19-CD66b+ (ANOVA).

**Figure S2 Validations of our grouping strategy in external cohorts.** (A,B) Heatmap with indicated immunogenomic features (A, left panel), Kaplan–Meier curves (A, right panel) and multivariate Cox proportional hazards model (B) for overall survival based on our subgrouping standard in Jusakul et al.’s cohort. H.R., Hazard ratio. C.I., confidence interval. (C,D) Heatmap with indicated immunogenomic features (C, left panel) and Kaplan–Meier curves (C, right panel) and multivariate Cox proportional hazards model (D) for overall survival based on our subgrouping standard in Job et al.’s cohort. H.R., Hazard ratio. C.I., confidence interval. (E-H) Comparisons of immune subgroups to the subgroups resulted from previously reported standards (E) and the heatmap of clustering analysis of Anderson et al (F), Oiishi et al (G), and Job et al (H). The P values were calculated by Pearson’s Chi-square-test.

**Figure S3 Potential different immunogenomic features of the three immune subgroups.** (A) Principal-component feature loadings (magnitude and direction) shown in the variables factor map. Vectors were colored according to a major biological classification of cancer hallmark gene sets and ESTIMATE tumor purity. (B,C) The means of loadings per hallmark gene set for PC1 (B) and PC2 (C),
respectively. The \( P \) values were calculated by Wilcoxon rank sum test. (D) Hierarchically-clustered heatmap of normalized enrichment scores (NES). Rows indicated hallmark gene sets and columns indicated tumor samples. Normalized tumor purity values and immune subgroups shown on top. (E) Hallmark pathway enrichment analysis after controlling for tumor purity (adjusted ANOVA). (F-H) Comparisons of COX-IS (F), cytolytic score (G), and immunophenoscore (H) among the three immune subgroups (Wilcoxon rank sum test).

**Figure S4 Association of the immune subgroups with genomic alterations.** (A) Comparison of total SCNA levels among the three immune subgroups (Wilcoxon rank sum test). (B) Correlation analysis between ESTIMATE immune score and SCNA burden. (C) The expression heatmaps depicted mRNA and protein levels of various markers involved in cancer-promoting or cancer-inhibitory inflammation (Wilcoxon rank sum test). (D,E) Comparisons of COX-IS (D) and cytolytic score (E) in tumor samples with and without \( KRAS \) mutation (Wilcoxon rank sum test). (F) GSEA analysis based on the MDSC signature up gene set in tumor samples with and without \( KRAS \) mutation. (G-I) Comparisons of neutrophil infiltration (G), COX-2 mRNA level (H), and S100A8 mRNA level (I) of patients with and without \( KRAS \) mutations in Jusakul’s cohort (Wilcoxon rank sum test). (J) Pathway enrichment analysis based on differentially expressed genes that associated with \( KRAS \) mutations in Jusakul’s cohort (Wilcoxon rank sum test). (K-O) HuCCT1 and HCCC9810 cells ectopically expressing Myc-tag \( KRAS^{WT} \) or Myc-tag \( KRAS^{G12D} \) were generated and then
qRT-PCR (K,L), western blot (M), and ELISA (O) were performed to detect the indicated molecules. Quantification of three independent western blotting assays were shown (N). The NC group indicated cell lines without transfection. Representative data of triplicate experiments were shown. Data are expressed as mean ± SEM (ANOVA). nd, not determined. (P) Transwell assay showing that the culture supernatants from KRAS$^{WT}$ and KRAS$^{G12D}$ overexpressing HuCCT1 and HCCC9810 significantly enhanced the migration of dHL60. Representative data of triplicate experiments were shown. Data are expressed as mean ± SEM (ANOVA). (Q) Culture supernatants from KRAS$^{G12D}$ overexpressing HuCCT1 and HCCC9810 treated with DMSO or 15 μM COX-2 inhibitor Celecoxib for 48 h were collected. Then the DMSO treated supernatant (DMSO group), DMSO treated supernatant plus 15 μM Celecoxib (CELE untreated group), and Celecoxib treated supernatant (CELE treated group) were used to perform the transwell assay. Representative data of triplicate experiments were shown. Data are expressed as mean ± SEM (ANOVA). (R,S) The effects of pre-treatment dHL60 by indicated culture supernatants on anti-CD3/CD28 activated Jurkat T cells. The apoptosis (R) and IL-2 (S) secretion of Jurkat T cells were detected. Representative data of triplicate experiments were shown. Data are expressed as mean ± SEM (ANOVA).

Figure S5 The KRAS mutation-associated tumor microenvironment of mouse iCCA model. (A) Schematic of vectors used in the hydrodynamic injections. Luc, luciferase. (B,C) Immunohistochemical analysis for CD8 and CD19 (B) and
quantification of staining intensities for the indicated markers are shown (C). Representative data of triplicate experiments were shown. Data are expressed as mean ± SEM. (D) UMAP plot showing the annotation and color codes for cell types in the FU-iCCA ecosystem. (E) Violin plot showing the expression of marker genes in the indicated cell types. The top dots label the clusters corresponding to specific cell types in (D) and the bottom shows indicated marker genes. (F) UMAP plot showing the group origin in AY and AYK mouse iCCA samples. (G) Violin plots showing $Ptgs2$ and $Il1b$ expression of indicated cell composition in AY and AYK samples, respectively. The $P$ values were calculated by ANOVA. (H) Overview of KRAS mutation-dominated myeloid inflammation.

**Figure S6 Immunogenomic features of IG2.** (A) Comparison of TNB among the three immune subgroups (Wilcoxon rank sum test). (B-D) Comparison of TCR diversity (B), BCR diversity (C), and RNA-seq reads mapping to VDJ loci (D) among the three immune subgroups (Wilcoxon rank sum test). (E) Heatmap showed the relative distribution of co-inhibitors (left panel) and co-stimulators (right panel) among the three immune subgroups (adjusted ANOVA). (F) Associations of HLA LOH with immune subgroups, clinicopathologic factors, and multi-omics profiles (Pearson’s Chi-square-test, Fisher’s exact test, or ANOVA).

**Figure S7 Multi-omics features related to TLSs.** (A) Associations of intra-tumoral TLSs with TLS score, immune subgroups, clinicopathologic factors, and multi-omics
profiles (ANOVA, Pearson’s chi-square test, or Fisher’s exact test). (B) Comparisons of xCELL enrichment scores of indicated immune subsets between TLSs- and TLSs+ subgroups (ANOVA). (C) Comparisons of calculated scores of co-inhibitors or co-stimulators between TLSs- and TLSs+ subgroups (ANOVA). (D,E) Comparison of cytolytic scores (D) and COX-IS (E) between TLSs- and TLSs+ subgroups (Wilcoxon rank sum test).

Figure S8 Influences of HBV infection on the immune microenvironment of iCCA. (A) Comparisons of HBV infection in patients of IG1 and IG2/IG3 (Fisher's exact test). (B) UMAP plot, showing the annotation and color codes for cell types in the FU-iCCA ecosystem. (C) PCA analysis of all FU-iCCA samples based on the genes exclusively enriched in IG1 and IG3. The 10 scRNA-seq samples were also projected onto the plot. (D) Dot plot of cell types assignment for indicated genes in scRNA-seq. Dot size indicated fractions of expressing cells, colored according to z-score normalized expression levels. (E,F) Bubble heatmap showing marker genes across T/NK subgroups (E) and myeloid subgroups (F) from tumor samples. Dot size indicated fraction of expressing cells, colored according to z-score normalized expression levels. (G) Immune clusters showing significant differences in the comparison among the three immune subgroups. The color saturation represented the ration to T/NK or myeloid subgroups. The P values were calculated by ANOVA. (H-J) Volcano plot showed differently expressed genes of CD4+ T cells (H), NK cells (I), and macrophages (J) in HBV positive versus HBV negative tumor samples.
Fig. S2

A

B

C

D

E

F

G

H

J Immunother Cancer

BMJ Publishing Group Limited (BMJ) disclaims all liability and responsibility arising from any reliance placed on this supplemental material which has been supplied by the author(s).

Fig. S7

A

TLS score ($P = 2.9e-50$)
Immune subgroup ($P = 4.5e-04$)
TNM stage ($P = 0.018$)
Tumor necrosis ($P = 6.5e-03$)
Microvascular invasion ($P = 0.033$)
Intrahepatic metastasis ($P = 0.011$)
Tumor size ($\geq 5\text{ cm}$) ($P = 8.1e-04$)

mRNA

G2/MI checkpoint
Hypoxia
Epithelial mesenchymal transition
Xenobiotic metabolism
Bile acid metabolism
Complement and coagulation cascades
Drug metabolism cytochrome p450
Peroxisome

Protein

TNF-α signaling via NF-κB
Glycolysis
Coagulation
Complement
Inflammatory response
Metabolism of xenobiotics by cytochrome p450
Oxidative phosphorylation
Fatty acid metabolism

B

TLS type

TLSs-
TLSs+

Enrichment score

P = 0.03, P = 1.2e-03, P = 0.15, P = 1.7e-04, P = 8.3e-03, P = 0.006, P = 0.01, P = 6.1e-03, P = 0.057

C

TLS type

TLSs-
TLSs+

Relative score

P = 0.068
P = 0.27

D

TLSs-
TLSs+

P = 0.032

E

TLSs-
TLSs+

P = 0.027