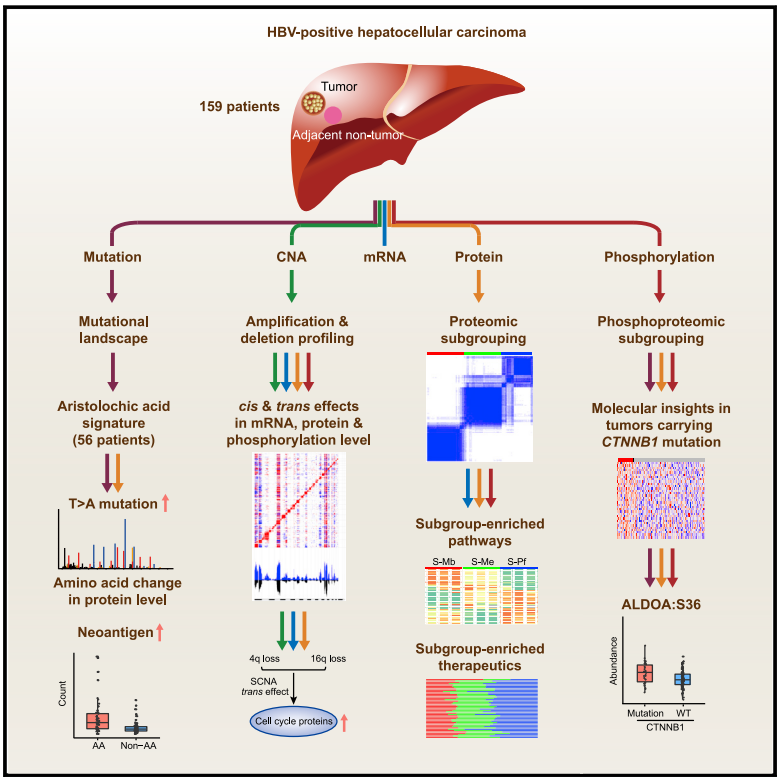
Resource

Cell

Integrated Proteogenomic Characterization of HBV-Related Hepatocellular Carcinoma

Graphical Abstract



Highlights

- Proteomic subgroups stratify patient survival and allocate specific treatments
- Alterations of the liver-specific proteome and metabolism in HCC are identified
- Multi-omics profile of key signaling and metabolic pathways in HCC is depicted
- *CTNNB1* mutation-associated ALDOA phosphorylation promotes HCC cell proliferation

Authors

Qiang Gao, Hongwen Zhu, Liangqing Dong, ..., Daming Gao, Hu Zhou, Jia Fan

Correspondence

dgao@sibcb.ac.cn (D.G.), zhouhu@simm.ac.cn (H.Z.), fan.jia@zs-hospital.sh.cn (J.F.)

In Brief

Proteogenomic characterization of HBV-related hepatocellular carcinoma (HCC) using paired tumor and adjacent liver tissues identifies three subgroups with distinct features in metabolic reprogramming, microenvironment dysregulation, cell proliferation, and potential therapeutics.

Gao et al., 2019, Cell *179*, 561–577
October 3, 2019
https://doi.org/10.1016/j.cell.2019.08.052

CellPress

Resource

Cell

Integrated Proteogenomic Characterization of HBV-Related Hepatocellular Carcinoma

Qiang Gao,[1,12] Hongwen Zhu,[2,12] Liangqing Dong,[1,12] Weiwei Shi,[3,12] Ran Chen,[4,5,12] Zhijian Song,[3] Chen Huang,[6] Junqiang Li,[3] Xiaowei Dong,[3] Yanting Zhou,[2] Qian Liu,[2,5] Lijie Ma,[1] Xiaoying Wang,[1] Jian Zhou,[1,7] Yansheng Liu,[8] Emily Boja,[9] Ana I. Robles,[9] Weiping Ma,[10] Pei Wang,[10] Yize Li,[11] Li Ding,[11] Bo Wen,[6] Bing Zhang,[6] Henry Rodriguez,[9] Daming Gao,[4,5,*] Hu Zhou,[2,5,*] and Jia Fan[1,7,13,*]

[1]Department of Liver Surgery and Transplantation, Liver Cancer Institute, Zhongshan Hospital, Fudan University, and Key Laboratory of Carcinogenesis and Cancer Invasion of Ministry of Education, 180 Fenglin Road, Shanghai 200032, China

[2]Department 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

[3]OrigiMed, Shanghai 201114, China

[4]State Key Laboratory of Cell Biology, CAS Center for Excellence in Molecular Cell Science, Shanghai Institute of Materia Medica, Chinese Academy of Sciences, 555 Zuchongzhi Road, Shanghai 201203, China

[5]University of Chinese Academy of Sciences, Number 19A Yuquan Road, Beijing 100049, China

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

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

[8]Department of Pharmacology, Cancer Biology Institute, Yale University School of Medicine, West Haven, CT 06516, USA

[9]Office of Cancer Clinical Proteomics Research, Center for Strategic Scientific Initiatives, National Cancer Institutel, NIH, Bethesda, MD 20892, USA

[10]Department of Genetics and Genomics Sciences, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA

[11]Department of Medicine, McDonnell Genome Institute, Siteman Cancer Center, Washington University, St. Louis, MO 63108, USA

[12]These authors contributed equally

[13]Lead Contact

*Correspondence: dgao@sibcb.ac.cn (D.G.), zhouhu@simm.ac.cn (H.Z.), fan.jia@zs-hospital.sh.cn (J.F.)

https://doi.org/10.1016/j.cell.2019.08.052

SUMMARY

We performed the first proteogenomic characterization of hepatitis B virus (HBV)-related hepatocellular carcinoma (HCC) using paired tumor and adjacent liver tissues from 159 patients. Integrated proteogenomic analyses revealed consistency and discordance among multi-omics, activation status of key signaling pathways, and liver-specific metabolic reprogramming in HBV-related HCC. Proteomic profiling identified three subgroups associated with clinical and molecular attributes including patient survival, tumor thrombus, genetic profile, and the liver-specific proteome. These proteomic subgroups have distinct features in metabolic reprogramming, microenvironment dysregulation, cell proliferation, and potential therapeutics. Two prognostic biomarkers, PYCR2 and ADH1A, related to proteomic subgrouping and involved in HCC metabolic reprogramming, were identified. *CTNNB1* and *TP53* mutation-associated signaling and metabolic profiles were revealed, among which mutated *CTNNB1*-associated ALDOA phosphorylation was validated to promote glycolysis and cell proliferation. Our study provides a valuable resource that significantly expands the knowledge of HBV-related HCC and may eventually benefit clinical practice.

INTRODUCTION

Liver cancer ranks the fourth leading cause of cancer-related death worldwide (Villanueva, 2019). Hepatocellular carcinoma (HCC) accounts for about 85%–90% of all primary liver malignancies, and the largest attributable causes are chronic infection by hepatitis B virus (HBV) and hepatitis C virus (HCV) (Sartorius et al., 2015), along with alcohol abuse and metabolic syndrome. Despite the success of direct-acting antiviral therapy on curing chronic HCV infection (Falade-Nwulia et al., 2017), current antiviral therapy could only reduce rather than eliminate HBV, which is estimated to affect 292,000,000 people globally (The Polaris Observatory Collaborators, 2018). Notably, HBV-related HCC accounts for about 85% of HCC cases in China (Prevention of Infection Related Cancer (PIRCA) Group, 2019), due to the high prevalence of HBV infection. Recent next-generation sequencing-based studies, including The Cancer Genome Atlas (TCGA) program, have uncovered the genetic landscape of HCC (Cancer Genome Atlas Research Network, 2017; Schulze et al., 2015; Totoki et al., 2014), revealing driver mutations in *TP53*, *CTNNB1*, *TERT* promoter, and other key gene loci. However, how genetic alterations drive cancer phenotypes in HBV-related HCC remains largely unknown.

Mass spectrometry (MS)-based proteomics can measure global protein abundance and post-translational modifications to provide additional biological insights, which may not be deciphered by genomic analysis alone. The combination of sequencing and MS provides a more comprehensive picture linking cancer "genotype" to "phenotype" through functional

proteomics and signaling networks (Zhang et al., 2019). As a partner of Clinical Proteomic Tumor Analysis Consortium (CPTAC) (Ellis et al., 2013; Mertins et al., 2016; Rudnick et al., 2016; Vasaikar et al., 2019; Zhang et al., 2014, 2016), we conducted a comprehensive proteogenomic analysis of HBV-related HCC from a Chinese cohort. Integrated analyses of genomic, transcriptomic, proteomic, and phosphoproteomic data from tumor and matched non-tumor liver tissues revealed the connection and discordance among multi-omics and alterations in key signaling and metabolic pathways. Proteomic clustering resulted in three distinct subgroups, which showed association with patient survival, personalized treatment, and HCC-specific features. Two prognostic proteins related to metabolic reprogramming (PYCR2 and ADH1A) were explored and depicted for associated multi-omics profiles. *CTNNB1* mutation-associated phosphorylation sites were identified on key metabolic enzymes including ALDOA, and the role of phospho-ALDOA in promoting metabolic reprogramming and cell proliferation was confirmed. Collectively, our study not only provides a high-quality proteogenomic resource of HBV-related HCC complementary to TCGA but also implicates promising prognostic and therapeutic significance and underlying regulatory mechanisms that may benefit clinical practice.

RESULTS

Comprehensive Proteogenomic Characterization of CHCC-HBV Samples

To obtain a comprehensive molecular understanding of Chinese HCC patients with HBV infection (CHCC-HBV), paired tumor and non-tumor liver tissues from 159 HCC patients were selected for proteogenomic analysis based on stringent criteria (Figures S1 and S2A-S2D; Table S1). Using whole-exome sequencing (WES) data, the tumor versus non-tumor liver comparison identified 10,235 mutated genes (about 64 genes per tumor) including 20,369 non-silent point mutations and 1,363 small insertions-deletions (Indels, Table S1). Blood samples were also available from 108 patients. The tumor versus blood comparison identified 14,103 somatic mutations, and 13,734 (96.6%) overlapped with tumor versus non-tumor comparison for the 108 patients having both non-tumor liver and blood samples (Figure S2C).

Isobaric tandem mass tags (TMT)-based global proteomics (Figures S2D–S2F) identified 10,783 proteins (encoded by 10,759 genes) with averagely 8,934 proteins per sample (Figure S2G). TMT-based phosphoproteomics identified 59,746 highly reliable phosphosites from 9,224 phosphoproteins with averagely 28,401 phosphosites per sample (Figure S2H). The MS data were of high quality as evaluated (Figures S2I–S2M). A total of 6,494 proteins (encoded by 6,478 genes, quantified across all 159 paired samples) and 26,418 phosphosites (quantified in at least half samples) were included in subsequent analyses (Table S1).

Pairing transcriptomic and proteomic data from the 159 patients created 6,203 mRNA-protein pairs (Table S1), which showed an overall positive correlation (median r = 0.54) with 90.3% (5,600/6,203) significant positive correlations (multiple-test adjusted $p < 0.01$, Figure S3A, top panel). Consistent with the previous studies (Mertins et al., 2016; Zhang et al., 2014, 2016), genes involved in metabolic processes had the strongest positive mRNA-protein correlations, while those involved in cell-cycle and mRNA processing had weaker correlations, indicating major post-transcriptional regulations (Figure S3A, bottom panel). Notably, discordance between mRNA and protein abundance was identified in 16 genes (NDUFS6, NDUFB9, NDUFB3, NDUFA12, NDUFS4, NDUFB4, NDUFC2, NDUFA7, NDUFS3, NDUFA2, NDUFB7, NDUFA13, NDUFS5, NDUFA11, NDUFB1, and MT-ND1). The 16 genes are mainly components of respiratory chain complex I, suggesting that the complex formation may attenuate genomic/transcriptomic variation-caused protein abundance change. Although co-expression network analysis largely identified the same two functional modules (metabolic pathway and tumor microenvironment) with transcriptomic and proteomic data, respectively (Figure S3B), the protein network increased prediction performance by above 10% compared with corresponding mRNA network for 37.5% (60/160) of the KEGG pathways, whereas the opposite trend was only observed for 6.3% (10/160) of the KEGG pathways (Figure S2M). These results affirmed the high quality of our proteomic data and the added value of proteomics for assessing gene functions.

Proteogenomic Landscape of CHCC-HBV

Among the 159 patients, five significantly mutated genes were identified (Figure 1A), including *TP53* (58%), *CTNNB1* (19%), *AXIN1* (18%), *KEAP1* (7%), and *RB1* (6%). Mutations of *AXIN1*, a negative regulator of WNT pathway, and *CTNNB1* were mutually exclusive (only 2 out of 56 cases with co-mutations), consistent with the previous report (Guichard et al., 2012). Mutation frequencies were relatively higher in the CHCC-HBV cohort than in the HBV-positive HCC subgroup from TCGA for several genes, including *AXIN1* (18% versus 8%), *TSC2* (7% versus 0%), *SMARCA2* (5% versus 0%), *ATRX* (5% versus 0%), and *KMT2C* (8% versus 0%), while mutation frequencies of *CTNNB1* (19% versus 35%), *ARID1A* (10% versus 16%), and *RB1* (6% versus 16%) were slightly lower (Figure 1B). Principal-component analysis (PCA) further identified a significant spatial separation of samples in our cohort and TCGA HCC samples with HCV infection but not HBV infection (Figures S4A and S4B). These results highlighted a potential impact of viral infection on the mutational signatures in hepatocarcinogenesis.

Then, a patient-specific database was constructed based on WES and RNA sequencing (RNA-seq) data (see STAR Methods). Single amino-acid variants were detected by searching tandem mass spectrometry (MS/MS) spectra against corresponding patient-specific database. Only a few proteomic variants (1,973, accounting for 1.75% of DNA and RNA variants) were confirmed by MS/MS at peptide level (Figure S4C; Table S1). Most of the peptide variants have been previously reported in dbSNP and COSMIC (Tate et al., 2019), while only 212 were new. In addition, 0.42% of novel junctions identified by RNA-seq were sparsely confirmed by proteomics (Figure S4D; Table S1).

Chinese herbal medicines containing aristolochic acids (AAs) were recently reported as a contributor to oncogenesis including HCC (Hoang et al., 2013; Kucab et al., 2019; Ng et al., 2017; Poon et al., 2013). As estimated up to 80% of the CHCC-HBV patients may have received Chinese herbal medicines for hepatitis

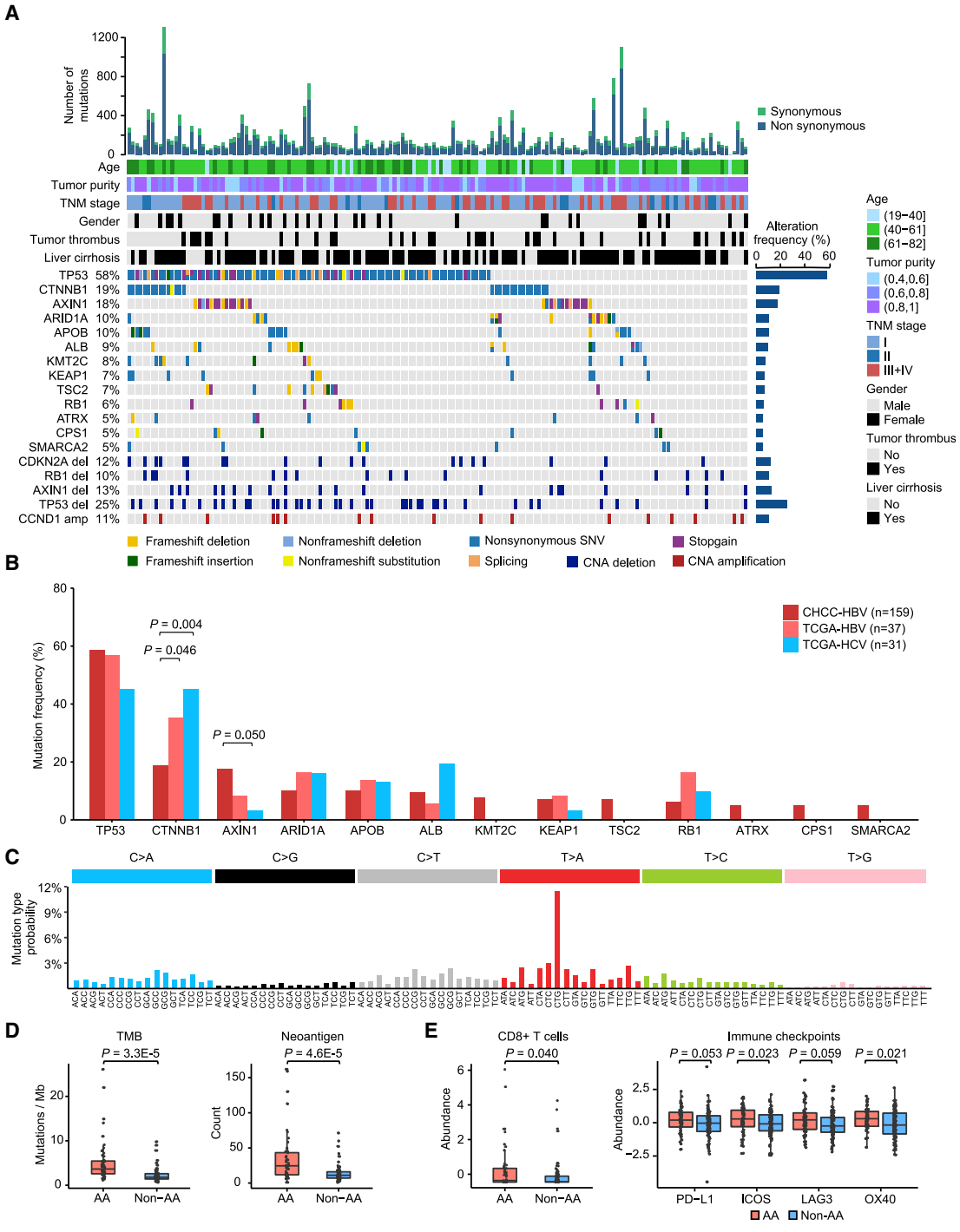


Figure 1. WES-Based Mutation Profile of the HBV-Related HCC Cohort

(A) Genetic profile and associated clinicopathologic features of all the 159 HCC patients.

(B) Comparisons of frequently mutated genes between CHCC-HBV cohort and TCGA HCC cohort (Fisher's exact test).

(C) AA signature mutations were identified in 56 of the 159 tumors. The relative mutation frequencies of all 96 tri-nucleotide mutation patterns are plotted with AA-like mutation patterns labeled in red.

(D) Comparisons of TMB and predicted neoantigens in tumors with and without AA signature (t test). The line and box represent median and upper and lower quartiles, respectively.

(E) Comparisons of $CD8^+$ T cells and immune regulatory molecules in tumors with and without AA signature (t test). The line and box represent median and upper and lower quartiles, respectively. The y axis is in $\log_2$ scale.

See also Figures S1 and S4 and Table S1.

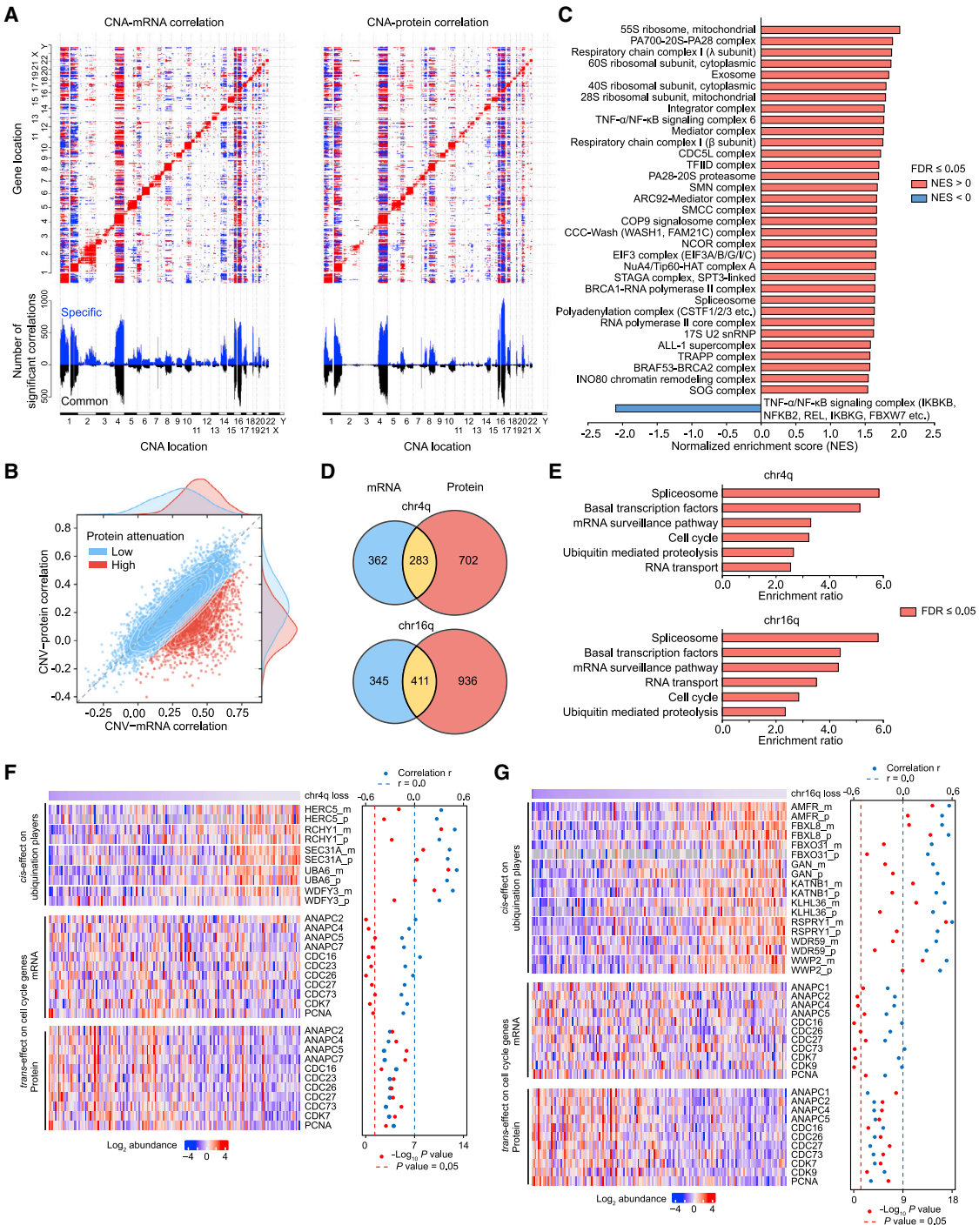


Figure 2. Effects of Copy-Number Alterations on mRNA and Protein Abundance

(A) Correlations of CNA (x axes) to mRNA (left) and protein (right) expression (y axes) with CNA *cis* and *trans* effects. Significant positive (red) and negative (blue) correlations (multiple-test adjusted $p < 0.01$, Spearman's correlation) between CNA and mRNA (left) or protein (right) are indicated in top panels. The numbers of mRNA and protein significantly associated with a particular CNA are presented as blue bars underneath the respective panels and those common to both are represented by black bars.

(B) Scatterplot of CNA correlation to mRNA and protein (Spearman's correlation). Each dot represents a transcript/protein. Attenuated proteins are represented in red using a Gaussian mixture model with two mixture components.

(C) GSEA analysis of the correlation differences between CNA-mRNA and CNA-protein (attenuation enrichment). NES, normalized enrichment score.

(D) Venn diagrams of mRNAs/proteins with negative CNA-mRNA and CNA-protein correlations in chromosomes 4q and 16q.

(legend continued on next page)

treatment (Zhang et al., 2010), WES data indeed identified a signature for AA exposure (AA signature), dominated by A:T > T:A transversions (Figure 1C). In total, 35.2% (56/159) of our patients harbored such AA signature (false discover rate [FDR] <0.05) (Table S1), with mutational bias toward non-transcribed strands (an approximate ratio of 2:1). Based on proteomic data, three peptides derived from this signature were identified (Figure S4E), suggesting that AA-related genomic alternations could indeed generate mutated proteins. Since AA could lead to DNA damage and somatic mutations, tumor mutational burden (TMB) in AA-signature-containing tumors (n = 56, median = 3.7 mutations/Mb) was found to be 2-fold higher than non-AA tumors (n = 103, median = 1.8 mutations/Mb) (p = 3.3E-5) (Figure 1D). Correspondingly, predicted neoantigen counts were above 2-fold higher in AA-signature-containing tumors (median = 24.5 neoantigens/tumor) than non-AA tumors (median = 11 neoantigens/tumor) (p = 4.6E-5) (Figure 1D). Increased TMB and neoantigen load indicated a potential benefit from an immunotherapy-like checkpoint blockade for these patients (Samstein et al., 2019). Indeed, tumors with an AA signature harbored significantly denser infiltrating $CD8^+$ T cells (p = 0.040), as well as higher expression of ICOS (p = 0.023), OX40 (p = 0.021), PD-L1 (p = 0.053), and LAG3 (p = 0.059) than those without (Figure 1E). Additionally, the AA signature negatively correlated with tumor thrombus (p = 0.002), serum albumin (ALB) level (p = 0.019), frameshift Indels (p = 8.4E-6), and Barcelona Clinic Liver Cancer (BCLC) stage (p = 0.005), while it positively correlated with splicing-site mutations (p = 0.004) (Figure S4F). However, this signature showed no significant association with patient prognosis. Overall, HBV-related HCC with AA signature represented a unique patient subgroup with special clinicopathologic and molecular features.

Effects of Copy-Number Alternations

Somatic copy-number alternations (SCNAs) based on WES data showed the most frequent gains in chromosomes 1q and 8q and losses in chromosomes 4q, 8p, 16p, 16q, and 17p (Figures S5A–S5C; Table S2), as previously described in HCC (Guichard et al., 2012; Wang et al., 2013). In addition, we identified amplifications in driver oncogenes including *CCND1* (11q13.3, 17 cases), *FGF19* (11q13.3, 17 cases), and *TERT* (5p15.33, 31 cases) (Figure S5A) and deletions of key tumor suppressors such as *AXIN1* (16p13.3, 19 cases), *CDKN2A/CDKN2B* (9p21.3, 19 cases), *RB1* (13q14.2, 19 cases), and *TP53* (17p13.1, 40 cases) (Figure S5C).

CNAs may affect mRNA, protein, and phosphoprotein abundance in either "*cis*" or "*trans*" modes, corresponding to the diagonal and vertical patterns in Figures 2A and S5D. The *cis* and *trans* associations between CNA-mRNA and CNA-protein were much more consistent in HCC than in breast cancer (Mertins et al., 2016) or colorectal cancer (Zhang et al., 2014). Quantitative analysis revealed strong phosphoprotein-level buffering of CNA-mRNA and CNA-protein associations. Specifically, *cis* association was observed for 48%, 46%, and 30% of the genes quantified at mRNA, protein, and phosphoprotein levels, respectively (Figure S5E). Comparing all the genes/proteins with both CNA-mRNA and CNA-protein correlations as previously described (Gonçalves et al. 2017) showed that mRNA abundance better correlated with CNA changes (median Spearman's r = 0.33) than protein abundance (median Spearman's r = 0.21). As shown in Figure 2B, 1,570 proteins were significantly attenuated, corresponding to 19.5% of all the 8,035 genes analyzed. These attenuated proteins were mainly enriched in large protein complexes, such as ribosome and spliceosome (Figure 2C). Possibly, protein complex assembly played an important role in post-translational modification and determining protein half-lives, resulting in decreased correlation between gene dosage and protein abundance.

Likewise, about 72% of the genes with CNAs exhibited significant *trans* effect on mRNA abundance of over 50 genes, whereas 44% and only 28% showed similar level of *trans* effect on protein and phosphoprotein abundance, respectively. The CNAs with *trans* effect were centered around chromosomes 1p, 1q, 4q and 16p, and 16q. Among them, 4q and 16q were predominantly co-deleted across the cohort (FDR <1E-6, Figures S5A–S5C), suggesting that they may co-regulate gene expression during hepatocarcinogenesis. Notably, both arms predominantly anticorrelated to the global proteomic abundance than transcriptomic abundance (Figure 2D; Table S2). Gene set enrichment analysis (GSEA) revealed that these 4q- and 16q-anticorrelated proteins were converged on the RNA-related (transcription, splicing, RNA transport), cell-cycle, and ubiquitin-mediated proteolysis pathways (Figure 2E). Many cell-cycle and key signaling regulators are known to be strictly regulated by the ubiquitin-proteasome system (UPS). We thus assumed that loss of UPS players in the two arms may contribute to global protein expression change and tumorigenesis. Indeed, 4q and 16q loss showed *cis* effect on many UPS genes and *trans* effect on cell-cycle master regulators (e.g., anaphase promoting complex [APC] components), whose protein expression, but not mRNA expression, significantly and inversely correlated with copy numbers of 4q and 16q (Figures 2F and 2G). Altogether, loss of 4q and 16q likely contributed to the global protein expression alteration and tumor progression via possible *cis* and *trans* effects.

Protein Abundance-Based Clustering of CHCC-HBV Tumors

Genomic and transcriptomic information have been used to cluster HCC into subgroups (Chaisaingmongkol et al., 2017; Chiang et al., 2008; Coulouarn et al., 2008; Hoshida et al., 2009; Lachenmayer et al., 2012; Lee et al., 2004). Proteomic data reflect gene function better than transcriptomic data (Wang et al., 2017), as exemplified by a recent study finding few or no overlapping genes among 30 sets of mRNA signatures from 25 studies in HCC (Cai et al., 2017). We then performed unsupervised clustering based on proteins differentially expressed

(E) Enrichment analysis of proteins with negative CNA-protein correlation in chromosomes 4q and 16q.
(F and G) Heatmaps of copy-number loss of chromosomes 4q (F) and 16q (G) and the mRNA(m)/protein(p) abundance of UPS and cell-cycle-related proteins.The Spearman's correlation coefficient between CNA and mRNA/protein is calculated with p values displayed in log10 scale. m, mRNA; p, protein.
See also Figures S3 and S5 and Table S2.

between tumor and non-tumor liver (see STAR Methods) and identified three subgroups among the 159 tumors (Figures 3A, S6A, and S6B; Table S3). Subgroup 1 was characterized by the highest level of metabolism-related proteins and liver function retention, such as ACAT1, ADH1A, G6PC, and PGM1 (denoted as metabolism subgroup, S-Mb). Subgroup 3 was featured by the increase in proliferative proteins, such as PARP1, TOP2A, PCNA, and MKI-67 (denoted as proliferation subgroup, S-Pf). Subgroup 2, with an intermediate expression of metabolic and proliferative proteins, predominantly downregulated immune, inflammatory, and stromal proteins, such as CD4, CD8A, S100A12, SPARC, and ITGB3 (denoted as microenvironment dysregulated subgroup, S-Me). Among the frequently mutated genes, *RB1* and *TSC2* exhibited significant enrichment in S-Pf and S-Me ($p = 0.004$ and 0.042, respectively, Fisher's exact test). Clinicopathologic factors such as larger tumor size ($p = 0.010$), tumor thrombus ($p = 2.86E-05$), and advanced TNM stages ($p = 0.022$) were more prominent in S-Pf versus other two subgroups (Fisher's exact test). CNA-based genome instability index (CGI) significantly differed among the three subgroups (median 2.71, 5.19, 4.12 for S-Mb, S-Me, S-Pf respectively; $p = 7.34E-8$, Kruskal-Wallis rank-sum test), with the most stable in S-Mb (Figure 3A). Of note, clustering our proteomic data with the transcriptomic signatures from previous studies (Chiang et al., 2008; Coulouarn et al., 2008; Hoshida et al., 2009; Lachenmayer et al., 2012; Lee et al., 2004) also resulted in a similar 3-subgroup allocation (Figure S6C), supporting the reliable subgrouping procedure in our study.

The proteomic subgroups significantly differed in survival ($p = 9.9E-06$, Figure 3B) and were authenticated as an independent prognosticator on multivariable analysis (hazard ratio [HR], 2.041; 95% confidence interval [CI], 1.425–2.922; $p = 9.8E-05$) (Table S4), after controlling for serum alpha-fetoprotein (AFP), tumor size, tumor thrombus, BCLC, and TNM stage. Despite that transcriptomic clustering also generated 3 subgroups with survival difference ($p = 1.4E-06$; Figure S6D) and correlated with proteomic clustering (Figure 3A), there was obvious discordance of patient allocation. Comparison of these discordant patients ($n = 36$) showed that proteomic subgroups (HR = 2.194) stratified these patients better than transcriptomic subgroups (HR = 1.065), indicating the superiority of proteomic clustering (Figure S6E). After stratifying patients according to TNM stage (Figure 3C), proteomic subgroups still strongly correlated with patient prognosis regardless of tumor stages, supporting the superior prognostic power of molecular features within our proteomic clustering. Clustering of TCGA HCC mRNA data with our proteomic signature also resulted in 3 subgroups with similar survival difference as ours (Figure 3D). Considering that tumor thrombus is one of the most important clinicopathologic features in HCC, we compared proteomic profiles between tumors with or without thrombus, revealing 82 differentially expressed proteins that regulated metabolic reprogramming, peroxisome, and liver function (Figures 3E and 3F). The results further supported a crucial role for dysregulated metabolism in HCC.

For the phosphoproteomic data, a total of 859 differentially expressed phosphoproteins were identified (Table S3). Pathway-based phosphoproteomic data also clustered tumors into three subgroups consistent with the proteomic subgroups (Figures S6F and S6G), with a concordance as high as 0.41. Likewise, the three phosphoproteomic subgroups differed in survival by both univariable ($p = 0.005$, Figure S6H) and multivariable analyses ($p = 0.092$, Table S4).

We attempted to identify individual genes associated with the proteomic subgroups across multi-omics data. We first compiled 9 HCC relevant genes based on literature. As showed in Figure S7A, AFP, PKM, RB1, SPP1 (osteopontin), CDK1, and CHEK2 were upregulated in S-Pf versus S-Mb and S-Me, with mRNA-protein correlation above 0.64. Contrarily, CTNNB1, CPS1, and GLYATL1 appeared downregulated in S-Pf. Then, focusing on clinically relevant HCC protein biomarkers, GPC3, CD90 (THY1), and GOLM1 were found significantly upregulated, while FUCA1 (AFU), CD44, EPCAM, and GLUL were in fact significantly downregulated, in tumors compared to non-tumor liver tissues, and correlated with the proteomic subgroups (Figure S7B). Drug target analysis based on drugBank (Wishart et al., 2018) showed distinct druggable target enrichment in each proteomic subgroup, indicating potential personalized therapies (Figure S7C). Most of the druggable targets were enriched in S-Pf and S-Me, such as proliferative or invasive regulators including TOP1/TOP2A/TOP2B, CDK1/CDK2, and MMP14. Meanwhile, microenvironment-associated proteins ATOX1 and STIP1 were enriched in S-Me, and metabolic regulators COQ8B and PLA2G2A were predominant in S-Mb. Since S-Mb had the highest TMB ($p = 0.028$) and neoantigen load ($p = 0.090$) compared to S-Pf and S-Me (Kruskal-Wallis test; Figure 3A), such patients may benefit more from immune checkpoint blockade (Samstein et al., 2019). In summary, these genes relevant or specific to HCC significantly correlated with the proteomic subgroups, further highlighting their clinical implications.

Identification and Validation of Prognostic Biomarkers

We performed supervised analysis to identify robust and representative prognostic proteins (Figure 4A). Three upregulated proteins and 39 downregulated proteins that mainly converged on amino acid metabolism and oxidoreductase activity were identified after stringent filtering (Table S5). PYCR2 (amino acid metabolism) and ADH1A (oxidoreductase activity) were selected and showed significantly differential expression across the proteomic subgroups (Figure 4B). Stratification of patients for survival differences using median as the cutoff was significant for PYCR2 and ADH1A, respectively (Figure 4C), which were confirmed by multivariable analyses (PYCR2 high versus low: HR, 1.792; 95% CI, 1.002–3.206; $p = 0.049$. ADH1A low versus high: HR, 2.703; 95% CI, 1.465–5.000; $p = 0.001$) (Table S4). Immunostaining on tissue microarrays from the current cohort of 155 cases (leaving 4 cases without qualified tissue blocks) validated the relative protein abundance and prognostic value of PYCR2 and ADH1A measured by MS/MS (Figures 4C and 4D). In an independent cohort of 243 HCC cases (Table S6), immunostaining of PYCR2 and ADH1A also significantly correlated with patient survival, further indicating their robust prognostic value for potential clinical application (Figure 4E).

PYCR2 is a crucial enzyme in proline biosynthesis, which was shown as the most significantly altered amino acid metabolism by metabolomics profiling in HCC (Tang et al., 2018). ADH1A belongs to the enzyme family that metabolizes a wide variety of

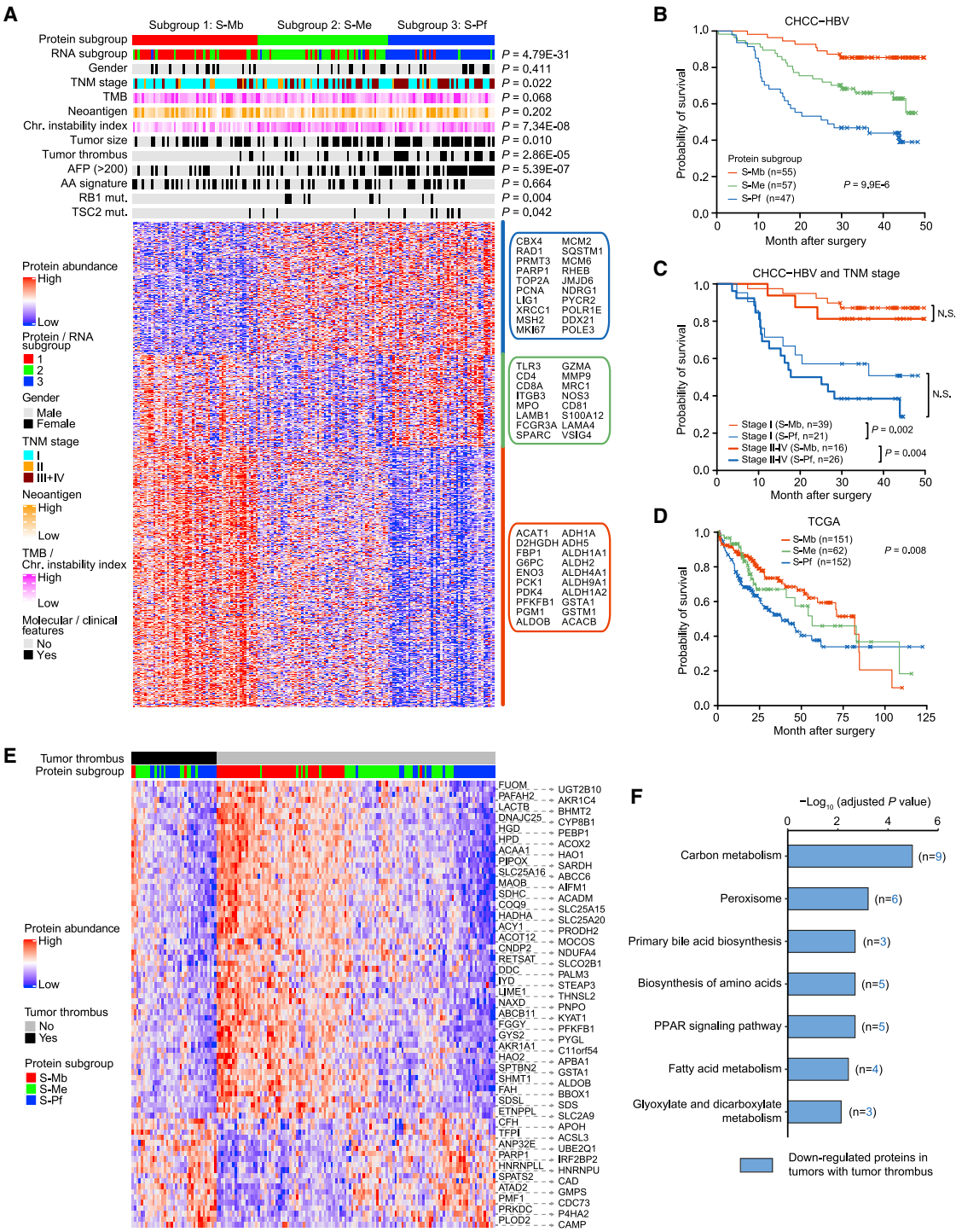


Figure 3. Proteomic Stratification of HBV-Related HCC and Their Clinicopathologic Correlations

(A) Patient subgrouping based on differentially expressed proteins (n = 1,274) between tumor and non-tumor tissues. Each column represents a patient sample and rows indicate proteins. Color of each cell shows *Z*-score (log_2 of relative abundance scaled by proteins' SD) of the protein in that sample.

(B) Kaplan-Meier curves for overall survival based on proteomic subgroups (log-rank test).

(C) Kaplan-Meier curves for overall survival of proteomic subgroups 1 (S-Mb) and 3 (S-Pf) at different TNM stages (stage I versus II–IV) (log-rank test).

(D) Survival difference of TCGA HCC cohort based on our proteomic subgrouping signature (log-rank test).

(E and F) The heatmap (E) and enriched pathways (F) of significantly differential expressed proteins (FDR q value < 0.05, t test) in tumors with or without tumor thrombus.

See also Figures S6 and S7 and Tables S3 and S4.

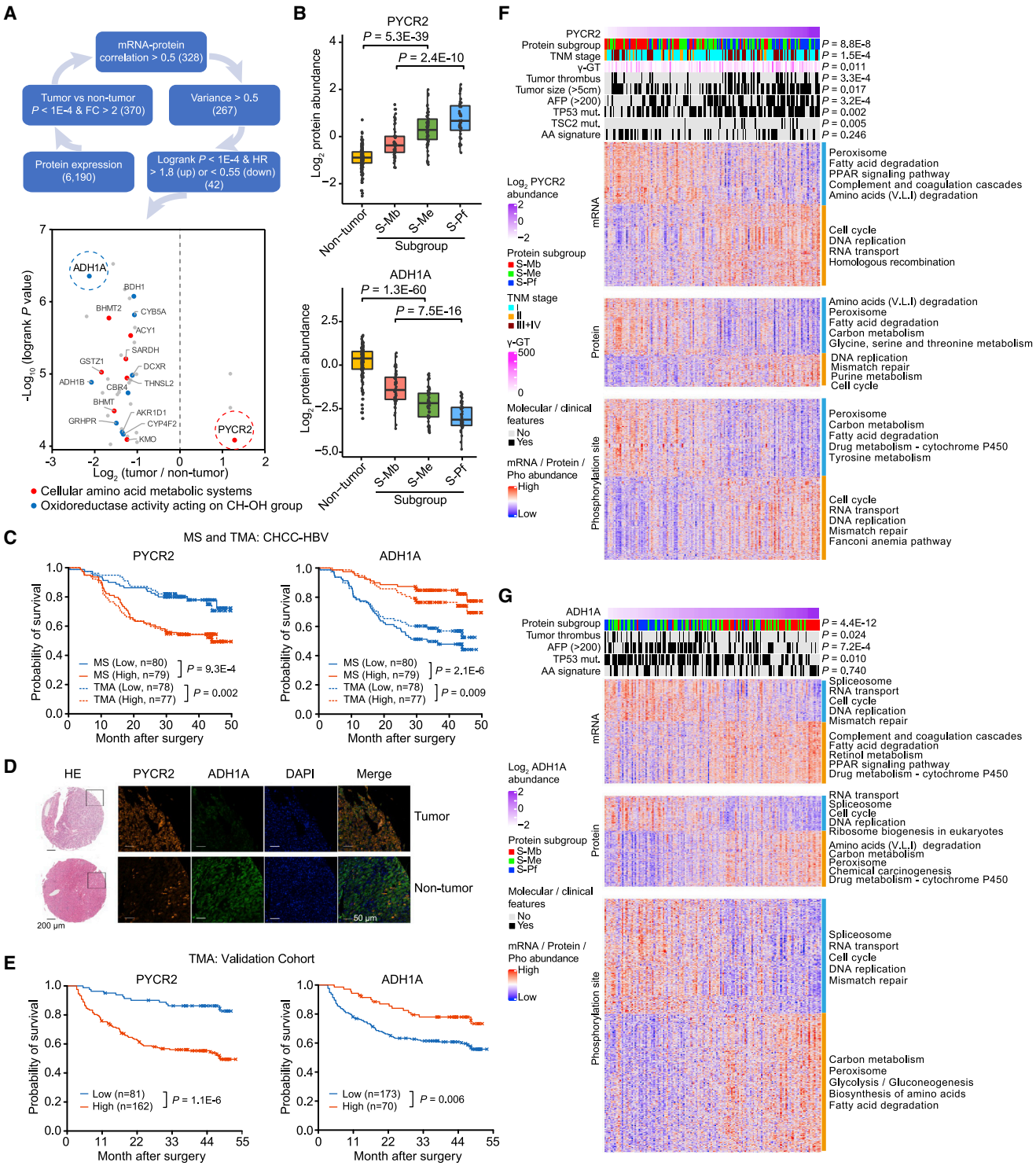


Figure 4. Identification and Validation of Proteomic Prognostic Biomarkers

(A) Workflow for selecting prognostic proteins with dot showing the 42 candidate proteins. FC, fold change; HR, harzad ratio.

(B) Relative abundance of PYCR2 and ADH1A between tumor and non-tumor tissues (t test) as well as across proteomic subgroups (ANOVA test). The line and box represent median and upper and lower quartiles, respectively.

(C) Kaplan-Meier curves for overall survival based on proteomic abundance (n = 159; solid lines) or immunostaining scores (n = 155; dotted lines) of PYCR2 and ADH1A (log-rank test).

(legend continued on next page)

xenobiotic compounds, including alcohol, retinol, aliphatic alcohols, hydroxysteroids, and lipid peroxidation products, which is a classic liver function (Molotkov et al., 2002a, 2002b). Data from the Human Protein Atlas (https://www.proteinatlas.org/) showed that ADH1A is largely a liver-specific protein. Thus, downregulation of ADH1A may contribute to dysregulated xenobiotic metabolism and facilitate HCC development.

Consistent with its prognostic value, patients with high PYCR2 were enriched in S-Me and S-Pf (p = 8.8E-08) and characterized by harboring tumor thrombus (p = 3.3E-04), large tumor (p = 0.017), more *TP53* (p = 0.002) or *TSC2* (p = 4.7E-03) mutations, and advanced TNM stages (p = 1.5E-04) (Fisher's exact test) (Figure 4F). Tumors with high PYCR2 showed specific downregulation of pathways relevant to metabolism and peroxisome and upregulated ones involving DNA replication, RNA transport and mismatch repair (Figure 4F; Table S5). Similarly, patients with high ADH1A featured fewer *TP53* mutations (p = 0.01), lower AFP level (p = 7.2E-04), and absence of tumor thrombus (p = 0.024) (Fisher's exact test) (Figure 4G). Tumors with high ADH1A showed specific elevation in pathways involving metabolism, peroxisome and liver function, and downregulation in pathways related to spliceosome, RNA transport, DNA replication, and cell cycle (Figure 4G; Table S5). Taken together, the two HCC-enriched prognostic biomarkers, displaying opposite regulatory directions, were consistently associated with key clinicopathologic features and biological pathways.

Dissection of HBV Proteins and Liver-Specific Proteome

Since all our patients were HBV positive, we investigated the clinical and biological relevance of HBV-related factors, including viral proteins and HBV receptor. Large envelope protein (S), external core antigen/capsid protein (E/C), and polymerase protein (P) were detected in both proteomic and RNA-seq data, while X protein (X) was only detected in RNA-seq data (Figures 5A and 5B). Although no significant associations between HBV proteins and patient survival were observed, there were less abundant HBV proteins and mRNAs (except for mRNA of protein C) in tumors than non-tumor liver tissues (Figures 5A and 5B). We also detected significantly lower protein and mRNA levels of HBV receptor SLC10A1 (also known as NCTP) (Yan et al., 2012) in tumors than non-tumor liver tissues (Figure 5C). Unlike HBV proteins, its decreased expression was significantly associated with S-Me and S-Pf (Figure 5C) as well as reduced survival (p = 2.1E-5; Figure 5E). Immunostaining confirmed the association between higher SLC10A1 and better prognosis in the current cohort (p = 0.010; Figures 5D and 5E), and in another independent HCC cohort (n = 243, p = 0.007; Figure 5F).

Since SLC10A1 is a liver-specific protein with primary function as a bile acid co-transporter (Hagenbuch and Meier, 1994), a general intrinsic correlation may exist between dysregulated bile acid metabolism and HBV-related HCC. Indeed, integrated analysis revealed the dramatic downregulation of most key proteins in bile acid metabolism (Russell, 2003), particularly in S-Me and S-Pf (Figure 5G). The results could be explained by the fact that bile acid metabolism and HBV transcription/replication are largely controlled by a same suite of transcriptional factors including FXR, RXR, and several others (Bar-Yishay et al., 2011). The impaired function of such transcriptional factors would contribute to the inhibition of bile acid metabolism and decrease HBV gene expression simultaneously (Figure 5H), a unique molecular feature of HBV-related HCC.

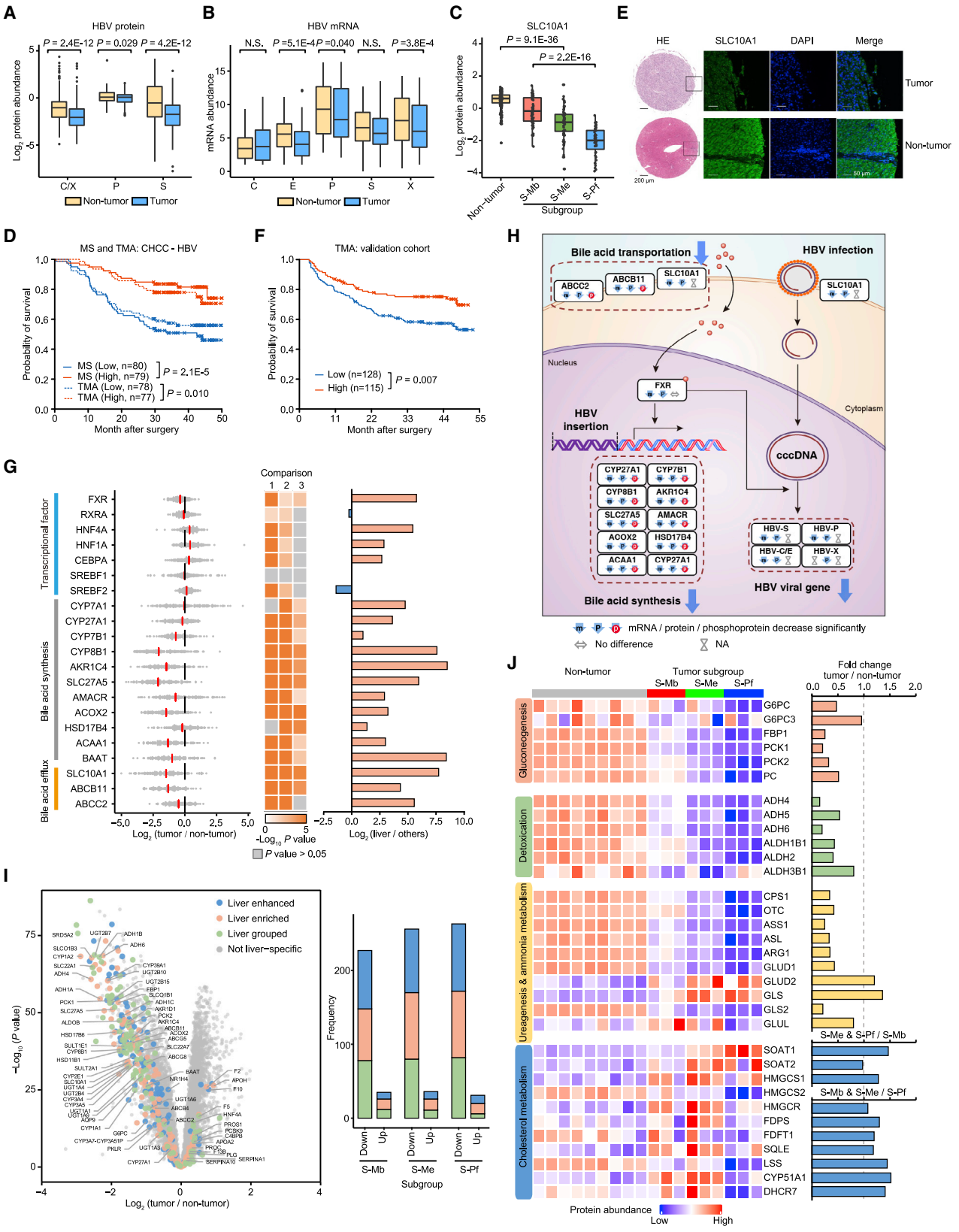
In addition, 80.3% (290/361) of the detected liver-specific proteins were downregulated in tumors (Figure 5I), and most proteins in liver-specific metabolic pathways such as gluconeogenesis, detoxication, and ureagenesis-ammonia were significantly attenuated in tumors. However, key enzymes in cholesterol metabolism (SOAT1, SOAT2, HMGCR, etc.) and ammonia/glutamine metabolism (GLS and GLUD2) were upregulated in tumors. SOAT1/SOAT2 were mainly upregulated in S-Me and S-Pf, suggesting that synthetic metabolism of fatty acid-cholesterol esters was enhanced in more proliferative tumors (Figure 5J). GLS/GLUD2 also showed higher expression in S-Me and S-Pf, indicating that glutamine metabolism was possibly more active in such tumors to meet their energetic demand (Figure 5J). Collectively, these data indicated a global reprogramming of liver-specific metabolism in HBV-related HCC.

Proteogenomic Analysis of Cellular Metabolic and Signaling Pathways

To obtain a general insight into dysregulation of cellular metabolic and signaling pathways, we integrated multi-omics data across all 159 cases (Figure 6A; Table S7). Significant upregulation of key enzymes of the glycolysis pathway (HK2, ALDOA, PKM2, etc.) was observed in tumors, indicating enhanced demands for glucose metabolism in HCC. However, there were no unified alterations for the TCA cycle-oxidative phosphorylation (OX-PHOS) system within mitochondria. Increased expression and phosphorylation of enzymes such as ACLY, ACSL3, and ACSL4 supported an overall activation of lipid biosynthesis in HCC. Dramatic downregulation of most key enzymes in the cholesterol-bile acid metabolic pathway was observed at transcriptomic, proteomic, and phosphoproteomic levels, indicating an impaired liver-specific metabolic function in HCC cells (Figures 5J and 6A).

Aberrant activation of WNT, Hippo-YAP, mTOR, and transforming growth factor β (TGF-β) pathways has been observed in HCC (Giannelli et al., 2014; Perugorria et al., 2019; Sohn et al., 2016; Villanueva et al., 2008). However, no general elevation of protein abundance or phosphorylation of WNT, Hippo-Yap, and mTOR pathways was observed across the whole cohort. General activation of TGF-β pathway was implied by increased SMAD2/3 phosphorylation in tumors, which was coherent with ADH1A downregulation, considering that TGF-β/SMAD2/3 could directly inhibit ADH1A expression (Ciuclan et al., 2010). Consistent with the common features found in

(D) Representative multiplex immunostaining images of PYCR2 and ADH1A on tumor and paired non-tumor liver tissues.
(E) Kaplan-Meier curves for overall survival based on immunostaining scores of PYCR2 and ADH1A in an independent HCC cohort (n = 243) (log-rank test).
(F and G) Associations of PYCR2 (F) and ADH1A (G) expression with proteomic subgroups, clinicopathologic factors, and multi-omics profiles.
See also Tables S5 and S6.



(legend on next page)

S-Me and S-Pf (Figure 3A), general upregulation/activation of key cell-cycle regulators was observed in tumors (Figure 6A).

Subgroup-specific pathway enrichment analysis clearly demonstrated distinct molecular features among the three proteomic subgroups (Figure 6B). S-Mb was high in TCA cycle, xenobiotic metabolism, fatty acid-lipid metabolism, and others, indicating tumors in this subgroup may be driven by elevated metabolic processes. S-Pf was more likely driven by proliferative signaling including WNT, Notch, Myc, and Hedgehog, and therefore S-Pf harbored the highest cell renewal pathways such as cell cycle, transcription, splicing, and ubiquitin-proteolysis. There were few pathways (Hippo, Circadian clock, and tight junctions) differentially activated in S-Me, which resembled a partial mixture of S-Mb and S-Pf with intermediate activation in certain metabolic and signaling pathways. In keeping with Figure 3, the results implied that HBV-related HCC included subgroups of tumors driven by distinct molecular pathways, thus providing unique therapeutic opportunities as depicted in Figure S7C.

To dig into HCC molecular features related to specific driver mutations, association of *TP53* and *CTNNB1* mutations with proteomic and phosphoproteomic data was explored. In addition to several metabolic pathways, proteins implicated in cell-cycle/DNA damage repair pathways were specifically enriched in *TP53*-mutated tumors (Figure 6C), supporting the classic role of *TP53* in cell-cycle regulation and guarding genome stability. Unexpectedly, phosphorylated peptides enriched in *TP53*-mutated tumors were much diversified and did not give a clear focused pattern (Figure 6D). Proteins enriched in *CTNNB1*-mutated tumors were associated with various metabolic processes including drug metabolism, glycolysis/gluconeogenesis, and amino acid metabolism (Figure 6E). Phosphorylation of key metabolic enzymes including ALDOA and ENO1 was upregulated in *CTNNB1*-mutated tumors, suggesting that their phosphorylation may contribute to metabolic reprogramming in *CTNNB1*-mutated tumors (Figure 6F). Therefore, *CTNNB1* activation may impact on metabolic reprogramming of HCC cells at both translational and post-translational levels.

CTNNB1 Mutation-Associated ALDOA Phosphorylation Promotes Glycolytic Metabolism and Cell Proliferation in HCC

Although *CTNNB1* mutations in HCC were reported to be related to glycolytic metabolism (Beyoğlu et al., 2013), the exact mechanisms remain undefined. Our proteogenomic analysis showed distinct metabolic alterations in tumors with *CTNNB1* mutations. Specifically, ALDOA Ser36 phosphorylation was significantly higher, despite slightly lower protein abundance of ALDOA, in *CTNNB1*-mutated tumors than wild-type tumors ($p = 0.015$) (Figures 6F, 7A, and 7B). Thus, ALDOA Ser36 phosphorylation (Figure 7C) may modulate glycolytic metabolism in *CTNNB1*-mutated HCC. Indeed, HepG2 cells ectopically expressing ALDOA (ALDOA-WT or ALDOA-S36E mutant) showed more potent glycolytic metabolism and proliferation than control cells, with the strongest effect in ALDOA-S36E-expressing cells (Figures 7D–7F), supporting an effective role of S36 phosphorylation in promoting ALDOA function and cell growth. ALDOA-S36E-expressing cells formed larger tumors than ALDOA-WT and control cells in xenograft models, further implying a strong tumor-promoting effect of ALDOA-S36 phosphorylation (Figures 7G and 7H).

We further constructed HepG2 cells expressing ΔN-CTNNB1 (mimicking a naturally occurring *CTNNB1* mutant) and depleted endogenous ALDOA (Figures 7I and 7J). Knockdown of ALDOA led to a more dramatic anti-proliferation effect in ΔN-CTNNB1-expressing cells than in control cells (Figures 7K and 7L), indicating the importance of ALDOA activity for *CTNNB1*-mutated HCC cells to support their growth. Altogether, phosphorylation of glycolytic enzymes including ALDOA may drive metabolic reprogramming and proliferation in *CTNNB1*-mutated HCC (Figure 7M).

DISCUSSION

Comprehensive genomic analysis of HCC has broadened our knowledge of the molecular events relevant to this fatal malignancy (Cancer Genome Atlas Research Network, 2017; Schulze et al., 2015; Totoki et al., 2014). Herein, global proteomic and phosphoproteomic data provided new insights into the clinical, biological, and therapeutic understanding of HCC. Although potential heterogeneous features within each tumor sample may be covered by the sample preparation process, integrated proteogenomic characterization of paired tumor and adjacent liver samples revealed the activation status of key signaling pathways, liver-specific metabolic reprogramming, clinically and therapeutically relevant subgroups, and HBV-specific features in HBV-related HCC.

Our integrated analysis revealed alterations of metabolic pathways among the most dramatic differences between tumor and

Figure 5. HBV Receptor and Bile Acid Metabolism Are Downregulated in HBV-Related HCC

(A) Expression of HBV viral proteins (E/C, n = 135; P, n = 50; S, n = 119, t test).

(B) Expression of HBV viral mRNAs (C, n = 14; E, n = 72; P, n = 155; S, n = 69; X, n = 153, t test).

(C) Expression of SLC10A1 protein is decreased in tumors (t test) and associated with proteomic subgroups (ANOVA test). The line and box represent median and upper and lower quartiles, respectively.

(D) Kaplan-Meier curves for overall survival based on proteomic abundance (n = 159; solid lines) or immunostaining scores of SLC10A1 in the current cohort (n = 155; dotted lines) (log-rank test).

(E) Representative immunostaining images of SLC10A1 expression on tumor and paired non-tumor liver tissues.

(F) Kaplan-Meier curves for overall survival based on immunostaining scores of SLC10A1 in an independent HCC cohort (n = 243) (log-rank test).

(G) Integrated analysis of key transcriptional factors, enzymes, and transporters in bile acid metabolism. The abundance of indicated proteins in tumor are plotted as compared with non-tumor liver. Heatmap represents the p values of comparisons between tumor/non-tumor (1), proteomic subgroups (2), and of prognostic significance by log-rank test (3). The bar plot indicates the liver enrichment of these proteins among all other organs based on the data from Human Protein Atlas.

(H) Diagram showing the multi-omics profiles of co-regulators of HBV and bile acid metabolism.

(I) Differential expression of the liver-specific proteins (list from The Human Protein Atlas) among the three proteomic subgroups.

(J) Heatmap and quantitative analysis of differentially expressed proteins in the liver-specific functions. The color of each cell represents average protein abundance.

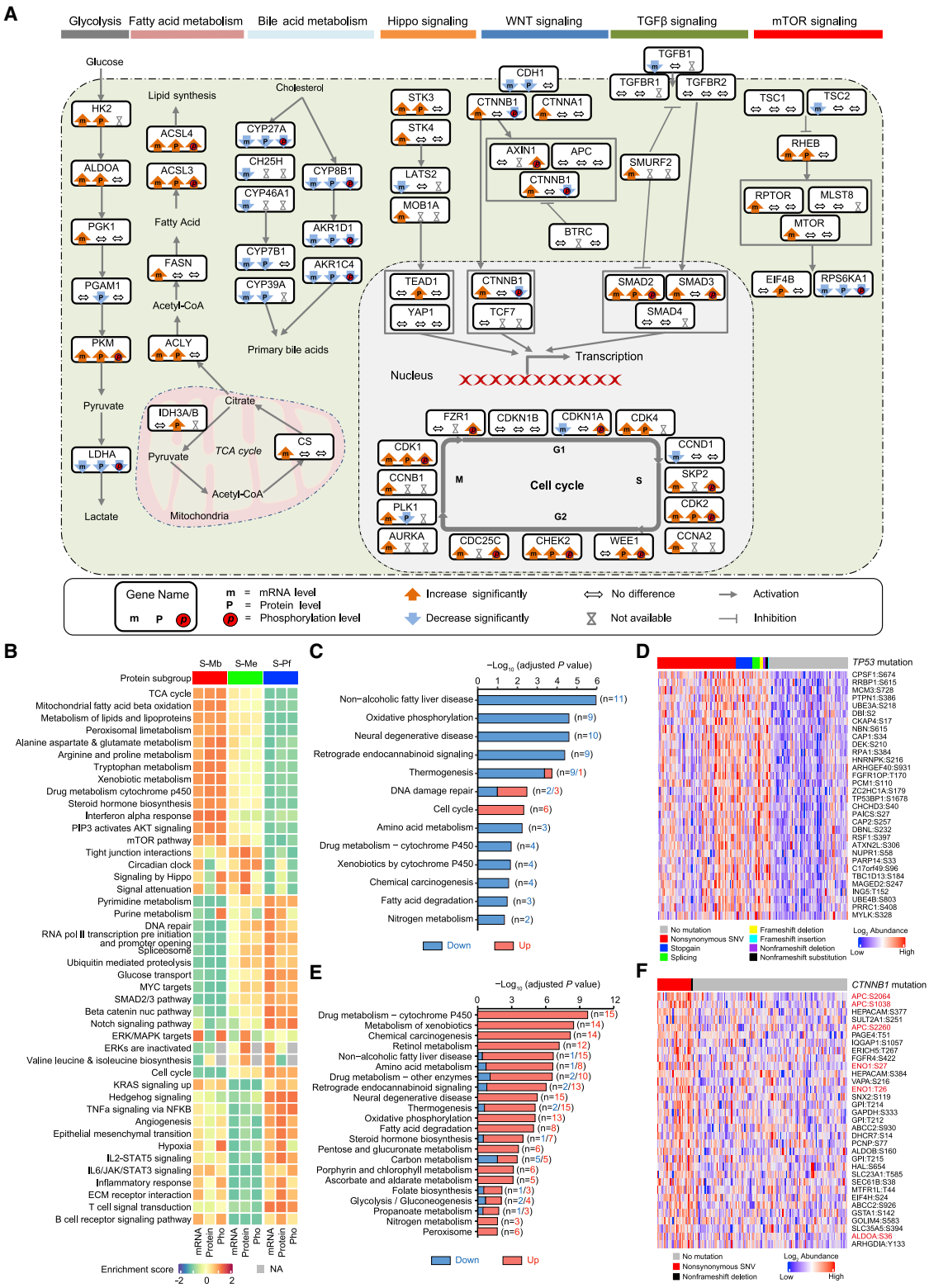


Figure 6. Metabolism and Signaling Pathways Are Altered in HBV-Related HCC

(A) Overview of metabolism and signaling pathways based on integrated proteogenomic analysis. The mRNA, protein, and phosphorylation abundance of tumors are indicated in comparison with non-tumor liver tissues.

(legend continued on next page)

non-tumor liver tissues. Together with metabolic alterations, proliferation and microenvironmental dysregulation stratified patients into three distinct subgroups. Surprisingly, differentially expressed proteins in tumors with or without thrombus were mainly cellular metabolic enzymes and regulators, indicating that metabolic reprogramming was associated with HCC aggressiveness. Notably, two metabolic enzymes, PYCR2 and ADH1A, were identified and validated as potential prognostic biomarkers. It has been reported that PYCR2 is upregulated in various cancer types and may drive cancer growth and metastasis (Ding et al., 2017; Liu et al., 2015; Sun et al., 2019). A previous study also revealed that loss of PYCR2 could lead to oxidative-stress-triggered apoptosis and result in microcephaly syndrome in an animal model (Nakayama et al., 2015). Therefore, high PYCR2 expression may enable tumor cells to cope with the excessive oxidative species derived from enhanced metabolism, representing a gain of function of tumor-specific metabolism. ADH1A is an enzyme involved in metabolizing various xenobiotic substrates (Molotkov et al., 2002a, 2002b). Downregulation of ADH1A may promote the transition from liver damage to hepatocarcinogenesis and enhance HCC progression on exposure to xenobiotic compounds. Moreover, ADH1A-low tumors may favor secondary metabolic pathways to promote proliferation, as they showed specific activation in DNA replication and cell-cycle pathways.

We found that both HBV proteins and HBV receptor (SLC10A1) were downregulated in HCC. SLC10A1 is a functional transporter that reabsorbs bile acid into hepatocytes. Decreased SLC10A1 expression is consistent with the overall downregulation of bile acid and liver-specific metabolism in HCC. These phenomena could be explained by the notion that hepatocarcinogenesis is partially a de-differentiation from functional hepatocytes to tumor cells (Kullak-Ublick et al., 1997), which involves loss of liver-specific function and metabolic reprogramming. Consistently, different and subgroup-specific activation of cellular metabolism and signaling pathways were observed in our cohort, indicating that HBV-related HCC are molecularly diversified.

So far, there is no evidence indicating that AA-signature mutations could be translated and detected at the protein level. Here, we identified 56 cases with AA signature and detected 3 mutant peptides encoded by AA-signature gene mutations, indicating that herbal medicine could indeed result in mutated protein products. Considering the association of AA signature with multiple clinicopathologic and molecular features (Figure S4F), it may have complex effects on HBV-related HCC. AA signature was associated with high TMB and neoantigen load, which may predict better responses to immunotherapy. Such an assumption still needs further investigation using experimental models or in clinical practice. Nevertheless, AA-containing herbs should be discouraged for clinical use, due to their HCC-promoting potential.

Taken together, our current work provides a comprehensive and integrated analysis of CHCC-HBV using multiple proteogenomic platforms. We revealed that metabolic alterations are possibly the most important factor that is associated with advanced disease stage and poor clinical outcome. Targeting cancer metabolism may provide a promising avenue to develop effective therapies for HCC. Our study not only generated a high-quality data resource that may benefit basic research but also provided additional biological insights underlying clinical features of HCC.

STAR★METHODS

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

- KEY RESOURCES TABLE
- LEAD CONTACT AND MATERIALS AVAILABILITY
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Clinical Sample Acquisition
  - Cell Line
- METHOD DETAILS
  - Proteogenomic Workflow
  - DNA/RNA Extraction, WES and RNA-seq
  - Peptides preparation for MS analysis
  - Nano-LC-MS/MS
  - Database Searching of MS Data
  - WES Data Analysis
  - RNA-seq Data analysis
  - Proteome and Phosphoproteome Data Analysis
  - mRNA, proteomic and phosphoproteomic subgrouping analysis
  - Multi-omics Data Analysis
  - Prognostic Biomarker Analysis for CHCC-HBV
  - Tissue MicroArray (TMA) Experiment
  - Drug Target Analysis
  - Functional Experiments
- QUANTIFICATION AND STATISTICAL ANALYSIS
- DATA AND CODE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.cell.2019.08.052.

ACKNOWLEDGMENTS

This work was supported by the Strategic Priority Research Program of the Chinese Academy of Sciences (grants XDA12010202, XDA12030203,

(B) Integrated analysis of differentially activated metabolic and signaling pathways at mRNA, protein, and phosphoprotein levels among the three proteomic subgroups.
(C) Pathway enrichment analysis based on differentially expressed proteins that associated with *TP53* mutations in tumors.
(D) Heatmap represents protein phosphorylation sites that associated with *TP53* mutations in tumors.
(E) Pathway enrichment analysis based on differentially expressed proteins that associated with *CTNNB1* mutations in tumors.
(F) Heatmap represents top protein phosphorylation sites that associated with *CTNNB1* mutations in tumors.
See also Tables S7.

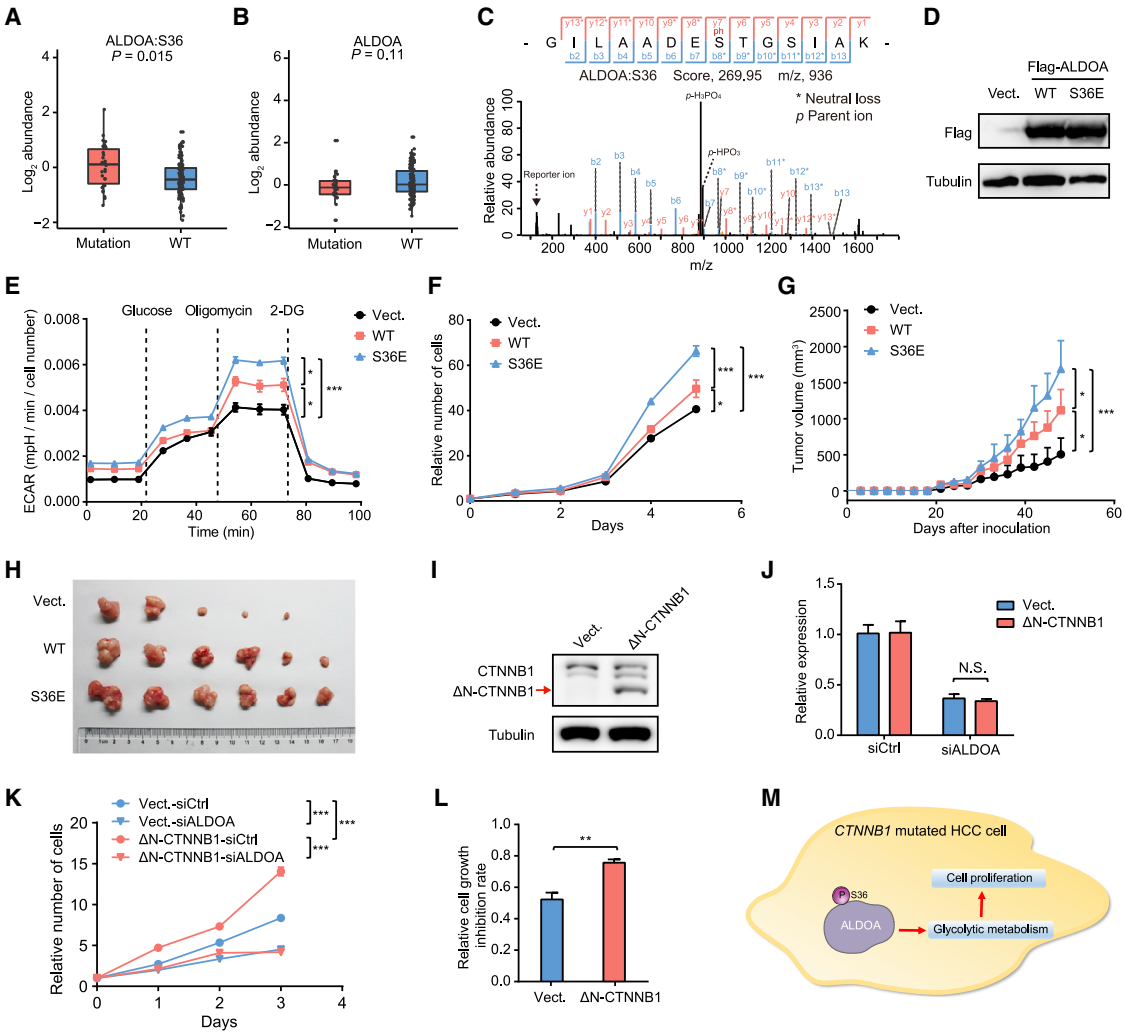


Figure 7. ALDOA-Ser36 Phosphorylation Drives HCC Cell Glycolysis and Proliferation

(A) ALDOA-Ser36 phosphorylation was upregulated in tumors with *CTNNB1* mutation (t test).

(B) Protein abundance of ALDOA was slightly lower in *CTNNB1*-mutated tumors (t test).

(C) Representative spectrum for peptide-containing S36 phosphorylation.

(D) HepG2 cells ectopically expressing FLAG-ALDOA-WT or FLAG-ALDOA-S36E mutants were generated and confirmed by western blot.

(E) Extracellular acidification rate (ECAR) was measured to determine glycolytic metabolism of indicated cells (two-way ANOVA followed by Tukey's multiple comparisons test). Data are represented as mean ± SEM, *p < 0.05; ***p < 0.001.

(F) Proliferation of indicated cells was measured by CCK8 method (two-way ANOVA followed by Tukey's multiple comparisons test). Data are represented as mean ± SEM, *p < 0.05; ***p < 0.001.

(G and H) Tumor growth curves (G) and xenograft tumor pictures (H) of indicated HepG2 cells subcutaneously injected into nude mice (two-way ANOVA followed by Tukey's multiple comparisons test). Data are represented as mean ± SEM, *p < 0.05; ***p < 0.001.

(I) HepG2 cells ectopically expressing FLAG-tag N-terminal deletion CTNNB1 (ΔN- CTNNB1) were generated and confirmed by western blot.

(J) Relative ALDOA mRNA levels in indicated HepG2 cells after ALDOA small interfering RNA (siRNA) transfection (t test). Data are represented as mean ± SEM.

(K) Proliferation of indicated HepG2 cells after ALDOA siRNA transfection (two-way ANOVA followed by Tukey's multiple comparisons test). Data are represented as mean ± SEM, *p < 0.05; ***p < 0.001.

(L) Relative cell growth inhibition rate of indicated HepG2 cells on the 3rd day after ALDOA siRNA transfection (t test). Data are represented as mean ± SEM.

(M) A brief model depicting functional impact of ALDOA phosphorylation in *CTNNB1*-mutated HCC cells.

XDA12020364, and XDB19020203); the National Key Research and Development Program from the Ministry of Science and Technology of China (grants 2017YFC1700200, 2015CB964502); National Natural Science Foundation of China (grants 91859105, 81790253, and 91853130); Basic Research Project from Technology Commission of Shanghai Municipality (grant 17JC1402200); and National Science and Technology Major Project (grant 2017ZX10203208-004). We thank Shisheng Wang (West China Hospital, Sichuan University) for help in spectrum filtering in variant peptide identification and suggestion on data analysis. We thank Dayun Lu and Han He (Shanghai Institute of Materia Medica, Chinese Academy of Sciences) for help in proteomic sample preparation. We thank Mr. Peng Li and Ms. Kuai Liu (Origimed) for the assistance of bioinformatics analysis. 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 U.S. National Cancer Institute's Office of Cancer Clinical Proteomics Research (Clinical Proteomic Tumor Analysis Consortium [CPTAC]). 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., H. Zhu, and L. Dong; Formal Analysis, H. Zhu, W.S., Z.S., C.H., B.W., W.M., Y. Li, Y. Liu, E.B., A.I.R., P.W., L. Ding, and B.Z.; Investigation, H. Zhu, L. Dong, R.C., L.M., X.W., Y.Z., and Q.L.; Resources, J.F., H. Zhou, D.G., Q.G., J.Z., J.L., and X.D.; Data Curation, Q.G. and H. Zhu; Writing, Q.G., H. Zhu, W.S., D.G., H. Zhou, and J.F.; Visualization, H. Zhu; Supervision, J.F., H. Zhou, D.G., Q.G., and H.R.; Funding Acquisition, J.F., H. Zhou, D.G., and Q.G.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: February 10, 2019
Revised: June 2, 2019
Accepted: August 26, 2019
Published: October 3, 2019

REFERENCES

Aran, D., Hu, Z., and Butte, A.J. (2017). xCell: digitally portraying the tissue cellular heterogeneity landscape. Genome Biol. *18*, 220.

Auton, A., Brooks, L.D., Durbin, R.M., Garrison, E.P., Kang, H.M., Korbel, J.O., Marchini, J.L., McCarthy, S., McVean, G.A., and Abecasis, G.R.; 1000 Genomes Project Consortium (2015). A global reference for human genetic variation. Nature *526*, 68–74.

Bambury, R.M., Bhatt, A.S., Riester, M., Pedamallu, C.S., Duke, F., Bellmunt, J., Stack, E.C., Werner, L., Park, R., Iyer, G., et al. (2015). DNA copy number analysis of metastatic urothelial carcinoma with comparison to primary tumors. BMC Cancer *15*, 242.

Bar-Yishay, I., Shaul, Y., and Shlomai, A. (2011). Hepatocyte metabolic signalling pathways and regulation of hepatitis B virus expression. Liver Int. *31*, 282–290.

Beyoğlu, D., Imbeaud, S., Maurhofer, O., Bioulac-Sage, P., Zucman-Rossi, J., Dufour, J.F., and Idle, J.R. (2013). Tissue metabolomics of hepatocellular carcinoma: tumor energy metabolism and the role of transcriptomic classification. Hepatology *58*, 229–238.

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.

Cai, J., Li, B., Zhu, Y., Fang, X., Zhu, M., Wang, M., Liu, S., Jiang, X., Zheng, J., Zhang, X., and Chen, P. (2017). Prognostic biomarker identification through integrating the gene signatures of hepatocellular carcinoma Properties. EBioMedicine *19*, 18–30.

Cancer Genome Atlas Research Network (2017). Comprehensive and integrative genomic characterization of hepatocellular carcinoma. Cell *169*, 1327–1341.e23.

Chaisaingmongkol, J., Budhu, A., Dang, H., Rabibhadana, S., Pupacdi, B., Kwon, S.M., Forgues, M., Pomyen, Y., Bhudhisawasdi, V., Lertprasertsuke, N., et al.; TIGER-LC Consortium (2017). Common molecular subtypes among Asian hepatocellular carcinoma and cholangiocarcinoma. Cancer Cell *32*, 57–70.e3.

Chalmers, Z.R., Connelly, C.F., Fabrizio, D., Gay, L., Ali, S.M., Ennis, R., Schrock, A., Campbell, B., Shlien, A., Chmielecki, J., et al. (2017). Analysis of 100,000 human cancer genomes reveals the landscape of tumor mutational burden. Genome Med. *9*, 34.

Chiang, D.Y., Villanueva, A., Hoshida, Y., Peix, J., Newell, P., Minguez, B., LeBlanc, A.C., Donovan, D.J., Thung, S.N., Solé, M., et al. (2008). Focal gains of VEGFA and molecular classification of hepatocellular carcinoma. Cancer Res. *68*, 6779–6788.

Ciuclan, L., Ehnert, S., Ilkavets, I., Weng, H.L., Gaitantzi, H., Tsukamoto, H., Ueberham, E., Meindl-Beinker, N.M., Singer, M.V., Breitkopf, K., and Dooley, S. (2010). TGF-beta enhances alcohol dependent hepatocyte damage via down-regulation of alcohol dehydrogenase I. J. Hepatol. *52*, 407–416.

Clark, M.J., Chen, R., Lam, H.Y., Karczewski, K.J., Chen, R., Euskirchen, G., Butte, A.J., and Snyder, M. (2011). Performance comparison of exome DNA sequencing technologies. Nat. Biotechnol. *29*, 908–914.

Coulouarn, C., Factor, V.M., and Thorgeirsson, S.S. (2008). Transforming growth factor-beta gene expression signature in mouse hepatocytes predicts clinical outcome in human cancer. Hepatology *47*, 2059–2067.

Cox, J., and Mann, M. (2008). MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat. Biotechnol. *26*, 1367–1372.

Díaz-Gay, M., Vila-Casadesús, M., Franch-Expósito, S., Hernández-Illán, E., Lozano, J.J., and Castellví-Bel, S. (2018). Mutational Signatures in Cancer (MuSiCa): a web application to implement mutational signatures analysis in cancer samples. BMC Bioinformatics *19*, 224.

Ding, J., Kuo, M.L., Su, L., Xue, L., Luh, F., Zhang, H., Wang, J., Lin, T.G., Zhang, K., Chu, P., et al. (2017). Human mitochondrial pyrroline-5-carboxylate reductase 1 promotes invasiveness and impacts survival in breast cancers. Carcinogenesis *38*, 519–531.

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.

Ellis, M.J., Gillette, M., Carr, S.A., Paulovich, A.G., Smith, R.D., Rodland, K.K., Townsend, R.R., Kinsinger, C., Mesri, M., Rodriguez, H., and Liebler, D.C.; Clinical Proteomic Tumor Analysis Consortium (CPTAC) (2013). Connecting genomic alterations to cancer biology with proteomics: the NCI Clinical Proteomic Tumor Analysis Consortium. Cancer Discov. *3*, 1108–1112.

Falade-Nwulia, O., Suarez-Cuervo, C., Nelson, D.R., Fried, M.W., Segal, J.B., and Sulkowski, M.S. (2017). Oral direct-acting agent therapy for hepatitis C virus infection: a systematic review. Ann. Intern. Med. *166*, 637–648.

Gao, Q., Zhao, Y.J., Wang, X.Y., Qiu, S.J., Shi, Y.H., Sun, J., Yi, Y., Shi, J.Y., Shi, G.M., Ding, Z.B., et al. (2012). CXCR6 upregulation contributes to a proinflammatory tumor microenvironment that drives metastasis and poor patient outcomes in hepatocellular carcinoma. Cancer Res. *72*, 3546–3556.

Giannelli, G., Villa, E., and Lahn, M. (2014). Transforming growth factor-β as a therapeutic target in hepatocellular carcinoma. Cancer Res. *74*, 1890–1894.

Gonçalves, E., Fragoulis, A., Garcia-Alonso, L., Cramer, T., Saez-Rodriguez, J., and Beltrao, P. (2017). Widespread post-transcriptional attenuation of genomic copy-number variation in cancer. Cell Syst. *5*, 386–398.e4.

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

Hagenbuch, B., and Meier, P.J. (1994). Molecular cloning, chromosomal localization, and functional characterization of a human liver Na+/bile acid cotransporter. J. Clin. Invest. *93*, 1326–1331.

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

Hoang, M.L., Chen, C.H., Sidorenko, V.S., He, J., Dickman, K.G., Yun, B.H., Moriya, M., Niknafs, N., Douville, C., Karchin, R., et al. (2013). Mutational signature of aristolochic acid exposure as revealed by whole-exome sequencing. Sci. Transl. Med. *5*, 197ra102.

Hoshida, Y., Nijman, S.M., Kobayashi, M., Chan, J.A., Brunet, J.P., Chiang, D.Y., Villanueva, A., Newell, P., Ikeda, K., Hashimoto, M., et al. (2009). Integrative transcriptome analysis reveals common molecular subclasses of human hepatocellular carcinoma. Cancer Res. *69*, 7385–7392.

Kucab, J.E., Zou, X., Morganella, S., Joel, M., Nanda, A.S., Nagy, E., Gomez, C., Degasperi, A., Harris, R., Jackson, S.P., et al. (2019). A compendium of mutational signatures of environmental agents. Cell *177*, 821–836.e16.

Kullak-Ublick, G.A., Glasa, J., Böker, C., Oswald, M., Grützner, U., Hagenbuch, B., Stieger, B., Meier, P.J., Beuers, U., Kramer, W., et al. (1997). Chlorambucil-taurocholate is transported by bile acid carriers expressed in human hepatocellular carcinomas. Gastroenterology *113*, 1295–1305.

Lachenmayer, A., Alsinet, C., Savic, R., Cabellos, L., Toffanin, S., Hoshida, Y., Villanueva, A., Minguez, B., Newell, P., Tsai, H.W., et al. (2012). Wnt-pathway activation in two molecular classes of hepatocellular carcinoma and experimental modulation by sorafenib. Clin. Cancer Res. *18*, 4997–5007.

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.

Lee, J.S., Chu, I.S., Heo, J., Calvisi, D.F., Sun, Z., Roskams, T., Durnez, A., Demetris, A.J., and Thorgeirsson, S.S. (2004). Classification and prediction of survival in hepatocellular carcinoma by gene expression profiling. Hepatology *40*, 667–676.

Li, B., and Dewey, C.N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics *12*, 323.

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

Li, J., and Tibshirani, R. (2013). Finding consistent patterns: a nonparametric approach for identifying differential expression in RNA-Seq data. Stat. Methods Med. Res. *22*, 519–536.

Liu, W., Hancock, C.N., Fischer, J.W., Harman, M., and Phang, J.M. (2015). Proline biosynthesis augments tumor cell growth and aerobic glycolysis: involvement of pyridine nucleotides. Sci. Rep. *5*, 17206.

Mermel, C.H., Schumacher, S.E., Hill, B., Meyerson, M.L., Beroukhim, 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.

Mertins, P., Mani, D.R., Ruggles, K.V., Gillette, M.A., Clauser, K.R., Wang, P., Wang, X., Qiao, J.W., Cao, S., Petralia, F., et al.; NCI CPTAC (2016). Proteogenomics connects somatic mutations to signalling in breast cancer. Nature *534*, 55–62.

Molotkov, A., Deltour, L., Foglio, M.H., Cuenca, A.E., and Duester, G. (2002a). Distinct retinoid metabolic functions for alcohol dehydrogenase genes Adh1 and Adh4 in protection against vitamin A toxicity or deficiency revealed in double null mutant mice. J. Biol. Chem. *277*, 13804–13811.

Molotkov, A., Fan, X., and Duester, G. (2002b). Excessive vitamin A toxicity in mice genetically deficient in either alcohol dehydrogenase Adh1 or Adh3. Eur. J. Biochem. *269*, 2607–2612.

Nakayama, T., Al-Maawali, A., El-Quessny, M., Rajab, A., Khalil, S., Stoler, J.M., Tan, W.H., Nasir, R., Schmitz-Abe, K., Hill, R.S., et al. (2015). Mutations in PYCR2, encoding pyrroline-5-carboxylate reductase 2, cause microcephaly and hypomyelination. Am. J. Hum. Genet. *96*, 709–719.

Ng, A.W.T., Poon, S.L., Huang, M.N., Lim, J.Q., Boot, A., Yu, W., Suzuki, Y., Thangaraju, S., Ng, C.C.Y., Tan, P., et al. (2017). Aristolochic acids and their derivatives are widely implicated in liver cancers in Taiwan and throughout Asia. Sci. Transl. Med. *9*, eaan6446.

Nielsen, M., and Andreatta, M. (2016). NetMHCpan-3.0; improved prediction of binding to MHC class I molecules integrating information from multiple receptor and peptide length datasets. Genome Med. *8*, 33.

Perugorria, M.J., Olaizola, P., Labiano, I., Esparza-Baquer, A., Marzioni, M., Marin, J.J.G., Bujanda, L., and Banales, J.M. (2019). Wnt-β-catenin signalling in liver development, health and disease. Nat. Rev. Gastroenterol. Hepatol. *16*, 121–136.

Polaris Observatory Collaborators (2018). Global prevalence, treatment, and prevention of hepatitis B virus infection in 2016: a modelling study. Lancet Gastroenterol. Hepatol. *3*, 383–403.

Poon, S.L., Pang, S.T., McPherson, J.R., Yu, W., Huang, K.K., Guan, P., Weng, W.H., Siew, E.Y., Liu, Y., Heng, H.L., et al. (2013). Genome-wide mutational signatures of aristolochic acid and its application as a screening tool. Sci. Transl. Med. *5*, 197ra101.

Prevention of Infection Related Cancer (PIRCA) Group, Specialized Committee of Cancer Prevention and Control, Chinese Preventive Medicine Association; Non-communicable & Chronic Disease Control and Prevention Society, Chinese Preventive Medicine Association; Health Communication Society, Chinese Preventive Medicine Association (2019). Strategies of primary prevention of liver cancer in China: expert consensus (2018). Zhonghua Yu Fang Yi Xue Za Zhi *53*, 36–44.

Rudnick, P.A., Markey, S.P., Roth, J., Mirokhin, Y., Yan, X., Tchekhovskoi, D.V., Edwards, N.J., Thangudu, R.R., Ketchum, K.A., Kinsinger, C.R., et al. (2016). A description of the Clinical Proteomic Tumor Analysis Consortium (CPTAC) common data analysis pipeline. J. Proteome Res. *15*, 1023–1032.

Russell, D.W. (2003). The enzymes, regulation, and genetics of bile acid synthesis. Annu. Rev. Biochem. *72*, 137–174.

Samstein, R.M., Lee, C.H., Shoushtari, A.N., Hellmann, M.D., Shen, R., Janjigian, Y.Y., Barron, D.A., Zehir, A., Jordan, E.J., Omuro, A., et al. (2019). Tumor mutational load predicts survival after immunotherapy across multiple cancer types. Nat. Genet. *51*, 202–206.

Sartorius, K., Sartorius, B., Aldous, C., Govender, P.S., and Madiba, T.E. (2015). Global and country underestimation of hepatocellular carcinoma (HCC) in 2012 and its implications. Cancer Epidemiol. *39*, 284–290.

Schulze, K., Imbeaud, S., Letouzé, E., Alexandrov, L.B., Calderaro, J., Rebouissou, S., Couchy, G., Meiller, C., Shinde, J., Soysouvanh, F., et al. (2015). Exome sequencing of hepatocellular carcinomas identifies new mutational signatures and potential therapeutic targets. Nat. Genet. *47*, 505–511.

Sims, D., Sudbery, I., Ilott, N.E., Heger, A., and Ponting, C.P. (2014). Sequencing depth and coverage: key considerations in genomic analyses. Nat. Rev. Genet. *15*, 121–132.

Sohn, B.H., Shim, J.J., Kim, S.B., Jang, K.Y., Kim, S.M., Kim, J.H., Hwang, J.E., Jang, H.J., Lee, H.S., Kim, S.C., et al. (2016). Inactivation of Hippo pathway is significantly associated with poor prognosis in hepatocellular carcinoma. Clin. Cancer Res. *22*, 1256–1264.

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., and Mesirov, J.P. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. USA *102*, 15545–15550.

Sun, C., Li, T., Song, X., Huang, L., Zang, Q., Xu, J., Bi, N., Jiao, G., Hao, Y., Chen, Y., et al. (2019). Spatially resolved metabolomics to discover tumor-associated metabolic alterations. Proc. Natl. Acad. Sci. USA *116*, 52–57.

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.

Talevich, E., Shain, A.H., Botton, T., and Bastian, B.C. (2016). CNVkit: genome-wide copy number detection and visualization from targeted DNA sequencing. PLoS Comput. Biol. *12*, e1004873.

Tang, L., Zeng, J., Geng, P., Fang, C., Wang, Y., Sun, M., Wang, C., Wang, J., Yin, P., Hu, C., et al. (2018). Global metabolic profiling identifies a pivotal role of proline and hydroxyproline metabolism in supporting hypoxic response in hepatocellular carcinoma. Clin. Cancer Res. *24*, 474–485.

Tate, J.G., Bamford, S., Jubb, H.C., Sondka, Z., Beare, D.M., Bindal, N., Boutselakis, H., Cole, C.G., Creatore, C., Dawson, E., et al. (2019). COSMIC: the catalogue of somatic mutations in cancer. Nucleic Acids Res. *47* (D1), D941–D947.

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.

Totoki, Y., Tatsuno, K., Covington, K.R., Ueda, H., Creighton, C.J., Kato, M., Tsuji, S., Donehower, L.A., Slagle, B.L., Nakamura, H., et al. (2014).

Trans-ancestry mutational landscape of hepatocellular carcinoma genomes. Nat. Genet. *46*, 1267–1273.

Trapnell, C., Pachter, L., and Salzberg, S.L. (2009). TopHat: discovering splice junctions with RNA-Seq. Bioinformatics *25*, 1105–1111.

Tyanova, S., Temu, T., and Cox, J. (2016). The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. Nat. Protoc. *11*, 2301–2319.

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.; Clinical Proteomic Tumor Analysis Consortium (2019). Proteogenomic analysis of human colon cancer reveals new therapeutic opportunities. Cell *177*, 1035–1049.e19.

Villanueva, A. (2019). Hepatocellular Carcinoma. N. Engl. J. Med. *380*, 1450–1462.

Villanueva, A., Chiang, D.Y., Newell, P., Peix, J., Thung, S., Alsinet, C., Tovar, V., Roayaie, S., Minguez, B., Sole, M., et al. (2008). Pivotal role of mTOR signaling in hepatocellular carcinoma. Gastroenterology *135*, 1972–1983, 1983.e1–1983.e11.

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, K., Li, M., and Hakonarson, H. (2010). ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. *38*, e164.

Wang, K., Lim, H.Y., Shi, S., Lee, J., Deng, S., Xie, T., Zhu, Z., Wang, Y., Pocalyko, D., Yang, W.J., et al. (2013). Genomic landscape of copy number aberrations enables the identification of oncogenic drivers in hepatocellular carcinoma. Hepatology *58*, 706–717.

Wang, J., Ma, Z., Carr, S.A., Mertins, P., Zhang, H., Zhang, Z., Chan, D.W., Ellis, M.J., Townsend, R.R., Smith, R.D., et al. (2017). Proteome profiling outperforms transcriptome profiling for coexpression based gene function prediction. Mol. Cell. Proteomics *16*, 121–134.

Wilkerson, M.D., and Hayes, D.N. (2010). ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. Bioinformatics *26*, 1572–1573.

Wishart, D.S., Feunang, Y.D., Guo, A.C., Lo, E.J., Marcu, A., Grant, J.R., Sajed, T., Johnson, D., Li, C., Sayeeda, Z., et al. (2018). DrugBank 5.0: a major update to the DrugBank database for 2018. Nucleic Acids Res. *46* (D1), D1074–D1082.

Wiśniewski, J.R., Zougman, A., Nagaraj, N., and Mann, M. (2009). Universal sample preparation method for proteome analysis. Nat. Methods *6*, 359–362.

Xie, D.Y., Ren, Z.G., Zhou, J., Fan, J., and Gao, Q. (2017). Critical appraisal of Chinese 2017 guideline on the management of hepatocellular carcinoma. Hepatobiliary Surg. Nutr. *6*, 387–396.

Yan, H., Zhong, G., Xu, G., He, W., Jing, Z., Gao, Z., Huang, Y., Qi, Y., Peng, B., Wang, H., et al. (2012). Sodium taurocholate cotransporting polypeptide is a functional receptor for human hepatitis B and D virus. eLife *1*, e00049.

Zhang, L., Wang, G., Hou, W., Li, P., Dulin, A., and Bonkovsky, H.L. (2010). Contemporary clinical research of traditional Chinese medicines for chronic hepatitis B in China: an analytical review. Hepatology *51*, 690–698.

Zhang, B., Wang, J., Wang, X., Zhu, J., Liu, Q., Shi, Z., Chambers, M.C., Zimmerman, L.J., Shaddox, K.F., Kim, S., et al.; NCI CPTAC (2014). Proteogenomic characterization of human colon and rectal cancer. Nature *513*, 382–387.

Zhang, H., Liu, T., Zhang, Z., Payne, S.H., Zhang, B., McDermott, J.E., Zhou, J.Y., Petyuk, V.A., Chen, L., Ray, D., et al.; CPTAC Investigators (2016). Integrated proteogenomic characterization of human high-grade serous ovarian cancer. Cell *166*, 755–765.

Zhang, B., Whiteaker, J.R., Hoofnagle, A.N., Baird, G.S., Rodland, K.D., and Paulovich, A.G. (2019). Clinical potential of mass spectrometry-based proteogenomics. Nat. Rev. Clin. Oncol. *16*, 256–268.

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.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Rabbit monoclonal anti-ADH1A	Abcam	Cat# ab108203; RRID: AB_10891950
Rabbit polyclonal anti-PYCR2	Proteintech	Cat# 17146-1-AP; RRID: AB_2253344
Rabbit polyclonal anti-SLC10A1	Abcam	Cat# ab131084; RRID: AB_11155311
Rabbit polyclonal anti-CTNNB1	ABclonal	Cat# A11932; RRID: AB_2758875
Rabbit polyclonal anti-FLAG	Sigma Aoldrich	Cat# F7425; RRID: AB_439687
Mouse monoclonal anti-Tubulin	Santa Cruz Biotechnology	Cat# sc-134237; RRID: AB_2212295
Anti-Rabbit Immunoglobulins/HRP	Dako	Cat# P0217; RRID: AB_2728719
Anti-Mouse Immunoglobulins/HRP	Dako	Cat# P0260; RRID: AB_2636929
Biological Samples		
Paired tumor, adjacent non-tumor liver tissues and blood samples from a cohort of 316 HBV-related HCC patients	Zhongshan Hospital, Fudan University	This paper
Critical Commercial Assays		
QIAamp Fast DNA tissue kit	QIAGEN	Cat# 51404
QIAamp DNA blood Midi Kit	QIAGEN	Cat# 51185
SureSelect Human All Exon V6 kit	Agilent Technologies	Cat# 5190-8865
RNAlater Reagent	Invitrogen	Cat# AM7020
TRIzol Reagent	Invitrogen	Cat# 15596026
TMT10plex Isobaric Label Reagent	Thermo Scientific	Cat# 90111
TMT11-131C Label Reagent	Thermo Scientific	Cat# A34807
High-Select Fe-NTA kit	Thermo Scientific	Cat# A32992
Deposited Data		
Proteogenomic data of the CHCC-HBV cohort (n = 159)	This paper	NODE database: OEP000321
Database of the 1000 Genomes	Auton et al., 2015	PMID: 26432245
NHLBI Exome Sequencing Project	N/A	NHLBI Exome Sequencing Project (ESP); RRID: SCR_012761
Exome Aggregation Consortium (EXAC)	N/A	http://exac.broadinstitute.org/
Genome Aggregation Database (gnomAD)	N/A	Genome Aggregation Database; RRID: SCR_014964
TCGA-LIHC somatic MAF data	N/A	http://gdac.broadinstitute.org
HBV infection status data	Cancer Genome Atlas Research Network, 2017	PMID: 28622513
GENCODE (version 19)	N/A	https://www.gencodegenes.org
MSigDB c2 gene sets (version 6.2)	N/A	http://software.broadinstitute.org/gsea/msigdb/index.jsp
Drugbank database (version 5.1.1, released 2018-07-03)	N/A	https://www.drugbank.ca/
Experimental Models: Cell Lines		
HepG2, male origin	Prof. Lei Zhang (Shanghai Institute of Biochemistry and Cell Biology)	N/A
Recombinant DNA		
pLEX-MCS-CMV-puro	Addgene, USA	N/A

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Software and Algorithms		
MaxQuant 1.6.1.0	Cox and Mann, 2008	http://www.coxdocs.org/doku.php?id=maxquant:start
Burrows-Wheeler Aligner (BWA) (version 0.7.15)	Li and Durbin, 2009	http://bio-bwa.sourceforge.net/
Picard (version 2.0.1)	GitHub	http://broadinstitute.github.io/picard/
Genome Analysis Toolkit (GATK) (version 4.0.6.0)	Broad Institute	GATK; RRID: SCR_001876
Mutect (version 2)	Broad Institute	https://software.broadinstitute.org/gatk/
ANNOVAR (version 2017 Jul 17)	Wang et al., 2010	http://annovar.openbioinformatics.org/en/latest/
MutSigCV (version 1.4)	Lawrence et al., 2013	https://software.broadinstitute.org/cancer/cga/mutsig
Mutational Signatures in Cancer (MuSiCa)	Díaz-Gay et al., 2018	PMID: 29898651
Mutational signature activity (mSigAct) (version 0.9)	Ng et al., 2017	PMID: 29046434
OptiType (version 1.2.1)	Szolek et al., 2014	PMID: 25143287
NetMHCpan (version 3.0)	Nielsen and Andreatta, 2016	PMID: 27029192
xCell	Aran et al., 2017	PMID: 29141660
CNVkit (version 0.9.5)	Talevich et al., 2016	PMID: 27100738
STAR2 (version 2.4.2a)	Dobin et al., 2013	PMID: 23104886
RSEM (version 1.3.0)	Li and Dewey, 2011	PMID: 21816040
OmicsEV	GitHub	https://github.com/bzhanglab/OmicsEV/
Samr R package (version 2.0)	Li and Tibshirani, 2013	PMID: 22127579
ConsensusClusterPlus R package (version 1.42.0)	Wilkerson and Hayes, 2010	RRID: SCR_016954; PMID: 20427518
GSVA R package (version v1.26.0)	Hänzelmann et al., 2013	http://www.bioconductor.org/packages/release/bioc/html/GSVA.html
GSEA software (version 2-2.2.3)	Subramanian et al., 2005	http://software.broadinstitute.org/gsea/index.jsp
TopHat (version 2.1.1)	Trapnell et al., 2009	PMID:19289445
CustomProDB (version 1.20.2)	Wang and Zhang, 2013	PMID: 24058055
Vectra (version 3.0)	PerkinElmer	https://www.perkinelmer.com.cn/
InForm Cell Analysis (version 2.4)	PerkinElmer	https://www.perkinelmer.com.cn/
Others		
Tissue Microarrays (TMAs) of 155 paired tumor and non-tumor liver tissues from the CHCC-HBV cohort	Zhongshan Hospital, Fudan University	This paper
TMAs of an independent cohort of 243 HCC patients	Zhongshan Hospital, Fudan University	This paper

LEAD CONTACT AND MATERIALS AVAILABILITY

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

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Clinical Sample Acquisition

Paired tumor, adjacent non-tumor liver tissues and blood samples from a cohort of 316 HBV-related HCC patients were initially enrolled for the current Clinical Proteomic Tumor Analysis Consortium (CPTAC) project (designated as CHCC-HBV patients). All the patients underwent primary curative resection from June 2010 to December 2014 at Zhongshan Hospital and received no prior anticancer treatments. Tissue samples were collected within 30 min after operation and snap-frozen in liquid nitrogen. Peripheral blood samples were collected the day before surgery. Postoperative surveillance and treatment were conducted according to our consensus guideline as described previously (Xie et al., 2017; Zhou et al., 2018). Tumor differentiation was graded according to the Edmondson system. 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 Line

HepG2 cells were a kind gift from Prof. Lei Zhang (Shanghai Institute of Biochemistry and Cell Biology, China). HepG2 cells were cultured in DMEM with 10% FBS, 100 units of penicillin and 100 μg/mL streptomycin.

METHOD DETAILS

Proteogenomic Workflow

The proteogenomic analysis of the samples was performed according to the following procedures (Figure S1). For tumor cellularity analysis, the middle section of each tissue block (< 10 mm) was resected and subjected to hematoxylin and eosin (H&E) staining. The histological assessment of all tumor samples was accomplished by two experienced pathologists separately. To reduce the impact of intra-tumor heterogeneity on multi-omics analysis, the remaining liver tissue was pulverized using the CryoPrepTM CP02 (Covaris) and then divided into three parts. For each case, ~30 mg tissue sample was used for DNA extraction and whole exome sequencing (WES); ~200 mg tissue sample was immediately transferred into a 1.5 mL EP tube and then added 1 mL RNAlater reagent (Invitrogen) for RNA sequencing (RNA-seq); ~200 mg tissue sample was lysed with SDS lysis buffer (4% SDS, 100 mM Tris-HCl, 0.1 M DTT, pH 7.6) and kept in −80°C for the following proteomic and phosphoproteomic analyses.

According to CPTAC clinical sample collection procedures, the following criteria were used for sample selection on the 316 paired samples, 1) successful extraction of DNA from both tumor and adjacent non-tumor liver tissues for WES (159 pairs); 2) qualified RNA from both tumor and adjacent non-tumor liver tissues for RNA-seq (165 pairs); 3) qualified protein extraction from both tumor and adjacent non-tumor liver tissues as revealed from the SDS-PAGE image without obvious protein degradation and aberrant pattern (165 pairs for proteomic/phosphoproteomic analysis); 4) no tumor cells were observed in the adjacent non-tumor liver tissues. Finally, 159 high-quality paired samples were selected for the integrative proteogenomic analysis, and the average tumor cellularity of these samples was 81% (±14% SD, Table S1). The median age of these patients was 54, with 128 males and 31 females. In total, 91, 14 and 54 patients were classified as TNM stages I, II and III-IVA respectively. Detailed clinicopathologic features were summarized in Table S1. Among them, qualified blood samples from 108 patients were also available and used for WES as germline genomic reference.

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. Matched blood DNA was extracted using the QIAamp DNA blood Midi Kit (QIAGEN). DNA was quantified by the Qubit 3.0 (Invitrogen) and NanoDrop 2000 (Thermo Scientific) and the integrity was assessed by TapeStation (Agilent Technologies). WES libraries were prepared and captured using the SureSelect Human All Exon V6 kit (Agilent Technologies) following manufacturer's instructions. The DNA library with 150 bp paired-end reads was sequenced with Illumina HiSeq X Ten system. WES was conducted with a mean coverage depths of 187X (range: 108-344X) for tumor samples and 191X (range: 128-356X) for adjacent non-tumor liver samples (Figure S2A), consistent with the recommendations for WES (Clark et al., 2011; Sims et al., 2014).

Total RNA was extracted and purified from fresh frozen tissues using the Trizol reagent (Invitrogen). RNA integrity was measured on an Agilent 2100 Bioanalyzer (Agilent Technologies). Paired samples with high RNA integrity (RNA integrity number > 5), no contaminants and enough amount of RNA were used to prepare the transcriptome library. RNA was isolated using Sera-Mag oligo (dT) beads (Thermo Scientific) and fragmented with a NEB Fragmentation Reagents kit (NEB). The cDNA synthesis, end-repair, A-base addition, and ligation of the Illumina index adapters were performed according to Illumina's TruSeq RNA protocol (Illumina). Library quality was measured on an Agilent 2100 Bioanalyzer for product size and concentration. Paired-end libraries were sequenced by an Illumina HiSeq X Ten (2 × 150-nucleotide read length), with a sequence coverage of 40 million paired reads. For each tumor and its adjacent non-tumor sample, RNA-seq resulted in an average of 40.1 M and 39.1 M high-quality reads, respectively (Figure S2B). RNA-seq data analysis identified 19,860 protein-coding genes with an average of 18,839 genes per sample, covering the majority of all the genes in proteomic identification.

Peptides preparation for MS analysis

Protein Extraction and Digestion

For protein extraction, the SDS lysis buffer was added into the powdered tissues, and sonicated at 15% amplitude for 5 s on and 5 s off with the total working time of 2 min (JY92-IIDN, Ningblio Scientz Biotechnology Co., LTD, China). The proteins were then denatured and reduced at 95°C for 5 min. The insoluble debris was removed by centrifugation at 12,000 g for 10 min and the supernatant was retained for proteomic experiment. The protein concentration was determined using tryptophan-based fluorescence quantification method (Thakur et al., 2011).

Filter-aided sample preparation (FASP) procedure was used for protein digestion (Wiśniewski et al., 2009). Briefly, proteins were loaded in 10 kDa centrifugal filter tubes (Millipore), washed twice with 200 μL UA buffer (8 M urea in 0.1 M Tris-HCl, pH 8.5), alkylated with 50 mM iodoacetamide in 200 μL UA buffer for 30 min in the darkness, washed thrice with 100 μL UA buffer again and finally washed thrice with 100 μL 50 mM triethyl ammonium bicarbonate (TEAB). All above steps were centrifuged at 12,000 g at 25°C. Proteins were digested at 37°C for 18 hr with trypsin (Promega) at a concentration of 1:50 (w/w) in 50 mM TEAB. After digestion, peptides were eluted by centrifugation. The peptide concentration was determined by BCA protein quantification kit. For each sample, 400 μg peptides were prepared by vacuum centrifugation dryness for the following TMT labeling experiment.

For the "internal reference" mixed sample used in TMT labeling, 50 pairs of tumor and adjacent non-tumor liver samples were randomly selected and mixed in equal protein amount. The peptides of mixed samples were also prepared by FASP and divided into 400 μg per EP tube for each set of TMT labeling experiment as the internal reference.

TMT 11-plex Labeling

The isobaric labeling experiment was conducted according to the TMT kit instructions. For each set of TMT 11-plex labeling experiment, each channel was labeled with 400 μg peptides. The mixed peptides were labeled with channel 126 as the internal reference, and five pairs of tumor and adjacent non-tumor liver samples were labeled with the other ten channels (Tumor labeled with 127N, 128N, 129N, 130N and 131N; adjacent non-tumor liver tissue labeled with 127C, 128C, 129C, 130C and 131C). Two sets of TMT reagents (0.8 mg) were dissolved in anhydrous acetonitrile (41 μL × 2) and added to 400 μg peptides (dissolved in 200 μL 100 mM TEAB) to achieve a final acetonitrile concentration of approximately 30% (v/v). Following incubation for 1 hr at room temperature, 16 μL 5% hydroxylamine was added to the samples and incubated for 15 min to quench the labeling reaction. The labeled peptides were pooled and then subjected to vacuum centrifugation dryness and C18 solid-phase extraction desalting (3M Empore). The 165 pairs of tumor and adjacent non-tumor liver tissue samples were eventually labeled in 33 sets of TMT 11-plex experiments for the nanoLC-MS/MS analysis.

High-pH RPLC Fractionation

To increase the depth of protein identification, high-pH reverse phase liquid chromatography was used for peptide fractionation. A total of 4.4 mg TMT 11-plex labeled peptides were fractionated using a Waters XBridge BEH300 C18 column (250 × 4.6 mm, OD 5 μm) at a flow rate of 0.7 mL/min on Agilent 1100 LC instrument. Solvent A (10 mM NH_4COOH, adjusted to pH 10.0 with $NH_3 \cdot H_2O$) and a nonlinear increasing concentration of solvent B (90% ACN, 10 mM NH_4COOH, adjusted to pH 10.0 with $NH_3 \cdot H_2O$) were used for peptide separation. A 110-min gradient was set as follows, 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. The eluate was collected every 1 min into 96 fractions from 3 min to 99 min. The 96 fractions were combined by a concatenation strategy into 48 fractions (1&49, 2&50…48&96). 5% of the 48 fractions were taken out and dried by vacuum centrifugation for proteome analysis. The other 95% of the 48 fractions were further combined into 24 fractions for phosphopeptide enrichment and phosphoproteome analysis.

Phosphopeptide Enrichment

The phosphopeptide enrichment was performed using High-Select Fe-NTA kit (Thermo Scientific, A32992) according to the kit instructions with minor modifications. Briefly, the 24 fractions were dissolved with 200 μL loading buffer (80% ACN, 0.1% TFA). The resins of one spin column in the kit were divided into 24 equal parts and mixed with each peptide fractions. The peptide-resin mixture was incubated for 15 min at room temperature and then transferred into the filter tip (Axygen, TF-20-L-R-S). The supernatant was removed after centrifugation. Then the resins adsorbed with phosphopeptides were washed sequentially with 200 μL × 3 washing buffer (80% ACN, 0.1% TFA) and 200 μL × 3 H_2O to remove nonspecifically adsorbed peptides. The phosphopeptides were eluted off the resins by 100 μL × 2 elution buffer (50% ACN, 5% $NH_3 \cdot H_2O$). All centrifugation steps above were conducted at 50 g. The eluates were collected for speed-vac and dried for mass spectrometry analysis.

Benchmark Sample Preparation

The benchmark samples were prepared for longitudinal quality control of mass spectrometry performance (Figures S2E and S2F). Five pairs of hepatobiliary carcinoma tissue and their adjacent non-tumor liver samples were used for protein extraction and digestion, TMT 10-plex labeling (tumor tissues labeled with 126, 127C, 128C, 129C and 130C; adjacent non-tumor liver tissues labeled with 127N, 128N, 129N, 130N and 131) and peptide fractionation (24 fractions). The benchmark samples (~1 μg each fraction) were analyzed before every four sets of HCC proteomic samples on a Q Exactive HF mass spectrometer. The diluted benchmark samples (~100 ng each fraction) were analyzed before every four sets of HCC phosphoproteomic samples on an Orbitrap Fusion mass spectrometer.

Nano-LC-MS/MS

Proteomic Analysis

For proteomic analysis, the fractionated peptides (~1 μg each fraction) were resolved using 0.1% formic acid and a quarter of each fraction was separated using a home-made micro-tip C18 column (75 μm × 200 mm) packed with ReproSil-Pur C18-AQ, 3.0 μm resin (Dr. Maisch GmbH, Germany) on a nanoflow HPLC Easy-nLC 1000 system (Thermo Fisher Scientific), using a 70 min LC gradient at 300 nL/min. Buffer A consisted of 0.1% (v/v) formic acid in H_2O and Buffer B consisted of 0.1% (v/v) formic acid in acetonitrile. The gradient was set as follows: 2%–5% B in 1 min; 5%–27% B in 53 min; 27%–40% B in 10 min; 40%–90% B in 2 min; 90% B in 4 min. Proteomic analyses were performed on a Q Exactive HF mass spectrometer (Thermo Fisher Scientific). The spray voltage was set at 2,500 V in positive ion mode and the ion transfer tube temperature was set at 275°C. Data-dependent acquisition was performed using Xcalibur software in profile spectrum data type. The MS1 full scan was set at a resolution of 120,000 @ m/z 200, AGC target 3e6 and maximum IT 50 ms by orbitrap mass analyzer (350-1700 m/z), followed by 'top 15' MS2 scans generated by HCD fragmentation at a resolution of 60,000 @ m/z 200, AGC target 1e5 and maximum IT 120 ms. The fixed first mass of MS2 spectrum was set 105.0 m/z. Isolation window was set at 1.0 m/z. The normalized collision energy (NCE) was set at NCE 32%, and the dynamic exclusion time was 30 s. Precursors with charge 1, 7, 8 and > 8 were excluded for MS2 analysis. The 24 benchmark fractions were analyzed using a 90 min LC gradient. The gradient was set as follows: 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%–90% in 1 min; 90% B in 5 min. MS parameters were set the same as HCC proteomic samples.

Phosphoproteomic Analysis

For phosphoproteomic analysis, the enriched phosphopeptides were resolved using 0.1% formic acid and half of each fraction was loaded for LC separation on a nanoflow HPLC Easy-nLC 1200 system (Thermo Fisher Scientific), using a 70 min LC gradient at 300 nL/min. The RP chromatographic column was the same as above. Buffer A consisted of 0.1% (v/v) formic acid in H_2O and Buffer B consisted of 0.1% (v/v) formic acid in 80% acetonitrile. The gradient was set as followings: 5%–8% B in 4 min; 8%–30% B in 46 min; 30%–44% B in 10 min; 44%–100% B in 3 min; 100% B in 7 min. Enriched phosphopeptides were analyzed on an Orbitrap Fusion mass spectrometer (Thermo Fisher Scientific). The spray voltage was set at 2,500V in positive ion mode and the ion transfer tube temperature was set at 275°C. Data-dependent acquisition was performed using Xcalibur software in profile spectrum data type. The MS1 full scan was set at a resolution of 120,000 @ m/z 200, RF lens 60%, AGC target 4e5 and maximum IT 50 ms by orbitrap mass analyzer (350-1500 m/z), followed by top-speed MS2 scans generated by HCD fragmentation at a resolution of 50,000 @ m/z 200, AGC target 1e5, inject ions for all available parallelizable time and maximum IT 100 ms in a 3 s cycle time. The fixed first mass of MS2 spectrum was set 105.0 m/z. Isolation window was set at 1.2 m/z. The HCD collision energy was set at 38%, and the dynamic exclusion time was 30 s. Precursors with charge state 2-6 were selected for MS2 analysis. The 24 benchmark fractions were analyzed using a 70 min LC gradient. The gradient was set as followings: 5%–8% B in 4 min; 8%–35% B in 46 min; 35%–50% B in 10 min; 50%–100% B in 2 min; 100% B in 8 min. MS parameters were set the same as HCC phosphoproteomic samples.

Database Searching of MS Data

All mass spectrometric data were analyzed using MaxQuant 1.6.1.0 against the human Swiss-Prot database containing 20,231 sequences (downloaded in December, 2017) plus 269 Swiss-Prot HBV protein sequences (Tyanova et al., 2016). TMT 11-plex (HCC and liver samples) or TMT 10-plex (Benchmark samples)-based MS2 reporter ion quantification was chosen with reporter mass tolerance set at 0.003 Da. The PIF (precursor intensity fraction) filter value was set at 0.5 to reduce the interference of precursor co-fragmentation. Carbamidomethyl cysteine was searched as a fixed modification. Oxidized methionine, protein N-term acetylation, lysine acetylation, asparagine and glutamine (NQ) deamidation were set as variable modifications. In HCC phosphorylation data analysis, phospho (STY) was also chosen as a variable modification. A maximum number of 5 modifications per peptide were allowed for each peptide. Enzyme specificity was set as trypsin. The maximum missing cleavage site was set as 2. The tolerances of first search and main search for peptides were set at 20 ppm and 4.5 ppm, respectively. The minimal peptide length was set at 7. The false discovery rates (FDR) of peptide, protein and site were all < 0.01. For each set of TMT labeling data, the purities of TMT labeling channels were corrected according to the kit LOT number. The Class I (a localization probability filter > 0.75) phosphorylation sites were considered as highly reliable sites.

WES Data Analysis

Somatic Mutation Calling and Filtering

WES sequencing reads after exclusion of low-quality reads were mapped to the UCSC hg19 reference sequence with BWA (version 0.7.15, http://bio-bwa.sourceforge.net/) (Li and Durbin, 2009). PCR duplicates were removed by Picard (version 2.0.1, http://broadinstitute.github.io/picard/), and recalibrated by the BaseRecalibrator tool from GATK (version 4.0.6.0, https://software.broadinstitute.org/gatk/). Somatic variants were detected using Mutect (version 2) on exome data of tumor and matched non-tumor pairs. Annotation of variants was performed by Annovar (version 2017 Jul 17, http://annovar.openbioinformatics.org/en/latest/) (Wang et al., 2010) on Refseq gene models (version 2017/06/01). Germline variants were filtered from database of the 1000 Genomes (Auton et al., 2015), NHLBI Exome Sequencing Project (ESP6500), Exome Aggregation Consortium (EXAC), and Genome Aggregation Database (gnomAD). A stringent downstream filter comprised of the following criteria was used to obtain high quality somatic variants: a minimum of 8X coverage; Variant Allele Fraction (VAF) $\geq$ 5% and at least 5 variant supporting reads in the tumor sample, and VAF $< 1\%$ in the non-tumor sample; strand bias ≤ 0.95; After that, mutations in the non-coding regions (3′UTR, 5′UTR, Intron, gene intergenic etc.) were removed. This finally resulted in 20,369 non-silent somatic SNV calls and 1,363 indel calls in total for tumor samples in comparison to matched non-tumor liver samples (159 pairs).

Analysis of Significantly Mutated Genes

The filtered mutations (including SNV and indel) were further used to identify significantly mutated genes by MutSigCV (https://software.broadinstitute.org/cancer/cga/mutsig, version 1.4) with default parameters. The final MutSigCV *P value*s were converted to q-values with the method of Benjamini and Hochberg (Lawrence et al., 2013), and genes with $q \leq 0.1$ were declared to be significantly mutated.

Mutual Exclusivity Analysis of Mutations

To detect mutual exclusivity of significantly mutated genes in our mutational dataset, Fisher's exact test was used to detect mutually exclusively mutated genes.

Germline Variants Calling and Filtering

Germline variants were identified using the HaplotypeCaller tool from GATK (version 4.0.6.0) with the genotyping_mode DISCOVERY -stand_call_conf 30; post-processing filter was performed by QD < 2.0, FS > 60.0, MQ < 40.0, MappingQualityRankSum < -12.5, ReadPosRankSum < -8.0 as the GATK best practice recommended.

Comparisons of Frequently Mutated Genes between CHCC-HBV and TCGA cohort

The somatic MAF data called by Mutect from The Cancer Genome Atlas Liver Hepatocellular Carcinoma (TCGA-LIHC) were obtained from the GDC database (http://gdac.broadinstitute.org), and HBV infection status was retrieved from TCGA HCC paper (Cancer Genome Atlas Research Network, 2017). The most frequently mutated genes within CHCC-HBV cohort and TCGA cohort were compared (Figure 1B).

Mutational Signature Analysis

Mutation signatures were jointly inferred for 159 tumors with the software of Mutational Signatures in Cancer (MuSiCa) (Díaz-Gay et al., 2018). The 96 mutational vectors (or contexts) generated by somatic SNVs based on six base substitutions (C > A, C > G, C > T, T > A, T > C, and T > G) within 16 possible combinations of neighboring bases for each substitution were used as input data to infer their contributions to observed mutations. MuSiCa using non-negative matrix factorization (NMF) approach was implemented to decipher the 96 × 159 (i.e., mutational context-by-sample) matrix by 30 known COSMIC cancer signatures (https://cancer.sanger.ac.uk/cosmic/signatures) and infer their exposure contributions. Moreover, the frequencies of substitutions in 96 possible mutation types of CHCC-HBV and TCGA-HBV cohorts as well as CHCC-HBV and TCGA-HCV cohorts were compared by PCA analysis with sklearn PCA and visualized in 3-demension by matplotlib of Python v2.7 (Figures S4A and S4B). Wilks' λ test was used to assess the significance of the mean vector differences in different cohorts.

Aristolochic Acid (AA) Signature Analysis in HCC

Mutational signature activity (mSigAct, version 0.9) (Ng et al., 2017) was applied to assess presence of the aristolochic acid (AA) signature in 159 tumor samples. The mSigAct software provided a signature presence test to infer whether the observed mutation spectrum could be better explained with a contribution from the AA mutational signature (COSMIC signature 22) than without it and compared them with a likelihood ratio test. The null hypothesis was that the mutational counts were generated without the AA signature and the alternative hypothesis was that they were from the AA signature. The test was then carried out by a standard likelihood ratio test on these two hypotheses. The mSigAct revealed strong evidence of AA exposure with FDR < 0.05.

Tumor Mutational Burden (TMB)

TMB was defined as the number of somatic mutations (including base substitutions and indels) in the coding region. To reduce sampling noise, synonymous alterations were also counted (Chalmers et al., 2017). In order to calculate the TMB, the total number of mutations counted was divided by the size of the coding sequence region of the Agilent SureSelect Human All Exon V6.

Neoantigen Prediction

For neoantigen prediction, HLA class I types (HLA-A, HLA-B, HLA-C) for each sample were identified using OptiType (version 1.2.1) (Szolek et al., 2014). NetMHCpan (version 3.0) (Nielsen and Andreatta, 2016) was applied to predict the binding affinity of peptides and identify MHC ligands. Predicted binding affinity parameters > 1,000 nM were considered as weak binding, and those with strong binding affinity ($IC_{50} \leq 500$ nM) were enrolled as predicted neoantigen.

The association analysis of AA signature with immune signatures

The abundance of CD8+ T cells was inferred by xCell webtool (Aran et al., 2017). *P value*s for signature distribution between AA and non-AA cohorts were calculated using t test.

Exome-Based Somatic Copy Number Alteration (SCNA) Analysis

For each tumor, SCNAs were inferred by CNVkit (version 0.9.5, https://cnvkit.readthedocs.io/en/stable/pipeline.html) (Talevich et al., 2016) using Circular Binary Segmentation algorithm with default parameters. Segment-level ratios were calculated and $\log_2$ transformed. Significant focal SCNAs across all samples were identified by Genomic Identification of Significant Targets in Cancer (GISTIC, version 2.0) (Mermel et al., 2011) to determine which SCNA regions were significantly gained or lost than expected by chance with q value ≤ 0.1. Based on the published literature (Bambury et al., 2015), a $\log_2$ ratio cut-off of ± 0.8 was used to define SCNA amplification and deletion. To further summarize the arm-level copy number change (i.e., chromosomal instability), we used a weighted sum approach (Vasaikar et al., 2019), in which the segment-level $\log_2$ copy ratios for all the segments located in the given arm were added up with the length of each segment being weighted.

RNA-seq Data analysis

RNA-seq data analysis with RSEM

After removal of adaptor contamination, polyA and polyC, sequencing reads were aligned using STAR2 (version 2.4.2a) (Dobin et al., 2013) to human reference sequence (UCSC hg19 assembly). Gene expression values were quantitated with RSEM (version 1.3.0) (Li and Dewey, 2011) against the GENCODE (version 19) (https://www.gencodegenes.org) transcript models, and then were normalized within each sample to upper quartile. The RNA-seq experiments were performed in 4 batches due to the 4 different sample delivery times. The batch effect of RNA data was evaluated by PCA and corrected with ComBat in SVA (R package, https://cran.r-project.org/web/packages/COMBAT/index.html). Finally, the $\log_2$ transformed upper-quartile normalized RSEM counts were used for the following analysis. RNA-seq data of four non-tumor liver samples (ID numbers: N127, N431, N777 and N813) were excluded in the subsequent analysis because they didn't pass the sample gender check and tumor/normal status check procedures.

RNA-seq Variant Calling

For RNA-seq variant calling, QC passed data were realigned using STAR2 (version 2.4.2a, https://github.com/alexdobin/STAR) in two-pass mode. Pre-processing steps of deduplication, splitting reads into exon segments, hard-clipping any sequences overhanging into intronic regions as well as local realignment and recalibration were performed. Variants calling was implemented by

'HaplotypeCaller' mode with parameters of –genotyping_mode DISCOVERY -recoverDanglingHeads -dontUseSoftClippedBases –dbsnp dbsnp_137.hg19.*vcf.* -stand_emit_conf 20). Highly accurate variants were filtered by applying 'VariantFiltration' (with parameters: -window 35 -cluster 3 -filterName FS –filter ''FS >30.0'' -filterName QD -filter ''QD <2.0'') and filtered with depths $\geq$ 6 and allelic depths for the alt alleles $\geq$ 3.

Proteome and Phosphoproteome Data Analysis

Data Normalization

Data were normalized using the median centering method across total proteins or phosphorylation sites to correct sample loading differences. In normalized samples, these proteins or phosphorylation sites should have a log TMT ratio value centered at zero. Normalized Proteins/phosphorylation sites with SwissProt ID were converted to Human Genome Nomenclature Committee's HUGO symbols provided by HGNC (https://www.genenames.org).

Missing Value Imputation

K-nearest neighbor (k-NN) imputation was applied to impute the missing values. Before missing value imputation, proteins and phosphorylation sites having more than 50% missing data were excluded to ensure that each sample had enough data for imputation. The imputation method was implemented in the pamr package in R.

Batch effect and data quality analysis for proteomic data

The batch effect due to TMT multiplexes was assessed by performed performing unsupervised PCA on the proteomic data. The leading PCs of the global proteomic data clearly separated the tumor from normal samples, and no obvious batch effect was observed among the 33 TMT batches. In addition, the quality of proteomics data was examined based on protein complex correlation analysis and co-expression network-based function prediction (Wang et al., 2017), and compared with RNA-Seq data. The analysis was performed using OmicsEV (https://github.com/bzhanglab/OmicsEV/).

Differential Expression Analysis

The proteomic data filtered with no missing values (n = 6,478 genes) and with imputation values (n = 8,958 genes) were both used as input data for differential expression analysis. Samr R package (Li and Tibshirani, 2013) was used to identify proteins that were differentially expressed in tumor and non-tumor liver tissues using paired two-class of Samr with 1,000 permutations and an FDR threshold of 0.05. For proteomic data with no missing values, a total of 1,274 proteins were identified by differential analysis with a fold change > 1.5.

The parameters for differential analysis of RNA-seq data were as follows: the upper-quartile normalized RSEM counts data (n = 19,860 genes) were used as input data for differential expression analysis. Samr R package was used to identify proteins that were differentially expressed between tumor and non-tumor liver tissues using paired two-class of Samr with 1,000 permutations and an FDR threshold of 0.05. A total of 3,697 genes were thus identified by differential analysis with a fold change > 2.

mRNA, proteomic and phosphoproteomic subgrouping analysis

Consensus Clustering for mRNA and Proteomic Data

We chose differentially expressed proteins with no missing values for subgrouping. 1,274 proteins expressed differentially between tumor and non-tumor liver tissues were first selected by SAM (significance analysis of microarray) with statistically significance (FDR q value < 0.05 and fold change > 1.5). Among them, 1,126 proteins (88.4%) were also among the top 50% most varied proteins within tumors (Figure S6B). K-means consensus clustering was then performed on the selected proteins to generate subgroups. Consensus clustering was implemented on these 1,274 differentially expressed proteins using the ConsensusClusterPlus R package (Wilkerson and Hayes, 2010), and the following detail settings were used for clustering: number of repetitions = 1,000 bootstraps; pItem = 0.8 (resampling 80% of any sample); pFeature = 0.8 (resampling 80% of any protein); and k-means clustering with up to 6 clusters. The number of clustering was determined by three factors, the average pairwise consensus matrix within consensus clusters, the delta plot of the relative change in the area under the cumulative distribution function (CDF) curve, and the average silhouette distance for consensus clusters. We selected a 3-cluster as the best solution for the consensus matrix with k = 3 or k = 4 deemed to be a cleanest separation among clusters, but the consensus CDF and delta plot exhibited that there was little increase in area for k = 3 compared to k = 4. Moreover, the average silhouette distance for k = 3 was larger than k = 4 or k = 5 and did not have significant negative values. Based on the evidence above, the HCC proteomic data were clustered into 3 groups (Figure S6A). As summarized in Figure 3A, the clustering analysis of the tumors (vertical column) by protein abundance (horizontal rows) divided all tumors into three proteomic subgroups defined by silhouette analyses (Figure S6A). The decision was finally attributed to (i) the average silhouette distance for 3 clusters (0.79) and (ii) no silhouette widths with significant negative values observed for 3 clusters.

A total of 3,697 genes identified by differential RNA-seq data analysis were performed with the following parameters used for consensus clustering: number of repetitions = 1,000 bootstraps; pItem = 0.8 (resampling 80% of any sample); pFeature = 0.8 (resampling 80% of any genes); and k-means clustering with up to 6 clusters, and 3-cluster as the optimized solution for clustering.

Comparison of the CHCC-HBV Subgrouping with Previous HCC Subgroupings

Five reported HCC gene expression signatures, including Chiang HCC signature (Chiang et al., 2008), Late TGF-β responsive genes signature (Coulouarn et al., 2008), Hoshida sub-classes (S1, S2, and S3) signatures (Hoshida et al., 2009), WNT-pathway activation signatures (Lachenmayer et al., 2012), and NCI proliferation (NCIP) signature (Lee et al., 2004), were collected for comparison. Different from those previous HCC classifications, our three subgroups were based on HBV-infected HCC patients at proteomic

level. For the patient stratification, consensus clustering was also employed on HCC protein abundance with previously defined signature genes, and the results showed the concordance with previous gene expression-based classifications.

Association between Proteomic Subgroup and Clinical Outcome

We performed survival analysis of patient stratification in different subgroups from Consensus Clustering. The Log-rank test was used to compare survival outcomes among these three subgroups generated by proteomics and mRNA-based clustering, respectively, and Kaplan-Meier survival curves were plotted by R ggsurvplot package.

Defining Phosphoproteomic Clusters in Pathway Level

The phosphorylation sites from the same phosphoproteins were collapsed by calculating the median ratio, and then samples with any missing values in phosphoprotein were excluded from subsequent analysis, resulting in a clean dataset of 3,836 phosphoproteins. In addition, a total of 859 differentially expressed phosphoproteins were identified by SAM analysis with statistical significance (FDR q value < 0.05 and fold change > 1.5). The phosphoprotein dataset of these differential proteins across tumor samples was subjected to single-sample gene set enrichment (ssGSEA) analysis (GSVA R package, http://www.bioconductor.org/packages/release/bioc/html/GSVA.html) (Hänzelmann et al., 2013) to obtain enrichment scores over MSigDB c2 (canonical gene sets, https://software.broadinstitute.org/gsea/msigdb/in dex.jsp) pathway database with at least 10 overlapping genes. Next, the pathway-mapped phosphoprotein data were clustered into 3 robust groups, using k-means consensus clustering and evaluated by the consensus CDF, delta area plot as well as silhouette plots, which was consistent with proteomics stratification.

Multi-omics Data Analysis

mRNA-Protein Correlation

Spearman correlation coefficient was applied to measure the correlation between mRNA expression and protein abundance for each gene-protein pair across all 159 CHCC-HBV samples. In addition, *P value* corresponding to the correlation coefficient was computed and adjusted by the FDR correction. Significance of the correlation pair was determined, based on an adjusted *P value* cut-off of 0.01. A total of 6,203 mRNA-protein matched genes were calculated with a median Spearman correlation of r = 0.54. Moreover, mRNA and protein were positively correlated for most (98.6%) mRNA-protein pairs, and 90.3% showed significant positive correlation (multiple-test adjusted *P value* < 0.01).

Joint Random Forest (JRF) Co-expression Network Analysis

Co-expression network construction analysis was performed to study the interaction patterns among genes and proteins, based on the global proteomic and RNA-seq data across the 159 CHCC-HBV samples. The top 15% expressed mRNAs and proteins with the largest interquartile range were chosen respectively, and generated into two 629 × 159 data matrices in both sets, in which one was for gene expression and the other was for protein expression. Joint Random Forest (JRF) method was utilized to join two co-expression networks and enable information to be shared both in proteomic and transcriptomic data leading to give an accurate estimation. The settings used for JRF were as following: the total number of trees was set to 1,000, the number of variables sampled at each node was set to sqrt (p-1) with p = 629. FDR of importance scores was calculated with 400 permutations. Genes in both existing mRNA and protein co-expression network edges were mapped to pathways and imported into Cytoscape (https://cytoscape.org/) to create co-expression networks.

Effects of Copy Number Alternations

SCNAs affecting mRNA and protein/ phosphoprotein abundance in either "*cis*" (within the same aberrant locus) or "*trans*" (remote locus) mode were visualized by multiOmicsViz (R package). Spearman's correlation coefficients and associated multiple-test adjusted *P value*s were calculated for all CNA-mRNA pairs for 18,054 genes, resulting in CNA-protein pairs for 8,284 genes and CNA-phosphoprotein pairs for 6,104 genes, respectively.

Patient Specific Database Construction and Variant Peptide Identification

For each of the 159 CHCC-HBV patients analyzed, DNA-variants (somatic/germline) by WES and RNA-variants by RNA-seq were obtained with the method described above. RNA junction files were generated by aligning clean reads to the highly reliable human reference genome (version hg19) using TopHat (version 2.1.1, http://ccb.jhu.edu/software/tophat/index.shtml) with parameters of –g 1,–bowtie2 (version 2.3.3.3), -M, -x 1, and–fusion-search settings. 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. First, 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 analysis of CHCC-HBV samples with different phenotypes was analyzed with t test, including differential proteins in tumors with versus without tumor thrombus, differential proteins, and phosphorylation sites in tumors carrying mutated versus non-mutated *TP53* and *CTNNB1*, or differential mRNAs, proteins and phosphorylation sites in tumors with high or low expression of PYCR2 and ADH1A. Genes with FDR < 0.1 and a fold change > 1.5 or other thresholds were visualized by ComplexHeatmap (R package). Pathway enrichment analysis of the significant genes was performed using clusterProfiler (R package). Pathways with an FDR threshold of 0.05 were regarded to be significantly regulated. The proteome and phosphoproteome samples with < 50% missing

values were imputed (see missing value imputation section above) and used for subsequent analyses except for the subgrouping analysis (with only no missing values).

Gene Set Enrichment Analysis (GSEA)

GSEA was performed by the GSEA software (http://software.broadinstitute.org/gsea/index.jsp) (Subramanian et al., 2005). Gene sets used in this article were c2.cp.kegg.v6.2.symbols.gmt downloaded from the Molecular Signatures Database (MSigDB, http://software.broadinstitute.org/gsea/msigdb/index.jsp).

Functional enrichment analysis of multi-omics level data in three subgroups

To further analyze biological characteristics of three subgroups, we performed single-sample gene set enrichment (ssGSEA) analysis to identify the pathway alterations that underlie the HCC subgroups. Gene expression data of mRNA, proteome and phosphoproteome levels across 159 tumor samples were used to achieve enrichment scores over MSigDB database v.6.2 with at least 10 overlapping genes and the R/Bioconductor package GSVA. The significance of the pathway enrichment scores (PES) over the three subgroups was estimated by linear model and moderated with the F-statistic using the R/Bioconductor package limma. The resulting significant PES among three subgroups were corrected by the Benjamini–Hochberg method, which used an adjusted *P value* cut-off of 0.05.

Prognostic Biomarker Analysis for CHCC-HBV

Cox proportional hazard model for overall survival data was implemented to identify biomarkers for HCC prognosis. We stratified the CHCC-HBV patients into two groups and used the median as cutoff to define high and low protein expression. Kaplan-Meier curve (Log-rank test) was used to visualize survival difference. The filter criteria for survival analysis were as follows: tumor versus non-tumor with t test *P value* < 0.0001 and fold change > 2, correlation between mRNA and protein expression > 0.5, variance in tumor > 0.5, Log-rank *P value* < 0.0001, and HR > 1.8 for upregulated or < 0.55 for downregulated proteins.

We first used TMAs to validate the proteins' abundance by immunofluorescence staining in the same CHCC-HBV cohort with 155 samples as discovery set (leaving 4 patients without high-quality paraffin tissue blocks). Next, a second cohort consisting of 243 HCC samples was used as an independent validation.

Tissue MicroArray (TMA) Experiment

TMA Construction

TMAs were constructed using 155 paired tumor and non-tumor liver tissues from the CHCC-HBV cohort using the method as we previously described (Gao et al., 2012). In brief, all cases were histologically inspected by H&E staining and representative areas were pre-marked on the paraffin blocks, away from necrotic and hemorrhagic regions. Duplicates of 1.5-mm-diameter cylinders from two different areas, tumor center and non-tumor liver, were included in each case, along with different controls, to ensure reproducibility and homogeneous staining of the slides.

For validation, the TMAs from an independent cohort consisting of 243 HCC patients were used. These 243 HCC patients received curative surgery from January to December 2007 at Zhongshan Hospital and received no prior anticancer treatments. In this cohort, the median age was 51, with 97.9% HBV-positive. 134, 20 and 89 patients were classified as TNM stage I, II and III-IVA respectively. Detailed clinicopathologic features were summarized in Tables S7.

Multiplexed Immunostaining

Multiplex staining of ADH1A (clone EPR4439, No. ab108203 Abcam) and PYCR2 (Polyclonal, No. 17146-1-AP, Protein Tech) was performed by the Vectra Automated Quantitative Pathology Imaging and Analysis platform through multispectral imaging system and inForm^{TM} 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-PYCR2/Opal570, anti-ADH1A/Opal520. After two sequential reactions, slides were counterstained with DAPI (D9542, Sigma) and mounted with fluorescence mounting medium (S3023, Dako). Following similar procedures, the staining of SLC10A1 (Polyclonal, No. Ab131084, Abcam) was performed on TMAs.

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 comprising four combined 200 multispectral image cubes. Multispectral images for each TMA core was 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 (PYCR2/ADH1A/SLC10A1), 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 modified H-scores for PYCR2, ADH1A and SLC10A1 (percentage of tumor cells with positive staining multiplied by the average intensity of positive staining) were divided into two equally sized groups (median as cutoff).

Drug Target Analysis

Drug targets either approved by FDA or under clinical trials were retrieved from Drugbank database (version 5.1.1, released 2018-07-03) (https://www.drugbank.ca/). Target proteins that were upregulated in tumor compared to non-tumor with potential curative drugs (antagonist and inhibitor) were chosen.

Functional Experiments

Plasmids

The Flag-tagged coding sequence of human ALDOA wild-type or the relevant ALDOA-S36E mutant; human CTNNB1 delta N-terminal (ΔN-CTNNB1) were cloned into the lentiviral vector pLEX-MCS-CMV-puro (Addgene, USA) to generate corresponding expression plasmids.

Construction of Stable Cell Lines

pLEX-MCS-CMV-puro lentiviral virus packaging and subsequent generation of stable cell lines by infection were performed according to the protocol previously described (Boehm et al., 2005).

Cell Proliferation Assay

For cell growth assays, HepG2 cells were plated in 96-well plate (2×10^3 cells/well). CCK-8 solution (C0039, Beyotime Biotechnology) was added to the wells for 2 hr and the absorbance was measured at 450 nm.

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 ALDOA RNAi experiment, three different ALDOA siRNA were equally mixed together to transfect into indicated cells at 50 nM final concentration. The siRNA transfected cells were harvested for qPCR assay 48 hr after transfection. Oligonucleotide sequences are as following:

siALDOA-1 Sense: 5′-GCGGUGUUGUGGGCAUCAAdTdT-3′, Antisense: 5′-UUGAUGCCCACAACACCGCdTdT-3′
siALDOA-2 Sense: 5′-GGCGUUGUGUGCUGAAGAUdTdT-3′, Antisense: 5′-AUCUUCAGCACACAACGCCdTdT-3′
siALDOA-3 Sense: 5′-GGAGGAGUAUGUCAAGCGAdTdT-3′, Antisense: 5′-UCGCUUGACAUACUCCUCCdTdT-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 mg total proteins were separated by 10% SDS-PAGE gel and blotted with indicated primary antibodies. Primary antibodies used for western blot analysis were as follows: anti-Flag (1:2000; F7425; Sigma Aldrich), anti-CTNNB1 (1:1000; A11932; ABclonal), anti-Tubulin (1:5000; sc-134237; Santa Cruz Biotechnology). Peroxidase-labeled anti-mouse (1:5000; P0217; DAKO) or anti-rabbit (1:5000; P0260; DAKO) IgG secondary antibody were used. The western blot gel image was obtained with an Minichemi 610 chemiluminescent imager (Sagecreation, Beijing, China).

Real-time Quantitative PCR

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

ALDOA Forward: 5′-CAGGGACAAATGGCGAGACTA-3′, Reverse: 5′-GGGGTGTGTTCCCCAATCTT-3′
β-actin Forward: 5′-AGAGCTACGAGCTGCCTGAC-3′, Reverse:5′-AGCACTGTGTTGGCGTACAG-3′

Assessment of Cellular Metabolism

Cellular respiration of HepG2 cells was measured by a Seahorse XF24 analyzer (Seahorse Bioscience, North Billerica, MA). 1×10^5 cells/well were seeded in Seahorse XF 24-well culture plates (Bucher Biotech AG, Basel, Switzerland) in growth medium and incubated at 37°C/5% CO_2 overnight. Cells were changed to the XF Glycolysis Stress Test Assay Medium (Sigma-Aldrich; D5030) and placed in a 37°C non-CO_2 incubator for 1 hr prior to the assay after they were washed two times with the same assay medium. An assay template was created on the XF Controller and allowed to calibrate and equilibrate, which consisted of 3-minute mix, 2-minute wait, and 3-minute measure cycles. Three basal rate measurements were conducted prior to the first injection, and then glucose (10 mM), oligomycin (1 μM), and 2-deoxy-d-glucose (2-DG, 100 mM) were injected into each well at the indicated time. After injection, the oxygen consumption rate and extracellular acidification rate were closely monitored until the rates stabilized, and then the experiment was terminated. After each measurement, the living cell number of each well was calculated using a EnSight Multimode Plate Reader (PerkinElmer, Germany) after Hochest/PI double staining for further normalization.

Xenograft Tumorigenesis Assay

5-week-old male BALB/c nude mice were purchased from the SLAC Company (Shanghai, China) and maintained in pathogen-free conditions. All animals were acclimated for 1 week before experiments. 1×10^7 HepG2 cells (wild-type or overexpressed with ALDOA or S36E mutant) in 100 μL PBS were subcutaneously inoculated at the flanks of randomly grouped (6 mice per group) nude mice. Tumor sizes were measured every 3 days with a caliper and tumor volumes were calculated by the formula: volume = $(\text{width})^2 \times$ length $\times$ 0.52. All mice were euthanized and tumors were harvested 8 weeks after inoculation, followed by photography. All animals received human care and all animal experiments were performed in accordance with the guidelines of the Institutional Animal Care and Use Committee of Shanghai Institutes for Biological Sciences.

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. For categorical variables versus categorical variables, Fisher's exact test was used in a 2 × 2 table, otherwise Chi-square 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. All statistical tests were two-sided, and statistical significance was considered when $P\ value < 0.05$. To account for multiple-testing, the *P value*s were adjusted using the Benjamini-Hochberg FDR correction. 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 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 ± standard error of the mean (SEM). The statistical significance of differences was determined by two-way ANOVA for CCK8 and Seahorse results. Statistical analysis was performed using GraphPad Prism (version 5.01).

DATA AND CODE AVAILABILITY

The data of WES, transcriptome sequencing, proteome, and phosphoproteome generated in this study can be viewed in NODE (https://www.biosino.org/node) by pasting the accession (OEP000321) into the text search box or through the URL: https://www.biosino.org/node/project/detail/OEP000321.

Softwares used for single-omic and multi-omic analyses were described and referenced in the respective Method Details subsections and listed in the Key Resources Table.

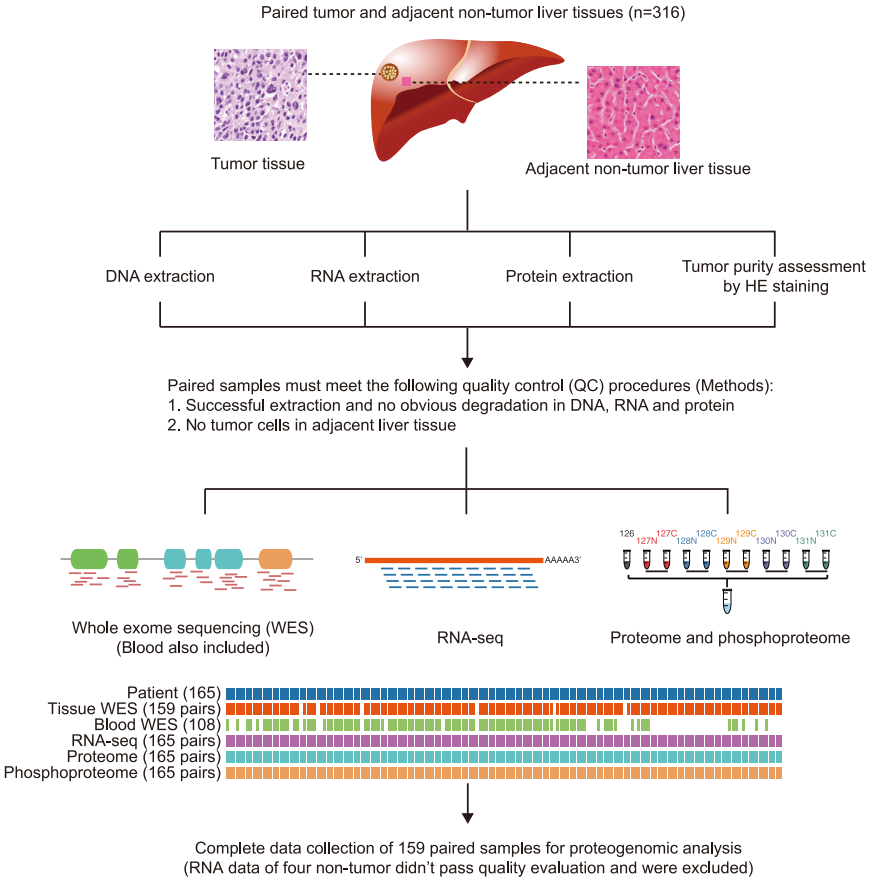
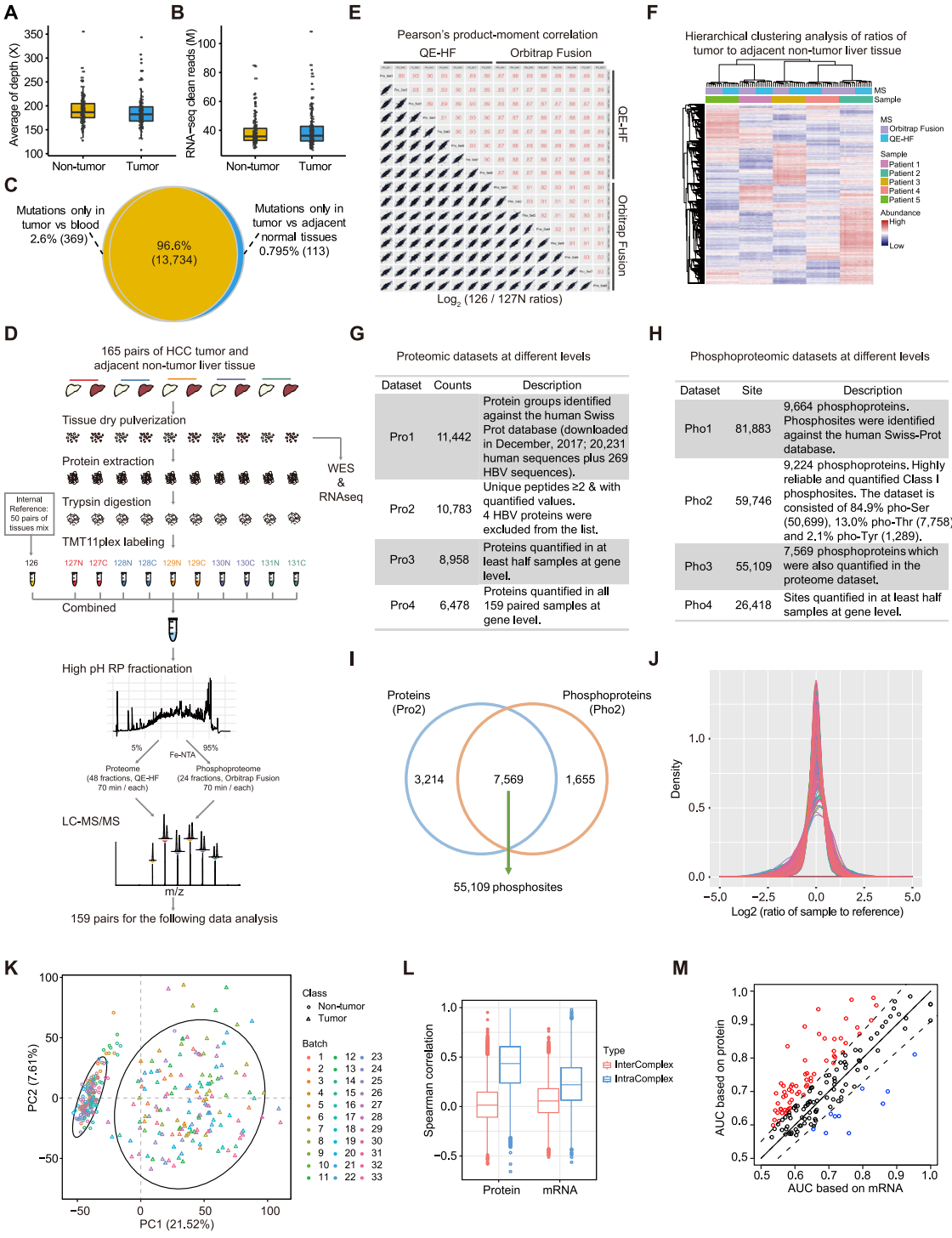


Figure S1. The Workflow of the CHCC-HBV Proteogenomic Study, Related to STAR Methods

Paired tumor and adjacent non-tumor liver tissues from a consecutive cohort of 316 patients were obtained for WES (blood also included), RNA-seq, proteomics, and phosphoproteomics analyses. The tumor purities were assessed by Hematoxylin-Eosin (HE) staining. The sample filtering criteria include 1) successful extraction and no obvious degradation in DNA, RNA and protein and 2) no tumor cells in adjacent liver tissue. Eventually, complete data of WES, RNA-seq, proteome, and phosphoproteome were collected for the 159 paired samples and used in the following proteogenomic analysis. RNA-seq data of 4 non-tumor liver samples (N127, N431, N777 and N813) were excluded due to unqualified RNA-seq data.



Dataset	Counts	Description
Pro1	11,442	Protein groups identified against the human Swiss Prot database (downloaded in December, 2017; 20,231 human sequences plus 269 HBV sequences).
Pro2	10,783	Unique peptides ≥2 & with quantified values. 4 HBV proteins were excluded from the list.
Pro3	8,958	Proteins quantified in at least half samples at gene level.
Pro4	6,478	Proteins quantified in all 159 paired samples at gene level.

Dataset	Site	Description
Pho1	81,883	9,664 phosphoproteins. Phosphosites were identified against the human Swiss-Prot database.
Pho2	59,746	9,224 phosphoproteins. Highly reliable and quantified Class I phosphosites. The dataset is consisted of 84.9% pho-Ser (50,699), 13.0% pho-Thr (7,758) and 2.1% pho-Tyr (1,289).
Pho3	55,109	7,569 phosphoproteins which were also quantified in the proteome dataset.
Pho4	26,418	Sites quantified in at least half samples at gene level.

Figure S2. Quality Assessments for WES, RNA-Seq, and MS Data, Related to STAR Methods

(A) Sequencing depths of WES for tumors and adjacent non-tumor liver tissues.

(B) QC passed reads in RNA-seq for tumors and adjacent non-tumor liver tissues.

(C) Venn diagram showing comparison of somatic alternations called using controls from matched blood samples versus adjacent non-tumor liver tissues from the 108 patients with both controls available.

(D) The TMT 11-plex proteomic and phosphoproteomic workflow. A total of 330 tumor and adjacent non-tumor liver tissues from 165 patients were subjected for dry pulverization, protein extraction, trypsin digestion and analyzed in 33 TMT 11-plex experiments with 5 paired tumor and adjacent non-tumor tissues and the internal reference sample. The reference sample contained 50 pairs of tumor and adjacent non-tumor liver samples mixed in equal protein amount. The labeled

(legend continued on next page)

peptides were combined for high pH RP fractionation. 5% of each fraction was used for proteome analysis and 95% was used for phosphopeptides enrichment and analysis. MaxQuant software and human Swiss-Prot protein database were used for database searching. The proteome dataset resulted in 202,690 peptide sequences identified and quantified from 8,679,925 MS/MS spectra. MS data from the 159 paired samples were used for the following data analysis.

(E) The quantification repeatability of longitudinal benchmark samples showing the robust and accurate proteome/phosphoproteome platform (To take the ratios of 126 to 127N as an example, Pearson's correlation coefficients, 0.86-0.93).

(F) Hierarchical clustering analysis for the ratios of tumor to adjacent non-tumor liver sample (126/127N, 127C/128N, 128C/129N, 129C/130N, 130C/131) in benchmark samples. The data from the same patient samples can be clustered into the same groups, suggesting robust and accurate proteome/phosphoproteome quantification platform.

(G) Summary of proteomic datasets at different levels. 10,783 proteins were quantified with at least 2 unique peptides. Four HBV proteins with $\geq$ 2 unique peptides were identified.

(H) Summary of phosphoproteomic datasets at different levels. 59,746 highly reliable Class I phosphosites (i.e., a localization probability filter >0.75 in MaxQuant) on 9,224 phosphoproteins were quantified, consisting of 50,699 pho-Serine, 7,758 pho-Threonine and 1,289 pho-tyrosine.

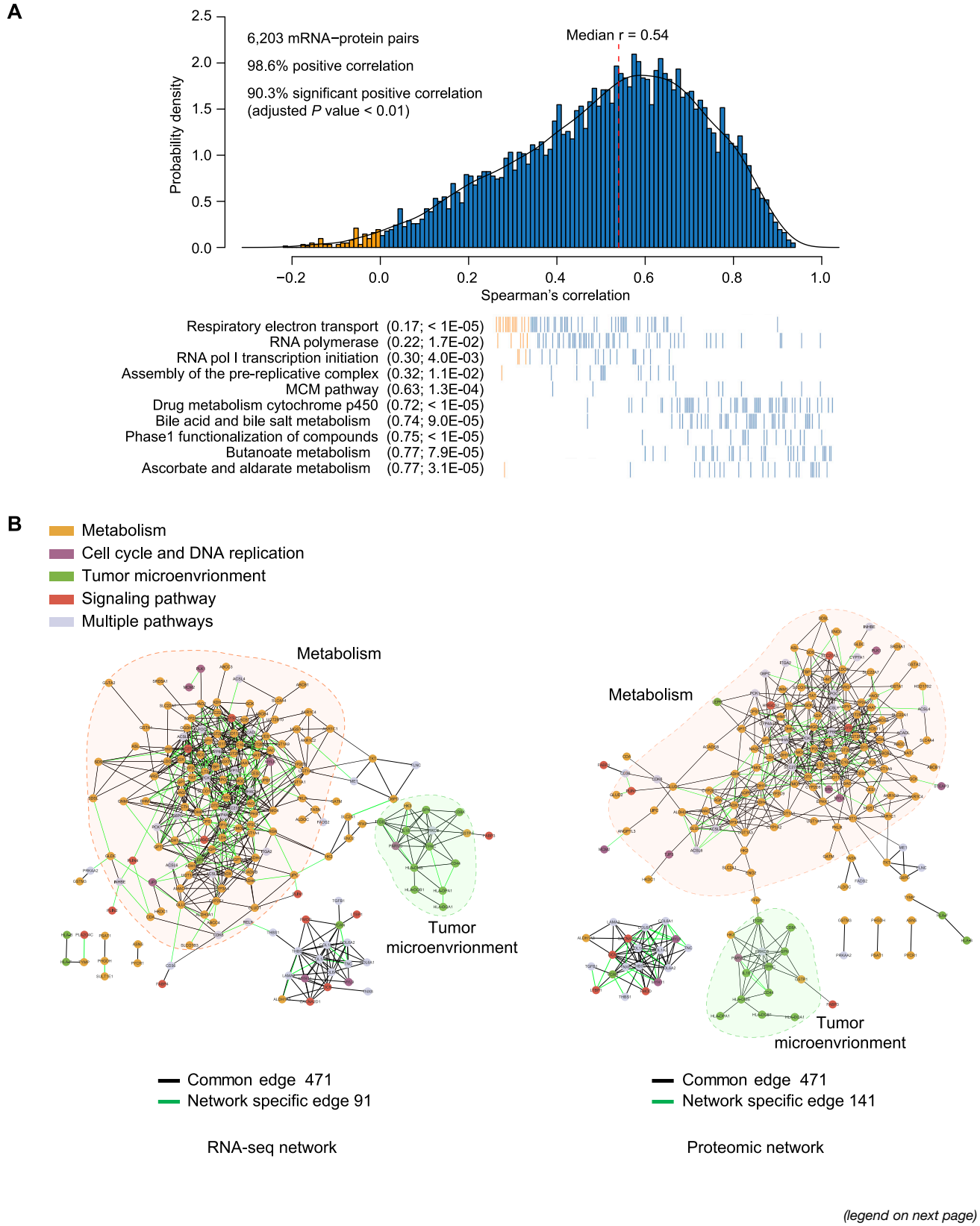
(I) The overlap of proteins and phosphoproteins. 7,569 proteins were identified with 55,109 highly reliable Class I phosphosites (92% of all Class I sites). 3,214 proteins were identified with only their non-phosphorylated forms (29.8% of all proteins). 1,655 proteins were identified with only their phosphorylated forms (17.9% of all phosphoproteins).

(J) The unimodal distributions of the ratios of sample to internal reference (Hartigans' dip test *P value* > 0.05) suggests no obvious degradation in tumor and adjacent non-tumor liver samples.

(K) Principal-component analysis clearly separated the tumor and normal samples based on the 33 TMT global data, and no batch effects were observed.

(L) Quality assessment of proteomic and transcriptomic data based on protein complex correlation analysis.

(M) Comparison of gene function prediction accuracy using co-expression networks based on mRNA and protein profiles for the 160 KEGG pathways. Network-based gene function prediction was based on the random walk-based algorithm. KEGG pathways with at least 20 quantified genes were compared. Prediction performance was evaluated using five-fold cross validation and quantified based on the area under the receiver operating characteristic curve (AUC). Dotted lines indicate 10% increase or decrease of prediction performance. Both protein complex correlation analysis (L) and co-expression network-based function prediction analysis (M) indicated that proteomic data were superior to transcriptomics for accurately predicting gene function.



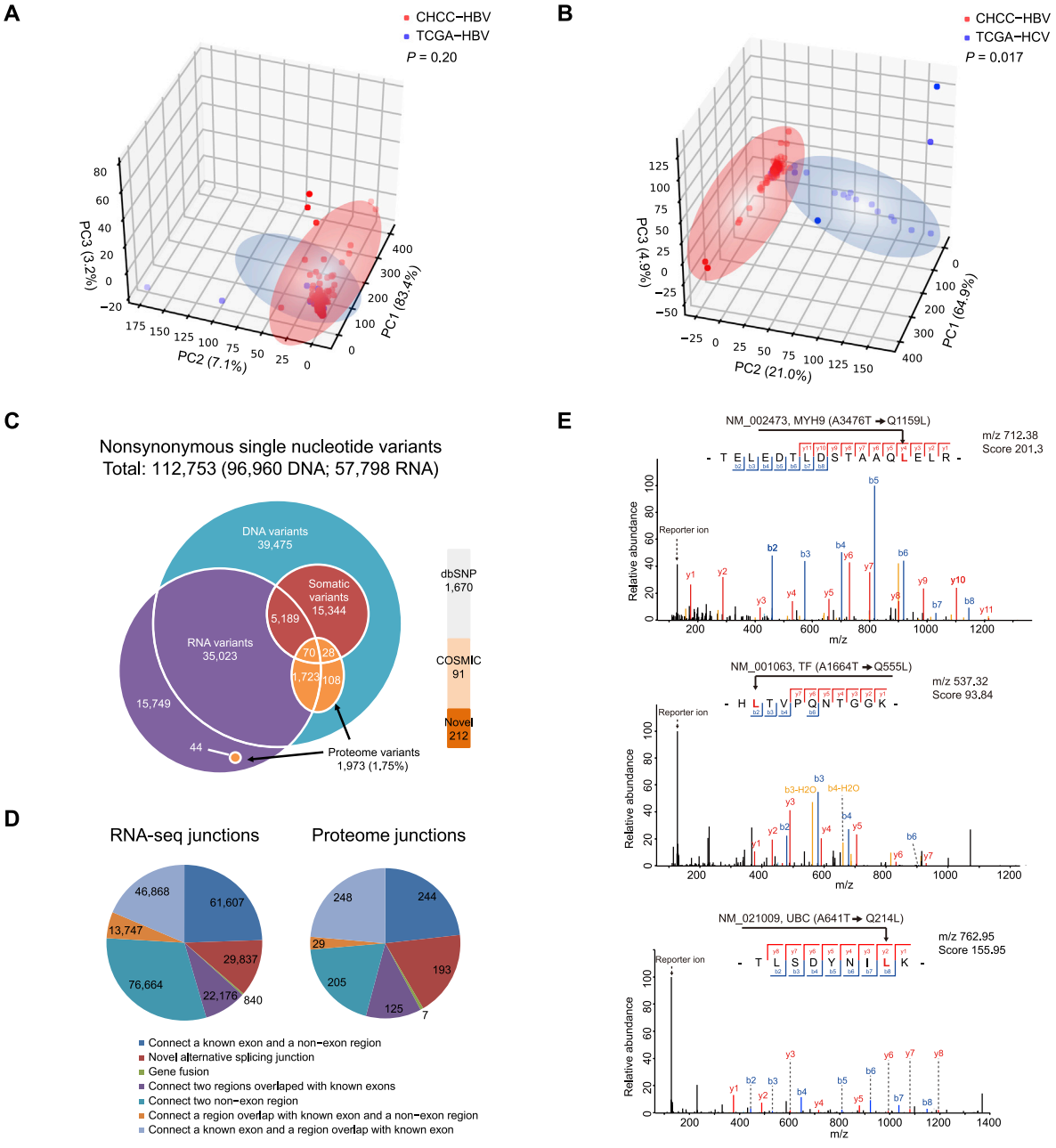
(legend on next page)

Figure S3. The Overall Correlation, Co-clustering, and Co-expression Network Analyses between mRNA and Protein Data, Related to STAR Methods

(A) Top panel: mRNA and protein were positively correlated for most (98.6%) mRNA-protein pairs across the 159 samples, and 90.3% showed significant positive correlation (multiple-test adjusted $p < 0.01$) with a median Spearman's correlation coefficient of 0.54 in 6,203 mRNA-protein pairs.

Bottom panel: Different GSEA enrichment pathways showed significantly different levels of correlation. Metabolic pathways of cytochrome p450, bile acid, bile salt, and butanoate displayed high mRNA-protein correlation, while respiratory electron transport, RNA polymerase and assembly of the pre-replicative complex were poorly correlated. The mean correlation was shown in parentheses, followed by Benjamini-Hochberg adjusted *P* values calculated by using a Kolmogorov–Smirnov test following the names from MSigDB (the Molecular Signature Database). Blue bars indicate positive correlations, and yellow ones indicate negative correlations. Individual proteins in each pathway (represented as bars on the x axis) were sorted by correlation values from low to high.

(B) Co-expression networks of protein and mRNA data respectively, based on Joint Random Forest (JRF) method.



(legend on next page)

Figure S4. The Mutational Signature Analyses, Direct Effects of Genomic Alterations on Protein Level, and Analysis of AA Mutational Signature in CHCC-HBV, Related to Figure 1

(A-B) Principal-component analysis of the 96 substitution patterns in the exonic regions by comparing CHCC-HBV cohort with TCGA-HBV subgroup (A) or TCGA-HCV subgroup (B) (Wilks test).

(C-D) Overlap of protein coding single amino acid variants (C) and RNA splice junctions (D) not present in UCSC hg19 RefSeq (Release 92) detected by WES, RNA-seq, and LC-MS/MS. Proportions of novel variants are noted.

(E) Three peptides carrying mutations with AA signature are validated by mass spectrometry. The MS/MS spectrums and mutated amino acids are manually validated.

(F) Association of AA signature with clinicopathologic and mutational features. Mutations in splicing sites and frameshift Indels are calculated by dividing mutation counts in each sample. Statistical test methods: tumor thrombus, Fisher's exact test; ALB level, Wilcoxon test; frameshift Indels, Fisher's exact test; BCLC stage, Fisher's exact test; splicing-site mutations, Wilcoxon test.

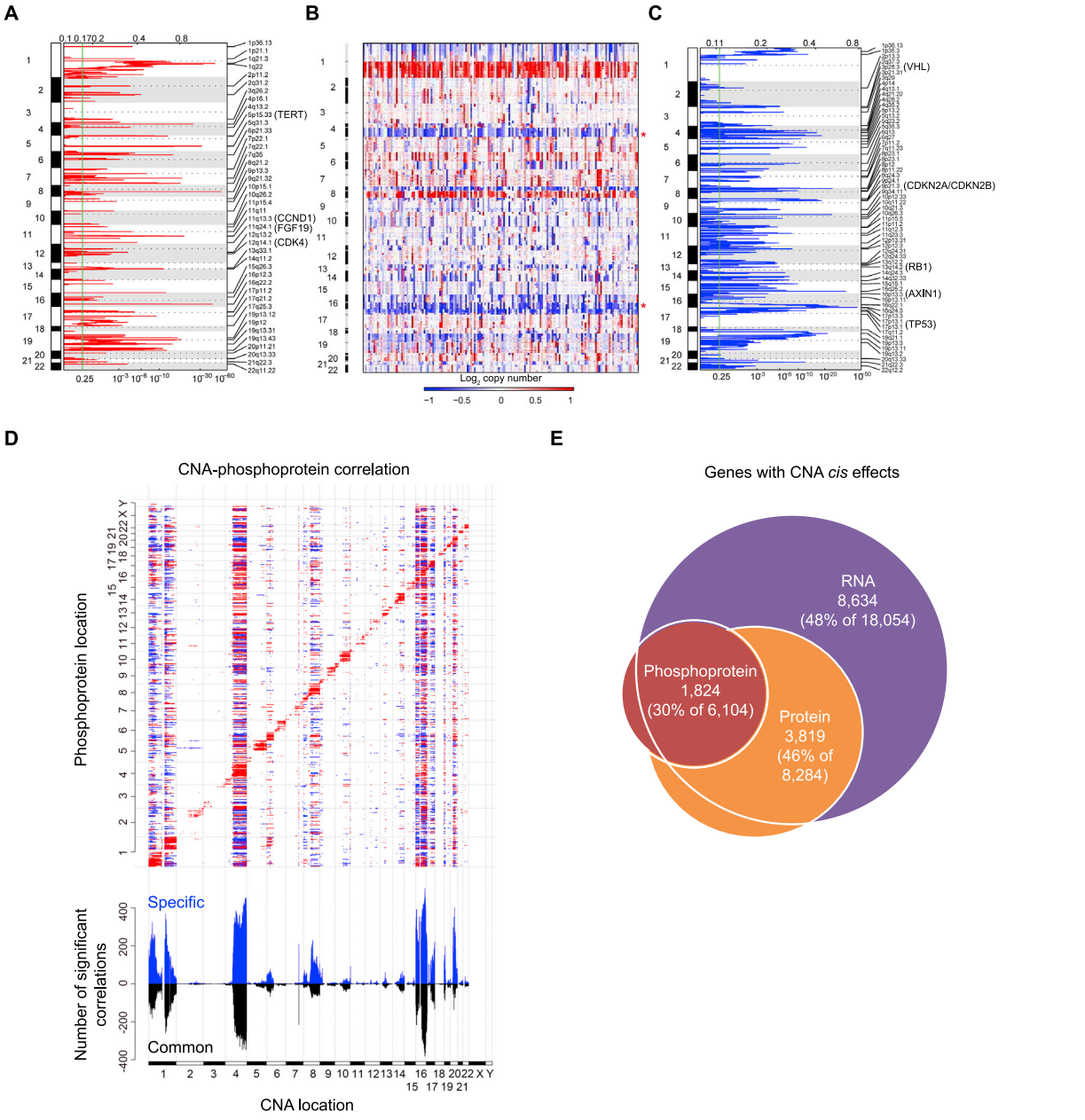


Figure S5. Profiles of Copy-Number Alterations and Correlations of CNA to Phosphoprotein Level in CHCC-HBV Cohort, Related to Figure 2

(A) Genome-wide focal amplification.

(B) Heatmap of the CNAs of 159 HCC tumor samples: Red and blue represents copy gain and loss, respectively, in units of $\log_2$ (tumor/adjacent non-tumor). It is indicated that 4q and 16q are predominantly co-deleted across the cohort as showed by asterisks. The x axis represents the 159 tumor samples.

(C) Genome-wide focal deletions. Chromosomal locations of peaks of significantly recurring focal amplifications (red) and deletions (blue) were filtered by FDRs. Peaks were annotated with candidate driver oncogenes or tumor suppressors by Cytoband (5p15.33 (*TERT*), 11q13.3 (*CCND1*), 11q24.1 (*FGF19*), 12q14.1 (*CDK4*), 3p25.3 (*VHL*), 9p21.3 (*CDKN2A*/*CDKN2B*), 13q14.2 (*RB1*), 16p13.3 (*AXIN1*), 17p13.1 (*TP53*).

(D) Significant positive (red) and negative (blue) correlations between CNA and phosphoprotein are indicated (multiple-test adjusted $p < 0.01$, Spearman's correlation). CNA *cis* effects appear as a red diagonal line, CNA *trans* effects as vertical stripes. CNA regions exhibiting the most *trans* associations at the phosphoprotein level are found on chromosomes 1q, 4q, 16 and 21q. (FDR < 0.01).

(E) Genes with CNA *cis* effects (adjusted *P* value < 0.01, Spearman's correlation) among mRNA, protein, and phosphoprotein levels.

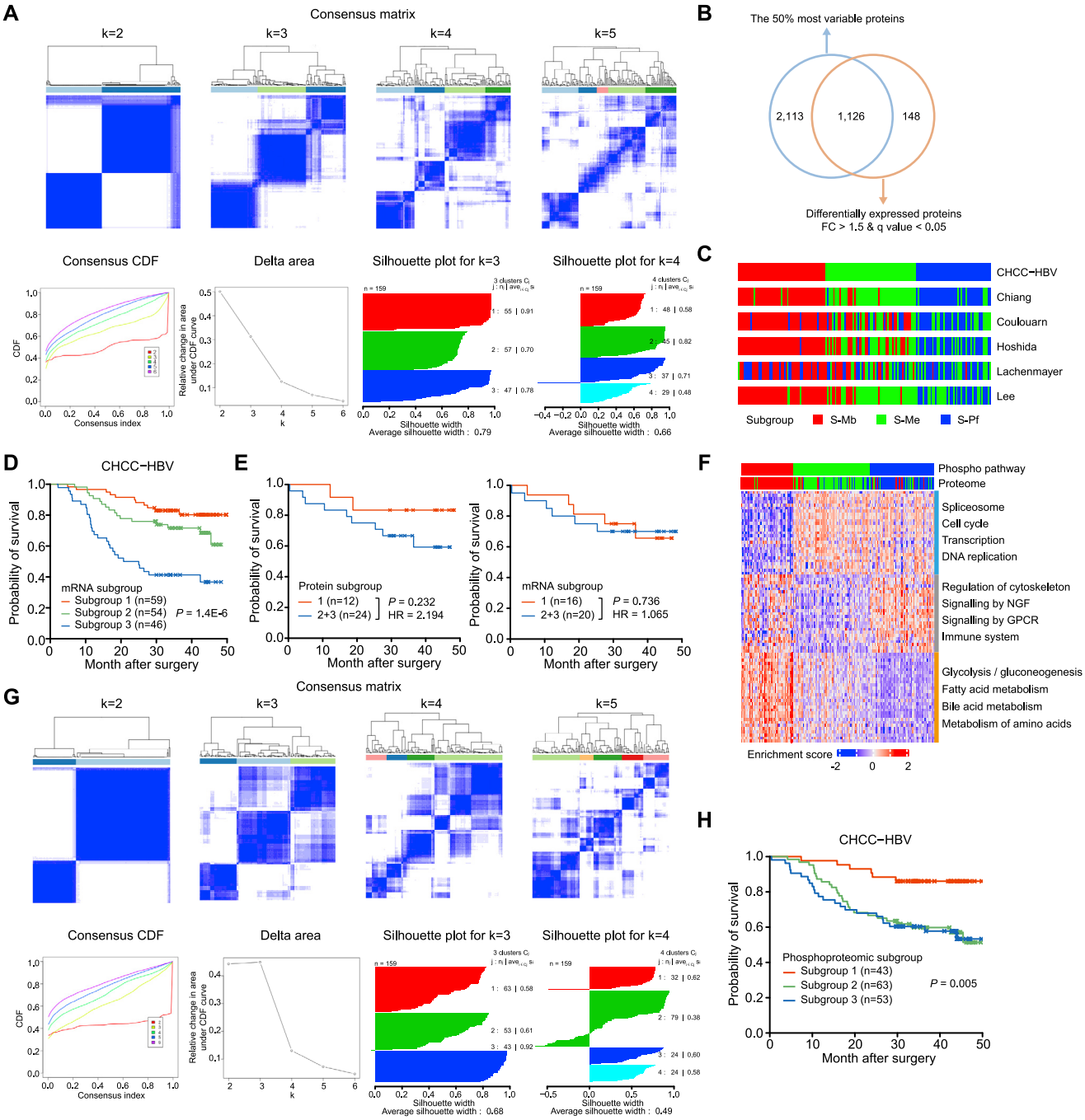


Figure S6. Consensus Clustering for Proteomics and Phosphoproteomics Data in CHCC-HBV Cohort, Related to Figure 3

(A) Subgroups are identified based on proteomic data of CHCC-HBV cohort (n = 159) by K-means consensus clustering upon their abundance (STAR Methods). k was tested from 2 to 5 and consensus clustering was based on 1,000 resampled datasets. Consensus matrices, as well as consensus cumulative distribution function (CDF) plot, delta area (change in CDF area) plot and silhouette plots (k = 3&4) are shown.

(B) The overlap between top 50% most variable proteins and differentially expressed proteins with FC > 1.5 & q value < 0.05.

(C) Comparisons of the CHCC-HBV proteomic subgrouping to the subgrouping resulted from previously reported standards.

(D) Kaplan-Meier curves for overall survival based on mRNA subgroups (Log-rank test).

(E) Prognostic difference of the discordant patients (n = 36) based on protein subgroup or mRNA subgroup.

(legend continued on next page)

(F) Phosphoproteomic subgroups (the top row) are identified (n = 159) by K-means consensus clustering upon pathway ssGSEA scores (STAR Methods). The resulted samples are also labeled by their proteomic subgroups (the second row). The heatmap with column representing samples and rows representing pathways was plotted. Color of each cell indicates Z score (log_2 of relative abundance scaled by ssGSEA score' SD) in that sample.
(G) The consensus matrix, consensus CDF and delta area (change in CDF area) plots, as well as the silhouette plots, were shown.
(H) Kaplan-Meier curves for overall survival based on the three phospho-subgroups (Log-rank test).

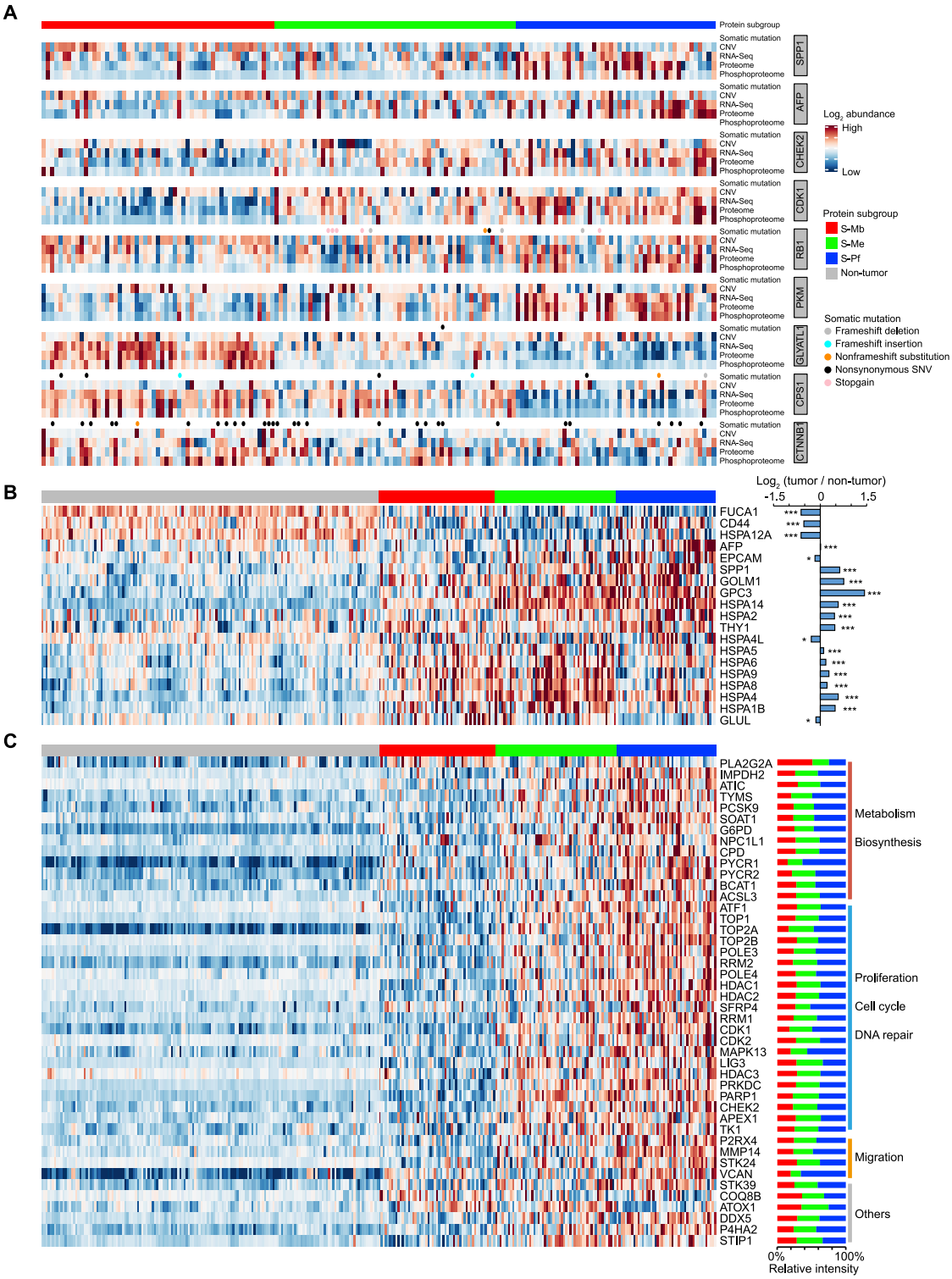


Figure S7. HCC Relevant Genes, Clinical Biomarkers, and Potential Drug Targets, Related to Figure 3

(A) Heatmap of 9 HCC relevant genes and their associations with the subgroups across somatic mutation, CNA, RNA-seq, proteome, and phosphoproteome.

(B) Heatmap of clinically relevant HCC biomarkers in tumor compared to non-tumor liver tissues.

(C) Heatmap of potential drug targets based on proteomic data and bar plot showing the expression ratio among proteomic subgroups. The heatmaps with column representing samples and rows representing proteins were plotted. Color of each cell indicates Z score (log_2 of relative abundance scaled by score' SD) in each sample.

Update

Cell

Volume 179, Issue 5, 14 November 2019, Page 1240

DOI: https://doi.org/10.1016/j.cell.2019.10.038

Cell

Correction

Integrated Proteogenomic Characterization of HBV-Related Hepatocellular Carcinoma

Qiang Gao, Hongwen Zhu, Liangqing Dong, Weiwei Shi, Ran Chen, Zhijian Song, Chen Huang, Junqiang Li, Xiaowei Dong, Yanting Zhou, Qian Liu, Lijie Ma, Xiaoying Wang, Jian Zhou, Yansheng Liu, Emily Boja, Ana I. Robles, Weiping Ma, Pei Wang, Yize Li, Li Ding, Bo Wen, Bing Zhang, Henry Rodriguez, Daming Gao,* Hu Zhou,* and Jia Fan*
*Correspondence: dgao@sibcb.ac.cn (D.G.), zhouhu@simm.ac.cn (H.Z.), fan.jia@zs-hospital.sh.cn (J.F.)
https://doi.org/10.1016/j.cell.2019.10.038

(Cell *179*, 561–577.e1–e22; October 3, 2019)

It has come to our attention that a recent and important reference concerning proteomic analysis of hepatocellular carcinoma was inadvertently omitted during the preparation of our paper. The missing reference is: "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." In the corrected version of this article, this key citation is now referenced in the Introduction.

Additionally, an institutional affiliation of our paper was mis-typed during the submission of final manuscript. It should be "[4]State 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."

We apologize for any confusion and inconvenience that these errors may have caused.

