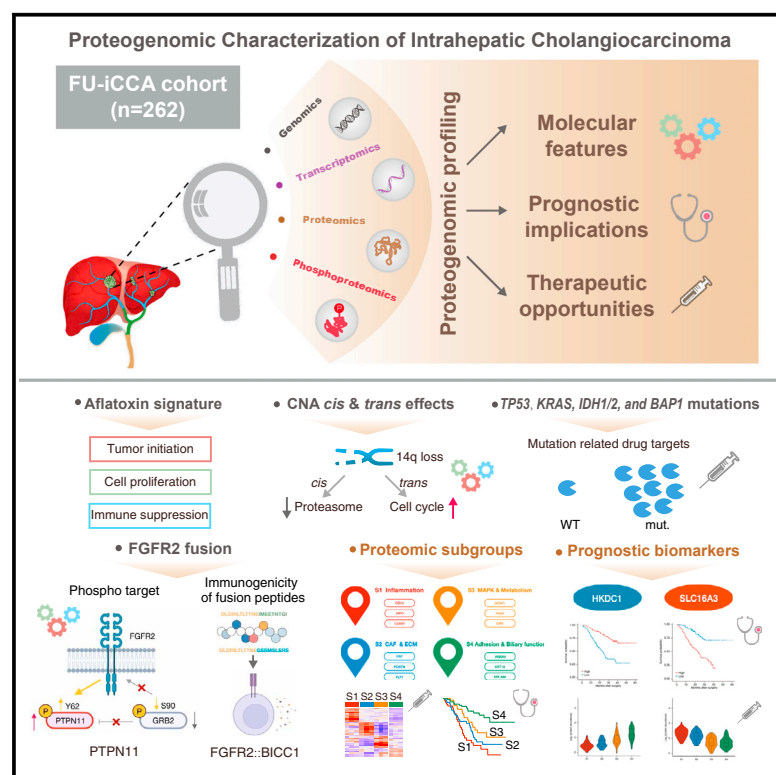


Proteogenomic characterization identifies clinically relevant subgroups of intrahepatic cholangiocarcinoma

Graphical abstract



Authors

Liangqing Dong, Dayun Lu, Ran Chen, ..., Daming Gao, Hu Zhou, Jia Fan

Correspondence

gaoqiang@fudan.edu.cn (Q.G.),
dgao@sibcb.ac.cn (D.G.),
zhouhu@simmm.ac.cn (H.Z.),
fan.jia@zs-hospital.sh.cn (J.F.)

In brief

Dong et al. yield insights into cancer etiology and taxonomy of intrahepatic cholangiocarcinoma (iCCA) through large-scale proteogenomics. They identify four iCCA subgroups with distinct clinical and multi-omics features, as well as prognostic biomarkers and potential therapeutic targets

Highlights

- Proteomic subgroups linked to iCCA patient survival, treatment, and molecular features
- *TP53*, *KRAS*, *FGFR2*, *IDH1/2*, and *BAP1* showed diverse pathways and therapeutic potential
- Immunogenomics revealed the peptide immunogenicity derived from *FGFR2* fusion
- *HKDC1* and *SLC16A3* as significant prognostic indicators with biological functions



Article

Proteogenomic characterization identifies clinically relevant subgroups of intrahepatic cholangiocarcinoma

Liangqing Dong,^{1,18} Dayun Lu,^{2,3,18} Ran Chen,^{3,4,18} Youpei Lin,^{1,18} Hongwen Zhu,^{2,18} Zhou Zhang,^{5,18} Shangli Cai,^{5,18} Peng Cui,⁵ Guohe Song,¹ Dongning Rao,¹ Xinpei Yi,¹⁴ Yingcheng Wu,¹ Nixue Song,^{2,3} Fen Liu,^{3,4} Yunhao Zou,^{3,4} Shu Zhang,¹ Xiaoming Zhang,⁶ Xiaoying Wang,¹ Shuangjian Qiu,¹ Jian Zhou,^{1,7} Shisheng Wang,⁸ Xu Zhang,^{9,10} Yongyong Shi,¹¹ Daniel Figeys,^{9,10} Li Ding,¹² Pei Wang,¹³ Bing Zhang,¹⁴ Henry Rodriguez,¹⁵ Qiang Gao,^{1,7,16,*} Daming Gao,^{3,4,17,*} Hu Zhou,^{2,3,10,*} and Jia Fan^{1,7,19,*}

¹Department of Liver Surgery and Transplantation, Liver Cancer Institute, Zhongshan Hospital, Key Laboratory of Carcinogenesis and Cancer Invasion of Ministry of Education, Fudan University, Shanghai 200032, China

²Department of Analytical Chemistry and CAS Key Laboratory of Receptor Research, Shanghai Institute of Materia Medica, Chinese Academy of Sciences, 555 Zuchongzhi Road, Shanghai 201203, China

³University of Chinese Academy of Sciences, Number 19A Yuquan Road, Beijing 100049, China

⁴State Key Laboratory of Cell Biology, CAS Center for Excellence in Molecular Cell Science, Shanghai Institute of Biochemistry and Cell Biology, Chinese Academy of Sciences, 320 Yueyang Road, Shanghai 200031, China

⁵Burning Rock Biotech, Shanghai 201114, China

⁶Key Laboratory of Molecular Virology and Immunology, Institut Pasteur of Shanghai, Chinese Academy of Sciences, Shanghai, China

⁷Key Laboratory of Medical Epigenetics and Metabolism, Institutes of Biomedical Sciences, Fudan University, Shanghai 200032, China

⁸Frontiers Science Center for Disease-related Molecular Network, Institutes for Systems Genetics, Key Lab of Transplant Engineering and Immunology, MOH, West China Hospital, Sichuan University, Chengdu 610041, China

⁹Ottawa Institute of Systems Biology, Department of Biochemistry, Microbiology and Immunology, Faculty of Medicine, University of Ottawa, Ottawa, ON, Canada

¹⁰Shanghai Institute of Materia Medica-University of Ottawa Joint Research Center in Systems and Personalized Pharmacology, 555 Zuchongzhi Road, Shanghai 201203, China

¹¹Key Laboratory for the Genetics of Developmental and Neuropsychiatric Disorders (Ministry of Education), The Collaborative Innovation Center for Brain Science, Bio-X Institutes, Shanghai Jiao Tong University, Shanghai, China

¹²Department of Medicine, McDonnell Genome Institute, Siteman Cancer Center, Washington University, St. Louis, MI 63108, USA

¹³Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA

¹⁴Lester and Sue Smith Breast Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

¹⁵Office of Cancer Clinical Proteomics Research, Division of Cancer Treatment and Diagnosis, National Cancer Institute, National Institutes of Health, Bethesda, MD 20892, USA

¹⁶State Key Laboratory of Genetic Engineering, Fudan University, Shanghai 200433, China

¹⁷School of Life Science, Hangzhou Institute for Advanced Study, University of Chinese Academy of Sciences, Hangzhou 310024, China

¹⁸These authors contributed equally

¹⁹Lead contact

*Correspondence: gaoqiang@fudan.edu.cn (Q.G.), dgao@sibcb.ac.cn (D.G.), zhouhu@sim.ac.cn (H.Z.), fan.jia@zs-hospital.sh.cn (J.F.)
<https://doi.org/10.1016/j.ccell.2021.12.006>

SUMMARY

We performed proteogenomic characterization of intrahepatic cholangiocarcinoma (iCCA) using paired tumor and adjacent liver tissues from 262 patients. Integrated proteogenomic analyses prioritized genetic aberrations and revealed hallmarks of iCCA pathogenesis. Aflatoxin signature was associated with tumor initiation, proliferation, and immune suppression. Mutation-associated signaling profiles revealed that *TP53* and *KRAS* co-mutations may contribute to iCCA metastasis via the integrin-FAK-SRC pathway. *FGFR2* fusions activated the Rho GTPase pathway and could be a potential source of neoantigens. Proteomic profiling identified four patient subgroups (S1–S4) with subgroup-specific biomarkers. These proteomic subgroups had distinct features in prognosis, genetic alterations, microenvironment dysregulation, tumor microbiota composition, and potential therapeutics. *SLC16A3* and *HKDC1* were further identified as potential prognostic biomarkers associated with metabolic reprogramming of iCCA cells. This study provides a valuable resource for researchers and clinicians to further identify molecular pathogenesis and therapeutic opportunities in iCCA.

INTRODUCTION

Intrahepatic cholangiocarcinoma (iCCA) is the second most common primary hepatic malignancy. The incidence of iCCA is expected to increase worldwide by up to 10-fold during the next two to three decades (Rizvi et al., 2018), with treatment strategies still limited. Most patients are typically diagnosed at advanced stages, when the available systemic therapies are of limited effectiveness (Banales et al., 2020). Therefore, deeper mechanistic insights into the pathogenesis of iCCA are urgently needed to facilitate the identification of suitable targets for therapy.

Recent progress in genomic and transcriptomic sequencing analysis has revealed the genetic landscape of iCCA (Farshidfar et al., 2017; Jusakul et al., 2017; Lowery et al., 2018; Nakamura et al., 2015). Accordingly, precision medicine developed for promising molecular targets, including FGFR2 inhibitors and IDH1/IDH2 inhibitors, is becoming a reality (Abou-Alfa et al., 2020a, 2020b). Concurrently, multi-dimensional “omics” strategies encompassing proteomic and phosphoproteomic profiling in conjunction with genomic analysis have elucidated new disease subtypes and signaling pathways, and have discovered potential targets for therapeutic development in other cancer types (Clark et al., 2019; Dou et al., 2020; Gao et al., 2019; Jiang et al., 2019; Li et al., 2020; Vasaikar et al., 2019). This multi-omics strategy may help reveal new mechanisms and identify novel targets for developing therapies to provide additional treatment options for iCCA patients.

To this end, we performed an integrative genomic, transcriptomic, proteomic, phosphoproteomic, and microbiome analysis of 262 Chinese iCCA patients, under the auspices of the International Cancer Proteogenome Consortium (Rodríguez and Pennington, 2018). The underlying data serve as an important resource for further biological, diagnostic, and drug discovery efforts. Our integrated proteogenomic analysis illustrated that genetic alterations may drive phenotypic perturbations, and revealed disease subgroups with distinct features and prognosis, which delineated the complex mechanisms of iCCA pathogenesis for prospective exploration of precision management.

RESULTS

Proteogenomic landscape of Chinese iCCA

To characterize the proteogenomic landscape of iCCA in China, 262 prospectively collected treatment-naïve iCCA tumors and paired adjacent non-tumor liver tissues (Fudan University [FU]-iCCA cohort) were obtained (Figure S1A). In total, 12,197 (range, 1–495, averaging 48 per sample) nonsynonymous somatic point mutations and 603 indels (range, 0–43, averaging 2.4 per sample) were identified (Table S1). Isobaric tandem mass tag-based global proteomic/phosphoproteomic profiling quantified 10,529 unique proteins (average 7,385) and 62,879 phosphosites (average 21,431) from 9,832 phosphoproteins, with 5,690 proteins and 2,761 phosphoproteins quantified across all 214 samples (Table S1; Figures S1B–S1D). Totally, 1,413 (1.60%) nonsynonymous somatic variants were confirmed at the protein level (Figure S1E).

The landscape of genetic alterations in FU-iCCA was generally consistent with previous reports (Farshidfar et al., 2017; Jusakul

et al., 2017; Lowery et al., 2018; Nakamura et al., 2015), identifying 16 significantly altered driver genes, including *TP53*, *KRAS*, *FGFR2*, *IDH1/2*, *BAP1*, *ARID1A*, and *PBRM1* (false discovery rate [FDR] < 0.01, MutSigCV; Figures 1A and S1F). Nine (3.5%) samples were revealed to have a relatively high tumor mutation burden (TMB) (>199 mutations/exome) (Allgauer et al., 2018), including one sample defined as microsatellite instability-high. Further analysis showed that *KRAS* mutation was mutually exclusive with *IDH1/2*, *BAP1*, and *FGFR2* alterations. *FGFR2* alterations were mutually exclusive with *TP53*, *KRAS*, and *IDH1/2* mutations (Figure S1G; Table S1). This lack of overlap indicates that those alterations may be independent driver events of biliary carcinogenesis.

Of note, different mutational distribution and mutation signature were observed among FU-iCCA and other iCCA cohorts (Hoadley et al., 2018; Jiao et al., 2013; Jusakul et al., 2017; Zou et al., 2014), showing geographically different patterns. The FU-iCCA cohort showed higher *KRAS* mutation frequency and lower *IDH1*, *ARID1A*, and *TERT* mutation frequencies than the Memorial Sloan Kettering Cancer Center (MSKCC) cohort, which is comprised of a western iCCA population (Figure 1B). An aflatoxin signature (signature 24; Figures 1C, S1H, and S1I) was observed in 7.1% of the FU-iCCA cohort and 9.7% of Zou et al.’s cohort, both as Chinese cohorts, but was absent in the three western cohorts. A similar trend was observed for aristolochic acid (AA) signature (signature 22; Figure 1C). The samples with aflatoxin and AA signatures showed significantly higher TMB ($p = 1.0 \times 10^{-4}$ and 7.2×10^{-3}) and neoantigen load ($p = 1.9 \times 10^{-3}$ and 2.1×10^{-3} , Wilcoxon test; Figure 1D). The tumors harboring AA signature showed significantly increased infiltration of activated NK cells and enhanced CD160 expression (Figure S1J), similar to previous findings in hepatocellular carcinoma (Gao et al., 2019). The tumors with aflatoxin signature were locally immunosuppressed without any specific enrichment of immune cells or immune checkpoints (data not shown). By comparing the multi-omics data of all tumors with or without aflatoxin signature, we observed that tumors with aflatoxin signature showed upregulation of DNA repair and cell-cycle pathways and downregulation of apoptosis, infection inflammation, and immune pathways (Figure 1E; Table S1). Further analysis revealed that *TP53* mutations, especially R249S, were significantly associated with aflatoxin signature ($p = 7.3 \times 10^{-4}$, $p = 3.7 \times 10^{-4}$, chi-square test; Figures 1F and 1G), consistent with a previous report that aflatoxin could induce R249S alteration (Hsu et al., 1991), implicating the contributing role of aflatoxin in iCCA carcinogenicity.

Multi-omic analysis of somatic copy number alterations

The most frequent gains occurred in chromosomes 6p21.33 (43.5%) and 1q21.3 (33.2%), with the most frequent losses occurring in chromosomes 5q23.3 (20.2%) and 1p36.11 (20.2%) (Figure S2A; Table S2), consistent with previous studies (Farshidfar et al., 2017; Wardell et al., 2018). We also observed amplifications in known oncogenes including *STK19* (42.0%), *HIST1H1E* (35.9%), *MCL1* (32.1%), and *MDM4* (26.0%), and deletions of tumor suppressors including *ARID1A* (19.5%), *MAP2K4* (14.5%), and *MLH1* (13.0%). Totally, 39.3% (101/253) of FU-iCCA harboring at least one actionable genetic alteration, including 61 mutant cases, 9 cases with high TMB, 32 cases with fusions, and 4 cases with amplification.

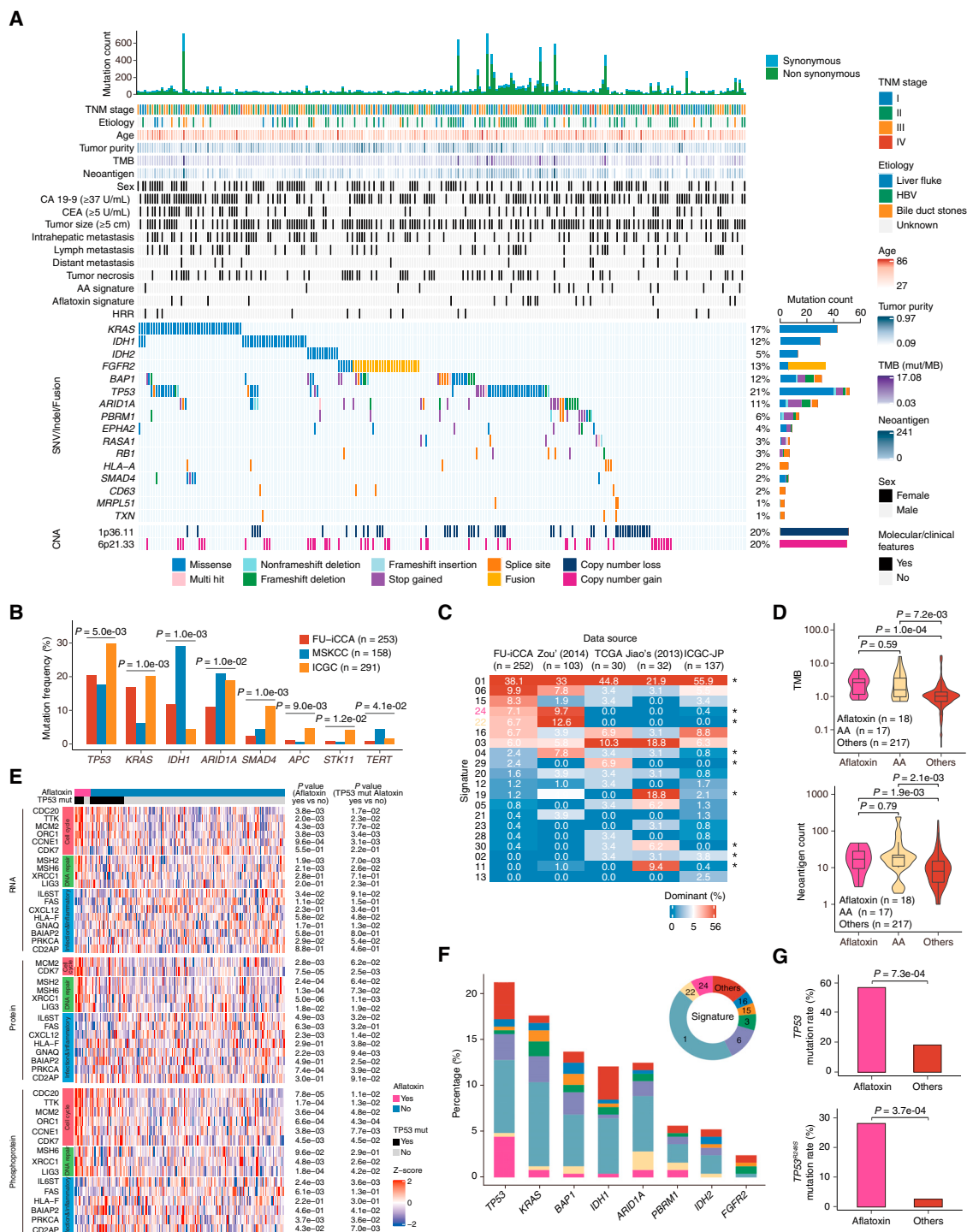


Figure 1. Genomic landscape of the FU-iCCA cohort

(A) Genetic profile and associated clinicopathologic features of all 253 iCCA patients.

(B) Mutation frequency of 8 significantly mutated genes in the FU-iCCA, MSKCC (Lowery et al., 2018), and International Cancer Genome Consortium (ICGC) cohorts (Jusakul et al., 2017) (chi-square test).

(C) Comparison of mutation signatures between the FU-iCCA cohort and other published iCCA cohorts. Asterisks represent $p < 0.05$ by Fisher's exact test.

(D) Comparison of TMB and tumor neoantigen burden between iCCA tumors with dominant aflatoxin signature, dominant AA signature, and others (Wilcoxon test). Boxes indicate interquartile range (IQR) with a line at the median. Whiskers represent the minimum and maximum values at most $1.5 \times$ IQR from the quartiles. The width of violin plots represents probability that members of the population will take on the given value.

(legend continued on next page)

These copy number alterations (CNAs) had *cis* and *trans* impacts on mRNA, protein, and phosphoprotein abundance (Figures 2A and S2B). Totals of 3,981, 1,081, and 408 significant *cis* correlations were observed for mRNA, proteins, and phosphoproteins, respectively, with only 194 significant *cis* effects overlapping across all 3 omics (Figure 2B). In total, 963 proteins showed both CNA-mRNA and CNA-protein *cis* effects, which were mainly enriched in metabolic, biosynthetic, and protein-processing pathways (Figure 2C; Table S2). Among the 723 cancer gene census (CGC) (Sondka et al., 2018), 15 CGC genes showed significant overlapping CNA-mRNA and CNA-protein *cis* regulations, including *AKT1*, *ARHGAP5*, *ARHGEF12*, *DICER1*, *EGFR*, *EPS15*, *ERCC5*, *EZR*, *GOLGA5*, *KTN1*, *MAP2K4*, *PCM1*, *PRCC*, *RB1*, and *USP8* (FDR < 0.01; Figure S2C). The CGCs with top 10 *trans* events are shown, among which *MDM4* amplification has been linked with the hyperprogression in immunotherapy (Figure 2D) (Kato et al., 2017). *MDM4* amplification correlated with proteins enriched in spliceosome and mRNA surveillance pathways, and anti-correlated with proteins enriched in protein export, VEGF signaling pathway, and FcγR-mediated phagocytosis (Figure S2D), providing clues for hyperprogression in immunotherapy.

The genes showing the most divergent number of CNA-protein and CNA-mRNA *trans* events fell in chromosomes 1p and 14q (Figure 2A). Notably, 14q loss showed differential correlations with global proteomic and transcriptomic abundance (Figure 2E), suggesting potential protein-level regulation in these regions. For the 585 genes whose protein abundances increased with 14q loss, enrichment analysis revealed pathways of spliceosome, mismatch repair, DNA replication, cell cycle, and metabolism (Figure 2F; Table S2). Besides, 14q loss showed *cis* and *trans* effects on protein expression (significant inverse correlation), rather than mRNA expression, of proteasome genes and cell-cycle regulators, respectively (Figure 2G). Altogether, loss of 14q-associated global *cis* and *trans* correlations may provide mechanistic underpinnings for iCCA tumorigenesis (Figure 2H). Correlation analysis of paired transcriptomic and proteomic data showed that 97.15% of 5,434 mRNA-protein pairs had positive correlations (Figure S2E; Table S2), and 2.85% had negative correlations. These strong mRNA-protein positive correlations were enriched in metabolism and hematopoietic cell lineage pathways, while those with weaker or negative correlations were involved in ribosome, spliceosome, and oxidative phosphorylation processes (Figure S2E).

Proteogenomic association analysis of somatic driver mutations

The impacts of somatic driver mutations on proteome and phosphoprotein abundance were analyzed, emphasizing on the top four significantly mutated drivers (*TP53*, *KRAS*, *IDH1/2*, and *BAP1*). Generally, *TP53* and *BAP1* mutations caused loss of function of the original tumor suppressors (Aubrey et al., 2016;

Carbone et al., 2020), while *KRAS* and *IDH1/2* mutations were oncogenic (Aguirre and Hahn, 2018; Waitkus et al., 2018). Among them, mutations in *TP53* ($p = 1.1 \times 10^{-2}$) and *KRAS* ($p = 5.2 \times 10^{-4}$) were significantly associated with poor survival, while *BAP1* ($p = 0.20$) or *IDH1/2* ($p = 0.90$) mutations were not associated with patient survival in the FU-iCCA cohort (Figures S3A–S3C). Proteogenomic analysis by comparing tumors harboring each of the four mutations with the rest was then conducted (Table S3). Here, *TP53* mutations were associated with the upregulation of cell-cycle, drug metabolism, phagosome, and carbon metabolism pathways, and the downregulation of ECM-focal adhesion, PI3K-AKT, and Hippo-YAP signaling pathways (Figures 3A and S3D; Table S3). *KRAS* mutations were associated with the increased proteins in inflammation-infection and ECM-focal adhesion pathways, and the decreased proteins in cell-cycle pathway (Figures 3B and S3D; Table S3).

Interestingly, 10/215 iCCA patients in our cohort harbored *TP53*/*KRAS* co-mutations, and had significantly worse survival than patients with *TP53* or *KRAS* single mutation ($p = 3.0 \times 10^{-5}$, log-rank test; Figure 3C), suggesting their unique molecular features. Cell adhesion-related molecules, including *ITGA6*, *ITGB4*, *ITGB6*, *CDH3*, and *CLDN18*, were most highly expressed in *TP53*^{Mut}/*KRAS*^{Mut} tumors, compared with *TP53*^{Mut}/*KRAS*^{WT}, *TP53*^{WT}/*KRAS*^{Mut}, and *TP53*^{WT}/*KRAS*^{WT} tumors (Figures 3D and 3E). Integrins can mediate cell-cell/cell-matrix interactions, and transduce extracellular signals by recruiting focal adhesion kinases (FAKs), SRC family kinases, and other signaling proteins (Hynes, 1992; Schlaepfer et al., 1999), thus contributing to cell mobility and metastasis (Desgrosellier and Cheresch, 2010). Notably, it has been reported that *KRAS* and *TP53* mutations could promote *ITGA6* expression through transcription factor FOSL1 (found elevated at mRNA level in *TP53*^{Mut}/*KRAS*^{Mut} tumors) and miR-30e-5p, respectively (Laudato et al., 2017; Zhang et al., 2017). Consistently, *TP53*^{Mut}/*KRAS*^{Mut} tumors had the highest percentage of regional lymph node metastasis (40%), compared with *TP53*^{WT}/*KRAS*^{Mut} tumors (29.6%), *TP53*^{Mut}/*KRAS*^{WT} tumors (24.1%), and *TP53*^{WT}/*KRAS*^{WT} tumors (16.7%). Therefore, *TP53*/*KRAS* co-mutations may contribute to iCCA metastasis and progression via the integrin-FAK-SRC pathway (Figure 3F). Further analysis indicated that ECM and bile secretion pathways were mutually active in both *BAP1*^{Mut} tumors and *IDH1/2*^{Mut} tumors, while phagosome, inflammation, and MAPK pathways were only relatively active in either of them (Figures 3G, 3H, and S3D). This was in contrast to *TP53*^{Mut} and *KRAS*^{Mut} tumors, in which bile secretion pathway was relatively less active than other tumors (Figure S3D).

Next, we analyzed the potential therapeutic targets elevated in tumors with each mutation. Cell-cycle-related factors, such as TOP2A, CDK1, and PCNA, were identified in *TP53*-mutated tumors, while ECM-interaction and kinase signaling factors, including *ITGA2B*, *ITGB6*, and *RIPK3*, stood out in tumors with

(E) Heatmap showing mRNA, protein, and phosphorylation relative Z score abundance of aflatoxin- and *TP53* mutation-influenced pathways in the FU-iCCA cohort. p values were calculated by Wilcoxon test (mRNA) or limma (protein and phosphoprotein).

(F) Distribution of mutational signatures associated with driver mutations. The probability of each driver mutation due to each mutational process were estimated and summed by genes (barplot). The cumulative probabilities across all driver mutations are shown in the pie chart.

(G) Comparison of *TP53* mutation rate between iCCA tumors with or without dominant aflatoxin signature (Fisher's exact test). The frequency of *TP53* and *TP53* R249S occurrence in both groups are shown as histograms.

See also Figure S1 and Table S1.

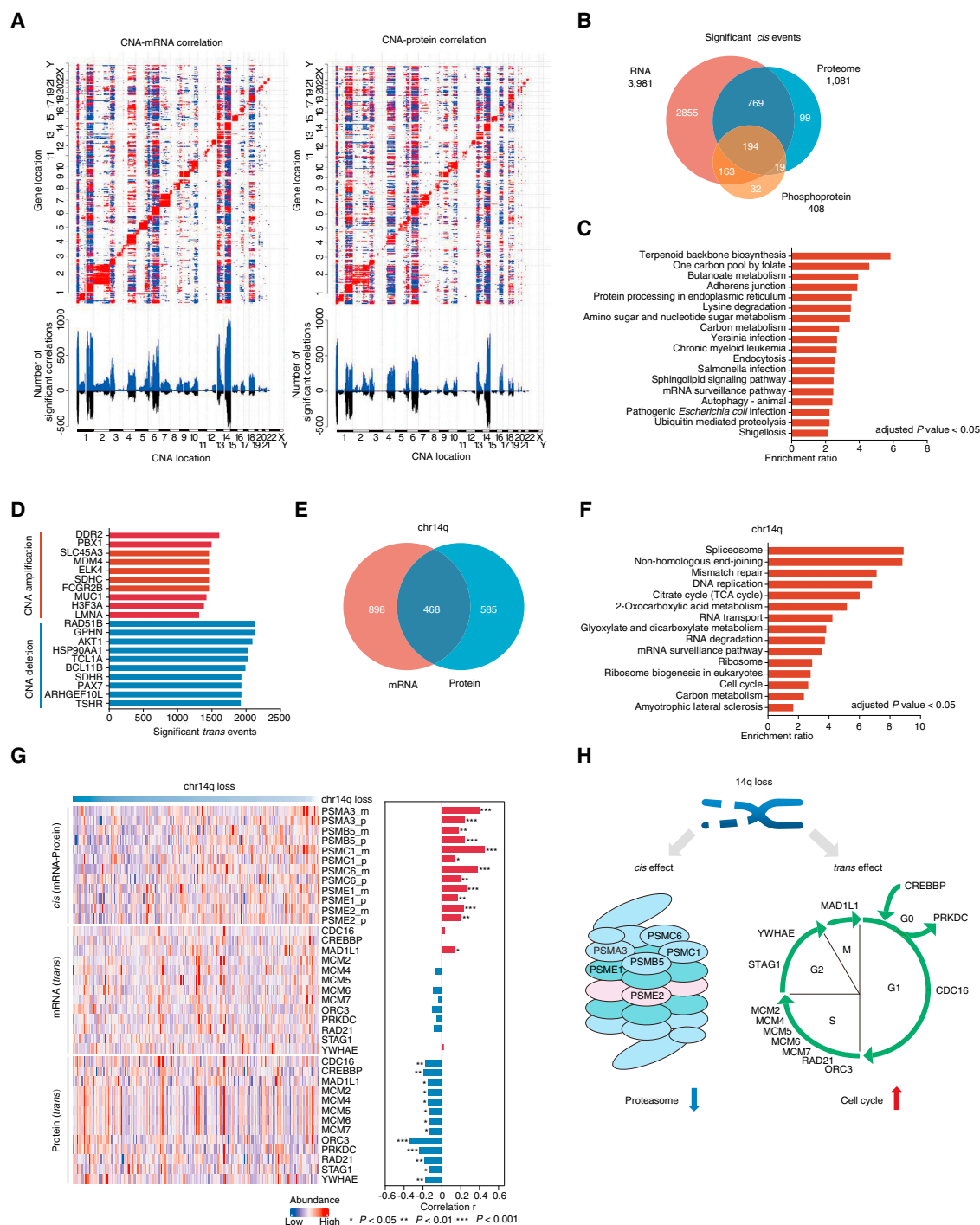


Figure 2. Effects of copy number alterations on mRNA and protein abundance

(A) Correlations of CNA (x axes) to mRNA (left) and protein expression (right) levels (y axes) highlight CNA *cis* and *trans* effects. Significant (FDR < 0.01, Spearman's correlation) positive (red) and negative (blue) correlations between CNA and mRNAs or proteins are indicated. The numbers of significant *cis* and *trans* events corresponding to the indicated genomic loci specific to mRNAs or proteins are represented by blue bars, and overlap between CNA-mRNA and CNA-protein events are represented by black bars.

(B) Venn diagram depicting significant *cis* CNA-mRNA, CNA-protein, and CNA-phosphoprotein correlations (FDR < 0.01, Spearman's correlation).

(C) Enrichment analysis of overlapped genes of *cis* CNA-mRNA and CNA-protein correlations (n = 963 genes, FDR < 0.01, Spearman's correlation).

(D) *trans* effect events of CNA-protein (FDR < 0.01, Spearman's correlation).

(E) Venn diagram of mRNAs/proteins having significant negative *trans* CNA-mRNA and/or CNA-protein correlations with 14q.

(legend continued on next page)

KRAS mutation. RNAi and chemical inhibitor treatment were performed with patient-derived primary cancer cell lines (PDPs: ICC1370 [*TP53* p.M246I], ICC880 [*KRAS* p.G12D], and ICC772 [*TP53*^{WT}; *KRAS*^{WT}]) established previously (Dong et al., 2018), and successfully validated those therapeutic targets for iCCA with *TP53* or *KRAS* mutations (Figures 3I and S3E–S3G). As expected, *TP53*, *KRAS*, *BAP1*, or *IDH1* mutations had impacts on their own mRNA, protein, and phosphorylation levels. *KRAS* mutation led to significantly elevated expression, while *TP53* and *BAP1* mutants had lower expression. Besides, mutated *TP53* had elevated phosphorylation level, but mutated *BAP1* had reduced phosphorylation level ($p < 0.05$, Wilcoxon test). However, no significant expression changes in *IDH1* mutation were observed (Figure S3H).

Phosphoproteomic aberrations associated with *FGFR2* alterations and *KRAS* mutations

Enhanced *FGFR2* signaling in iCCA is mediated by mutations and chromosomal translocations. Our results showed that all the *FGFR2* mutations located outside the kinase domain (Figure 4A), in contrast to the activating mutations in *EGFR* (Gillette et al., 2020). The S372C, Y375C, and C382R mutations were gain of function through promoting constitutive *FGFR2* dimerization (Brown et al., 2014; Pollock et al., 2007; Wang et al., 2012). *FGFR2* fusions found in 11.9% (30/253) of FU-iCCA were mostly generated with fusion partners containing dimerization domains to induce ligand-independent receptor dimerization. The fusion partners included *BICC1* ($n = 11$), *AHCYL1* ($n = 2$), *ETV6* ($n = 2$), *INA* ($n = 2$), and others, all of which fused with the same *FGFR2* break-apart point (p.E767) and retained its kinase domain (Figure 4B). Each of these fusions was validated by fluorescent *in situ* hybridization (Figure 4C), and the identical mRNA sequence around the breakpoint was confirmed (Figure S4A). Nearly all the tumor cells of each case with *FGFR2* fusion displayed the separated probe signals, indicating *FGFR2* fusions as the early clonal events in iCCA. Both mutations and fusions were categorized into *FGFR2* alterations.

As *FGFR2* alterations and *KRAS* mutations may affect the MAPK, PI3K-AKT-mTOR, and Rho GTPase pathways (Arai et al., 2014; Gillette et al., 2020), we explored the significantly altered phosphosites and their corresponding protein expression in tumors with and without these alterations, revealing phosphosite outliers in both (Figure 4D). *KRAS*-mutated tumors showed significant cascade activation of the MAPK pathway, such as MAPK14 phosphorylation at T180 and Y182, MAP3K2 phosphorylation at S514, and RPS6KA4 phosphorylation at S343 and T687 ($p < 0.01$, Wilcoxon test). In contrast, the MAPK pathway was slightly downregulated in *FGFR2*-altered tumors, while the Rho GTPase pathway was significantly upregulated (Figure 4E). In addition, *FGFR2*-altered tumors showed remarkable tyrosine phosphorylation of PTPN11 at Y62 and TLN1 at Y70, with a similar trend observed at the protein level ($p < 0.01$) (Figure 4F). Heatmaps showed simultaneous upregulation of *FGFR2* and PTPN11 proteins and PTPN11 Y62 phos-

phorylation, with downregulation of GRB2 expression and S90 phosphorylation in *FGFR2*-altered tumors (Figure 4G). Usually, dimeric GRB2 can bind to *FGFR2* under physiological conditions, resulting in partial phosphorylation of *FGFR2* and inhibiting both dephosphorylation of *FGFR2* by PTPN11 and phosphorylation of PTPN11 by *FGFR2* (Ahmed et al., 2013). When *FGFR2* fusion occurs, GRB2-*FGFR2* interaction may be affected, without which PTPN11 can interact with *FGFR2* to increase baseline phosphorylation, eliciting a significant downstream kinase response (Lin et al., 2012). Notably, the outlier residue Y62 lies in the interface between the N-SH2 and PTP domains, where phosphorylation was thought to stabilize the active protein conformation (Ren et al., 2010) (Figures 4G and 4H). Therefore, inhibition of PTPN11 may be a promising therapeutic strategy for *FGFR2*-altered populations beyond *FGFR2* inhibitors (Chen et al., 2016).

Alternative splicing within the third Ig domain of *FGFR2* creates two main isoforms, of which the dynamic distribution will affect the corresponding FGF signals (Figure S4B) (Ishiwata et al., 2012; Yan et al., 1993). *FGFR2*-altered tumors had a relatively high *FGFR2* expression along with high *Il1b* expression, while the *Il1c* was quite limited (Figure S4C). However, the expression of classical *Il1b*-related ligands showed no differences between the two groups, suggesting that *FGFR2* signaling was activated mainly through high expression of mutant autophosphorylation-activated protein, rather than ligand signals, in *FGFR2*-altered tumors (Figure S4D). Conversely, tumors without *FGFR2* alterations had higher *Il1c* expression, implying a splicing switch from *Il1b* to *Il1c* during malignant transformation (Ishiwata et al., 2012).

Identification of potential neoepitopes derived from *FGFR2::BICC1* fusion and corresponding T cell responses

Gene fusions could generate neo-peptide candidates for personalized immunotherapy (Gao et al., 2018; Yang et al., 2019). For the two most frequent *FGFR2::BICC1* fusion proteins, the sequences spanning through breakpoint residues (DLDRILTLTTNEIMEETNTQI and DLDRILTLTTNEGSSMSLSRS) were used to build the sequence database containing all the potential eight to ten amino acid peptides (Figure 5A). By tetramer exchange assay, a group of these synthetic peptides were successfully exchanged to the HLA-A02:01 and HLA-A24:02 tetramers (Figure 5B). Furthermore, we generated iCCA cell lines, HuCCT1 and HCCC-9810, that ectopically expressed *FGFR2::BICC1* fusion protein identified in our cohort (Figures S4F and S4G). MHC-I immunopeptidome directly detected the fusion peptide LTLTTNEIME from *FGFR2::BICC1*-overexpressed HuCCT1 cells (Figure S4H; Table S4D). These results support that fusion peptides could be presented by corresponding HLA alleles.

T cells from healthy donors may provide a rich source of T cell repertoires that could specifically recognize neoantigens presented on tumors (Stronen et al., 2016; van der Lee et al.,

(F) Enrichment analysis of proteins with only negative *trans* CNA-protein correlations in 14q ($n = 585$ genes, FDR < 0.01, Spearman's correlation).

(G) Heatmaps depicting copy-number loss of 14q and the mRNA/protein abundance of proteasome and cell cycle.

(H) Schematic diagram showing the *cis* and *trans* effects of 14q on proteasome and cell cycle, respectively.

See also Figure S2 and Table S2.

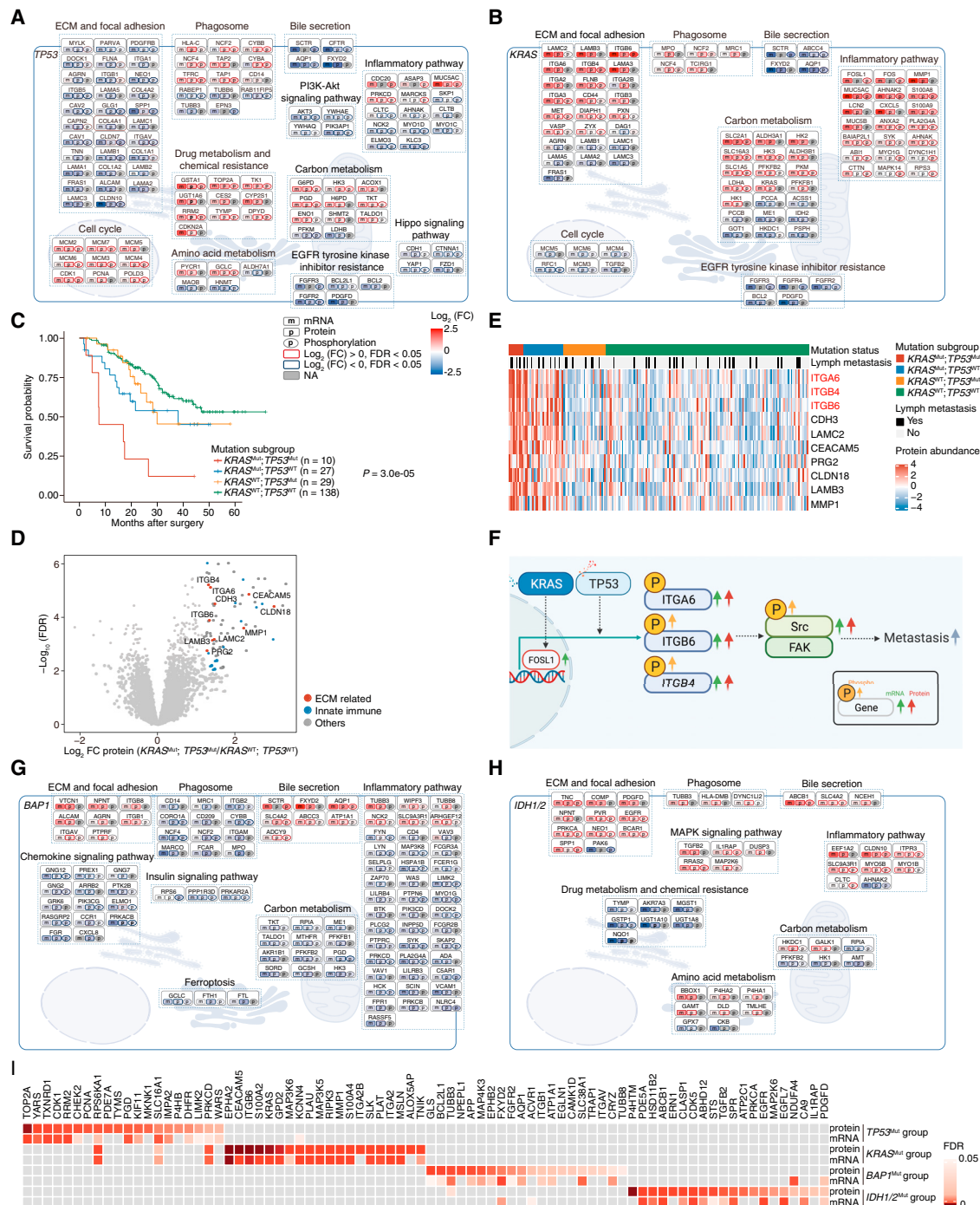


Figure 3. Impact of driver mutations on the proteogenomic landscape

(A and B) Overview of major signaling pathways associated with *TP53* mutation (A) or *KRAS* mutation (B) based on integrated proteogenomic analysis.

(C) Kaplan-Meier curves for overall survival based on the mutation status of *KRAS* and *TP53* (log-rank test).

(D) Volcano plot comparing protein abundance between the *KRAS*^{Mut}, *TP53*^{Mut} group and the *KRAS*^{WT}, *TP53*^{WT} group. Genes in ECM pathway are significantly upregulated in the *KRAS*^{Mut}, *TP53*^{Mut} group.

(E) Heatmap showing global abundance of enriched proteins in the ECM pathway that differentially expressed among *KRAS*^{Mut}, *TP53*^{Mut}, *KRAS*^{Mut}, *TP53*^{WT}, *KRAS*^{WT}, *TP53*^{Mut}, and *KRAS*^{WT}, *TP53*^{WT} groups.

(F) Schematic diagram illustrating the molecular mechanism for the increased intrahepatic metastasis caused by *KRAS* and *TP53* co-mutation.

(G and H) Overview of major signaling pathways associated with *BAP1* mutation (G) or *IDH1/2* mutation (H) based on integrated proteogenomic analysis.

(l) Heatmap showing top druggable genes enriched (FC > 1.2, FDR < 0.05) at the mRNA and protein levels that associated with *TP53*, *KRAS*, *BAP1*, and *IDH1/2* mutations.

See also [Figure S3](#) and [Table S3](#).

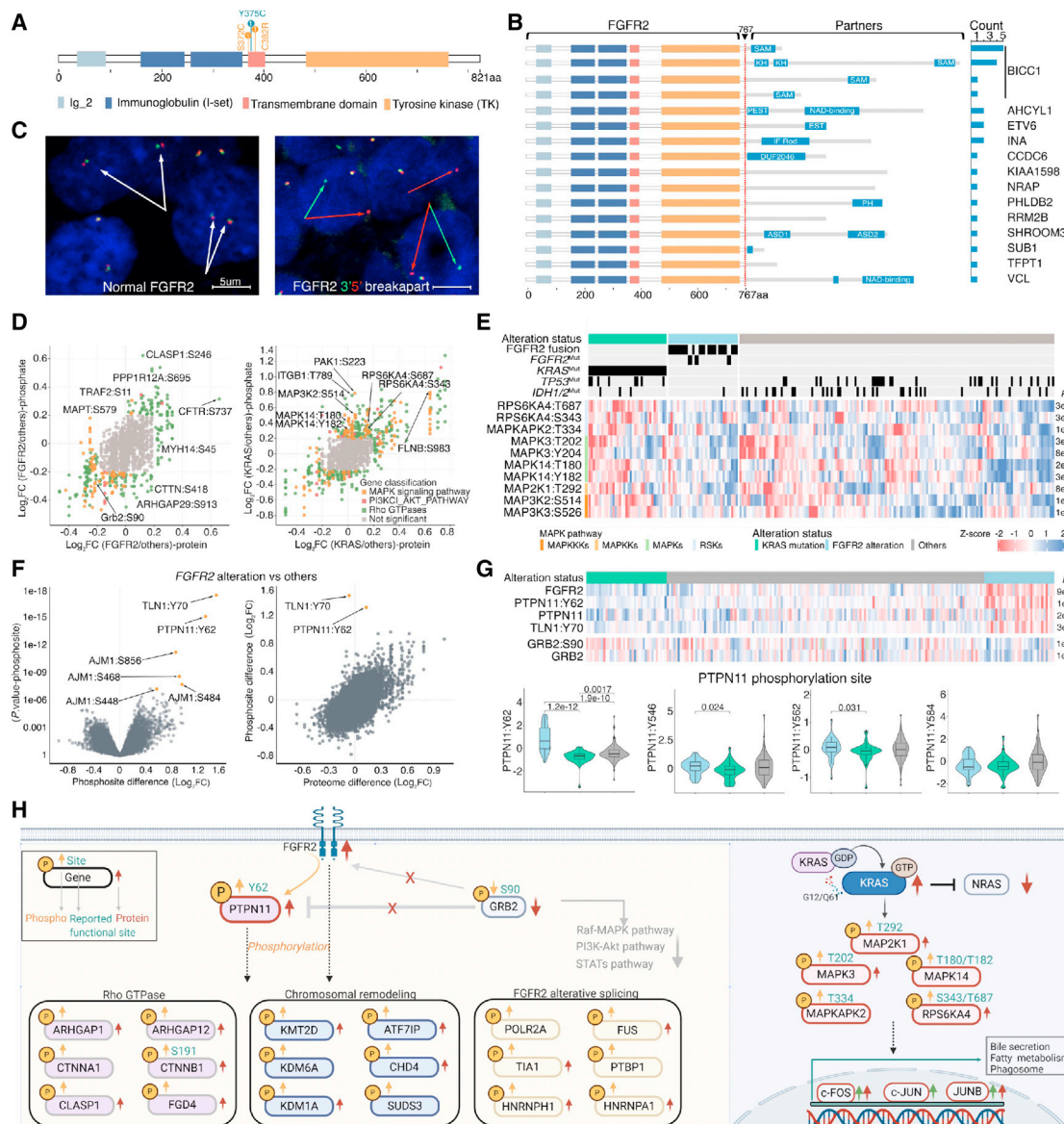


Figure 4. Impact of *FGFR2* alteration and *KRAS* mutation on proteogenomic landscape

(A) Somatic mutations of *FGFR2* in the FU-ICCA cohort. The number in the circle indicates the frequency.

(B) Schematic representation of the all *FGFR2* fusions identified by RNA-seq.

(C) *FGFR2* fusions validated by fluorescent *in situ* hybridization using two specific probes for the 5' (red) and 3' (green) part of *FGFR2*.

(D) Scatterplot showing significantly regulated phosphosites and their corresponding protein expression in tumors with and without *FGFR2* alterations (left) or *KRAS* mutations (right). Altered genes included three groups: the MAPK signaling pathway (yellow), the PI3K/AKT/mTOR pathway (red), and the Rho GTPases pathway (green).

(E) Phosphorylation status of the MAPK cascade aligned with the alteration status (Wilcoxon test).

(F) Scatterplot showing significantly regulated phosphosites and their corresponding protein expression in tumors with and without *FGFR2* alterations (Wilcoxon test).

(G) Heatmaps showing the protein and phosphorylation changes among *FGFR2*- and *KRAS*-altered groups (upper) and boxplots showing the elevated phosphopeptide expression (Y62) in tumors with *FGFR2* alteration (lower) (Wilcoxon test). Boxes indicate interquartile range (IQR) with a line at the median. Whiskers represent the minimum and maximum values at most 1.5* IQR from the quartiles.

(H) Overview of significantly enriched pathways associated with *FGFR2* alteration and *KRAS* mutations.

See also Figure S4 and Table S4.

2019). Here, peripheral blood mononuclear cells (PBMCs) from three healthy donors were used to perform *in vitro* stimulation for neoantigen-specific T cell response. 3 PBMCs were co-

cultured with the above fusion peptides and followed with interferon- γ (IFN- γ) ELISPOT assay. IFN- γ secretion was significantly upregulated when PBMCs from donor 1 and donor 2 were

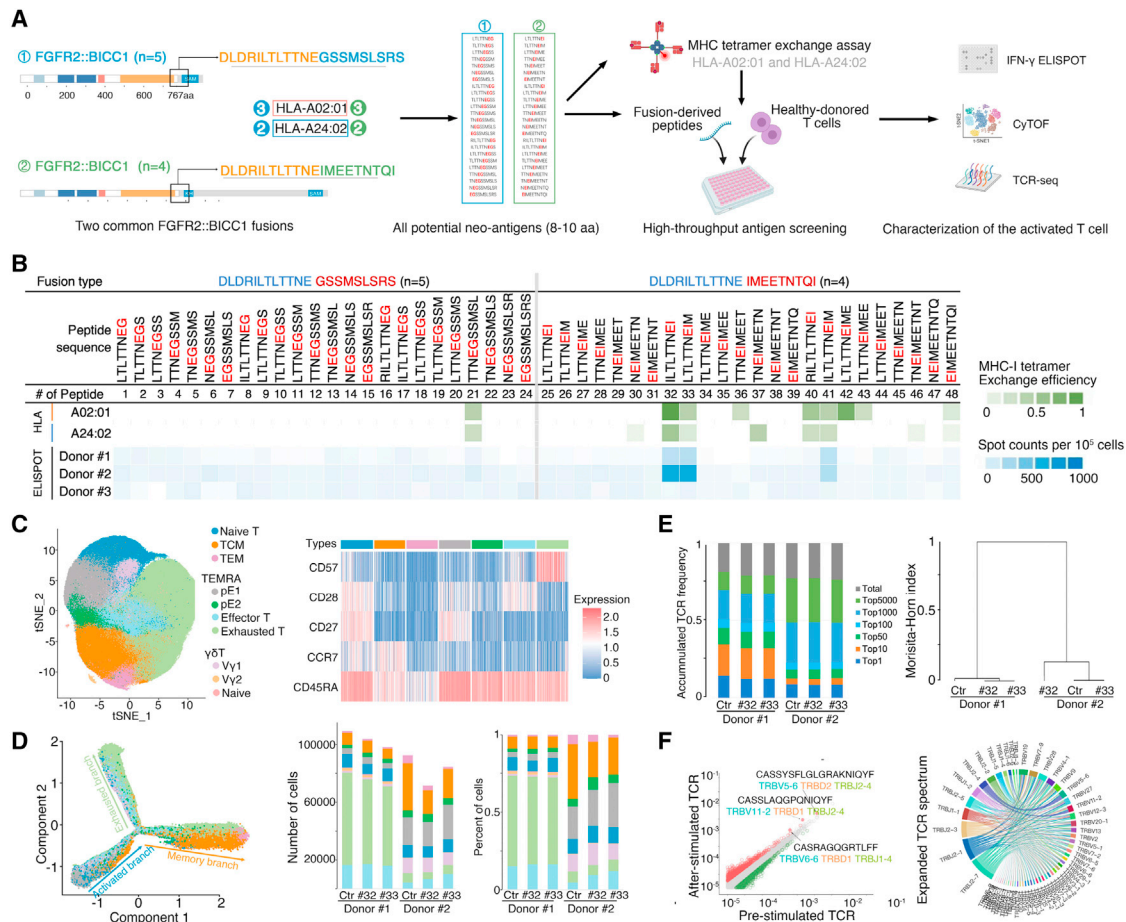


Figure 5. Identification of potential neoepitopes derived from FGFR2:BICC1 fusion and their neoepitope-reactive T cell phenotypes

(A) Schematic diagram of immunogenicity analysis for FGFR2:BICC1 fusion-derived neo-epitopes.

(B) The exchange efficiency of HLA-A02:01 and HLA-A24:02 tetramers with all the fusion-derived peptides (upper panel) and the spot counts of IFN-γ ELISPOT assay when PBMCs of three donors were stimulated (lower panel).

(C) CyTOF tSNE plot of T cells from all groups merged (left panel) and heatmap displaying normalized expression of select genes in indicated cell types (right panel).

(D) Trajectory reconstruction of all stimulated T cells reveals three branches: activated branch, memory branch, and exhausted branch (left panel). Histogram plot displaying the dynamic changes of T cell population after peptide stimulation (right panel). Ctr, control.

(E) The distribution of TCR clonotypes and the similarity of the TCR spectrum estimated using the Morisita-Horn index.

(F) Scatterplots comparing the T cell spectrum changes when donor 2 PBMC was stimulated by peptide 32 (left panel). The red and green dots represented the expanded and shrunk TCR clonotypes, respectively. The right panel displays the usage of V and J genes of all the expanded TCRs.

See also Figure S4 and Table S4.

co-cultured with the peptides 32 and 33, while donor 3 was reactive to none of these peptides (Figures 5B and S4E). Next, PBMCs from donor 1 and donor 2 were stimulated for 7 days using peptides 32 and 33, followed by cytometry by time of flight (CyTOF) and TCR sequencing. CyTOF data showed that the phenotypes of T cells evolved along with the peptide stimulation (Figure 5C). Pseudo-time analysis revealed that T cells progressed toward activation and bifurcated into memory and exhausted branches thereafter, representing two lineages at late maturation stages (Figure 5D). Meanwhile, an increase of a subset of effector memory T cells re-expressing CD45RA (termed TEMRA), especially effector T cell (T_{eff} , $CD45RA^{+}CCR7^{-}CD27^{-}CD28^{-}$) and pE2 ($CD45RA^{+}CCR7^{-}CD27^{+}CD28^{-}$), was seen, whereas naive T cells ($CD45RA^{+}CCR7^{+}$) and central

memory T_{CM} cells ($CD45RA^{-}CCR7^{+}$) were reduced (Figure 5D). Thus, T cell cultures with peptide stimulation were characterized by strong expansion of $CD8^{+}$ T cells with effector phenotype. Further TCR-seq revealed the distribution of donor 2's TCR repertoire underwent significant transformation after peptide 32 stimulation in donor 2, and multiple CDR3s with relatively low frequency were expanded, such as CASSYSFLGLGRAKNIQYF and CASSLAQGPQNIQYF (Figure 5F). Despite the inter-donor heterogeneity, the low frequency of neo-epitope T cell reactivity represented a valuable external source to overcome immunoediting and resistance for iCCA fusion patients.

Proteomic subgroups with distinct biological and clinical features

We identified 4 distinct proteomic subgroups (S1–S4) by clustering using the 1,376 most variable proteins (Figure 6A; Table S5), which had diverse clinical, genomic, immunologic, and microenvironmental features. S1 showed the most abundant expression of inflammatory proteins, such as CD14, MPO, and C5AR1. S2 had the highest level of proteins related to cancer-associated fibroblasts and ECM, including FAP, POSTN, and FLT1. S3 was characterized by elevated MAPK and metabolic proteins, such as ACAT1, FASN, and IDH1. S4 retained maximum expression of adhesion and biliary-specific proteins, such as ANXA4, KRT18, and EPCAM (Figures 6B and S5A).

Clinically, S1 mostly comprised tumors with increased CA19-9 level, tumor necrosis, and intrahepatic metastasis. S2 had higher lymph node metastasis. S3 was featured by patients with HBV infection. S4 was mainly enriched of patients with lower CA19-9 level and less metastasis. The four proteomic subgroups significantly differed in overall survival (OS), with a median OS of 17.2, 25.3, 36.7, and >60 months for S1–S4, respectively ($p = 2.0 \times 10^{-5}$, Figure 6C). Multivariable Cox analysis confirmed the proteomic classification as a significant prognosticator, independent of clinicopathologic features (hazard ratio = 0.786; 95% CI, 0.630, 0.981; $p = 3.3 \times 10^{-2}$; Table S5). Importantly, the proteomic subgroups could also stratify patient survival for those at early TNM stages ($p = 4.0 \times 10^{-3}$, Figure 6C) and we also found that the prognostic stratification of proteomics subgroups was independent of CA19-9 and CEA levels (Figure S5B). Besides, we observed a comparable overlap between our proteomics subgroup and previously described molecular subtypes of biliary cancer (Andersen et al., 2012; Job et al., 2019; Nakamura et al., 2015; Oishi et al., 2012) (Figure S5C).

Each subgroup showed distinct profiles of the recurrently altered genes (Figure 6D). *KRAS* mutations were significantly enriched in S1, together with upregulated inflammatory pathway. *TP53* mutations predominated in S3, accompanied by the enriched pathways of cell-cycle and MAPK signaling. *FGFR2* alterations, *BAP1* mutations, and *IDH1/2* mutations were most frequent in S4, partially coinciding with elevation of the bile secretion.

Based on xCell analysis (Aran, 2020), S1 was dominant by neutrophil, monocyte, macrophage, dendritic, Th2, and Treg infiltration, most of which may be involved in immunosuppression (Figure 6E). S2 showed the highest intensity of fibroblasts and endothelial cells, with increased stromal score and angiogenesis score. S4 had the most abundant basophils and CD8+ naive T cells. The immune checkpoints also showed a subgroup-specific pattern, with IDO, CEACAM1, CD47, VSIG4, and CD274 elevated in S1, CD276, CD40, and NT5E in S2, as well as VTCN1, HMGB1, and PVRL2 in S4, showing the potential for immunotherapy benefits by proteomic stratification. Meanwhile, some subgroup-specific targets were identified, such as *FGFR2* and *FXRD2* in S4 (Cotto et al., 2018) (Figure S5D).

Four subgroup-specific markers, MPO (neutrophil), POSTN (fibroblast), ALDOB (metabolism), and EPCAM (biliary), were selected to represent inflammatory (S1), mesenchymal (S2), metabolic (S3), and differentiated (S4), respectively (Figure 6F). In both FU-iCCA cohort ($n = 84$) and validation cohort ($n = 144$), expression of the 4 markers measured by multiplex immunostaining could stratify patient survival ($p = 2.3 \times 10^{-3}$ and $p =$

2.8×10^{-3} , respectively) (Figure 6G). Meanwhile, by clustering of mRNA and phosphoprotein, we also identified three transcriptomic subgroups and three phosphoproteomic subgroups with significant prognostic values, respectively (Figures S5E and S5F). The transcriptomic and phosphoproteomic classification showed partial overlaps and different molecular features with the proteomic clustering (Figure S5G).

Next, we performed proteogenomic analysis for 15 iCCA cell lines to match the proteomic stratification, finding that three cell lines were concordant with S1, 5 with S2, 4 with S3, and 3 with S4 (Tables S5I and S5J). High-throughput drug screening was conducted on 10/15 cell lines (Table S5K). We found that most of cell lines were generally resistant to first-line platinum-based drugs, while Gemcitabine and Paclitaxel were sensitive for most cell lines, except those in S1 and S3. Both topoisomerase I/II inhibitors and microtubule-associated inhibitors had spectral anti-iCCA activity for all subgroups. Besides, S1 and S2 were more sensitive to EGFR inhibitors than S3 and S4 (Figures S5H and S5I). Together, these data suggest that different subgroups have variable drug responses and subgroup-specific treatment may be needed.

Identification and validation of prognostic proteomic biomarkers

We performed supervised analysis to identify prognostic biomarkers, revealing 19 proteins with good prognosis and 15 proteins with poor prognosis by stringent filtering processes (STAR Methods; Figure 7A; Table S6). Among them, HKDC1 and SLC16A3 ranked as the most significant ones that positively ($p = 1.8 \times 10^{-8}$, log-rank test) or negatively ($p = 3.6 \times 10^{-9}$, log-rank test) correlated with patient survival in the FU-iCCA cohort, respectively (Figure 7B), independent of conventional clinicopathologic features (Table S6). Moreover, the expression of HKDC1 ($p = 2.9 \times 10^{-19}$) and SLC16A3 ($p = 1.3 \times 10^{-16}$, Kruskal-Wallis test) was significantly different across the four proteomic subgroups, with SLC16A3 upregulated in S1 and HKDC1 elevated in S4 (Figure 7C). Likewise, high expression of HKDC1 and SLC16A3 showed opposite associations with aggressive clinicopathologic features (Table S6).

Correlation analysis based on multi-omics data revealed that tumors with high HKDC1 were significantly associated with lower expression of factors in central carbon metabolism, IL-17 signaling, cytoskeleton regulation/ECM-interaction, and PI3K-AKT pathways, but higher expression of insulin/bile biosynthesis pathway (Figure S6A). On the other hand, tumors with high SLC16A3 had higher expression of proteins involved in active carbon metabolism, cytoskeleton regulation/ECM-interaction, and complement signaling pathways, but lower steroid hormone biosynthesis and drug metabolism pathways (Figure S6B). In another independent iCCA cohort (Figure 7D, $n = 222$), low expression of HKDC1 ($p = 2.3 \times 10^{-2}$) and high expression of SLC16A3 ($p = 4.4 \times 10^{-2}$) detected by immunostaining predicted a worse clinical outcome, respectively, confirming their prognostic value based on proteomics and potential clinical utility.

HKDC1 contains a hexokinase domain, and could catalyze the first step of the glycolytic metabolism by generating hexose 6-phosphate (Guo et al., 2015; Irwin and Tan, 2008). Overexpression of HKDC1 has been observed in many cancers, including

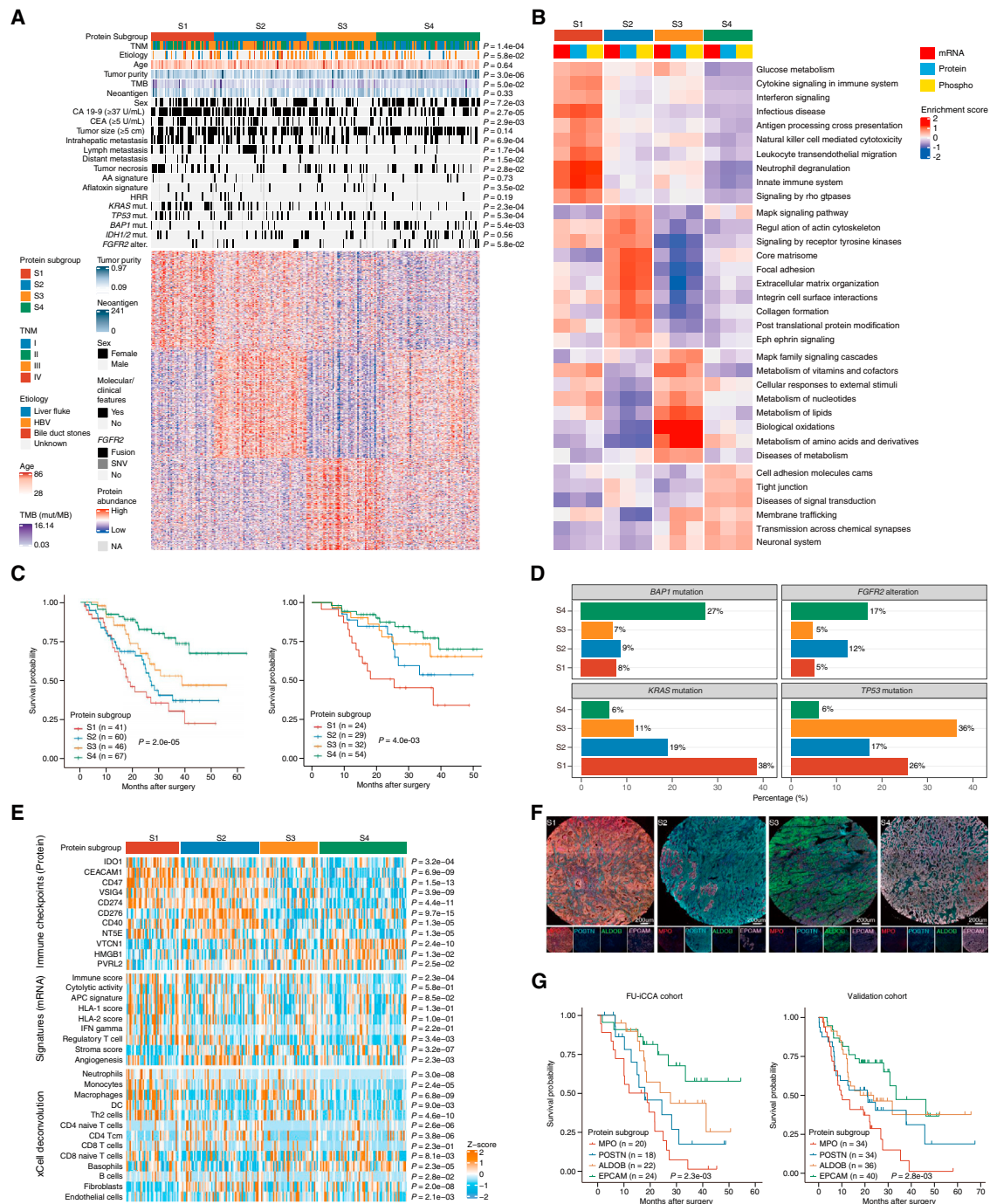


Figure 6. Proteomic stratification of FU-iCCA cohort and the corresponding molecular and pathway features

(A) Patient subgrouping based on the mostly varied proteins ($n = 1,376$). Samples and proteins are displayed as columns and rows, respectively, and the color of each cell shows the Z score of the relative abundance of the protein (logarithmic scale in base 2). The significance of clinical and molecular features across proteomic subgroups is calculated using Kruskal-Wallis rank-sum test and Fisher's exact test for continuous and categorical variables separately.

(B) Integrated analysis of altered pathways at transcriptome, proteome, and phosphoproteome levels among the four proteomic subgroups.

(C) Kaplan-Meier curves for overall survival based on the proteomic subgroups for the whole FU-ICCA cohort (left panel) and early-stage tumors (stage I/II, right panel) (log-rank test).

(D) The percentages of significant mutated genes in the proteomic subgroups.

(E) Heatmap of immune checkpoints, immune signatures, and immune cell abundance among proteomic subgroups (Kruskal-Wallis rank-sum test).

(F) Representative immunostaining images illustrating the molecular features of four proteomic subgroups.

(G) Kaplan-Meier curves for overall survival based on the immunostaining of four subgroup-specific biomarkers on our FU-iCCA cohort (left panel) and validation cohort (right panel).

See also [Figure S5](#) and [Table S5](#).

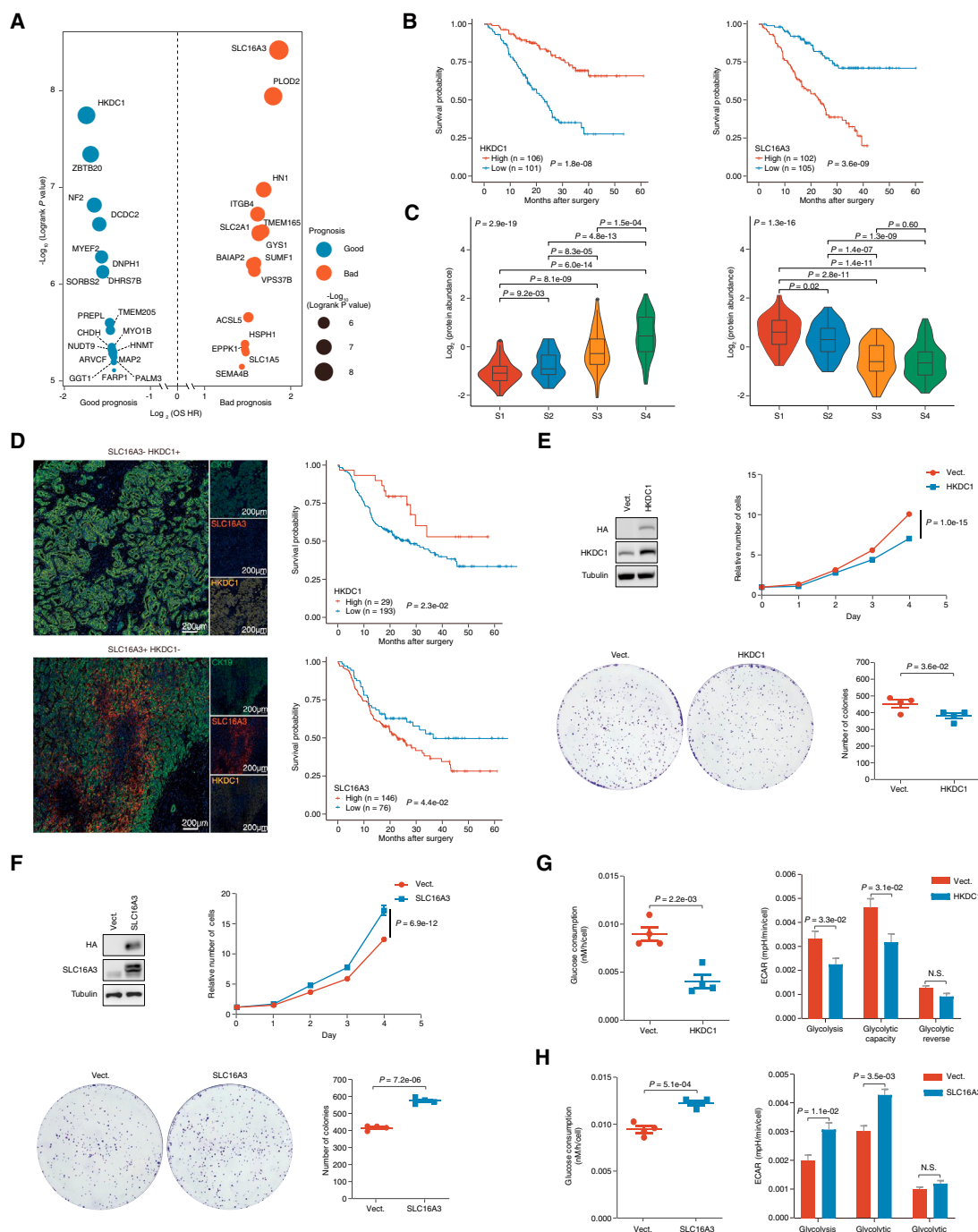


Figure 7. Identification and validation of proteomic prognostic biomarkers

(A) The dot plot showing the 34 candidate prognostic proteins. OS, overall survival; HR, hazard ratio.
 (B) Kaplan-Meier curves for overall survival based on proteomic abundance of HKDC1 and SLC16A3 (log-rank test).
 (C) Relative abundance of HKDC1 and SLC16A3 among proteomic subgroups (Kruskal-Wallis test). Boxes indicate interquartile range (IQR) with a line at the median. Whiskers represent the minimum and maximum values at most 1.5* IQR from the quartiles.
 (D) Representative immunostaining images showing tumors with high or low expression of HKDC1 and SLC16A3, and Kaplan-Meier curves for overall survival based on staining intensity of HKDC1 and SLC16A3 in an independent icCA cohort (n = 222, log-rank test).
 (E and F) The impacts of HKDC1 (E) or SLC16A3 (F) overexpression on cholangiocarcinoma cell growth. HuCCT1 cells ectopically expressing HA tag empty vector or HKDC1 (or SLC16A3) were generated and validated by western blot (upper left panel). CCK8 proliferation (upper right panel) and colony formation (lower panel) assays of indicated cells were measured. Data are represented as mean $\pm$ SEM (by two-way ANOVA and t test).
 (G and H) Glucose consumption and ECAR analyses of HuCCT1 cells overexpressing HKDC1 (G) or SLC16A3 (H) compared with control cells. Data are represented as mean $\pm$ SEM (t test).
 See also Figure S6 and Table S6.

hepatocellular carcinoma, and correlates with poor clinical outcome (Go et al., 2014; Zhang et al., 2016). However, its expression or function in iCCA has not been studied. Unlike previous reports, our survival and pathway association analyses suggested that HKDC1 may function as a tumor suppressor in iCCA. Indeed, overexpression of HKDC1 decreased cell proliferation, colony formation, glucose consumption, and glycolytic capacity of the HuCCT1 cells, supporting its anti-glycolysis/proliferation role in iCCA cells (Figures 7E, 7G, and S6C). SLC16A3 belongs to the proton-linked monocarboxylate transporter family, and mainly carries acidic metabolites, such as lactate and pyruvate, across the plasma membrane to mediate metabolic exchange between cells and the microenvironment (Contreras-Baeza et al., 2019). High expression of SLC16A3 has been observed as evidence of elevated glycolytic metabolism in various cancers (Baek et al., 2014; Kim et al., 2018). As expected, SLC16A3-overexpressing HuCCT1 cells exhibited faster proliferation and higher colony formation (Figure 7F), possibly due to the more potent glycolytic capacity (Figures 7H and S6D). Consistently, SLC16A3 knockdown significantly reduced HuCCT1 cell proliferation, colony formation, and glycolytic capacity (Figures S6E and S6F). Thus, SLC16A3 mainly functioned as an oncogenic factor in iCCA by promoting glycolysis metabolism. Moreover, SLC16A3-overexpressing or knockdown cells were more sensitive or resistant to a canonical MCT inhibitor, α -cyano-4-hydroxycinnamic acid, respectively (Figure S6G), indicating that MCTs inhibitors are promising therapeutic drugs for iCCA with high SLC16A3. Therefore, our results indicated that HKDC1 and SLC16A3, as robust prognostic biomarkers, functioned as a tumor suppressor and a strong oncogenic factor in iCCA, respectively.

Integration of microbiome analysis into proteogenomic profiling

To investigate the possible relationship between iCCA tissue bacteria and cancer progression, we conducted 16S rRNA sequencing to obtain the microbiota compositions in 80 pairs of iCCA randomly and equally selected from the 4 proteomic subgroups (Table S7). Several different methods (Kurilshikov et al., 2017) to measure the tumor microbial alpha diversity uniformly found a downward trend in alpha diversity from the highest in S1 to the lowest in S4 (Figure S7A). However, no significant differences were observed in microbial alpha diversity between tumors and adjacent liver tissues (Figure S7B). Further analysis revealed that the tumor microbial community presented phylogenetic proximity in each subgroup (Lozupone et al., 2011; Wong et al., 2016) (Figure S7C). Analyzing tumor microbial compositions at genus in each proteomic subgroup found different communities among them (Figure S7D). A total of 12 genera were identified as differentially abundant in the 4 subgroups (Figure S7E). The “transkingdom” correlation analysis showed strong positive correlations between OTUs and subgroup signature proteins in S1 and S3, but negative correlations in S2 and S4 (Figure S7F) (Liu et al., 2019). Metagenome function analysis showed enriched KEGG pathway modules from the four subgroups, such as biosynthesis and inflammation pathway (Lanigille et al., 2013) (Figures S7G and S7H). Thus, the four proteomic subgroups had distinct tumor microbiota diversity, composition, and functions, which may play important roles in immune and metabolic remodeling.

DISCUSSION

In this study, we report a large-scale proteogenomic characterization of 262 iCCAs, involving WES, RNA sequencing (RNA-seq), proteomics, and phosphoproteomics analyses. These data may provide a valuable resource for further discovery and precise oncological practice. Based on WES data, the aflatoxin signature was identified as the major mutational pattern potentially contributing to iCCA pathogenesis via impairing genomic stability, accelerating cell cycle, and dampening immune and inflammation responses. Consistent with previous studies (Lamarca et al., 2020), about 40% of iCCAs in this study harbored currently druggable mutations, amplifications or fusions, including *KRAS*, *IDH1/2*, *FGFR2*, *MET*, and others. Proteomic and phosphoproteomic data provide deeper insights into the downstream signaling consequences of the above genetic alterations to identify putative regulatory mechanisms and potential therapeutic targets. In particular, iCCA with *TP53* and *KRAS* co-mutation had extremely poor prognosis and increased metastasis, probably acting via the integrin-FAK/SRC pathway. *FGFR2* fusions and mutations may promote iCCA through activating Rho GTPase pathway rather than conventional MAPK pathway. These findings may open new therapeutic opportunities to target these specific genetic alterations. In addition, the immunogenic peptides derived from FGFR2:BICC1 fusion, which occurred as an early clonal event, may potentially serve as immunogenic targets.

Usually, iCCA is characterized by a desmoplastic microenvironment with a large number of cancer-associated fibroblasts and suppressive innate immune cells (Hogdall et al., 2018). Interestingly, our proteomic clustering identified an inflammatory subgroup and a mesenchymal subgroup, providing a better stratification of these microenvironmental alterations that may facilitate the design of subgroup-specific immunotherapy and ECM-oriented treatments. Both hepatocytes and cholangiocytes are considered as the potential origin of iCCA cells (Zhu and Kwong, 2020). The identified metabolic subgroup and the differentiated subgroup retained partial features of hepatocytes and cholangiocytes, respectively, possibly representing two opposite differentiation directions of iCCA. Moreover, the two subgroups had relatively better prognosis than the inflammatory and mesenchymal subgroups, possibly as they kept partial normal cell features and were therefore less aggressive. In addition to specific enrichment of molecular targets, the four subgroups showed different expression patterns of immune checkpoints, underscoring the need for designing personalized checkpoint inhibition and providing clues to explain the diverse response to anti-PD-1 monotherapy in advanced iCCA (Kim et al., 2020). In addition, microbiome analysis of iCCAs revealed that the tumor microbiota composition was distinct in the four proteomic subgroups, suggesting that the tumor bacteria might also contribute to the tumor immune and metabolic responses.

We identified HKDC1 and SLC16A3 as putative biomarkers that may predict iCCA clinical outcome. The tumor-suppressing role of HKDC1 in iCCA contrasted with previous reports in other cancer types where a tumor-promoting function was observed (Li and Huang, 2014; Zhang et al., 2016). Multi-omics data showed that HKDC1 expression negatively correlated with metabolic, IL-17 signaling, and ECM-interaction pathways.

Considering that inflammation and stromal responses significantly contributed to ICCA progression, HKDC1 may act to suppress ICCA by regulating those pathways. Such contrasting roles have also been reported in other molecules. For example, SHP2 suppresses liver cancer via downregulation of inflammatory signaling (Bard-Chapeau et al., 2011), but promotes tumorigenesis and progression in other malignancies. Similarly, the pro- and anti-oncogenic roles of HKDC1 could also be cell type specific and context dependent. As a major trans-membrane transporter of glycolytic products, SLC16A3 has been indicated in many cancer types (Baek et al., 2014; Kim et al., 2018). Notably, high SLC16A3 expression will more effectively transport lactate to the extracellular environment, change the acidic property and, therefore, affect the recruitment of certain immune cells, which may, in turn, promote tumor growth. Thus, the two biomarkers, although both are components of the glycolytic metabolism, indicate two opposite directions of prognosis and may guide distinct therapeutics for ICCA patients.

In summary, the integration of comprehensive genomic, transcriptomic, proteomic, phosphoproteomic, and microbiome data identified the underlying molecular mechanisms and disease subgroups that were not fully captured by genomic and transcriptomic analysis, and the information may open new paths for the development of personalized targeted therapy and immunotherapy to eventually benefit clinical practice. We hope that the observations and analysis described here, and the data underlying them, will be a rich resource for the in-depth study of the carcinogenesis, progression and therapeutics of ICCA.

STAR★METHODS

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

- **KEY RESOURCES TABLE**
- **RESOURCE AVAILABILITY**
 - Lead contact
 - Materials availability
 - Data and code availability
- **EXPERIMENTAL MODEL AND SUBJECT DETAILS**
 - Clinical specimens of FU-ICCA
 - Cell lines
- **METHOD DETAILS**
 - Specimen acquisition and preparation
 - DNA/RNA extraction, WES, and RNA-seq
 - TCR-sequencing
 - Proteomic sample preparation
 - Cell line sample preparation for TMTpro 16plex labeling
 - Nano-LC-MS/MS
 - Database searching of MS data
 - WES data analysis
 - HLA typing and neoantigen prediction
 - RNA-seq data analysis
 - Proteome and phosphoproteome data analysis
 - mRNA, proteomic and phosphoproteomic subgrouping analysis
 - Multi-omics data analysis

- Potential prognostic protein biomarker detection
- ICCA cell line subgrouping based on existed molecular model of FU-ICCA
- Tissue MicroArray (TMA) experiment
- Production and identification of eluted HLA-I peptides
- Peptide-MHC tetramer exchange assays
- Enzyme-linked immunospot (ELISPOT) assay
- Cytometry by time of flight (CyTOF)
- CyTOF data analysis
- Drug target analysis
- cDNA conversion and RT-PCR reaction for *FGFR2* fusion validation
- Functional experiments
- The workflow for microbiome analysis
- DNA extraction and bacterial 16S rRNA sequencing
- Microbiome pipeline analysis

● QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.ccell.2021.12.006>.

ACKNOWLEDGMENTS

This work was supported by the National Key Research and Development Program of China (grants 2020YFE0202200, 2020YFA0803203, 2017YFC1700200, 2020YFA0509000, and 2019YFA0802102), the National Natural Science Foundation of China (grants 82130077, 81961128025, 91859105, 81925029, 91853130, 81790253, and 82002514), the Strategic Priority Research Program of the Chinese Academy of Sciences (grant XDB19020203), Research Projects from the Science and Technology Commission of Shanghai Municipality (grants 21JC1410100, 20JC1418900, 19XD1420700, 20XD1424900, and 20YF1407400), the Shanghai Municipal Health Commission Collaborative Innovation Cluster Project (2019CXJQ02), Shanghai Municipal Science and Technology Major Project, CAS project for young scientists in basic research (YSBR-014). We thank Dr. Fuchu He (Beijing Institute of Lifeomics, Beijing) for expert guidance and insightful suggestions to our study. We also thank Mr. Bing Li, Guoqiang Wang, and Xuan Lin (Burning Rock Biotech) for the assistance of bioinformatics analysis. We thank Mr. Xu Ding (TinyGene Bio-Tech Co., Ltd., Shanghai) for the 16S rRNA sequencing. This work was done under the auspices of a Memorandum of Understanding between the Shanghai Institute of Materia Medica, Chinese Academy of Science, Fudan University, and the US National Cancer Institute's Office of Cancer Clinical Proteomics Research (Clinical Proteomic Tumor Analysis Consortium—CPTAC), as well as under the auspices of the US National Cancer Institute's International Cancer Proteogenome Consortium (ICPC). CPTAC collaborates with international organizations/institutions to accelerate the understanding of the molecular basis of cancer through the application of proteogenomics, standards development, and publicly available datasets.

AUTHOR CONTRIBUTIONS

Conceptualization, J.F., H. Zhou, D.G., Q.G., and H.R.; methodology, Q.G., L. Dong, D.L., R.C., Y.L., and Z.Z.; formal analysis, L. Dong, D.L., Q.G., Y.L., H. Zhu, N.S., S.W., S.C., P.C., G.S., D.R., X.Y., S.Z., S.Q., X.Z., Y.S., D.F., P.W., L. Ding, and B.Z.; investigation, L. Dong, D.L., R.C., Y.L., F.L., Y.Z., and Q.G.; resources, J.F., H. Zhou, D.G., Q.G., X.W., and J.Z.; data curation, Q.G., L. Dong, and D.L.; writing, Q.G., L. Dong, D.L., R.C., Z.Z., D.G., H. Zhou, and J.F.; visualization, L. Dong, D.L., Y.W., P.C., and Z.Z.; supervision, J.F., H. Zhou, D.G., Q.G., and H.R.; funding acquisition, J.F., H. Zhou, D.G., L. Dong, and Q.G.

DECLARATION OF INTERESTS

Z.Z., S.C., and P.C. are employees of Burning Rock Biotech. Other authors declare no potential conflict of interest.

Received: January 12, 2021
 Revised: July 19, 2021
 Accepted: December 8, 2021
 Published: December 30, 2021

REFERENCES

- Abou-Alfa, G.K., Macarulla, T., Javle, M.M., Kelley, R.K., Lubner, S.J., Adeva, J., Cleary, J.M., Catenacci, D.V., Borad, M.J., Bridgewater, J., et al. (2020a). Ivosidenib in IDH1-mutant, chemotherapy-refractory cholangiocarcinoma (ClarIDHy): a multicentre, randomised, double-blind, placebo-controlled, phase 3 study. *Lancet Oncol.* 21, 796–807.
- Abou-Alfa, G.K., Sahai, V., Hollebecque, A., Vaccaro, G., Melisi, D., Al-Rajabi, R., Paulson, A.S., Borad, M.J., Gallinson, D., Murphy, A.G., et al. (2020b). Pemigatinib for previously treated, locally advanced or metastatic cholangiocarcinoma: a multicentre, open-label, phase 2 study. *Lancet Oncol.* 21, 671–684.
- Aguirre, A.J., and Hahn, W.C. (2018). Synthetic lethal vulnerabilities in KRAS-mutant cancers. *Cold Spring Harb. Perspect. Med.* 8, a031518.
- Ahmed, Z., Lin, C.C., Suen, K.M., Melo, F.A., Levitt, J.A., Suhling, K., and Ladbury, J.E. (2013). Grb2 controls phosphorylation of FGFR2 by inhibiting receptor kinase and Shp2 phosphatase activity. *J. Cell Biol.* 200, 493–504.
- Allgauer, M., Budczies, J., Christopoulos, P., Endris, V., Lier, A., Rempel, E., Volckmar, A.L., Kirchner, M., von Winterfeld, M., Leichsenring, J., et al. (2018). Implementing tumor mutational burden (TMB) analysis in routine diagnostics—a primer for molecular pathologists and clinicians. *Transl. Lung Cancer Res.* 7, 703–715.
- Andersen, J.B., Spee, B., Blechacz, B.R., Avital, I., Komuta, M., Barbour, A., Conner, E.A., Gillen, M.C., Roskams, T., Roberts, L.R., et al. (2012). Genomic and genetic characterization of cholangiocarcinoma identifies therapeutic targets for tyrosine kinase inhibitors. *Gastroenterology* 142, 1021–1031.e15.
- Arai, Y., Totoki, Y., Hosoda, F., Shiota, T., Hama, N., Nakamura, H., Ojima, H., Furuta, K., Shimada, K., Okusaka, T., et al. (2014). Fibroblast growth factor receptor 2 tyrosine kinase fusions define a unique molecular subtype of cholangiocarcinoma. *Hepatology* 59, 1427–1434.
- Aran, D. (2020). Cell-type enrichment analysis of bulk transcriptomes using xCell. *Methods Mol. Biol.* 2120, 263–276.
- Aran, D., Hu, Z., and Butte, A.J. (2017). xCell: digitally portraying the tissue cellular heterogeneity landscape. *Genome Biol.* 18, 220.
- Aubrey, B.J., Strasser, A., and Kelly, G.L. (2016). Tumor-suppressor functions of the TP53 pathway. *Cold Spring Harb. Perspect. Med.* 6, a026062.
- Baek, G., Tse, Y.F., Hu, Z., Cox, D., Buboltz, N., McCue, P., Yeo, C.J., White, M.A., DeBerardinis, R.J., Knudsen, E.S., et al. (2014). MCT4 defines a glycolytic subtype of pancreatic cancer with poor prognosis and unique metabolic dependencies. *Cell Rep.* 9, 2233–2249.
- Banales, J.M., Marin, J.J.G., Lamarca, A., Rodrigues, P.M., Khan, S.A., Roberts, L.R., Cardinale, V., Carpino, G., Andersen, J.B., Braconi, C., et al. (2020). Cholangiocarcinoma 2020: the next horizon in mechanisms and management. *Nat. Rev. Gastroenterol. Hepatol.* 17, 557–588.
- Bard-Chapeau, E.A., Li, S., Ding, J., Zhang, S.S., Zhu, H.H., Princen, F., Fang, D.D., Han, T., Bailly-Maitre, B., Poli, V., et al. (2011). Ptpn11/Shp2 acts as a tumor suppressor in hepatocellular carcinogenesis. *Cancer Cell* 19, 629–639.
- Blokzijl, F., Janssen, R., van Boxtel, R., and Cuppen, E. (2018). MutationalPatterns: comprehensive genome-wide analysis of mutational processes. *Genome Med.* 10, 33.
- Boehm, J.S., Hession, M.T., Bulmer, S.E., and Hahn, W.C. (2005). Transformation of human and murine fibroblasts without viral oncoproteins. *Mol. Cell Biol.* 25, 6464–6474.
- Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120.
- Bolotin, D.A., Poslavsky, S., Mitrophanov, I., Shugay, M., Mamedov, I.Z., Putintseva, E.V., and Chudakov, D.M. (2015). MiXCR: software for comprehensive adaptive immunity profiling. *Nat. Methods* 12, 380–381.
- Brown, N.A., Rolland, D., McHugh, J.B., Weigelin, H.C., Zhao, L., Lim, M.S., Elenitoba-Johnson, K.S., and Betz, B.L. (2014). Activating FGFR2-RAS-BRAF mutations in ameloblastoma. *Clin. Cancer Res.* 20, 5517–5526.
- Bunis, D.G., Andrews, J., Fragiadakis, G.K., Burt, T.D., and Sirota, M. (2020). dittoSeq: universal user-friendly single-cell and bulk RNA sequencing visualization toolkit. *Bioinformatics* 36, 5535–5536.
- Butler, A., Hoffman, P., Smibert, P., Papalexi, E., and Satija, R. (2018). Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* 36, 411–420.
- Camp, R.L., Dolled-Filhart, M., and Rimm, D.L. (2004). X-tile: a new bioinformatics tool for biomarker assessment and outcome-based cut-point optimization. *Clin. Cancer Res.* 10, 7252–7259.
- Caporaso, J.G., Lauber, C.L., Walters, W.A., Berg-Lyons, D., Huntley, J., Fierer, N., Owens, S.M., Betley, J., Fraser, L., Bauer, M., et al. (2012). Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J.* 6, 1621–1624.
- Carbone, M., Harbour, J.W., Brugarolas, J., Bononi, A., Pagano, I., Dey, A., Krausz, T., Pass, H.I., Yang, H., and Gaudino, G. (2020). Biological mechanisms and clinical significance of BAP1 mutations in human cancer. *Cancer Discov.* 10, 1103–1120.
- Cerami, E., Gao, J., Dogrusoz, U., Gross, B.E., Sumer, S.O., Aksoy, B.A., Jacobsen, A., Byrne, C.J., Heuer, M.L., Larsson, E., et al. (2012). The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov.* 2, 401–404.
- Chen, Y.N., LaMarche, M.J., Chan, H.M., Fekkes, P., Garcia-Fortanet, J., Acker, M.G., Antonakos, B., Chen, C.H., Chen, Z., Cooke, V.G., et al. (2016). Allosteric inhibition of SHP2 phosphatase inhibits cancers driven by receptor tyrosine kinases. *Nature* 535, 148–152.
- Cingolani, P., Platts, A., Wang, L., Coon, M., Nguyen, T., Wang, L., Land, S.J., Lu, X., and Ruden, D.M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly (Austin)* 6, 80–92.
- Clark, D.J., Dhanasekaran, S.M., Petralia, F., Pan, J., Song, X., Hu, Y., da Veiga Leprevost, F., Reva, B., Lih, T.M., Chang, H.Y., et al. (2019). Integrated proteogenomic characterization of clear cell renal cell carcinoma. *Cell* 179, 964–983.
- Consortium, G.T. (2013). The genotype-tissue expression (GTEx) project. *Nat. Genet.* 45, 580–585.
- Contreras-Baeza, Y., Sandoval, P.Y., Alarcon, R., Galaz, A., Cortes-Molina, F., Alegria, K., Baeza-Lehnert, F., Arce-Molina, R., Guequen, A., Flores, C.A., et al. (2019). Monocarboxylate transporter 4 (MCT4) is a high affinity transporter capable of exporting lactate in high-lactate microenvironments. *J. Biol. Chem.* 294, 20135–20147.
- Cotto, K.C., Wagner, A.H., Feng, Y.Y., Kiwala, S., Coffman, A.C., Spies, G., Wollam, A., Spies, N.C., Griffith, O.L., and Griffith, M. (2018). DGIdb 3.0: a redesign and expansion of the drug-gene interaction database. *Nucleic Acids Res.* 46, D1068–D1073.
- Desgrosellier, J.S., and Cheresch, D.A. (2010). Integrins in cancer: biological implications and therapeutic opportunities. *Nat. Rev. Cancer* 10, 9–22.
- Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29, 15–21.
- Dong, L.Q., Peng, L.H., Ma, L.J., Liu, D.B., Zhang, S., Luo, S.Z., Rao, J.H., Zhu, H.W., Yang, S.X., Xi, S.J., et al. (2020). Heterogeneous immunogenomic features and distinct escape mechanisms in multifocal hepatocellular carcinoma. *J. Hepatol.* 72, 896–908.
- Dong, L.Q., Shi, Y., Ma, L.J., Yang, L.X., Wang, X.Y., Zhang, S., Wang, Z.C., Duan, M., Zhang, Z., Liu, L.Z., et al. (2018). Spatial and temporal clonal evolution of intrahepatic cholangiocarcinoma. *J. Hepatol.* 69, 89–98.
- Dou, Y., Kawaler, E.A., Cui Zhou, D., Gritsenko, M.A., Huang, C., Blumenberg, L., Karpova, A., Petyuk, V.A., Savage, S.R., Satpathy, S., et al. (2020). Proteogenomic characterization of endometrial carcinoma. *Cell* 180, 729–748.
- Farshidfar, F., Zheng, S., Gingras, M.C., Newton, Y., Shih, J., Robertson, A.G., Hinoue, T., Hoadley, K.A., Gibb, E.A., Roszik, J., et al. (2017). Integrative

genomic analysis of cholangiocarcinoma identifies distinct IDH-mutant molecular profiles. *Cell Rep.* 18, 2780–2794.

Gao, Q., Liang, W.W., Foltz, S.M., Mutharasu, G., Jayasinghe, R.G., Cao, S., Liao, W.W., Reynolds, S.M., Wyczalkowski, M.A., Yao, L., et al. (2018). Driver fusions and their implications in the development and treatment of human cancers. *Cell Rep.* 23, 227–238.e3.

Gao, Q., Zhu, H., Dong, L., Shi, W., Chen, R., Song, Z., Huang, C., Li, J., Dong, X., Zhou, Y., et al. (2019). Integrated proteogenomic characterization of HBV-related hepatocellular carcinoma. *Cell* 179, 561–577.e22.

Gillette, M.A., Satpathy, S., Cao, S., Dhanasekaran, S.M., Vasaikar, S.V., Krug, K., Petralia, F., Li, Y., Liang, W.W., Reva, B., et al. (2020). Proteogenomic characterization reveals therapeutic vulnerabilities in lung adenocarcinoma. *Cell* 182, 200–225.e35.

Go, A.S., Mozaffarian, D., Roger, V.L., Benjamin, E.J., Berry, J.D., Blaha, M.J., Dai, S., Ford, E.S., Fox, C.S., Franco, S., et al. (2014). Executive summary: heart disease and stroke statistics—2014 update: a report from the American Heart Association. *Circulation* 129, 399–410.

Gu, Z., Eils, R., and Schlesner, M. (2016). Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* 32, 2847–2849.

Guo, C., Ludvik, A.E., Arlotto, M.E., Hayes, M.G., Armstrong, L.L., Scholtens, D.M., Brown, C.D., Newgard, C.B., Becker, T.C., Layden, B.T., et al. (2015). Coordinated regulatory variation associated with gestational hyperglycaemia regulates expression of the novel hexokinase HKDC1. *Nat. Commun.* 6, 6069.

Guo, S.S., Wang, Y., and Fan, Q.X. (2019). Raddeanin A promotes apoptosis and ameliorates 5-fluorouracil resistance in cholangiocarcinoma cells. *World J. Gastroenterol.* 25, 3380–3391.

Haas, B.J., Dobin, A., Li, B., Stransky, N., Pochet, N., and Regev, A. (2019). Accuracy assessment of fusion transcript detection via read-mapping and de novo fusion transcript assembly-based methods. *Genome Biol.* 20, 213.

Hanzelmann, S., Castelo, R., and Guinney, J. (2013). GSEA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinformatics* 14, 7.

Hoadley, K.A., Yau, C., Hinoue, T., Wolf, D.M., Lazar, A.J., Drill, E., Shen, R., Taylor, A.M., Cherniack, A.D., Thorsson, V., et al. (2018). Cell-of-origin patterns dominate the molecular classification of 10,000 tumors from 33 types of cancer. *Cell* 173, 291–304.e6.

Hogdall, D., Lewinska, M., and Andersen, J.B. (2018). Desmoplastic tumor microenvironment and immunotherapy in cholangiocarcinoma. *Trends Cancer* 4, 239–255.

Hsu, I.C., Metcalf, R.A., Sun, T., Welsh, J.A., Wang, N.J., and Harris, C.C. (1991). Mutational hotspot in the p53 gene in human hepatocellular carcinomas. *Nature* 350, 427–428.

Hynes, R.O. (1992). Integrins: versatility, modulation, and signaling in cell adhesion. *Cell* 69, 11–25.

Irwin, D.M., and Tan, H. (2008). Molecular evolution of the vertebrate hexokinase gene family: identification of a conserved fifth vertebrate hexokinase gene. *Comp. Biochem. Physiol. Part D Genomics Proteomics* 3, 96–107.

Ishiwata, T., Matsuda, Y., Yamamoto, T., Uchida, E., Korc, M., and Naito, Z. (2012). Enhanced expression of fibroblast growth factor receptor 2 Ilc promotes human pancreatic cancer cell proliferation. *Am. J. Pathol.* 180, 1928–1941.

Jiang, Y., Sun, A., Zhao, Y., Ying, W., Sun, H., Yang, X., Xing, B., Sun, W., Ren, L., Hu, B., et al. (2019). Proteomics identifies new therapeutic targets of early-stage hepatocellular carcinoma. *Nature* 567, 257–261.

Jiao, Y., Pawlik, T.M., Anders, R.A., Selaru, F.M., Streppel, M.M., Lucas, D.J., Niknafs, N., Guthrie, V.B., Maitra, A., Argani, P., et al. (2013). Exome sequencing identifies frequent inactivating mutations in BAP1, ARID1A and PBRM1 in intrahepatic cholangiocarcinomas. *Nat. Genet.* 45, 1470–1473.

Job, S., Rapoud, D., Dos Santos, A., Gonzalez, P., Desterke, C., Pascal, G., Elarouci, N., Ayadi, M., Adam, R., Azoulay, D., et al. (2019). Identification of four immune subtypes characterized by distinct composition and functions of tumor microenvironment in intrahepatic cholangiocarcinoma. *Hepatology* 72, 965–981.

Jurtz, V., Paul, S., Andreatta, M., Marcatili, P., Peters, B., and Nielsen, M. (2017). NetMHCpan-4.0: improved peptide-MHC class I interaction predictions integrating eluted ligand and peptide binding affinity data. *J. Immunol.* 199, 3360–3368.

Jusakul, A., Cutcutache, I., Yong, C.H., Lim, J.Q., Huang, M.N., Padmanabhan, N., Nellore, V., Kongpetch, S., Ng, A.W.T., Ng, L.M., et al. (2017). Whole-genome and epigenomic landscapes of etiologically distinct subtypes of cholangiocarcinoma. *Cancer Discov.* 7, 1116–1135.

Kanehisa, M., and Goto, S. (2000). KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* 28, 27–30.

Kato, S., Goodman, A., Walavalkar, V., Barkauskas, D.A., Sharabi, A., and Kurzrock, R. (2017). Hyperprogressors after immunotherapy: analysis of genomic alterations associated with accelerated growth rate. *Clin. Cancer Res.* 23, 4242–4250.

Kim, H.K., Lee, I., Bang, H., Kim, H.C., Lee, W.Y., Yun, S.H., Lee, J., Lee, S.J., Park, Y.S., Kim, K.M., et al. (2018). MCT4 expression is a potential therapeutic target in colorectal cancer with peritoneal carcinomatosis. *Mol. Cancer Ther.* 17, 838–848.

Kim, R.D., Chung, V., Alese, O.B., El-Rayes, B.F., Li, D., Al-Toubah, T.E., Schell, M.J., Zhou, J.M., Mahipal, A., Kim, B.H., et al. (2020). A phase 2 multi-institutional study of nivolumab for patients with advanced refractory biliary tract cancer. *JAMA Oncol.* 6, 888–894.

Kuilman, T., Velds, A., Kemper, K., Ranzani, M., Bombardelli, L., Hoogstraat, M., Nevedomskaya, E., Xu, G., de Ruiter, J., Lolkema, M.P., et al. (2015). Copywriter: DNA copy number detection from off-target sequence data. *Genome Biol.* 16, 49.

Kurilshikov, A., Wijmenga, C., Fu, J., and Zhernakova, A. (2017). Host genetics and gut microbiome: challenges and perspectives. *Trends Immunol.* 38, 633–647.

Lamarca, A., Barriuso, J., McNamara, M.G., and Valle, J.W. (2020). Molecular targeted therapies: ready for "prime time" in biliary tract cancer. *J. Hepatol.* 73, 170–185.

Langille, M.G., Zaneveld, J., Caporaso, J.G., McDonald, D., Knights, D., Reyes, J.A., Clemente, J.C., Burkepille, D.E., Vega Thurber, R.L., Knight, R., et al. (2013). Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nat. Biotechnol.* 31, 814–821.

Laudato, S., Patil, N., Abba, M.L., Leupold, J.H., Benner, A., Gaiser, T., Marx, A., and Allgayer, H. (2017). P53-induced miR-30e-5p inhibits colorectal cancer invasion and metastasis by targeting ITGA6 and ITGB1. *Int. J. Cancer* 141, 1879–1890.

Lawrence, M.S., Stojanov, P., Polak, P., Kryukov, G.V., Cibulskis, K., Sivachenko, A., Carter, S.L., Stewart, C., Mermel, C.H., Roberts, S.A., et al. (2013). Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* 499, 214–218.

Lefranc, M.P., Giudicelli, V., Duroux, P., Jabado-Michaloud, J., Folch, G., Aouinti, S., Carillon, E., Duvergey, H., Houles, A., Paysan-Lafosse, T., et al. (2015). IMGT(R), the International Immunogenetics Information System(R) 25 years on. *Nucleic Acids Res.* 43, D413–D422.

Li, G.H., and Huang, J.F. (2014). Inferring therapeutic targets from heterogeneous data: HKDC1 is a novel potential therapeutic target for cancer. *Bioinformatics* 30, 748–752.

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754–1760.

Lin, C.C., Melo, F.A., Ghosh, R., Suen, K.M., Stagg, L.J., Kirkpatrick, J., Arold, S.T., Ahmed, Z., and Ladbury, J.E. (2012). Inhibition of basal FGF receptor signaling by dimeric Grb2. *Cell* 149, 1514–1524.

Liu, F., Liu, A., Lu, X., Zhang, Z., Xue, Y., Xu, J., Zeng, S., Xiong, Q., Tan, H., He, X., et al. (2019). Dysbiosis signatures of the microbial profile in tissue from bladder cancer. *Cancer Med.* 8, 6904–6914.

Li, C., Sun, Y.D., Yu, G.Y., Cui, J.R., Lou, Z., Zhang, H., Huang, Y., Bai, C.G., Deng, L.L., Liu, P., et al. (2020). Integrated Omics of Metastatic Colorectal Cancer (Cancer Cell).

Lowery, M.A., Ptashkin, R., Jordan, E., Berger, M.F., Zehir, A., Capanu, M., Kemeny, N.E., O'Reilly, E.M., El-Dika, I., Jarnagin, W.R., et al. (2018).

- Comprehensive molecular profiling of intrahepatic and extrahepatic cholangiocarcinomas: potential targets for intervention. *Clin. Cancer Res.* 24, 4154–4161.
- Lozupone, C., Lladser, M.E., Knights, D., Stombaugh, J., and Knight, R. (2011). UniFrac: an effective distance metric for microbial community comparison. *ISME J.* 5, 169–172.
- Magoc, T., and Salzberg, S.L. (2011). FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* 27, 2957–2963.
- Mermel, C.H., Schumacher, S.E., Hill, B., Meyerson, M.L., Beroukhi, R., and Getz, G. (2011). GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome Biol.* 12, R41.
- Nakamura, H., Arai, Y., Totoki, Y., Shirota, T., Elzawahry, A., Kato, M., Hama, N., Hosoda, F., Urushidate, T., Ohashi, S., et al. (2015). Genomic spectra of biliary tract cancer. *Nat. Genet.* 47, 1003–1010.
- Oishi, N., Kumar, M.R., Roessler, S., Ji, J., Forgues, M., Budhu, A., Zhao, X., Andersen, J.B., Ye, Q.H., Jia, H.L., et al. (2012). Transcriptomic profiling reveals hepatic stem-like gene signatures and interplay of miR-200c and epithelial-mesenchymal transition in intrahepatic cholangiocarcinoma. *Hepatology* 56, 1792–1803.
- Patro, R., Duggal, G., Love, M.I., Irizarry, R.A., and Kingsford, C. (2017). Salmon provides fast and bias-aware quantification of transcript expression. *Nat. Methods* 14, 417–419.
- Petitprez, F., de Reynies, A., Keung, E.Z., Chen, T.W., Sun, C.M., Calderaro, J., Jeng, Y.M., Hsiao, L.P., Lacroix, L., Bougouin, A., et al. (2020). B cells are associated with survival and immunotherapy response in sarcoma. *Nature* 577, 556–560.
- Pollock, P.M., Gartside, M.G., Dejeza, L.C., Powell, M.A., Mallon, M.A., Davies, H., Mohammadi, M., Futreal, P.A., Stratton, M.R., Trent, J.M., et al. (2007). Frequent activating FGFR2 mutations in endometrial carcinomas parallel germline mutations associated with craniosynostosis and skeletal dysplasia syndromes. *Oncogene* 26, 7158–7162.
- Qiu, X., Mao, Q., Tang, Y., Wang, L., Chawla, R., Pliner, H.A., and Trapnell, C. (2017). Reversed graph embedding resolves complex single-cell trajectories. *Nat. Methods* 14, 979–982.
- Quast, C., Priesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., and Glockner, F.O. (2013). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* 41, D590–D596.
- Ren, Y., Chen, Z., Chen, L., Fang, B., Win-Piazza, H., Haura, E., Koomen, J.M., and Wu, J. (2010). Critical role of Shp2 in tumor growth involving regulation of c-Myc. *Genes Cancer* 1, 994–1007.
- Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). Limma powers differential expression analyses for RNA-seq and microarray studies. *Nucleic Acids Res.* 43, e47.
- Rizvi, S., Khan, S.A., Hallemeier, C.L., Kelley, R.K., and Gores, G.J. (2018). Cholangiocarcinoma—evolving concepts and therapeutic strategies. *Nat. Rev. Clin. Oncol.* 15, 95–111.
- Rodriguez, H., and Pennington, S.R. (2018). Revolutionizing precision oncology through collaborative proteogenomics and data sharing. *Cell* 173, 535–539.
- Schlaepfer, D.D., Hauck, C.R., and Sieg, D.J. (1999). Signaling through focal adhesion kinase. *Prog. Biophys. Mol. Biol.* 71, 435–478.
- Shen, R., and Seshan, V.E. (2016). FACETS: allele-specific copy number and clonal heterogeneity analysis tool for high-throughput DNA sequencing. *Nucleic Acids Res.* 44, e131.
- Sia, D., Losic, B., Moeini, A., Cabellos, L., Hao, K., Revill, K., Bonal, D., Miltiadous, O., Zhang, Z., Hoshida, Y., et al. (2015). Massive parallel sequencing uncovers actionable FGFR2-PPH1N1 fusion and ARAF mutations in intrahepatic cholangiocarcinoma. *Nat. Commun.* 6, 6087.
- Sondka, Z., Bamford, S., Cole, C.G., Ward, S.A., Dunham, I., and Forbes, S.A. (2018). The COSMIC cancer gene census: describing genetic dysfunction across all human cancers. *Nat. Rev. Cancer* 18, 696–705.
- Stronen, E., Toebes, M., Kelderman, S., van Buuren, M.M., Yang, W., van Rooij, N., Donia, M., Bösch, M.L., Lund-Johansen, F., Olweus, J., and Schumacher, T.N. (2016). Targeting of cancer neoantigens with donor-derived T cell receptor repertoires. *Science* 352, 1337–1341.
- Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., et al. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U S A* 102, 15545–15550.
- Szolek, A., Schubert, B., Mohr, C., Sturm, M., Feldhahn, M., and Kohlbacher, O. (2014). OptiType: precision HLA typing from next-generation sequencing data. *Bioinformatics* 30, 3310–3316.
- Thakur, S.S., Geiger, T., Chatterjee, B., Bandilla, P., Frohlich, F., Cox, J., and Mann, M. (2011). Deep and highly sensitive proteome coverage by LC-MS/MS without prefractionation. *Mol. Cell Proteomics* 10, M110 003699.
- Tyanova, S., Temu, T., and Cox, J. (2016). The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. *Nat. Protoc.* 11, 2301–2319.
- Van der Auwera, G.A., Carneiro, M.O., Hartl, C., Poplin, R., Del Angel, G., Levy-Moonshine, A., Jordan, T., Shakir, K., Roazen, D., Thibault, J., et al. (2013). From fastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr. Protoc. Bioinformatics* 43, 11 10 11–11 10 33.
- van der Lee, D.I., Reijmers, R.M., Honders, M.W., Hagedoorn, R.S., de Jong, R.C., Kester, M.G., van der Steen, D.M., de Ru, A.H., Kweekel, C., Bijen, H.M., et al. (2019). Mutated nucleophosmin 1 as immunotherapy target in acute myeloid leukemia. *J. Clin. Invest.* 129, 774–785.
- Vasaikar, S., Huang, C., Wang, X., Petyuk, V.A., Savage, S.R., Wen, B., Dou, Y., Zhang, Y., Shi, Z., Arshad, O.A., et al. (2019). Proteogenomic Analysis of Human Colon Cancer Reveals New Therapeutic Opportunities. *Cell* 177, 1035–1049 e1019.
- Waitkus, M.S., DiPas, B.H., and Yan, H. (2018). Biological role and therapeutic potential of IDH mutations in cancer. *Cancer Cell* 34, 186–195.
- Wang, K., Li, M., and Hakonarson, H. (2010). ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* 38, e164.
- Wang, X., and Zhang, B. (2013). customProDB: an R package to generate customized protein databases from RNA-seq data for proteomics search. *Bioinformatics* 29, 3235–3237.
- Wang, Y., Zhou, X., Oberoi, K., Phelps, R., Couwenhoven, R., Sun, M., Rezza, A., Holmes, G., Percival, C.J., Friedenthal, J., et al. (2012). p38 inhibition ameliorates skin and skull abnormalities in Fgfr2 Beare-Stevenson mice. *J. Clin. Invest.* 122, 2153–2164.
- Wardell, C.P., Fujita, M., Yamada, T., Simbolo, M., Fassan, M., Karic, R., Polak, P., Kim, J., Hatanaka, Y., Maejima, K., et al. (2018). Genomic characterization of biliary tract cancers identifies driver genes and predisposing mutations. *J. Hepatol.* 68, 959–969.
- Wisniewski, J.R., Zougman, A., Nagaraj, N., and Mann, M. (2009). Universal sample preparation method for proteome analysis. *Nat. Methods* 6, 359–362.
- Wong, R.G., Wu, J.R., and Gloor, G.B. (2016). Expanding the UniFrac toolbox. *PLoS One* 11, e0161196.
- Xu, T., Le, T.D., Liu, L., Su, N., Wang, R., Sun, B., Colaprico, A., Bontempi, G., and Li, J. (2017). CancerSubtypes: an R/Bioconductor package for molecular cancer subtype identification, validation and visualization. *Bioinformatics* 33, 3131–3133.
- Yan, G., Fukabori, Y., McBride, G., Nikolaropoulos, S., and McKeehan, W.L. (1993). Exon switching and activation of stromal and embryonic fibroblast growth factor (FGF)—FGF receptor genes in prostate epithelial cells accompany stromal independence and malignancy. *Mol. Cell Biol.* 13, 4513–4522.
- Yang, W., Lee, K.W., Srivastava, R.M., Kuo, F., Krishna, C., Chowell, D., Makarov, V., Hoen, D., Dalin, M.G., Wexler, L., et al. (2019). Immunogenic neoantigens derived from gene fusions stimulate T cell responses. *Nat. Med.* 25, 767–775.

Yu, G., Wang, L.G., Han, Y., and He, Q.Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* 16, 284–287.

Zhang, K., Myllymaki, S.M., Gao, P., Devarajan, R., Kytola, V., Nykter, M., Wei, G.H., and Manninen, A. (2017). Oncogenic K-Ras upregulates ITGA6 expression via FOSL1 to induce anoikis resistance and synergizes with alphaV-Class integrins to promote EMT. *Oncogene* 36, 5681–5694.

Zhang, Z., Huang, S., Wang, H., Wu, J., Chen, D., Peng, B., and Zhou, Q. (2016). High expression of hexokinase domain containing 1 is associated with poor prognosis and aggressive phenotype in hepatocarcinoma. *Biochem. Biophys. Res. Commun.* 474, 673–679.

Zhou, J., Sun, H.C., Wang, Z., Cong, W.M., Wang, J.H., Zeng, M.S., Yang, J.M., Bie, P., Liu, L.X., Wen, T.F., et al. (2018). Guidelines for diagnosis and treatment of primary liver cancer in China (2017 Edition). *Liver Cancer* 7, 235–260.

Zhu, Y., and Kwong, L.N. (2020). Insights into the origin of intrahepatic cholangiocarcinoma from mouse models. *Hepatology* 72, 305–314.

Zou, S., Li, J., Zhou, H., Frech, C., Jiang, X., Chu, J.S., Zhao, X., Li, Y., Li, Q., Wang, H., et al. (2014). Mutational landscape of intrahepatic cholangiocarcinoma. *Nat. Commun.* 5, 5696.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Rabbit polyclonal anti-MPO	DAKO	Cat# A0398, RRID:AB_2335676
Rabbit monoclonal anti-Periostin	Abcam	Cat# ab215199
Rabbit Polyclonal anti-ALDOB	Protein tech	Cat# 18065-1-AP; RRID: AB_2273968
Rabbit monoclonal anti-EPCAM	Abcam	Cat# ab223582; RRID: AB_2762366
Rabbit monoclonal anti-Cytokeratin 19	Abcam	Cat# ab52625; RRID: AB_2281020
Mouse monoclonal anti-SLC16A3	Santa Cruz Biotechnology	Cat# sc-376465; RRID: AB_11150118
Rabbit Polyclonal anti-HKDC1	Abcam	Cat# ab228729
Mouse monoclonal anti-HA-probe (F-7)	Santa Cruz Biotechnology	Cat# sc-7392; RRID: AB_627809
Rabbit polyclonal anti-SLC16A3	ABclonal	Cat# A10548; RRID: AB_2758089
Rabbit monoclonal anti-TOP2A	ABclonal	Cat# A4389, RRID: AB_2863260
Rabbit polyclonal anti-CDK1	ABclonal	Cat# A0220, RRID: AB_2757034
Rabbit polyclonal anti-PCNA	ABclonal	Cat# A0264, RRID: AB_2757077
Rabbit polyclonal anti-CD41/ITGA2B	ABclonal	Cat# A5680, RRID: AB_2766440
Rabbit polyclonal anti-ITGB6	ABclonal	Cat# A16904, RRID: AB_2770011
Rabbit polyclonal anti-RIPK3	ABclonal	Cat# A5431, RRID: AB_2766237
Rabbit polyclonal anti-FGFR2	ABclonal	Cat# A2074, RRID: AB_2764096
Mouse monoclonal anti-alpha1a Tubulin (7-RY28)	Santa Cruz Biotechnology	Cat# sc-134237; RRID: AB_2212295
Polyclonal swine anti-rabbit immunoglobulins/HRP antibody	Agilent	Cat# P0217; RRID: AB_2728719
Polyclonal rabbit anti-mouse immunoglobulins/HRP antibody	Agilent	Cat# P0260; RRID: AB_2636929
Chemicals, peptides, and recombinant proteins		
Maxpar direct immune profiling assay	Fluidigm	Cat# 201325
α - cyano - 4 - hydroxycinnamic acid (α - CHCA)	Selleck	Cat# S8612
Ro-3306	Selleck	Cat# S7747
GSK'872 (GSK2399872A)	Selleck	Cat# S8465
Lapatinib	Selleck	Cat# S2111
Sorafenib tosylate	Selleck	Cat# S1040
Pazopanib HCl	Selleck	Cat# S1035
Etoposide	Selleck	Cat# S1225
Oxaliplatin	Selleck	Cat# S1224
Chloroquine diphosphate	Selleck	Cat# S4157
Gemcitabine	Selleck	Cat# S1224
Erlotinib	Selleck	Cat# S7786
Sunitinib Malate	Selleck	Cat# S1042
Paclitaxel	Selleck	Cat# S1150
Gimeracil	Selleck	Cat# S2055
Carboplatin	Selleck	Cat# S1215
Topotecan HCl	Selleck	Cat# S1231
Gefitinib	Selleck	Cat# S1025
Olaparib	Selleck	Cat# S1060
Mylosar	Selleck	Cat# S1782

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Adriamycin HCL	Selleck	Cat# S1208
Dacogen	Selleck	Cat# S1200
Teniposide	Selleck	Cat# S1787
Nilotinib	Selleck	Cat# S1033
Rapamycin	Selleck	Cat# S1039
Cisplatin	Selleck	Cat# S1166
Imatinib Mesylate	Selleck	Cat# S1026
Tandutinib	Selleck	Cat# S1043
Lenvatinib	Selleck	Cat# S1164
XL-184	Selleck	Cat# S4001
Everolimus	Selleck	Cat# S1120
Abiraterone	Selleck	Cat# S1123
Regorafenib	Selleck	Cat# S1178
Vemurafenib	Selleck	Cat# S1267
Carmofur	Selleck	Cat# S1289
Camptothecin	Selleck	Cat# S1288
MMAF	Selleck	Cat# S0328
MMAE	Selleck	Cat# S7721
SN-38	Selleck	Cat# S4908
FGFR2 fusion peptides	China Peptides Inc.	This paper

Biological samples

Paired tumor, adjacent non-tumor liver tissues from a cohort of 262 iCCA patients	Zhongshan Hospital, Fudan University	This paper
3 healthy donors' PBMC	Zhongshan Hospital, Fudan University	This paper

Critical commercial assays

RNAlater reagent	Invitrogen	Cat# AM7020
QIAamp Fast DNA tissue kit	QIAGEN	Cat# 51404
QuarPrep EZ DNA library kit	DynastyGene Biotechnologies	Cat#L1002A
QuarHyb reagent kit	DynastyGene Biotechnologies	Cat#C1001A
QuarPrep RNA library kit	DynastyGene Biotechnologies	Cat#L1004A
KC-digitalTM stranded TCR-seq library prep kit	Seqhealth Technology Co., Ltd	Cat. No. DT0813-02
FISH-tissue implementation kit	ZytoLight	Cat#Z-2028-20
SPEC <i>FGFR2</i> dual-color break-apart probe	ZytoLight	Cat#Z-2169-200
TMT10plex isobaric label reagent	Thermo Scientific	Cat# 90111
TMT11-131C label reagent	Thermo Scientific	Cat# A34807
IFN- γ ELISPOT kit	Abcam	Cat# ab62899
High-select Fe-NTA kit	Thermo Scientific	Cat# A32992
Fusion-derived peptides	China Peptides Inc.	This paper
HBV peptide pool	JPT.Inc	Cat# PM-HBV-CPULTRA
QuickSwitch quant HLA-A02:01 tetramers	MBL International	Cat# TB-7300-K1
QuickSwitch quant HLA-A24:02 tetramers	MBL International	Cat# TB-7302-K1
CellTiter-Glo™ luminescent cell viability assay kit	Promega	Cat# G7573

Deposited data

Proteogenomic data of the iCCA cohort (n = 262)	This paper	[NODE database]: OEP001105. https://www.biosino.org/node/project/detail/OEP001105
DGIdb (version 4.0.0)	(Cotto et al., 2018)	https://www.dgldb.org

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
GRCh37 genome assembly	N/A	https://www.ncbi.nlm.nih.gov/assembly/GCF_000001405.13/
SnEff build-in hg19 reference	(Cingolani et al., 2012)	https://pcingola.github.io/SnpEff/se_introduction/#databases
ANNOVAR filter-based annotation	(Wang et al., 2010)	https://annovar.openbioinformatics.org/en/latest/user-guide/download/
cBioportal	(Cerami et al., 2012)	https://www.cbioportal.org
GRCh37 gencode v19 CTAT lib (version Oct012019)		https://data.broadinstitute.org/Trinity/CTAT_RESOURCE_LIB/_genome_libs_StarFv1.8/
BSgenome.Hsapiens.UCSC.hg19	Hervé Pagès	https://bioconductor.org/packages/BSgenome.Hsapiens.UCSC.hg19/

Experimental models: Cell line

HuCTT1, male origin	Dr. Damin Gao (Shanghai Institute of Biochemistry and Cell Biology)	RRID:CVCL_0324
RBE	Dr. Damin Gao (Shanghai Institute of Biochemistry and Cell Biology)	Cat# RCB1292; RRID:CVCL_4896
HCCC-9810	Dr. Qiang Gao (Zhangshan hospital)	RRID:CVCL_6908
LIPF 178C	Eastern Hepatobiliary Surgery Institute, China	(Guo et al., 2019)
LIPF 155C	Eastern Hepatobiliary Surgery Institute, China	(Guo et al., 2019)
LICCF	Eastern Hepatobiliary Surgery Institute, China	(Guo et al., 2019)
ICC772	Dr. Qiang Gao (Zhangshan hospital)	(Dong et al., 2018)
ICC1370	Dr. Qiang Gao (Zhangshan hospital)	(Dong et al., 2018)
ICC1969	Dr. Qiang Gao (Zhangshan hospital)	(Dong et al., 2018)
ICC1239	Dr. Qiang Gao (Zhangshan hospital)	(Dong et al., 2018)
ICC880	Dr. Qiang Gao (Zhangshan hospital)	(Dong et al., 2018)
ICC892	Dr. Qiang Gao (Zhangshan hospital)	(Dong et al., 2018)
ICC2935	Dr. Qiang Gao (Zhangshan hospital)	This paper
ICC4715	Dr. Qiang Gao (Zhangshan hospital)	This paper
ICC5784	Dr. Qiang Gao (Zhangshan hospital)	This paper

Recombinant DNA

pLEX-MCS-CMV-puro	Thermo Scientific Open Biosystems	Cat# OHS4735
-------------------	-----------------------------------	--------------

Software and algorithms

MaxQuant 1.6.7.0	(Tyanova et al., 2016)	http://www.coxdocs.org/doku.php?id=maxquant:start
DreasAI R package (version 0.1.0)	Pei Wang Lab	https://github.com/WangLab-MSSM/DreamAI
ComplexHeatmap R package (version 2.2.0)	(Gu et al., 2016)	http://www.bioconductor.org/packages/release/bioc/html/ComplexHeatmap.html
GSEA software (version 4.0.3)	(Subramanian et al., 2005)	https://www.gsea-msigdb.org/gsea/index.jsp
Limma R package (version 3.42.2)	(Ritchie et al., 2015)	https://bioconductor.org/packages/release/bioc/html/limma.html
ClusterProfiler R package (version 3.14.3)	(Yu et al., 2012)	https://bioconductor.org/packages/release/bioc/html/clusterProfiler.html
CancerSubtypes R package (version 1.12.1)	(Xu et al., 2017)	https://www.bioconductor.org/packages/release/bioc/html/CancerSubtypes.html
GSVA R package (version 1.34.0)	(Hanzelmann et al., 2013)	http://www.bioconductor.org/packages/release/bioc/html/GSVA.html

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
multiOmicsViz (version 3.12)	Jing Wang	http://www.bioconductor.org/packages/release/bioc/html/multiOmicsViz.html
facets R package (version 0.6.1)	(Shen and Seshan, 2016)	https://github.com/mskcc/facets
facetsSuite R package (version 2.0.6)	Barry Taylor Lab	https://github.com/mskcc/facets-suite
GATK (version 4.1.7.0)	(Van der Auwera et al., 2013)	https://gatk.broadinstitute.org/hc/en-us
BWA (version 0.7.17)	(Li and Durbin, 2009)	http://bio-bwa.sourceforge.net/
SnEff (version 4.3T)	(Cingolani et al., 2012)	https://pcingola.github.io/SnpEff/
ANNOVAR (version 2019-10-24)	(Wang et al., 2010)	https://doc-openbio.readthedocs.io/
Optitype (version 1.3.3)	(Szolek et al., 2014)	https://github.com/FRED-2/OptiType
netMHCpan (version 4.0)	(Jurtz et al., 2017)	http://www.cbs.dtu.dk/services/NetMHCpan-4.0/
xCell R package (version 1.1.0)	(Aran et al., 2017)	https://xcell.ucsf.edu/
MutSigCV (version 1.3.01)	(Lawrence et al., 2013)	https://www.genepattern.org/modules/docs/MutSigCV
CopywriteR (version 3.12)	(Kuilmann et al., 2015)	https://www.bioconductor.org/packages/release/bioc/html/CopywriteR.html
MutationalPatterns (version 2.0.0)	(Blokzijl et al., 2018)	https://bioconductor.org/packages/release/bioc/html/MutationalPatterns.html
GISTIC2 (version 2.0.3)	(Mermel et al., 2011)	https://genomebiology.biomedcentral.com/articles/10.1186/gb-2011-12-4-r41
Trimmomatic (version 0.36)	(Bolger et al., 2014)	http://www.usadellab.org/cms/?page=trimmomatic
Salmon (version 1.1.0)	(Patro et al., 2017)	https://combine-lab.github.io/salmon
STAR (version 2.7.3a)	(Dobin et al., 2013)	https://github.com/alexdobin/STAR
X-tile (version 3.6.1)	(Camp et al., 2004)	https://medicine.yale.edu/lab/rimm/research/software/
STAR-Fusion (version 1.8.1)	(Haas et al., 2019)	https://github.com/STAR-Fusion/STAR-Fusion
Seurat R package (version 3.6)	(Butler et al., 2018)	https://satijalab.org/seurat/
dittoSeq R package (version 1.1.10)	(Bunis et al., 2020)	https://bioconductor.org/packages/release/bioc/html/dittoSeq.html
Monocle R package (version 2.14.0)	(Qiu et al., 2017)	http://cole-trapnell-lab.github.io/monocle-release/
MiXCR (version 3.0.3)	(Bolotin et al., 2015)	https://github.com/milaboratory/mixcr
Others		
TMA of an independent cohort of 222 iCCA patients	Zhongshan Hospital, Fudan University	This paper

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead author, Jia Fan (fan.jia@zs-hospital.sh.cn).

Materials availability

This study did not generate new unique reagents.

Data and code availability

The data of WES, RNA-seq, proteome, and phosphoproteome generated in this study can be viewed in biosino NODE database (NODE database: OEP001105, <https://www.biosino.org/node/project/detail/OEP001105>). Sample annotation, processed, and normalized data files are provided as Table S1.

The partial WES data of PDPCs from the previous cohort (Dong et al., 2018) can be accessed via the European Bioinformatics Institute (EBI EGA: EGAS00001002625). DGIdb (version 4.0.0) (Cotto et al., 2018) can be accessed via their portal at: <https://www.dgldb.org/>. GRCh37 genome assembly can be accessed via Genome Reference Consortium at https://www.ncbi.nlm.nih.gov/assembly/GCF_000001405.13. GRCh37 gencode v19 CTAT lib (version Oct012019) can be accessed via Broad institute website at https://data.broadinstitute.org/Trinity/CTAT_RESOURCE_LIB/_genome_libs_StarFv1.8/. SnpEff build-in hg19 reference (Cingolani et al., 2012) can be accessed via Pablo Cingolani website at https://pcingola.github.io/SnpEff/se_introduction/#databases. ANNOVAR filter-based annotation (Wang et al., 2010) can be found the ANNOVAR portal at <https://annovar.openbioinformatics.org/en/latest/user-guide/download/>. BSgenome.Hsapiens.UCSC.hg19 can be accessed via Bioconductor at <https://bioconductor.org/packages/BSgenome.Hsapiens.UCSC.hg19/>. Software and code used in this study are referenced in their corresponding STAR Methods sections and also the [key resources table](#).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Clinical specimens of FU-iCCA

The iCCA samples used for this study were collected from Zhongshan Hospital of Fudan University in China. Patients treated from December 2014 to October 2018 were randomly selected from our prospectively collected tissue bank. They all received primary resection for iCCA without any anti-cancer treatments prior to surgery. Tissues were obtained and stored in liquid nitrogen less than 30 min after resection. Paired adjacent non-tumor liver tissues were taken from the most distant site relative to the tumor border (usually >2 cm apart from tumor edge). Postoperative surveillance and treatment were conducted according to our consensus guideline as described previously (Zhou et al., 2018).

The clinical information of these enrolled patients was collected, including sex, age, CA19-9, etiology, date of surgical resection, tumor size, intrahepatic metastasis, lymph node metastasis, perineural invasion, vascular invasion, and TNM staging (AJCC cancer staging system 8th edition). The median age of these patients was 61, with 147 males and 115 females. In total, 83, 81 and 98 patients were classified as TNM stages I, II and III-IVA respectively. Detailed clinicopathologic features are summarized in [Table S1](#). Overall survival (OS) was defined as the interval between surgery and death. The study was approved by the Research Ethics Committee of Zhongshan Hospital, and written informed consent was obtained from each patient.

Cell lines

HuCC1 and RBE cells were kept by Dr. Daming Gao's lab and HCCC-9810 was kept by Qiang Gao's lab, respectively. LIPF 178C, LIPF 155C, and LICCF were gifts from Eastern Hepatobiliary Surgery Institute, China (Guo et al., 2019). In addition, we have also generated 9 in-house patient-derived primary cancer cell lines (PDPCs) of iCCA at low passage values (<20), 6 of which were randomly selected from multi-region sampling derived primary cell lines of previous study (Dong et al., 2018). The clinical information of 9 PDPCs including ICC772, ICC1370, ICC1969, ICC1239, ICC880, ICC892, ICC2935, ICC4715, ICC5784, is listed in [Table S51](#). All cell lines were cultured in RPMI.1640 medium supplemented with 10% FBS, 100 units of penicillin and 100 mg/mL streptomycin.

METHOD DETAILS

Specimen acquisition and preparation

To reduce the impact of intra-tumor heterogeneity on multi-omics analysis, the iCCA tumor tissue was pulverized using the CryoPrepTM CP02 (Covaris) and then divided into two parts: the first part was snap-frozen in liquid nitrogen and then stored in a –80°C refrigerator until being used for proteomics and phosphoproteomics as well as DNA extraction; the second part was processed with RNeasy Lysis Solution (Qiagen) and stored at –80°C until being used for the RNA extraction. Additional tissue samples were treated with formalin for HE staining and histological evaluation. Each sample was assigned a new research ID, and the patient's name or medical record number used during hospitalization was de-identified. Because of the difficulty of obtaining the paired intrahepatic biliary tract tissues, transcriptomic and proteomic analyses were only conducted on tumor tissues. WES analysis was applied on both tumor and paired adjacent non-tumor liver tissues. According to CPTAC clinical sample collection procedures, the following criteria were used for the 262 paired samples: 1) successful extraction of DNA from both tumor and adjacent non-tumor liver tissues for WES (253 pairs); 2) qualified RNA from tumor tissues for RNA-seq (255 samples); 3) qualified protein extraction from tumor tissues as revealed from the SDS-PAGE image without obvious protein degradation and aberrant pattern (214 samples for proteomic and phosphoproteomic analysis); 4) no tumor cells were observed in the adjacent non-tumor liver tissues. Finally, there were 204 high-quality paired samples with all four omics data, and the average tumor cellularity of these samples was 44% ([Table S1](#)).

DNA/RNA extraction, WES, and RNA-seq

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 QuirePrep EZ DNA Library Kit and QuireHyb 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.

Total RNA of tissues and PBMCs were extracted and purified from fresh frozen tissues using the Whole RNA extraction kit (TIANGEN). 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 QuirePrep 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 (Table S1), covering the majority of the genes in proteomics.

TCR-sequencing

About 2 µg total RNA of each sample were used for TCR sequencing library preparation using KC-Digital™ Stranded TCR-seq Library Prep Kit for Illumina® 150 (Seqhealth Technology Co., Ltd., Wuhan, China, Cat. No. DT0813-02) following the manufacturer's instruction. The library products corresponding to 250–500 bp were enriched, quantified and finally sequenced on NovaSeq (Illumina®).

Raw sequencing data was first filtered by SOAPnuke (version 1.6.0), followed by in-house scripts to eliminate duplication bias introduced in library preparation and sequencing. UMI was utilized to eliminate errors and biases introduced in sequencing process. Then The de-duplicated consensus sequences were mapped to the international ImMunoGeneTics (IMGT) database by Lefranc et al (Lefranc et al., 2015), using MiXCR software (version 3.0.3) to obtain V, D and J fragment, rearrangement and CDR3 sequences.

Proteomic sample preparation

Protein extraction and digestion

Samples for MS were prospectively collected as described above. 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, Ningbio 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 were digested by filter-aided sample preparation (FASP) procedure (Wisniewski et al., 2009) with 10 kDa centrifugal filter tubes (Millipore) and centrifuged at 12,000 g at 22°C. First, the protein was washed with 200 µL UA buffer (8 M urea in 0.1 M Tris-HCl, pH 8.5) three times and alkylated with 50 mM iodoacetamide for 30 min keep in dark place. Then, the iodoacetamide was removed by centrifugation. Continually, the protein was washed with 200 µL UA buffer three times and finally washed thrice with 200 µL 50 mM triethyl ammonium bicarbonate (TEAB). Trypsin (Promega) in 50 mM TEAB at a concentration of 1:50 (w/w) to digested proteins at 37°C for 18 hr. Finally, digested peptides were collected at 12,000 g centrifugation. 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.

TMT labeling of peptides

Each channel in TMT labeling experiment labeled 400 µg peptides. Ten channels (126, 127N, 127C, 128N, 128C, 129N, 129C, 130N, 130C, and 131N) labeled peptides for ten tumor tissue samples and the channel 131C labeled the peptides for the internal reference 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 mM TEAB) 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). The 214 ICCA tumor tissue samples were labeled in 22 sets of TMT 11-plex experiments for the following MS analysis.

Peptides fractionation

Peptides were fractionated through the high-pH reverse phase liquid chromatography to reduce the likelihood of peptides co-isolating and co-fragmenting in these highly complex samples. About 4 mg TMT labeled peptides were fractionated using a Waters XBridge BEH300 C18 column (250 x 4.6 mm, OD 5 mm) at a flow rate of 0.7 mL/min on Agilent 1100 LC instrument. Solvent A (10 mM NH₄COOH, adjusted to pH 10.0 with NH₃·H₂O) and a nonlinear increasing concentration of solvent B (90% ACN, 10 mM NH₄COOH, adjusted to pH 10.0 with NH₃·H₂O) were used for peptide separation. A 110-min gradient was set as 1%–5% B in 2 min; 5%–25% B in 35 min; 25%–40% B in 43 min; 40%–55% B in 6 min; 55%–95% B in 3 min; 95% B for 4 min; 95%–1% B in 1 min; 1% B for 16 min 96 fractions was collected every 1 min from 3 min to 99 min. Then the 96 fractions were combined into 24 fractions (1&25&49&73, 2&26&50&74...24&48&72&96). 5% of the peptides were separated from the 24 fractions and dried by vacuum freeze-drying for proteome analysis. The remaining part were prepared for phosphopeptide enrichment and phosphoproteome analysis.

Phosphopeptide enrichment

High-Select Fe-NTA kit (Thermo Scientific, A32992) was used for phosphopeptide enrichment according to the kit instructions. Briefly, 200 µL loading buffer (80% ACN, 0.1% TFA) was added into the 24 fractions and 1.2 mL loading buffer (80% ACN, 0.1% TFA) was added into the one spin column in the kit. 50 µL resin mixture was added into peptides and incubated for 15 min at room temperature, then the mixture was transferred into the filter tip (Axygen, TF-20-L-R-S). The mixture was washed with 200 µL washing buffer (80% ACN, 0.1% TFA) three times and 200 µL H₂O three times to remove nonspecifically adsorbed peptides.

100 μ L elution buffer (50% ACN, 5% $\text{NH}_3 \cdot \text{H}_2\text{O}$) was used to elute the phosphopeptides twice. All experiment was centrifuged at 50 g at 22°C and the eluates were dried by vacuum freeze-drying for proteome analysis.

Benchmark sample preparation

The longitudinal quality control of mass spectrometry performance was monitored by the test of benchmark samples. The benchmark samples protein was extracted and digested from five pairs of hepatobiliary carcinoma tissue and their adjacent non-tumor liver samples, and the peptides were labeled with TMT reagent (channels 126, 127C, 128C, 129C and 130C labeled the five tumor tissues sample, and the channels 127N, 128N, 129N, 130N and 131 labeled the five adjacent non-tumor liver tissues) and fractionation (24 fractions) through the high-pH reverse phase liquid chromatography. The samples (1 μ g each fraction) and the diluted samples (100 ng each fraction) were analyzed before every four sets of iCCA proteomic and phosphoproteomic samples on a Q Exactive HF-X mass spectrometer.

Cell line sample preparation for TMTpro 16plex labeling

The protein samples were extracted from 15 cell lines (ICC5784, ICC772, LIPF 178C, LIPF 155C, HuCCT1, ICC880, ICC4175, ICC1969, ICC1239, LICCF, ICC1370, ICC2935, HCCC-9810, mixed sample, ICC892, RBE), and then digested by trypsin. The resulting peptides were labeled with TMTpro 16plex reagent (channels 126, 127N, 127C, 128C, 129N, 129C, 130N, 130C, 131N, 131C, 132N, 132C, 133N, 133C and 134) and fractionated (32 fractions) by the high-pH reverse phase liquid chromatography. The samples (1 μ g from each fraction) were analyzed on a Q Exactive HF mass spectrometer.

Nano-LC-MS/MS

Proteomic analysis

0.1% formic acid was used to resolve the peptides and 1 μ g 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 $\times$ 200 mm diameter C18 column (ReproSil-Pur C18-AQ, 3.0 μ m 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. Parameters of Q Exactive HF-X mass spectrometer was set as follows: MS1: resolution – 120,000 @ m/z 200, mass range – 350 to 1700 m/z, AGC Target – 3e6, maximum IT – 50 ms, top 20 ions selected for MS2; MS2: resolution – 60,000 @ m/z 200, normalized collision energy (NCE) – 32%, isolation width (m/z) – 1.0 m/z, AGC Target – 1e5, maximum IT – 120 ms, fixed first mass – 105.0 m/z, charge state include – 2–6, dynamic exclusion – 40 s. 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.

The TMT-16plex labeling samples were analyzed with a 120 min LC gradient and the gradient was set as 2%–5% B in 1 min; 5%–27% B in 94 min; 27%–35% B in 15 min; 35%–52% B in 3 min; 52%–100% B in 1 min; 100% B in 6 min. Parameters of Q Exactive HF mass spectrometer was set as follows: MS1: resolution – 120,000 @ m/z 200, mass range – 300 to 1700 m/z, AGC Target – 3e6, maximum IT – 50 ms, top 20 ions selected for MS2; MS2: resolution – 60,000 @ m/z 200, normalized collision energy (NCE) – 30%, isolation width (m/z) – 1.0 m/z, AGC Target – 1e5, maximum IT – 100 ms, fixed first mass – 105.0 m/z, charge state include – 2–6, dynamic exclusion – 40 s.

Phosphoproteomic analysis

0.1% formic acid was used to resolve the peptides and half peptides of each fraction for phosphoproteomic analysis were separated on an Easy nLC 1200 UHPLC system (Thermo Fisher Scientific) with a 70 min LC gradient at 300 nL/min (Buffer A: 0.1% formic acid in water; Buffer B: 0.1% formic acid in 80% acetonitrile) of a home-made 75 mm $\times$ 200 mm diameter C18 column (ReproSil-Pur C18-AQ, 3.0 μ m 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%–32% B in 52 min; 32%–45% B in 8 min; 45%–60% B in 3 min; 60%–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. Parameters of Q Exactive HF-X mass spectrometer was set as follows: MS1: resolution – 120,000 @ m/z 200, mass range – 350 to 1700 m/z, AGC Target – 3e6, maximum IT – 50 ms, top 20 ions selected for MS2; MS2: resolution – 45,000 @ m/z 200, normalized collision energy (NCE) – 32%, isolation width (m/z) – 1.0 m/z, AGC Target – 1e5, maximum IT – 120 ms, fixed first mass – 105.0 m/z, charge state include – 2–6, dynamic exclusion – 30 s. The 24 benchmark fractions analyzed with a 70 min LC gradient and the gradient was set as 2%–5% B in 1 min; 5%–32% B in 52 min; 32%–45% B in 8 min; 45%–60% B in 3 min; 60%–100% B in 6 min. The parameters of Q Exactive HF-X mass spectrometer were set the same as iCCA phosphoproteomic samples.

Database searching of MS data

MaxQuant 1.6.7.0 was used for mass spectrometric data analyzing against the human Swiss-Prot database containing 20,656 sequences (downloaded in August, 2019).

Parameters of MaxQuant was set as follows: PIF (precursor intensity fraction) filter value – 0.75, reporter mass tolerance – 0.003 Da, fixed modification – Carbamidomethyl cysteine, variable modifications – Oxidized methionine, protein N-term acetylation, lysine

acetylation, asparagine and glutamine (NQ) deamidation (phospho (STY) added in iCCA phosphorylation data analysis), maximum number of modifications – 5, enzyme specificity – trypsin, maximum missing cleavage site – 2, tolerances of first search – 20 ppm, tolerances of main search – 4.5 ppm, minimal peptide length – 7, FDR (false discovery rates) of peptide, protein and site – 0.01. The kit LOT number was used to corrected the purities of TMT channels for each set of TMT labeling data. The highly reliable sites (The Class I phosphorylation sites) will be considered if the localization probability filter >0.75.

WES data analysis

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 (<https://broadinstitute.github.io/picard/>) and base quality recalibration with GATK BaseRecalibrator v4.1.7.0 (<https://gatk.broadinstitute.org/hc/en-us>). Following the GATK guidance on tumor-normal matching samples, short somatic variants were called and filtered using GATK Mutect2 and other components. These variants were annotated using SnpEff v4.3T (<https://pcingola.github.io/SnpEff/>) and Annovar v2019-10-24 (<https://doc-openbio.readthedocs.io/projects/annovar/en/latest/>).

Several filtering criteria along with the population frequency of germline mutations were taken into consideration in somatic variants filtering. The population frequency of germline mutations from different public databases were annotated before somatic variants filtering procedure, including the Catalogue Of Somatic Mutations In Cancer (COSMIC) (v91, 2020-04-07) (<https://cancer.sanger.ac.uk/cosmic>), the 1000 Genomes (v5b, 2015-08-24), NHLBI Exome Sequencing Project (ESP6500) (v2, 2014-12-22) (<https://evs.gs.washington.edu/EVS/>) and the Genome Aggregation Database (gnomAD) (v2.1.1 2019-04-09) (<https://gnomad.broadinstitute.org/>).

To obtain high quality somatic variants, stringent downstream filters were used. For SNVs, these filters included a base coverage of a minimum of 10 reads, variant allele fraction (VAF) over 5% and at least 5 reads supporting mutations in tumor. Indels observed in at least 10 reads, with a VAF of 10% or more were kept. Furthermore, variants identified at a frequency higher than 1% in 1000 Genomes Project, ESP6500 or gnomAD were removed.

Somatic Copy number alteration was called by CopyWriteR (version 3.12) and GISTIC (version 2.0.3). Paired BAM files from tumor and adjacent normal liver samples were processed by CopyWriteR with default parameters. Batch effects were checked by hclust method from ComplexHeatmap R package (version 2.2.0) and normal tissues showing batch effects were replaced by other stable normal samples to reduce the batch effects. Then, copy number segmentations derived from CopyWriteR were processed with to retrieve gene-level copy number values and focal peaks. We set up a threshold of $\alpha = 0.4$, $-\text{conf} = 0.9$, $-\text{brlen} = 0.5$. GISTIC2 was used to generate arm level and focal level SCNAs for the cohort with G-Score and FDR Q value indicating the significance and strength of the identified SCNAs.

HLA typing and neoantigen prediction

Briefly, four-digit HLA class I was identified for each sample using Optitype v1.3.3 (Szolek et al., 2014). Neoantigen prediction was performed with NetMHCpan v4.0 (Jurtz et al., 2017), testing all novel 9-mer mutant peptides containing the mutated amino acids for binding to the patient's germline HLA alleles. A peptide was defined as a putative neoantigen with a predicted binding affinity <500 nM.

RNA-seq data analysis

RNA-seq data preprocess

Sequence data were removed adaptors and low-quality reads using the Trimmomatic v 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: $-\text{chimSegmentMin } 30 -\text{chimJunctionOverhangMin } 10 -\text{alignSJDBoverhangMin } 10 -\text{alignIntronMax } 200000 -\text{alignSJstitchMismatchNmax } 5 -1 5 5$).

RNA expression quantification with RSEM

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 $-\text{seqBias} -\text{gcBias} -\text{posBias}$

RNA variants calling

Before variants calling, reads were pre-processed of deduplication with picard (2.20.4). Additionally, the tool of GATK's SplitNCigar-Reads was applied to the deduped reads. The detection of variants was conducted by GATK in mode of 'HaplotypeCaller', with parameters of $-\text{dont-use-soft-clipped-bases} -\text{standard-min-confidence-threshold-for-calling } 20.0$. 'VariantFiltration' was implemented in the following step (with parameters: $-\text{window } 35 -\text{cluster } 3 -\text{filter-name } \text{FS} -\text{filter-expression } \text{"FS"} > 30.0 -\text{filter-name } \text{QD} -\text{filter-expression } \text{"QD"} < 2.0$ "), and only highly accurate variants were remained.

Fusion and alternative splicing events detection

For searching gene fusion in transcriptome, we applied STAR-Fusion (v1.8.1) to chimeric-junction files generated in the previous RNA-seq variants calling procedure. The detected fusions were filtered if supporting reads were less than 20. To find out alternative splicing events, Tophat (2.1.0) was implemented to generate junction files using Bowtie2-indexed (2.4.1) reference genome (hg19), with parameters of $-\text{g } 1 -\text{M} -\text{x } 1$ in fusion search mode.

Tumor-infiltrating immune and stromal cells

Estimation of the proportions of tumor-infiltrating immune cells in iCCA tissues was performed using xCell (Aran, 2020). The immune score and stromal score were also calculated by the same algorithm. The signatures used to determine the tumor microenvironment were computed as the geometric mean signature RNA expression (Petitprez et al., 2020).

Proteome and phosphoproteome data analysis

Data normalization

We identified a total of 10,529 proteins (unique peptides ≥ 2) and 62,879 phosphosites. To correct sample loading differences, proteome and phosphosite data were normalized using median centering method, and then log2 transformed for the following analysis. In normalized samples, these proteins and phosphosites should have a log TMT ratio value centered at zero. Among the data, 5,690 proteins and 2,761 phosphosites were quantified among all samples. Except that, 8,320 proteins and 18,347 phosphosites which were quantified in at least half samples (Table S1) were imputed for the missing values using K-nearest neighbor (KNN) algorithm. The imputation method was implemented in the DreamAI R package. For the protein level phosphorylation data, we first collapsed the phosphorylation sites from the same protein with the median ratio and then conducted the same normalization process, resulting 5,624 phosphoproteins quantified in at least half samples. We also filtered out liver specific genes eliminating the potential confounding effects of high expression liver genes in the following patient subgrouping analysis, correlation analysis and differential analysis (Consortium, 2013; Farshidfar et al., 2017).

Batch effect and quality control

We provided comprehensive proteomic coverage in our 11-plex TMT-based LC-MS/MS analysis. Principal-component analysis and hierarchical clustering analysis clearly observed no batch effects in the 22 TMT plexes based on the TMT global proteome data for the FU-iCCA tumor tissues. The data from the same patient samples could be clustered into the adjoining place, suggesting robust and accurate proteome quantification platform (Figures S1B and S1C).

mRNA, proteomic and phosphoproteomic subgrouping analysis

Consensus clustering analysis

Unsupervised K-means consensus clustering was performed for proteome, transcriptome and phosphoproteome data respectively using the CancerSubtypes R package (Xu et al., 2017). For different omics datasets, 5,502 proteins with no missing values, 1,9803 mRNAs and 5,549 phosphoproteins quantified in at least half samples were involved in the analysis. Then the median absolute deviation (MAD) score of every mRNA, protein and phosphoprotein was calculated and the top 25% most varied genes was selected for the consensus analysis. For the detailed settings, a total number of subsampling was set to 1,000, and in each resampling, 80% samples and 80% genes were selected (pltem=0.8, pFeature=0.8). The number of clusters considered was 2 to 6. We compared the average pairwise consensus matrix for consensus clusters, the delta plot of the relative change in the area under the cumulative distribution function (CDF) curve, and the average silhouette distance within consensus clusters and selected $k = 4, 3, 3$ as the best sample subgrouping for proteome, transcriptome and phosphoproteome respectively. In addition, for the comparison of different patient subgrouping of omics data, we selected the over-represented transcriptome and phosphoproteome subgroups in each proteomic subgroup using Fisher's exact test ($p < 0.05$).

Comparison of the iCCA proteomic subgrouping with previous biliary cancer subgroupings

We collected four biliary cancer gene expression signatures from previous papers for comparison, including Nakamura biliary tract cancer signatures (Nakamura et al., 2015), Job immune signatures (Job et al., 2019), Oishi hepatic stem-like gene signatures (Oishi et al., 2012) and Andersen CCA signatures (Andersen et al., 2012). Hierarchical clustering was performed using these previously defined signatures and FU-iCCA mRNA data, and comparison results were aligned with our proteomic subgrouping.

Association between proteomic subgroups and clinical outcome

Survival analysis was performed for the identified patient subgroups from proteome, transcriptome and phosphoproteome respectively. The Kaplan-Meier survival curves was used to estimate the overall survival (OS) and compared by the log-rank test. We implemented these analyses using the 'survival' R package.

Multi-omics data analysis

mRNA-protein correlation

Spearman correlation coefficients and the corresponding p values of shared 5,434 mRNA/protein pairs were calculated across 208 FU-iCCA samples with RNA-Seq and MS data. In addition, the corresponding p values were adjusted by the Benjamini-Hochberg correction and a cut-off of 0.01 was determined as the significance of correlation pairs. In order to explore the biological functions of different expression correlations, gene set enrichment analysis (GSEA) was performed using the GSEA software (Subramanian et al., 2005). Molecular Signatures Database (MSigDB, c2.cp.kegg.v7.1.symbols.gmt) was used for enrichment analysis. We also classified samples into two TNM early or late groups based on stages I-II and stages III-IV, and ranked genes according to correlation changes. The GSEA analysis was then performed in the resulting ranked genes to explore the molecular dynamics during tumor progression.

Effects of copy number alterations

The "cis" (within the same aberrant locus) or "trans" (remote locus) regulation of mRNA, protein and phosphoprotein abundance by SCNAs were visualized by multiOmicsViz (R package). Spearman's correlation coefficients and associated multiple-test adjusted

p values using FDR were calculated for all CNA-mRNA, CNA-protein and CNA-phosphoprotein. FDR <0.01 was considered as statistically significant positive or negative correlations for CNA-mRNA, CNA-protein and CNA-phosphoprotein correlations.

Patient specific database construction and variant peptide identification

For each of the 262 iCCA patients analyzed, DNA-variants (somatic/germline) by WES and RNA-variants by RNA-seq were obtained with the method described above. The proteogenomic database tool CustomProDB (Wang and Zhang, 2013) was used to incorporate the germline and somatic SNVs, indels, and RNA-seq predicted junctions into a searchable protein specific database with default parameters for variant peptide identification. The identified MS2 spectra were removed from the original raw files using home-developed R script and the filtered spectra were transformed into mzxml files. For each set of TMT files, the mzxml files were searched against their corresponding customized databases. The database searching parameters were almost identical to those described above except that Oxidized methionine and protein N-term acetylation were set as variable modifications and the data were filtered at 1% PSM FDR.

Differential analysis and pathway enrichment analysis

Differential expression analysis of FU-iCCA samples with various phenotypes was performed, including differential proteins according to CA19-9, intrahepatic metastasis, lymph metastasis and TNM status; differential mRNAs, proteins and phosphoproteins in tumors with high or low expression of SLC16A3 and HKDC1, and absence or presence of driver mutations (*TP53*, *KRAS*, *BAP1*, *IDH1/2* and *TP53/KRAS* co-mutation) or fusions (*FGFR2*). Wilcoxon test was used for differential analysis with the transcriptome data and a cutoff of FDR <0.05 and fold change >2 was selected. For the proteome and phosphoproteome data, limma R package was implemented and moderated t-statistics was used. We selected the trend parameter to enhance the stability of the analysis and genes with FDR <0.05 and fold change >1.5 or other thresholds were considered as significantly changed genes. These significant genes were then performed KEGG pathway or GO enrichment analysis using clusterProfiler R package and an FDR value <0.05 or other thresholds were considered as the cutoffs of significantly regulated pathways.

Functional enrichment analysis of multi-omics data in four proteomic subgroups

To further elucidate the biological characteristics of the proteome subgrouping, ssGSEA analysis was performed to identify the pathway alterations among the four subgroups using GSVA package (Hanzelmann et al., 2013). Expression data from multi-omics levels of 594 overlapped genes among mRNA, proteins and phosphoprotein levels were selected to achieve enrichment scores over MsigDB database v.7.1 with at least 10 overlapping genes. The significance of the pathway enrichment scores (PES) among the four subgroups were estimated by a linear model and moderated with the F-statistic using limma package. We identified 110 pathways from the proteome ssGSEA analysis using the Benjamini-Hochberg corrected p values with the cutoff of 0.01, and 34 distinct pathways through knowledge-based selection were visualized by heatmap.

Potential prognostic protein biomarker detection

To identify potential protein biomarkers, we first stratified the FU-iCCA samples into two groups according to the median expression of each protein quantified at all samples. Kaplan-Meier curve with Log-rank test was used to visualize and compare the survival difference. We selected the Log-rank p value <1.0e-05 as biomarker candidates. Then, the Cox proportional hazard model was used to evaluate the hazard ratio for each gene, and the potential best or worst prognostic biomarkers were identified. We filtered out genes whose mRNA-protein correlations were less than 0.6 for better accordance of different omics.

iCCA cell line subgrouping based on existed molecular model of FU-iCCA

We used the caret R package (6.0.88) to classify all the 15 cell lines into four subtypes. The training dataset is the human proteomics dataset with 214 samples, and all these samples were classified into four subtypes as the classification label. The predicted dataset is the 15 cell line datasets. Filter methods built-in caret package were used to do feature selection with 10-fold cross-validation and repeated this process 5 times. Finally, the training set indicated that 181 genes with variance larger than 1 across all the samples were selected. Penalized multinomial regression model ('multinom') built-in caret package was selected to train the classification model based on the 181 features selected on the training dataset with 10-fold cross-validation and repeated this process 5 times. The model parameter with the best performance was selected during the training progress and applied for the 15 cell line datasets to predict the subtype label. Finally, three cell lines were classified into S1. Five cell lines were classified into S2. Four cell lines were classified into S3 and three cell lines were classified into S4.

Tissue MicroArray (TMA) experiment

Multiplexed immunostaining

Multiplex immunostaining of MPO (monoclonal, No. MAB-0779 Maxim), POSTN (clone EPR20806, No. ab215199 Abcam), ALDOB (polyclonal, No. 18065-1-AP Protein tech) and EPCAM (clone EPR20532-225, No. ab223582 Abcam) was performed by the Vectra Automated Quantitative Pathology Imaging and Analysis platform and inFormTM image analysis software (PerkinElmer). Slides were first deparaffinized and rehydrated, followed by microwave antigen retrieval (pH = 9.0, ADI-950-274-0500, ENZO). After blocking endogenous peroxidase and nonspecific binding sites (ZAE-ICT-6295- L100, ENZO), primary Abs and secondary HRP-conjugated polymers (MPX-2402, 5692 Vectorlabs) were applied. Each HRP-conjugated polymer covalently bound with a distinct fluorophore using tyramide signal amplification (Opal 7-color Fluorophore TSA plus Fluorescence Kit (NEL 797001KT; PerkinElmer). This covalent reaction was followed by additional antigen retrieval (pH = 6.0, ADI- 950-270-0500, ENZO) to remove background signal before next step. The process was conducted for the following antibodies/fluorescent dyes, in order: anti-EPCAM/Opal690, anti-ALDOB/Opal570, anti-MPO/Opal520, and

anti-POSTN/Opal650. After two sequential reactions, slides were counterstained with DAPI (D9542, Sigma, 1:500) and mounted with fluorescence mounting medium (S3023, Dako). Following similar procedures, the staining of KRT19 (clone EP1580Y, No. ab52625 Abcam), SLC16A3 (monoclonal No. sc-376465 SANTA CRUZ BIOTECHNOLOGY) and HKDC1 (polyclonal, No. ab228729 Abcam) were performed on an independent TMA cohort containing 222 iCCA patients to validate our proteomic prognostic biomarkers. These patients received curative surgery from January 2012 to December 2015 at Zhongshan Hospital and received no prior anticancer treatments. The median survival time of these patients is 19 months, with 112, 73, and 37 patients classified as TNM stages I, II, and III/IV, respectively.

Multispectral imaging, spectral unmixing, and analysis

A workflow enabling simultaneous evaluation of multiple biomarkers on TMAs was established. Briefly, multiplex stained TMA slides were scanned using the Vectra multispectral automated microscope (PerkinElmer), where original images comprised four combined 200 multispectral image cubes. Multispectral images for each TMA core were created by stitching images captured every 10 nm across the range of five filter cubes comprising DAPI (440–680 nm), FITC (520 nm–680 nm), Cy3 (570–690 nm), Texas Red (580–700 nm) and Cy5 (670–720 nm). A spectral library was produced by the supervised machine learning algorithms within Inform (version 2.4, PerkinElmer). Individual components were separated from each multispectral image by this spectral library (spectral unmixing). The spectrally unmixed and segmented images were subjected to a distinctive phenotyping algorithm for identification of each DAPI-stained cell according to each fluorophore expression and nuclear/cell morphological features. For each marker (anti-KRT19/Opal520, anti-SLC16A3/Opal650, anti-HKDC1/Opal570), the cutoff for positivity was decided according to the staining pattern and intensities on all images. All quantifications were evaluated blinded to patient clinical outcomes. The patients were divided into two groups according to the expression (number of positive staining cells per square millimeter per patient) of SLC16A3 and HKDC1 (cutoff was automatically set by the X-tile software, Version 3.6.1) (Camp et al., 2004).

Four subgroup-specific markers, including MPO, POSTN, ALDOB and EPCAM, were selected and performed multiplex immunostaining for both FU-iCCA cohort and validation cohort. The patient, who specifically have high expression of above protein, will be classified into the corresponding subgroups. Finally, 84/196 and 144/222 patients in FU-iCCA cohort and validation cohort fit the typical characteristics of subgroups, respectively.

Fluorescent in situ hybridization (FISH) analysis

The FISH was conducted as described in our previous study (Dong et al., 2018). The FISH-Tissue Implementation Kit (ZytoLight, Cat#Z-2028-20) was used in combination with SPEC *FGFR2* dual-color break-apart probe (ZytoLight, Cat#Z-2169-200) for the detection of *FGFR2* fusion events. The FFPE slides were fixed in 10% neutral buffered formalin and paraffin-embedded. Following dewaxing and rehydration using ethanol gradients, slides were placed in Heat Pretreatment Solution Citric buffer (ZytoLight, PT1) at 98°C for 15 min and washed. Digestion enzyme (pepsin solution, ZytoLight, ES1) was added to each slide and incubated for 15 min at 37°C. Slides were washed and dehydrated. 10 μ L *FGFR2* dual-color break-apart probe, containing 2 specific probes for the 5' (red) and 3' (green) part of *FGFR2*, was pipetted onto the pretreated slides and then denatured for 10 min at 75°C. Then, slides were incubated overnight at 37°C in a moisturized chamber. After washing with 0.5x SSC at 70°C, slides were stained and mounted using Vecta shield mounting media with DAPI (Vector Labs). Slides were scanned and images were captured using a 60x objective on confocal laser scanning microscope (Leica TCS SP8).

Production and identification of eluted HLA-I peptides

The immunopeptidomes of MHC-I peptides were performed as described in our previous study (Dong et al., 2020). The *FGFR2::BICC1*-overexpressed iCCA cell lines, HuCCT1 and HCCC-9810, were collected. Nearly 200mg cells were dissociated using 2 ml of protein lysis buffer (0.25% sodium deoxycholate, 0.2 mM iodoacetamide, 1 mM EDTA, 1:200 Protease Inhibitors Cocktail (Sigma-Aldrich, MO), 1 mM PMSF and 1% octyl- β -D glucopyranoside (Sigma-Aldrich, MO) in PBS). Lysates were pre-cleared using microfuge centrifugation for 20 minutes at 12,000rpm at 4°C.

We took immunoaffinity purification of pan-HLA-I molecules from cleared lysate with the W6/32 antibody (No. MA1-70111, Thermo Scientific) covalently bound to Pierce™ Protein A/G Magnetic beads (No. 88805, Thermo Scientific) using disuccinimidyl suberate (DSS) for 2 hours at room temperature. Next, the elution buffer was added to the beads to elute HLA-I protein. After immunoprecipitation, peptides were eluted from the purified MHC molecules using 10% acetic acid. Eluted peptides were further filtered with a 10 kDa centrifugal filter tubes (Millipore), followed by vacuum centrifugation dryness, de-salted on C18-based STAGE tips.

0.1% formic acid was used to resolve the eluted peptides which were separated on an Easy nLC 1200 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 $\times$ 200 mm diameter C18 column (ReproSil-Pur C18-AQ, 3.0 μ m resin, Germany). The gradient was set as 2%–5% B in 1 min; 5%–32% B in 97 min; 32%–45% B in 12 min; 45%–100% B in 3 min; 100% B in 1 min; 100% B in 7 min. Xcalibur software was used for data-dependent acquisition. Parameters of Q Exactive HF-X mass spectrometer was set as follows: MS1: resolution – 60,000 @ m/z 200, mass range – 300 to 1800 m/z, AGC Target – 5e5, maximum IT – 120 ms, top 20 ions selected for MS2; MS2: resolution – 30,000 @ m/z 200, normalized collision energy (NCE) – 27%, isolation width (m/z) – 1.2 m/z, AGC Target – 1e5, maximum IT – 120 ms, charge state include - 2–6, dynamic exclusion – 15 s.

The pFIND version 3.1.5 was used for mass spectrometric data analyzing against the mixed database containing the human Swiss-Prot database (20,656 sequences, downloaded in August, 2019) and the self-built database (48 sequences of the potential neo-antigens), with peptide false discovery rate (FDR) setting as <1% and enzyme specificity setting as unspecific.

Peptide–MHC tetramer exchange assays

All the fusion-derived peptides were synthesized according to standard procedure (China Peptides Inc.). The HBV-derived peptide pools were purchased from JPT.Inc (Cat. PM-HBV-CPULTRA). Peptides were loaded at 100 $\mu\text{g}/\text{ml}$ onto the QuickSwitch Quant HLA-A02:01 and HLA-A24:02 Tetramers (PE labeled) (MBL International) according to manufacturer's instructions. We used the HLA-A02:01 and HLA-A24:02 tetramer kits to validate peptide exchange efficiency of all the potential neo-peptides, and generate the peptide-specific tetramers by peptide exchange experiments. Each peptide was dissolved in DMSO to a 10mM solution to be assayed. Then, we pipetted 50 μL of HLA-A02:01 and HLA-A24:02 Tetramer into each well of the conical-bottom 96 well microtiter plate, added 1 μL of Peptide Exchange Factor plus 1 μL of peptide and mixed gently with pipetting. We repeated these steps for each additional peptide, including the Reference Peptide, and incubated the mixture overnight at room temperature protected from light. Next, the Magnetic Capture Beads were added to conjugated with the above tetramers, and the FITC-labeled Exiting Peptide Antibody was applied to the reaction against the Exiting Peptide. By measuring the percentage of original peptide replaced by a competing peptide, we may evaluate the exchange efficiency and determine whether the resulting tetramer is suitable for following staining.

In the analysis, we adjusted voltages, gains and threshold for FSC, SSC, FITC and second fluorochrome (PE) so that bead events, Control#1 (Beads only), Control#2 (Negative control) and Control#3 (Positive control) were within range.

Enzyme-linked immunospot (ELISPOT) assay

The ELISPOT plate was processed according to the manufacturer's instructions. Briefly, ELISPOT plates (Abcam, Cat# ab62899) were pretreated with 100 μL /well of PBS for 10 min. Then, we dispense into wells 100 μL of cell suspension containing nearly 10^5 cells. Cover the plate with a standard 96-well plate plastic lid and incubate cells at 37°C in a CO₂ incubator. After 20h of co-culture, the ELISPOT plates were washed with 100 μL of PBS-0.1% tween 20 (PBS-T) in wells, and then incubated for 1h at room temperature (RT) with 100 μL /well of anti-human IFN- γ detection antibody solution. The plate was then washed 3 times with PBS-T, followed by a 1 h incubation with 100 μL /well of Streptavidin-Alkaline phosphatase. The plate was then washed 3 $\times$ with PBS followed by development with 100 μL /well of ready-to-use BCIP/NBT buffer. The reaction was stopped by rinsing thoroughly with cold tap water. ELISPOT plates were scanned and counted using an ImmunoSpot plate reader and associated software (Cellular Technologies, Ltd).

Cytometry by time of flight (CyTOF)

Briefly, the cells were thawed and resuspend in Maxpar Cell Staining Buffer. We added 50 μL of the antibody cocktail, containing a panel of 30 antibodies (detailed antibody list in Table S4C), to each tube so the total staining volume was 100 μL . Each tube was gently pipetted to mix and they were incubated at RT for 15 minutes. Then we washed them by adding 2 mL Maxpar Cell Staining Buffer to each tube, centrifuged at 300 $\times$ g for 5 minutes, removed supernatant by aspiration, and repeated 3 times. Next, 1.6% FA solution was added to each tube to fix cells. 1 mL of the cell intercalation solution was added to each tube. They were incubated for 1 hour at RT. Then, sufficient volume of 0.1X EQ beads were prepared to re-suspend all samples, and cell concentration was adjusted to 1.0×10^6 cells/mL. The data was acquired on the HeliosTM platform.

CyTOF data analysis

Mass cytometry data were randomized with the Fluidigm acquisition algorithm (V6.0.626) and normalized by using MATLAB normalizer (v.7.14.0.739) in MATLAB R2012a. We splited the sample barcodes by utilizing a single cell debarcoder tool. Individual samples were manually gated using Cytobank to exclude normalization beads, cell debris, dead cells, and doublets for the identification of T cells for further downstream analyses. Separate viSNE plots were generated for subsets of total lymphoid cells. We further utilized Seurat (Butler et al., 2018) to perform the data normalization, feature selection, dimensionality reduction (parameter: RunTSNE (do.fast = T, dims = 1:10)), and clustering analysis (parameter: resolution = 1). To visualize the gene expression profile of immune cells, we performed the down-sampling analysis ($n = 5000$) and draw the heatmap by using Seurat (Butler et al., 2018). We used dittoSeq to visualize the proportion changes of immune cells (Bunis et al., 2020). To computationally model the differentiation state of selected immune cells, we utilized Monocle 2 (Qiu et al., 2017) to compute the cellular trajectory (parameter: method="DDRTress").

Drug target analysis

Drug target information was retrieved from DGIdb (<https://www.dgldb.org>) which involved over thirty trusted sources about drug-gene interactions such as drugbank, FDA, NCI, etc. Target proteins that were up regulated in protein subtypes or samples with driver mutations and had potential pharmacological action as inhibitors, antagonists and antibodies were chosen.

cDNA conversion and RT–PCR reaction for FGFR2 fusion validation

1 μg of total RNA was reverse transcribed into first-strand cDNA using the ABScript II RT Master Mix for qPCR Kit (ABclonal). The resulting cDNA was used as the template for semi-quantitative PCR amplification using the primers reported in Table S4A. To detect the presence of the fusion product, the PCR amplifications were performed using the following protocol: 95°C for 5 min, 40 cycles of 95°C for 30 s, 55°C for 30 s, 72°C for 2 min. PCR amplifications were performed in a volume of 25 μL reaction mixture containing 1.5 mM MgCl₂, 0.2 mM of each dNTP, 0.125 mM of each primer and 1 U of Platinum Taq DNA polymerase (Invitrogen). PCR products

were purified using the Qiaquick PCR purification kit (Qiagen) and sequenced using an Applied Biosystems 3700 DNA sequencer (Applied Biosystems). The samples were considered positive only if Sanger sequencing successfully confirmed the sequence of the fusion mRNA around the breakpoint (Sia et al., 2015).

Functional experiments

Plasmids and lentivirus-mediated construction of stable cell lines

The HA-tagged coding sequence of human HKDC1 and SLC16A3 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 stable cell lines by infection were performed according to the protocol before (Boehm et al., 2005).

RNA interference

The siRNAs were synthesized by Biotend Company. All siRNA transfections were performed with X-tremeGENE siRNA Transfection Reagent (Roche) at 50 nM final concentration according to the manufacturer's protocol. For drug targets validation experiments in Figure S3, three independent siRNAs against each gene were equally mixed up and transfected into indicated cells at 50 nM final concentration. The siRNA transfected cells were harvested for the followed assays 48 hr after transfection. Oligonucleotide sequences are as following:

siScr: 5'-UUCUCCGAACGUGUCACGUTT-3'
 siSLC16A3-1: 5'-GUCUACAUGUACGUGUUCAdTdT-3'
 siSLC16A3-2: 5'-GCAUUAGGAAGAAGCCCAAdTdT-3'
 siTOP2A-1: 5'-GCUCCAAAUCAUAUGUGAdTdT-3'
 siTOP2A-2: 5'-CCAACUUUGAUGUGCGUGAdTdT-3'
 siTOP2A-3: 5'-CUUCAACUAUCUUCUUGAUdTdT-3'
 siCDK1-1: 5'-CCUAUGGAGUUGUGUAUAAdTdT-3'
 siCDK1-2: 5'-GGUUAUAUCUCAUCUUUGAdTdT-3'
 siCDK1-3: 5'-GGGAAUUUCUCUAUUAAAdTdT-3'
 siPCNA-1: 5'-GAAGCACCAACCAGGAGAdTdT-3'
 siPCNA-2: 5'-GAUGCUGUUGUAAUCCUdTdT-3'
 siPCNA-3: 5'-CGUCUCUUGGUGCAGCUCdTdT-3'
 siITGA2B-1: 5'-CAAACUUUACAAACCUUCAdTdT-3'
 siITGA2B-2: 5'-CUGCUCUACAUCUGGAUAdTdT-3'
 siITGA2B-3: 5'-CUGGUGCAAGAUUCACUGAdTdT-3'
 siITGB6-1: 5'-GCCUCCAAACAUCUCCAUAdTdT-3'
 siITGB6-2: 5'-GAAACAUUUAUGGGCCUAdTdT-3'
 siITGB6-3: 5'-GCCAACCCUUGCAGUAGUAdTdT-3'
 siRIPK3-1: 5'-CUGAGAGACAAGGCAUGAAdTdT-3'
 siRIPK3-2: 5'-GUGGCUAAACAAACUGAAUdTdT-3'
 siRIPK3-3: 5'-CACAGGGUUGGUAAAUCAAdTdT-3'

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 μ g total proteins were separated by 10% SDS-PAGE gel and blotted with indicated primary antibodies. All Western blot gel images were obtained with an Minichemi 610 chemiluminescent imager (Sagecreation, Beijing, China).

High-throughput drug screening

The information of 35 screened drugs is listed in Table S5K. The cells were digested and seeded in 384-well plates at the pre-determined cell density with 50 μ l culture medium using Multidrop Combi Reagent Dispenser (Thermo Fisher Scientific). After overnight incubation, adherent cells were treated with drugs of 2-fold serial dilutions of 10 doses (from 40 μ M to 78 nM) using D300e (TECAN) and there was a replicate for each drug dose. The treated cells were incubated for 72 hours and then each well was added 20 μ l Cell-Titer-Glo reagent (Promega). The plates were centrifuged and then oscillated at 300rpm for 5 minutes. Finally, the luminescent signals were measured by EnSight (Perkin Elmer) to estimate the cell viabilities.

Cell proliferation and IC₅₀ assays

For cell proliferation assay, HuCCT1 cells were seeded into 96-well plate (1×10^3 cells per well). For each measurement, CCK-8 solution (C0039, Beyotime Biotechnology) at the final concentration of 10% was added to the wells, and absorbance at 450nm was measured 2 hr after incubation to represent the relative cell numbers. For IC₅₀ assay, 5×10^3 HuCCT1 cells per well were plated in 96-well plate, the cells were treated with different doses of α -CHCA (28166-41-8, Selleck Chemicals) for 72 hr, and absorbance at 450nm was measured 1 hr after 10% CCK-8 incubation. Each assay was repeated at least three times independently. All data were analyzed using GraphPad Prism 8 software.

Colony formation assay

1×10^3 HuCCT1 cells were seeded in 6-well plate. The cells were stained with crystal violet 7 to 10 days after routine culture, and the colonies were counted by using ImageJ (Version1.53). Each assay was repeated at least three times independently.

Glucose consumption analysis

The glucose consumption of HuCCT1 cells was analyzed by using FreeStyle Precision Pro Blood Glucose and β -Ketone Monitoring System (Abbott Diabetes Care Ltd.). 2.5×10^4 cells per well were seeded into 96-well plate, and the glucose concentration of each well was measured 12 hr after incubation. The glucose consumption was calculated as (measured glucose concentration value - glucose concentration value of RPMI.1640 medium supplemented with 10% FBS and 100 U/mL penicillin and 100 mg/mL streptomycin)/12 hr/ 2.5×10^4 cells. Each assay was repeated at least three times independently.

Assessment of extracellular acidification rate (ECAR)

ECAR of HuCCT1 cells was measured by a Seahorse XF 24 analyzer (Seahorse Bioscience). 2.5×10^4 cells per well were plated in Seahorse XF 24-well culture plates (Bucher Biotech AG) with RPMI.1640 medium supplemented with 10% FBS and 100 U/mL penicillin and 100 mg/mL streptomycin, and incubated at $37^\circ\text{C}/5\% \text{CO}_2$ incubator overnight. Cells were changed from RPMI.1640 medium to the XF Glycolysis Stress Test Assay Medium (D5030, Sigma-Aldrich) after washed twice with the same assay medium, and then placed in a $37^\circ\text{C}/\text{non-CO}_2$ incubator for 1 h. The basal rate was measured three times prior to the first injection, and then 10 mM glucose, 1 μM oligomycin, and 100mM 2-deoxy-D-glucose (2-DG) were injected into each well at the indicated time points. After injection, the ECAR of each well was monitored until the rates stabilized. After the whole measurement process accomplished, the living cell number of each well was calculated using a EnSight Multimode Plate Reader (PerkinElmer) after bright-field count and PI staining for further normalization. The corresponding glycolysis, glycolytic capacity, and glycolytic reverse parameter values of indicated cells were calculated according to rate measurement equations.

The workflow for microbiome analysis

Tumor tissues and adjacent non tumor liver tissues were selected from 80 pairs of ICCA patients for microbiological analysis. In the main steps of analysis: tissue processing, DNA extraction and DNA amplification, a comprehensive set of negative controls was tested to control for environmental sample contamination. The blank control (Blank-lab, $n = 6$) was set up to control contamination from tissue processing. Water samples of extraction kit and amplification kit were used respectively to control for labware, reagent and/or environmental contamination during DNA extraction (Blank-ext, $n = 6$) and amplification steps (Blank-amp, $n = 6$) for tissue 16S rRNA sequencing. After that, the negative control was thoroughly validated on a case-by-case basis and 16S rRNA quantification and sequencing data were used to discover tissue-specific bacterial signatures associated with ICCA. The abundance of OTUs common was calculated in blank samples but without a new source in experimental samples. Then, the OTU with the highest abundance was selected as the internal reference sequence. The theoretical abundance value (theoretical maximum value) of each contaminated OTU in each sample was calculated by using the internal reference sequence, and the abundance of OTUs that might contain background contamination was deducted from the above experimental samples.

DNA extraction and bacterial 16S rRNA sequencing

Approximately 100 μg tumor tissue and adjacent normal liver tissue sample from each case was pulverized and placed in 1.5ml Eppendorf tube. Tissue was incubated with Protease K and tissue DNA was isolated by phenol/chloroform method, and then detected on 1.2% agarose gels. The V3 (variable region 3) and V4 (variable region 4) regions of 16S rDNA were amplified by PCR and sequenced in the MiSeq platform (Illumina) using the paired-end protocol yielding pair-end reads that almost completely overlap.

The primers used for amplification contain adapters for MiSeq sequencing and single-index barcodes so that the PCR products may be pooled and sequenced directly (Caporaso et al., 2012), targeting about 40,000 reads per sample. Forward and reverse primers were 5'-ACTCCTACGGGAGGCAGCAG-3' (357F) and 5'-GGACTACHVGGGTWTCTAAT-3' (806R). To maximize data resolution, 16S rRNA gene pipeline data incorporated phylogenetic and alignment-based approaches.

The primers also contained the Illumina 5' overhang adapter sequences for two-step amplicon library building, following manufacturer's instructions for the overhang sequences and barcodes.

The initial 50 μL PCR reaction mix comprised 1–2 μL DNA template, 200 μM dNTPs, 0.2 μM of each primer, 5X reaction buffer 10 μL and 1U Phusion DNA Polymerase (New England Biolabs). The PCR conditions were an initial denaturation for 2 min at 94°C , followed by 25 cycles of denaturation for 30 s at 94°C , annealing for 30 s at 56°C and extension for 30 s at 72°C and a final extension for 5 min at 72°C . DNA gel extraction kit (Axygen) and the FTC -3000 TM real-time PCR (Funglyn Shanghai) method were used to purify and quantify the barcoded PCR products. The PCR products from different samples were mixed in equal proportion. In the second step, PCR with dual 8bp barcodes was used for multiplexing. Two unique barcodes were incorporated into either end of the amplicons by 8 cycle PCR reactions. Cycling conditions were as follows: a denaturation cycle of 3 min at 94°C , followed by eight cycles for 30 s at 94°C , then an incubation for 30 s at 56°C and 72°C for 30 s, and a final extension of 5 min at 72°C . The library was purified using a DNA gel extraction kit (Axygen) and sequenced using Novaseq 6000 SP 500 Cycle Reagent Kit (Illumina at TinyGen Bio-Tech (Shanghai) Co., Ltd) by 2 $\times$ 250 bp paired-end sequencing on the Novaseq platform. The raw fastq files were demultiplexed based on the barcode.

PE reads for all samples were run through trimmomatic (version 0.35) to remove low quality base pairs using these parameters (SLIDINGWINDOW: 50:20 MINLEN: 50)

Microbiome pipeline analysis

FLASH program (version 1.2.11) (Magoc and Salzberg, 2011) was used to merge the raw paired-end 16S rRNA reads (V3 and V4 region) into consensus fragments. With the following filtering parameters (maxambig=0, minlength = 200, maxlength = 485, maxhomop= 8), the low quality contigs were removed based on screen.seqs command. The 16S sequences data analyses were performed in mothur (version 1.33.3), UPARSE (search version v8.1.1756, <http://drive5.com/uparse/>), and R (version 3.6.0). The demultiplexed reads were clustered at 97% sequence identity into operational taxonomic units (OTUs) and the singleton OTUs were deleted using the UPARSE pipeline (http://drive5.com/usearch/manual/uparse_cmds.html). The OTU representative sequences were assigned for taxonomy against Silva 128 database with confidence score ≥ 0.6 by the classify.seqs command in mothur (Quast et al., 2013). OTUs were classified by NCBI (from Phylum to Species). Eighteen negative control samples (no tissues) were used to remove bacterial contaminant. Downstream comparative analysis used 2,885 sequences per sample subsampled from the contaminant-free 16S rRNA profiles. The alpha-diversity analysis, ACE index analysis, Shannon analysis, rarefaction curves analysis and so on were performed by mothur and the distribution of results were plotted by R. Kruskal-Wallis rank sum test that performed with “ggpubr::compare_means” function in “ggpubr” package of R was used to detect the significant changes of shannon index in four groups. For the beta-diversity metrics, the weighted and unweighted UniFrac distance matrix was calculated by mothur. Principal Coordinate Analysis (PCoA) and tree were generated based on the results using dendextend and ape in R.

Differential abundance analysis of alpha diversity features of interest evaluated differences using the nonparametric difference test. We filtered out OTUs which has significant correlations with at least five signature proteins and Spearman correlation was used. MAFFT (V7.471) was used for sequence alignment, FastTree (Version 2.1.8) was used for growth tree calculation, and MEGA5 was used for image visualization. High-quality non-contaminant 16S rRNA sequences were analyzed for functional gene content using PICRUSt (Langille et al., 2013) which provided proportional contributions of KEGG categories for each sample (Kanehisa and Goto, 2000). The heat map was generated based on differentially abundant functional categories (KEGG Level 2, FDR adjust. $p < 0.05$). In the four subgroups, the relative abundance values were centered on functional classification and colored according to enrichment or decrease (Magoc and Salzberg, 2011). The statistical annotations were added to indicate significant correlations with metadata, enabling rapid evaluation of many variables.

QUANTIFICATION AND STATISTICAL ANALYSIS

Quantification methods and statistical analysis methods for single-omic and multi-omic analyses were mainly described and referenced in the respective Method Details subsections.

Additionally, standard statistical tests were used to analyze the clinical data, including but not limited to Student's t test, Chi-square test, Fisher's exact test, Kruskal-Wallis test, Log-rank test and Cox proportional hazards regression model. For categorical variables versus categorical variables, Fisher's exact test was used; for categorical variables versus continuous variables, Kruskal-Wallis test was used to test if any of the differences between the subgroups were statistically significant; and for continuous variables versus continuous variables, spearman correlation was used. Kaplan-Meier plots (Log-rank test) were used to describe overall survival. 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. All statistical tests were two-sided, and statistical significance was considered when p value < 0.05 unless otherwise indicated. To account for multiple-testing, p values were adjusted using the Benjamini-Hochberg procedure. All the analyses of clinical data were performed in R (version 3.4.3).

For functional experiments, each was repeated at least three times independently, and results were expressed as mean $\pm$ standard error of the mean (SEM). The statistical significance of differences was determined by two-way ANOVA for the results of CCK8 and Seahorse experiment. Statistical analysis was performed using GraphPad Prism (version 8).