Scientific Reports

https://doi.org/10.1038/s41598-026-42856-7

Article in Press

Integrated multi-omics analysis reveals comprehensive metabolism-driven tumor heterogeneity and immune microenvironment in hepatocellular carcinomas

Received: 28 August 2025

Accepted: 27 February 2026

Published online: 04 March 2026

Cite this article as: Li Y., Luo Z., Zhang B. *et al.* Integrated multi-omics analysis reveals comprehensive metabolism-driven tumor heterogeneity and immune microenvironment in hepatocellular carcinomas. *Sci Rep* (2026). https://doi.org/10.1038/s41598-026-42856-7

Yu Li, Zihan Luo, Bingyu Zhang, Wenyao Zhu, Jinlian Zhang, Zizhao Li & Lingyu Zhang

Integrated multi-omics analysis reveals comprehensive metabolism-driven tumor heterogeneity and immune microenvironment in hepatocellular carcinomas

Yu Li [1,6], Zihan Luo[1,6], Bingyu Zhang[2,6], Wenyao Zhu[3], Jinlian Zhang[4], Zizhao Li[5], Lingyu Zhang[1]

Correspondence: Lingyu Zhang, Department of Laboratory Medicine, The First Affiliated Hospital of Bengbu Medical University, Bengbu, Anhui Province, 233004, People's Republic of China, Email lingyuzhang@zju.edu.cn.

[1]Department of Laboratory Medicine, The First Affiliated Hospital of Bengbu Medical University, Bengbu, Anhui Province, 233004, People's Republic of China;

[2]School of Public Health, China Medical University, Shenyang, Liaoning Province, 110112, People's Republic of China;

[3]Department of urology, the Central Hospital of Bengbu, Bengbu, Anhui Province, 233004, People's Republic of China;

[4]Department of pathology, the Second Affiliated Hospital of Bengbu Medical University, Bengbu, Anhui Province, 233004, People's Republic of China;

[5]Zongshan School of Medicine, Sun Yat-sen University, Shenzhen, 518107, People's Republic of China;

[6]Yu Li and Bingyu Zhang contributed equally to this work.

ABSTRACT

Background: Metabolic reprogramming is a well-recognized hallmark of cancer, characterized by its remarkable flexibility in activating alternative pathways in the absence of specific regulators or substrates.

Methods: Non-negative Matrix Factorization (NMF) was employed to identify tumor heterogeneity at the terms of metabolism, progression-related pathways and molecular subtypes of HCC based on gene set variation analysis (GSVA) and stratified survival analysis. The inherent heterogeneities of metabolism landscape which includes genomes, methelomic and transcriptomic data, proteomes, phosphoproteomes, mutational and immune microenvironment landscape between metabolic subtypes were performed based on multi-omics analyses.

Results: NMF has led three distinct survival-associated subtypes (NMF cluster 1, iHCC1; NMF cluster 2, iHCC2; NMF cluster 3, iHCC3) based on metabolic gene expression, and GSVA revealed remakeable metabolic differences. Additionally, multi-omics analysis further revealed the unique landscape of different metabolic subtypes, encompassing transcriptome, epigenetic and post-transcriptional modifications (PTMs) at both bulk and single cell seq-RNA level. Furthermore, 39 subtype-specific variables for identifying metabolic subtypes were screened using four feature selection algorithms and preliminarily validated on 8 machine learning models. We then built and verified a nomogram to guide the individualized strategy for HCC patients, utilizing a combination of metabolic signatures and clinical characteristics. Finally, we preliminarily identified the potential contribution of Aldolase B (ALDOB) in metabolic reprogramming triggered by epigenetic and PTMs.

Conclusions: Overall, the research defined robust subtypes and further revealed potential targets linking metabolism with immune microenvironment and non-mutational epigenetic modifications,

thereby advancing our understanding of metabolic heterogeneity for application in HCC diagnosis and clinical risk stratification.

Keywords: Hepatocellular carcinoma, Metabolic reprogramming, Epigenetic reprogramming, Post-translational modification, Multi-Omics Integration

1. INTRODUCTION

HCC is the predominant form of primary liver cancer worldwide, with cholangiocarcinoma being the subsequent type, accounting for over 90% of all cases and ranks as the fourth leading cause of cancer-derived mortality worldwide[1].The prevalence and risk factors for HCC include hepatitis B/ C virus infection, prolonged alcohol consumption, and nonalcoholic fatty liver disease (NAFLD) associated with metabolic syndrome (such as obesity) or dyslipidemia in individuals with type 2 diabetes mellitus (T2DM). The pathophysiology of HCC is a complex, multi-step, and multi-stage process driven by a intricate interplay of genetic susceptibility, viral and non-viral risk factors, microenvironmental substrates, immune cell activity, and chronic inflammation[2]. Metabolic abnormalities are a significant hallmark of cancer, as tumor cells undergo metabolic reprogramming to support their rapid proliferation. Simultaneously, metabolic alterations can further facilitate tumorigenesis and development through a specific interplay between metabolites and genes/ proteins[3,4]. Compared to normal tissues, cancer cells and tumors exhibit a diverse range of metabolic heterogeneity. This heterogeneity is characterized by varied metabolic activity within the tumor, as well as distinct metabolic preferences among tumor cells, resulting in disparities in drug response. The metabolic landscape consisting of genomes, transcriptomes, proteomes, metabolites, and their internal interplays between them is highly heterogeneous across cancer types, predominant histologic subtypes, and even single-cell levels under specific conditions[3,5,6].

The liver plays a crucial role in metabolizing three major nutrients in the body: namely glucose, lipids, and amino acids. During its progression, HCC exhibits various alterations in metabolic pathways, including heightened aerobic glycolysis and glucose consumption, augmented de novo fatty acid synthesis and pentose phosphate pathways, accelerated glutamine catabolism, and disrupted redox homeostasis. These metabolic changes provide the necessary energy and biological macromolecular raw materials for tumor cell growth and proliferation. Various enzymes and signaling molecules involved in multiple metabolic processes of HCC cells are indispensable[7-9]. Given that metabolic abnormalities precede the onset and progression of HCC, it is conceivable that these metabolic imbalances may be detectable in the early stages of HCC or even in precancerous lesions[6,10]. Consequently, changes in the activity of key enzymes within metabolism-related pathways and the metabolite content present opportunities for early diagnosis and novel therapeutic interventions for HCC. However, the metabolism of HCC tumor cells, such as genetic heterogeneity, exhibits obvious heterogeneity owing to different genetic backgrounds and tumor microenvironments. In essence, tumor cell metabolism does not exhibit universal alterations[11,12]. Instead, tumor cells generate diverse genetic variations during metastasis and treatment, leading to changes in their metabolic profiles in various directions[13,14]. HCC exhibits a complex network of metabolic pathways, including glucose metabolism, lipid metabolism, and amino acid metabolism. Targeting these pathways precisely is challenging due to their intricate interactions. HCC produces a variety of metabolic products during its progression, such as lactate, pyruvate, and fatty acids. These metabolites play pivotal roles in tumorigenesis and progression[15]. However, interventions targeting these metabolites are still in the early stages of research. The metabolic microenvironment of HCC is heterogeneous, with different tumor regions potentially exhibiting distinct metabolic features[16,17]. This heterogeneity poses significant challenges for treatment. Precision treatment targeting this heterogeneity is an essential research area. Yet, for HCC, precision treatment targeting its specific metabolic pathways remains a challenge[18]. Despite significant advancements in our understanding of

HCC metabolism, several gaps remain in previous research. One critical area that has been overlooked is the metabolic heterogeneity within HCC tumors. Previous studies have primarily focused on identifying common metabolic alterations in HCC, but have failed to address the significant variability in metabolic profiles between different patients and even within the same tumor. This metabolic heterogeneity can lead to variable treatment responses and the emergence of drug resistance, highlighting the need for more personalized and targeted therapeutic approaches. Consequently, a comprehensive exploration and comprehension of metabolic heterogeneity in HCC are critical for elucidating the molecular basis of metabolic abnormalities, accelerating the development of new metabolism-specific drugs, and promoting risk stratification and individualized management strategies for HCC patients.

Over the past decades, considerable efforts have been made to categorize HCC into several predominant molecular subtypes, each characterized by unique genomic alterations, mutational patterns, diverse prognoses, and/or activation of oncogenic pathways. To delve into the heterogeneity of HCC, numerous stratification strategies have been devised. Certain clinically significant subtypes have emerged as closely tied to specific pathways, histopathological features, and immune microenvironment characteristics[19-21]. Although previous studies have revealed several subtype-specific characteristics of HCC, it was difficult to reveal the comprehensive characterization of prognosis and progression with only a single or a small number of biomarkers. Therefore, as a result of the small sample size and a deficiency of multi-omics data, few studies have comprehensively investigated HCC from the perspective of metabolic disturbances and heterogeneity[22,23]. The present study aimed to comprehensively investigate which metabolic flow alters the overall metabolic preference and transcriptional signatures in HCC. Utilizing multi-omics data derived from 4 independent HCC cohorts, we successfully identified 3 metabolic subtypes. These 3 metabolic pathway-specific subtypes have well-marked metabolic preferences and gene expression patterns,

DNA methylation level, protein phosphorylation level, clinicopathologic features, and distinct prognosis. It is worthwhile to recognize the metabolic pattern associated with prognosis based on distinctly metabolic characteristics, which may facilitate biological insights in identifying potential therapeutic targets, as well as clinically unique subclasses that may affect the management strategies for HCC patients (**Graphical Abstract**).

2. MATERIALS AND METHODS

2.1 Data retrieval, pre-processing, and validation cohorts

RNA-seq gene expression profiling and associated phenotype data for HCC and non-cancerous liver tissue were obtained as transcripts per kilobase million values (TPM) from the Cancer Genome Atlas (TCGA, http://cancergenome.nih.gov/) and the International Cancer Genome Consortium (ICGC, www.icgc.org, http://lifeome.net/database/hccdb/download.html), and metabolic genes were selected based on MSigDB. To validate associations between metabolism-associated gene patterns and metabolic subtypes, an additional cohort holding of 221 HCC microarray samples and corresponding materials attained from GSE14520 were utilized[24]. The three HCC datasets were then combined into a single metadata set, and data were ComBat corrected for batch effects[25]. While overall survival (OS) and Recurrence-Free Survival (RFS) has been regarded as the standard primary clinical endpoint, and those without defined endpoints will be excluded. Figure S1 indicated principal component analysis (PCA) and t-distributed stochastic neighbor embedding (t-SNE) analyses before and after batch effect correction. Proteome and phosphoproteome data of 159 HCC patients were approved and provided by the researchers in supplementary materials[26]. The Random Forest model in multiple imputations by chained equations (MICE) was applied for the missing data imputation. Proteins and phosphorylation sites with more than 80% missing values were eliminated before imputation to ensure that each sample contained sufficient data for future processing. The imputation

method was implemented in the R package mice (version 3.16.0)[27,28]. A total of 983 HCC patients were enrolled in the study, and the baseline characteristics were summarized in Table S1.

2.2 DNA methylation data pre-processing and preliminary screening

TCGA provided a gene-specific DNA methylation profile of HCC patients. Illumina Human Methylation 450 Beadchip (450K array) was used to measure genome-wide methylation. Array quality assessment, data preprocessing, normalization, and filtration were performed in the R package Minfi (version 1.52.1). A total of 394201 CpG sites with applicable values throughout the genome were subjected to the following criteria: (i) probes with detection P-value ≥ 0.01 and < 3 beads in at least 5% of samples were eliminated; (ii) probes on the non-sex chromosomes and in CpG context; (iii) non-overlapping probes with single nucleotide polymorphisms; (iv) probes with stable genomic sites in the human genome[29,30]. We then executed data normalization and the β-value calculation in the R package minfi and ChAMP (version 2.36.0)[30-32]. The determination of DNA methylation levels (β-values) involved the calculation of the ratio of intensities between methylated and unmethylated alleles, which spanned from 0 (completely unmethylated) to 1 (completely methylated), respectively (Figure S2A-C). A total of 47474 probes from the DNA methylation chip were identified and annotated to 3124 gene signatures associated with metabolism (Figure S2D).

2.3 Non-negative matrix factorization (NMF) for subtype identification

Genes with a median absolute deviation (MAD) ≥ 0.5 were incorporated, resulting in a total of 3124 metabolism-associated genes were regarded as input matrix for clustering analysis in the R package NMF (version 0.27). To determine the optimal factorization rank (k), we performed NMF based on 50 random starting points for k values ranging from 2 to 10. Various separation qualities including silhouette width, cophenetic coefficients, and dispersion were evaluated for each factorization rank. The k values corresponding to a sudden decline in the magnitude of the cophenetic coefficient were

considered as the optimal number of clusters. The samples were then assigned to HCC subtype using the optimal k value on the coefficient matrix, and 100 iteration runs were performed from fixed random initial conditions. PCA and t-SNE plots were used to detect differences in expression between subgroups.

2.4 Gene set enrichment analysis (GSEA) and Gene set variation analysis (GSVA) analysis

GSEA analysis was performed in GSEA tool (https://www.gsea-msigdb.org/gsea/index.jsp, version: 4.3.2), and GSVA analysis was implemented in R package GSVA (version 1.52.3). Normalized enrichment scores (NES) were used to investigate the enrichment variation of biological processes in different phenotypes. Gene sets with a false discovery rate (FDR) q value and a nominal P value of < 0.05 were considered statistically significant in GSEA[33,34]. The normalized enrichment scores ranged from -1 to 1, corresponding to the relative enrichment score in each sample in GSVA. A Wilcoxon signed ranks test (P value < 0.05) was performed to compare enrichment scores in phenotype-matched samples for each gene set. A total of 140 metabolism-associated gene signatures, 31 cancer-relevant pathways, and 44 HCC subclass-relevant signatures were obtained from previously published studies using the MsigDB (Table S2)[5,26,35,36]. Subtype-specific pathways between at least two subgroups were included in further analysis.

2.5 Quantification of immune cell infiltration

The deconvolution algorithm took into account the transcriptome profile of heterogeneous samples by conceptualizing it as a convolution resulting from the presence of various cell types. It estimated the unknown cell abundance using a signature matrix describing cell type-specific expression profiles. Therefore, the relative proportions of 22 human immune cell subsets can be estimated quantitatively using a leukocyte gene signatures matrix (CIBERSORT deconvolution algorithm, https://cibersortx.stanford.edu/)[37,38]. The gene expression profile of HCC based on three RNA-seq

data portal was used as the input matrix for CIBERSORT, which calculated empirically defined deconvoluting global p-values for each sample.

2.6 Benefit prediction of immunotherapy and targeted therapy between HCC subtypes

Based on SubMap algorithm (Gene Pattern, https://www.genepattern.org/), available data from patients treated with immunotherapies were utilized to indirectly predict the potential benefits of three HCC subtypes by measuring the resemblance in gene expression patterns between subtypes and individuals who responded positively to anti-CTLA4, anti-PD1, Pembrolizumab, anti-PD-L1 or anti-MAGE-A3 therapy[39-42].

2.7 Identification of subtype-specific gene signatures and fuzzy C-means clustering

The R package Limma (version 3.60.3) was employed to identify differentially expressed genes (DEGs) within four subgroups, namely Normal (Control), iHCC1, iHCC2, and iHCC3, respectively. The samples are divided into two groups, for each subgroup, one containing sample from this subgroup and the other collecting the remaining samples from the data cohort. The Student's t-test was employed to conduct a comparative analysis between the two groups. Benjamini-Hochberg method was used for multiple inspections and correction of the obtained P values. Significant DEGs were assigned using a P value < 0.05 and a log2-transformed fold change of $\geq |\pm 1.5|$. Only genes exhibited significant differences across all 6 possible comparisons were considered as subgroup-specific DEGs. The fuzzy C-means algorithm of R package Mfuzz (version 2.66.0) was implemented to perform clustering analysis of DEGs according to HCC evolution orders, which concentrated genes with similar expression trends into one group. Normal (Control), iHCC1, iHCC2, and iHCC3 were sequentially assigned points 1, 2, 3, and 4, sequentially. Finally, DEGs were divided into 4 clusters with a membership threshold of 0.25[43,44].

2.8 Single-cell RNA sequencing data processing

The raw gene expression matrix was imported and subjected to processing in the R package Seurat (version 4.0.4). HCC scRNA-seq dataset was obtained from the GEO dataset[45]. HTseq calculated unique mapping reads and counts from the same UMI (Unique Molecular Identifier), which were then merged to obtain transcript counts for each gene from a single cell. Cells with < 8000 counts, 500 genes and > 20% of total UMIs mapping to the mitochondrial genome were excluded from the following analyses. After filtration, the scRNA-seq matrix was merged, and R package Harmony (version 1.2.0) (lambda = 1, max.iter.harmony = 20) was used to remove the batch effect[46,47]. The matrix was subsequently subjected to scaling and normalization using a library size factor so that every cell had 10,000 counts and was log2-transformed. The variance-to-mean ratio was then mployed to identify highly variable genes (HVGs) using the FindVariableFeatures function in R package Seurat (version 4.0.4). Significant genes (FDR $\leq 1e^{-3}$) were used to perform PCA analysis to reveal biologically significant variables, as determined by the application of. To ascertain the optimal cluster numbers, the FindClusters function (resolution = 0.8) was used[48,49]. Cells were projected into two dimensions using a Uniform Manifold Approximation and Projection (UMAP) dimensionality[50]. Cell types were identified using lineage-specific signatures[45]. The CellMarker (version 2.0) database was used for manual verification and correction[51]. The final annotation results were manually verified and visualized by UMAP.

2.9 Mutation landscape analysis

The DNA somatic mutation data for HCC patients were obtained from TCGA database. Additionally, three other mutation datasets are from cbioportal (Ulsan2014, FR2015) (https://www.cbioportal.org/)[52,53] and the Proteogenomic cohort. Somatic variant analysis was conducted to analyze and visualize the Mutation Annotation Format (MAF) using R package

maftools (version 2.20.0) [54]. A waterfall plot with variant frequencies depicted the mutation landscape.

2.10 Machine learning for subtype classifications and performance evaluation

R package caret (version 6.0-94) was utilized for model training and appraisal[55,56]. In brief, we initially determined ternary classification objectives based on metabolic subtypes, followed by conducting feature selection for each objective. Subtype-specific signatures were employed initially as an input matrix for iHCC identification. AdaBoost, XgBoost, Random Forest, and Boruta algorithms were adopted as the feature dimensionality reduction and elimination algorithms to identify the most significant features[57,58]. Venn diagrams were used to demonstrate the intersection of different algorithms and were regarded as detection features that should be incorporated into the final model construction. Subsequently, for each objective, we developed machine learning (ML) models based on the selected features. Furthermore, to evaluate the overall performance of the prediction models, a 5-fold cross-validation was conducted within the training set. Each fold underwent the aforementioned training procedure and was tested on the reserved portion of the training set. The specific algorithms were Decision Tree (DT), support vector machine (SVM), Random Forest (RF), linear discriminant analysis (LDA), k-nearest neighbor (KNN), extreme gradient boosting (XgBoost), Adaptive Boosting (AdaBoost), and MultiLR, respectively[59-61]. The HCC samples were randomly divided into 3 groups based on the subtype assignment collected from the NMF cluster (k = 3) so that each subtype was represented in 3 subgroups. The random segmentation was repeated ten times to obtain a reliable performance evaluation. The samples were randomly split into training (70%) and testing (30%) sets for each run, respectively. The parameters that yielded the highest cross-validation AUC for each classifier model were selected. Model performance comparison was employed based on the per-subtype confusion matrix and balanced accuracy (to adjust for the class imbalance

problem)[62,63]. Cohen's kappa was used to compare overall ML model performance for consistency inspection and classification accuracy measurement[61].

2.11 Development and external validation of metabolism-associated prognostic model

In the discovery cohort, whether the candidate signatures were relevant to the survival status was evaluated through univariate Cox proportional hazards regression analysis. Subsequently, the Elastic Net Regression was employed for further feature filtering. Stepwise Cox regression was utilized to identify an optimal subset for prognostic model construction. The risk score represented the weighted sum of each signature expression, considering their respective coefficients. The restricted cubic spline (RCS) was applied to visualize the optimal threshold, enabling the classification of all patients into high- and low-risk groups.

2.12 Human samples and immunohistochemistry assay

Resected HCC tissues and noncancerous liver tissues were obtained and granted by the clinical trial ethics committee (No. 2023023) diagnosed first by radiology or histology at the First Affiliated Hospital of Bengbu Medical University. The study has obtained written informed consent from participants or their legal guardians (for patients lacking autonomous capacity) and complies with the Declaration of Helsinki. Paraffin-embedded human HCC samples were cut in 5 μm sections, which were processed for histological staining and immunohistochemically (IHC) analysis was performed using a previously described method.

2.13 Statistics and reproducibility

Statistical analyses were conducted utilizing the R package (version 4.0.4, RStudio 2024.04.2 Build 764). The Fisher exact and Mann-Whitney U tests were used to compare categorical and continuous variables, respectively. Unless explicitly mentioned, all statistical tests (paired Wilcoxon rank-sum

test (Student's t-test, Mann-Whitney U-test), etc.) were two-sided, and adjusted P values were calculated using the Benjamini & Hochberg method for each comparison. Statistical significance was determined at a threshold of 0.05 or lower. Significance levels are as follows: *P < 0.05; ** P < 0.01; ***P < 0.001; N.S., insignificant.

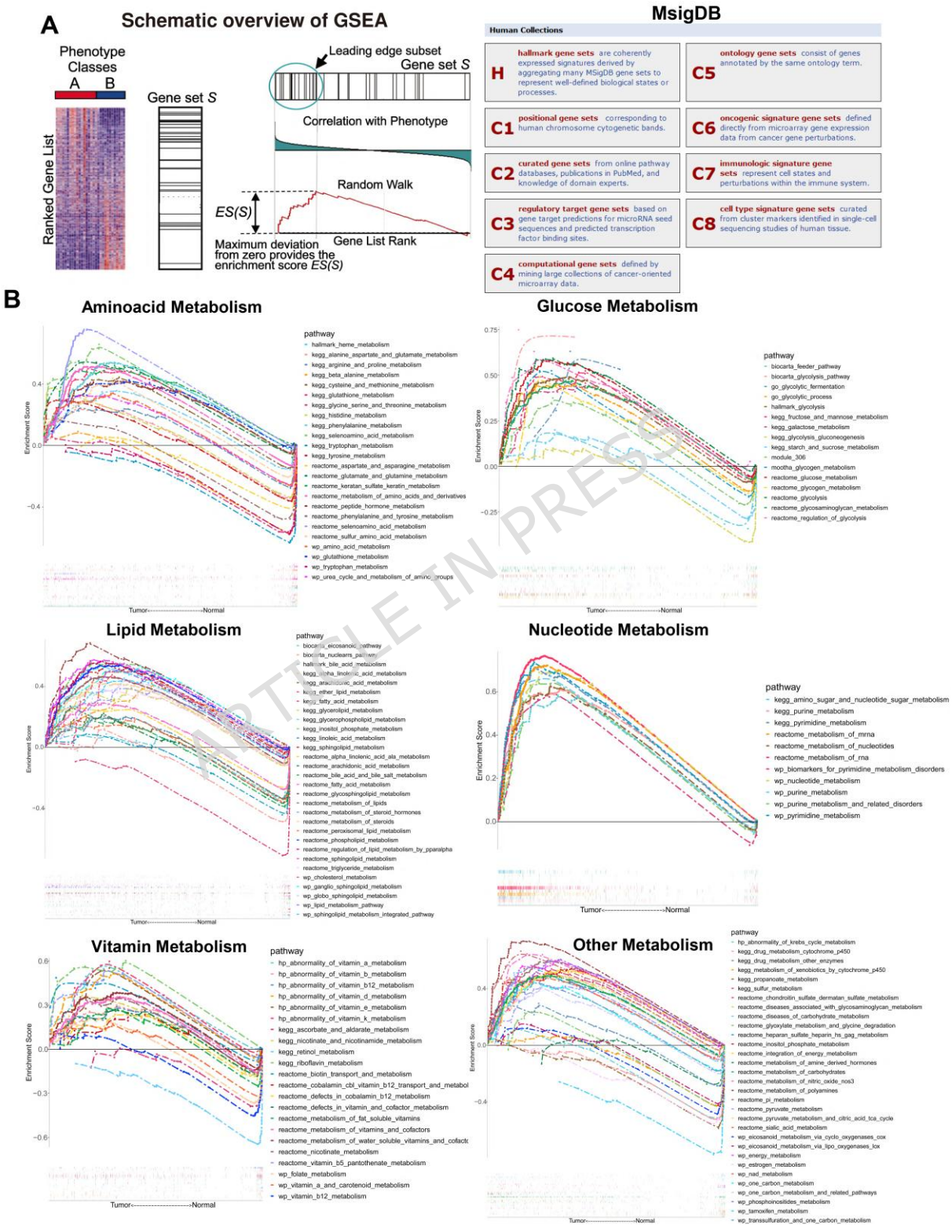
3. RESULTS

3.1 Gene set variation analysis (GSEA) displays metabolic dysregulation during HCC development

We obtained the gene expression profiles of HCC and normal liver tissues in TCGA, and searched metabolism-related genes in MsigDB (Figure 1A). Amino acid, glucose, lipids, nucleotide, vitamin metabolisms, and other related metabolic pathways were significantly dysregulated in HCC (Figure 1B, Table S3). Table S3 provided a comprehensive description of the difference in molecular pathways as determined by the GSEA analysis. NES and FDR q-value revealed an enrichment of metabolic pathways involved in different tissues. Activation of multiple metabolic pathways, particularly those related to nucleotides (such as the reactome metabolism of nucleotides), has been noted in HCC. The upregulation of these pathways is associated with increased levels of DNA synthesis and proliferation of tumor cells, suggesting that adaptive metabolic changes take place during the development of HCC. Anabolism and catabolism are two key biological processes involved in these metabolic pathways. A total of 3214 non-redundant metabolism-related genes were obtained through the extraction of genes from these gene sets. Furthermore, two other HCC RNA-seq profiles were obtained from ICGC and GEO databases, and HCC-related proteome/ phosphoproteome data were obtained from materials by other investigators. According to metabolic genes, expression profiles at gene and protein levels were extracted from these 4 HCC cohorts.

295 Heatmaps revealed the metabolic genes/ proteins difference in noncancerous liver tissues of patients

296 with HCC (Figure S3).



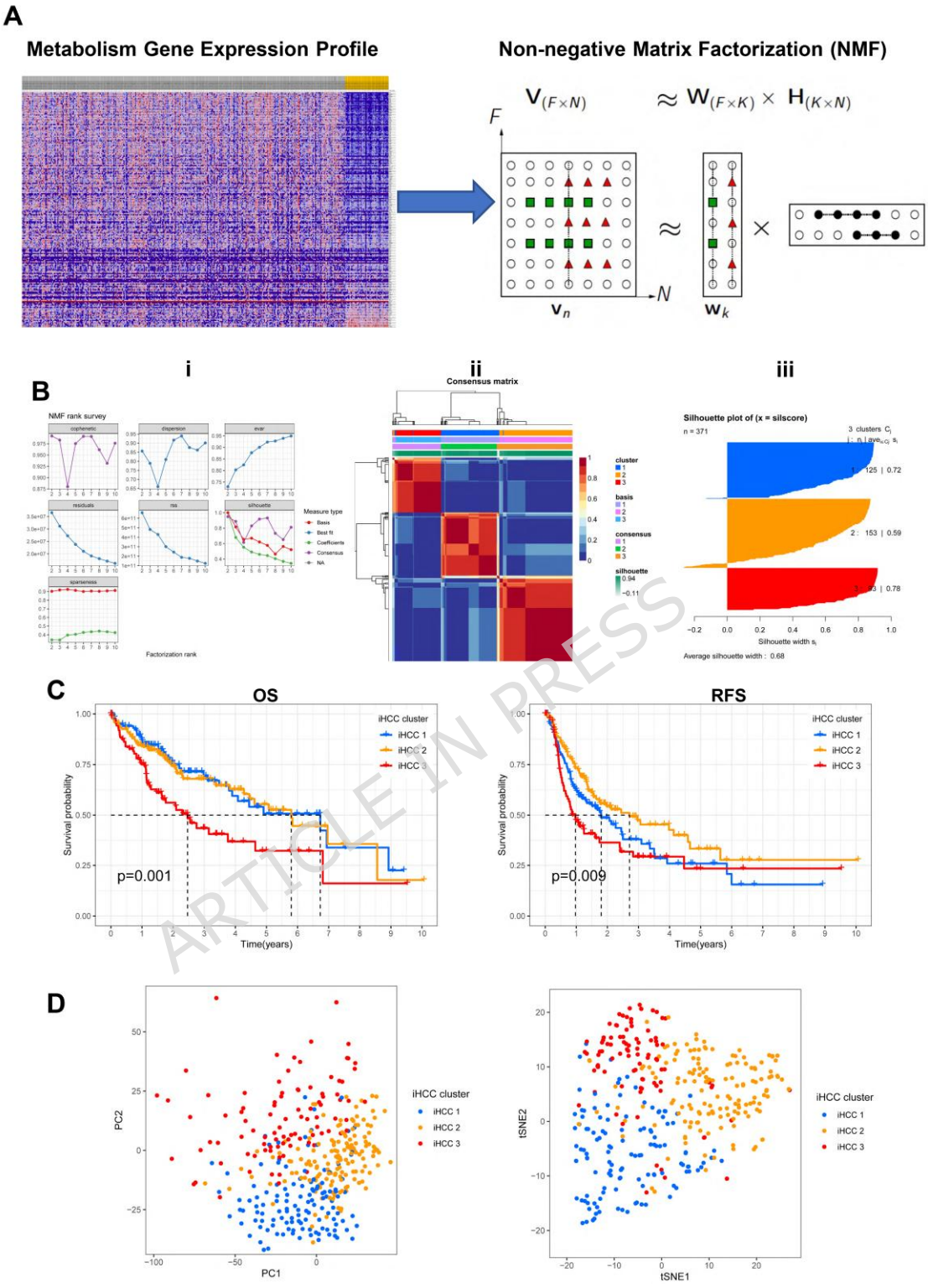
297

Figure 1. (**A**) Flow chart of Gene Set Enrichment Analysis (GSEA). (**B**) GSEA analysis showed that amino acid, glucose, lipids, nucleotide, vitamin metabolism and other related metabolic pathways were significantly dysregulated in HCC tissues.

3.2 NMF clustering identified distinct subtype-specific prognosis in HCC

In TCGA cohort, 371 HCC samples were categorized into 3 metabolic subtypes with different prognosis (iHCC1, iHCC2, and iHCC3) using unsupervised clustering based on the NMF algorithm (Figure 2A). The consensus matrix heatmap and silhouette coefficient revealed distinct boundaries between different subtypes, indicating that the clustering results of the samples were robust and reliable (Figure 2B). Among them, the iHCC3 subtype had the worst prognosis, with shorter overall survival (OS) (Figure 2C, left) and recurrence-free survival (RFS) (Figure 2C, right) than iHCC1 and iHCC2. The three metabolic subtypes were distinguished from one another using PCA and t-SNE dimensionality reduction techniques, revealing unique internal differences among the subtypes. (Figure 2D). We performed NMF clustering in two external RNA-seq cohorts (ICGC, and GEO), a Proteogenomic, and a merged cohort, to determine whether HCC can be robustly divided into three metabolic subtypes. These findings indicate that there are still three robust metabolic subtypes of HCC. Similarly, the iHCC3 subtype had the worst prognosis in different RNA-seq cohorts (Figure S4 and Figure S5) and the Proteogenomic cohort (Figure 3), with shorter OS and RFS times than the other subtypes. The Submap mapping algorithm was performed to compare metabolism-related gene (protein) expression patterns among the 3 metabolic subtypes. The iHCC2 and iHCC3 subtypes were significantly similar across different cohorts (Bonferroni corrected $P < 0.05$), indicating the robustness of iHCC2 and iHCC3 subtypes in HCC. In ICGC and Proteogenomic cohorts, iHCC1 was not significant (Bonferroni corrected $P > 0.05$) (Figure S6A). In addition, we investigated the

differential expression pattern of metabolism-related genes (proteins) among different metabolic subtypes. The heatmap illustrated clear disparities in the abundance of metabolic genes (proteins). Unique subtype-specific signatures in HCC, which may have driven the evolution of the different metabolic subtypes (Figure S6B).



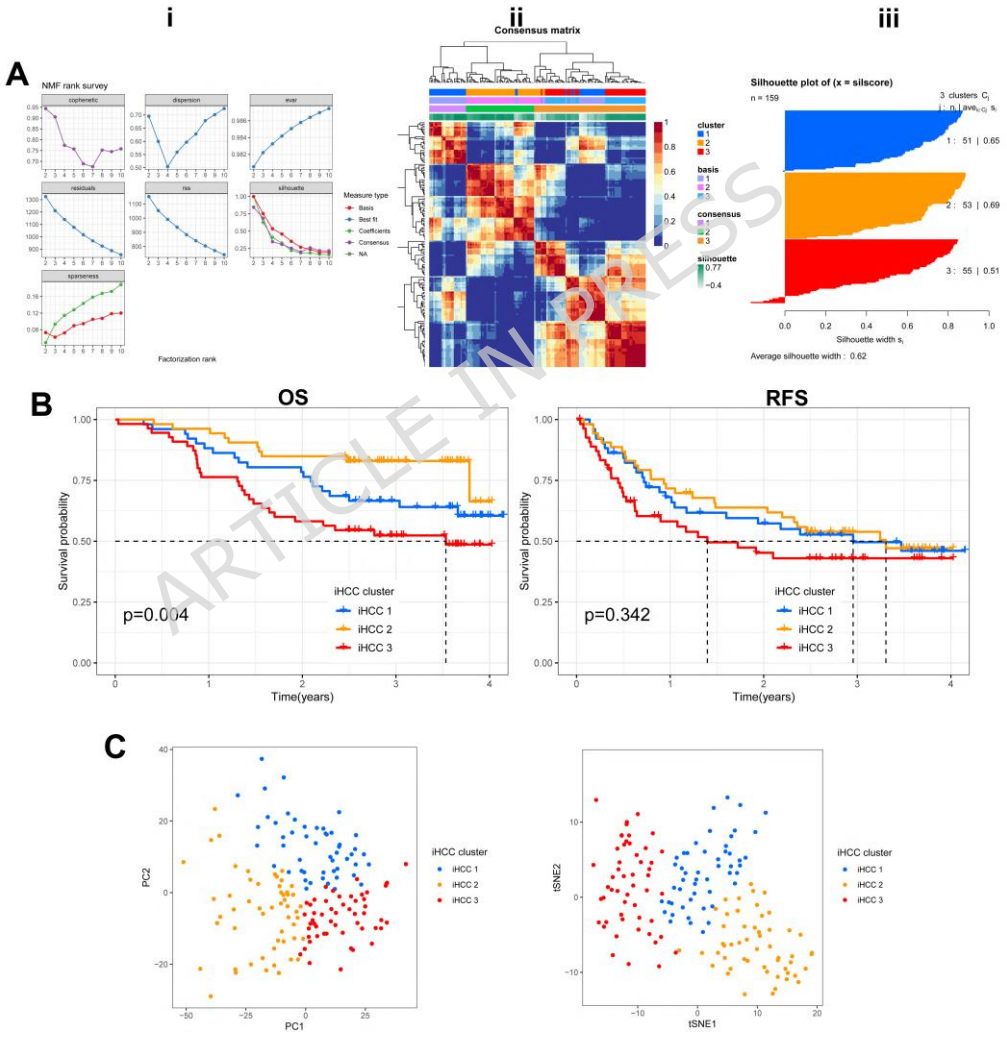
325

326 **Figure 2.** Identification of HCC subtypes using NMF consensus clustering in the TCGA cohort. (**A**)

327 Schematic of the NMF algorithm for unsupervised clustering of RNA-seq expression profiles. (i)

328 Strategy for selecting the optimal rank number (ii) Consensus matrix heatmap when NMF rank was 3

329 (iii) Silhouette plot displaying the composition (n = number of samples) and stability (average width)
330 of clustering (**B**) NMF clustering using 3124 metabolism-associated genes. The cophenetic
331 correlation coefficient for k =2-10 is shown (left); The consensus map (k = 3) of NMF classification.
332 Three HCC subtypes (iHCC 1, iHCC 2, iHCC 3) were identified in the TCGA cohort (middle); the
333 Silhouette Coefficient confirmed the robust subtype Classification (left). (**C**) OS and RFS of 3
334 subclasses (iHCC1, iHCC2, and iHCC3) in the TCGA cohort, with iHCC 3 having the worst
335 prognosis. (**D**) PCA and t-SNE analyses supported the stratification into three HCC subclasses.



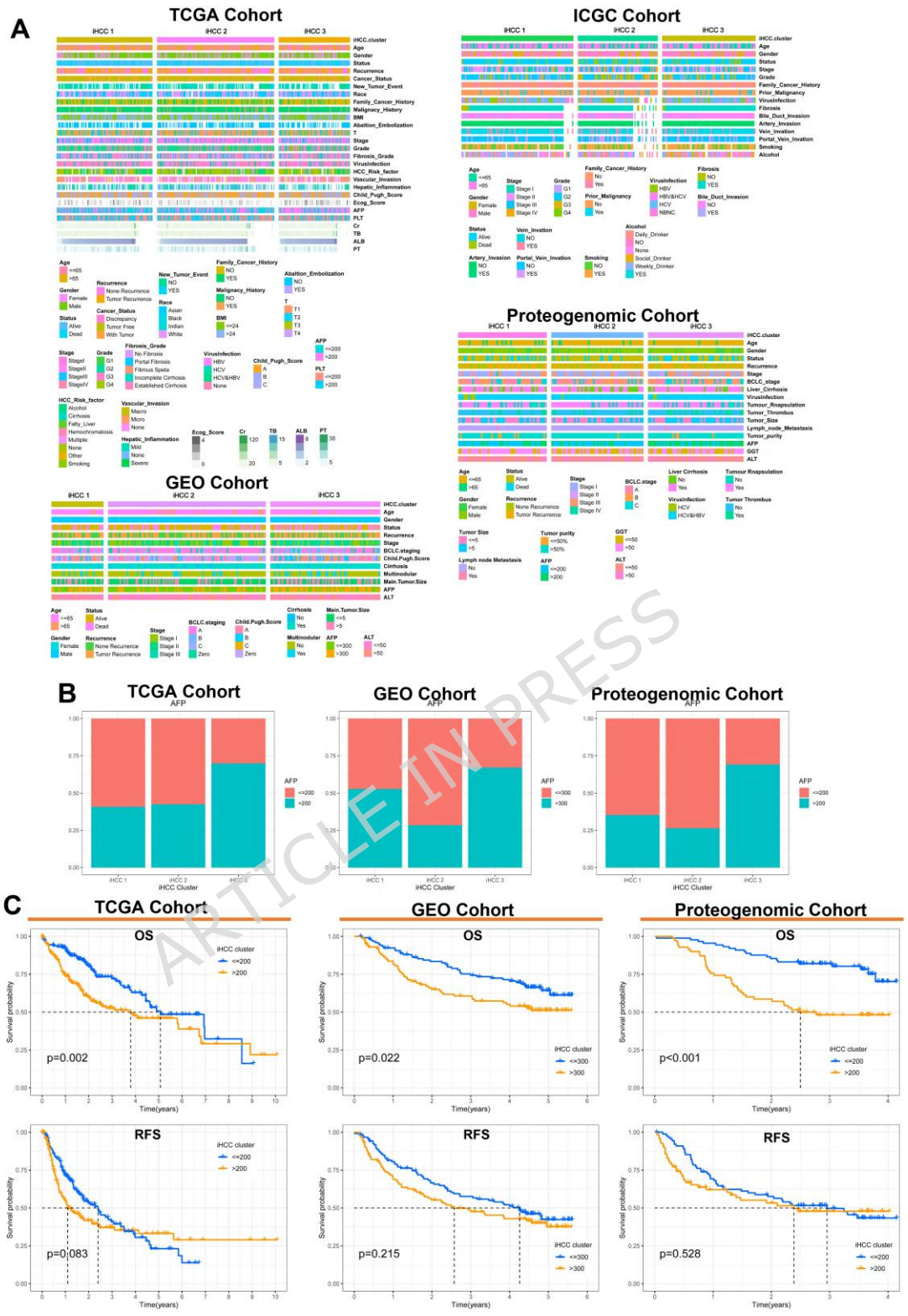
336

337 **Figure 3.** Consensus clustering for Proteogenomic data. (**A**) NMF clustering for k = 2-10 (left);
338 Consensus map for k = 3 (middle); the Silhouette Coefficient supported the robust 3 subtype

Classification (left). (**B**) OS and RFS curves based on 3 NMF clusters in the Proteogenomic cohort. (**C**) PCA and t-SNE projection of 3 HCC subtypes based on Proteogenomic data.

3.3 Characterizing different characteristics of HCC metabolic subtypes

The Chi-square test demonstrated a significant correlation between clinicopathological characteristics and HCC subtypes across different HCC cohorts. In TCGA cohort, iHCC3 had a higher histologic grade (G3/ G4) ($P < 0.001$), and a higher alpha-fetoprotein (AFP) level ($P < 0.001$) (Figure 4A, Table S4). In GEO cohort, iHCC3 was associated with a worse Child-Pugh score ($P < 0.001$), higher AFP level ($P < 0.001$), and higher pathologic stage (Stage III/ IV) ($P = 0.0379$) (Figure 4A, Table S4). Higher histologic grade (G3/ G4) ($P < 0.001$) and pathologic stage (Stage III/ IV) ($P = 0.00435$) were associated with iHCC3 in the ICGC cohort (Figure 4A, Table S4). AFP ($P < 0.001$), γ-Glutamyl Transferase (GGT) levels ($P = 0.00317$), and tumor purity ($P = 0.0354$) were higher in the iHCC3 in the Proteogenomic cohort (Figure 4A, Table S4). Consistently, higher AFP levels were strongly associated with the iHCC3 subtype (Figure 4B), and higher AFP generally predicted shorter OS or RFS survival (Figure 4C).



354

355 **Figure 4.** Clinical characteristics of 3 HCC subtypes. (**A**) Correlation of our classification (iHCC1,

356 iHCC2, and iHCC3) with clinical characteristics in the TCGA, GEO, ICGC, and Proteogenomic

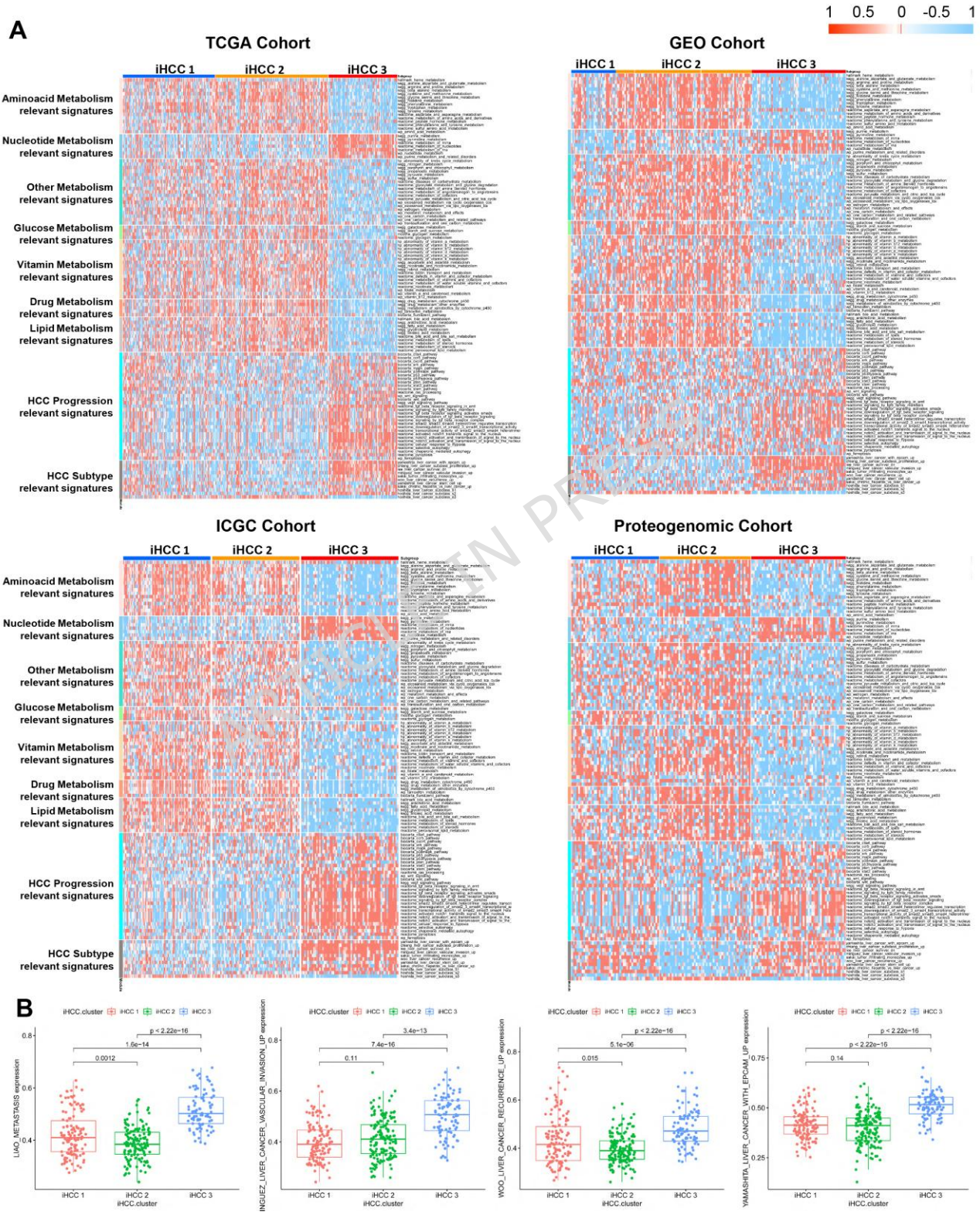
cohorts. (**B**) AFP distribution among different HCC subtypes, the iHCC 3 cluster has a higher level of AFP in the TCGA, GEO, and Proteogenomic cohort. (**C**) Kaplan-Meier curves show distinct prognoses among two AFP levels. Higher AFP levels usually predicted shorter OS or RFS.

3.4 Inter-tumor metabolic heterogeneity in HCC and distinct subtype-specific metabolism alterations

To further investigate the metabolic alterations between subtypes, we employed GSVA to quantify relevant metabolic pathways, followed by differential analysis to identify subtype-specific pathways. The heatmap revealed that iHCC3 had high nucleotide metabolic activity (for example, pyrimidine and purine metabolism) in the TCGA cohort, whereas pathways rather than nucleotide metabolism (Amino acid, glucose, lipids, vitamin metabolisms, and drug related metabolic pathways) were activated in iHCC2, a pattern consistent across other cohorts as well (Figure 5A). The overall landscape of iHCC1 was similar to that of iHCC2, but subtype-specific characteristics were preserved. In particular, ascorbate and aldarate, xenobiotics by cytochrome P450, heme and porphyrin, Krebs cycle, phenylalanine, porphyrins, pyruvate metabolism, TCA (Tricarboxylic acid) cycle, pentose and glucuronate interconversions, and fatty acid metabolism pathways were enriched in iHCC2 (Figure S7). Furthermore, we conducted a comprehensive review of the existing literatures to identify gene sets associated with HCC progression and previously reported HCC molecular subclasses. Through GSVA, we observed the activation of numerous pathways related to HCC progression and carcinogenicity in iHCC3, which aligned with the presence of subtypes with poorer prognosis. In detail, pathways associated with tumor metastasis, vascular invasion, recurrence, and EPCAM expression were up-regulated in the iHCC3 subtype (Figure 5B). The Wnt, Notch/PI3K-

379 AKT/ mTOR, TGF-β/ Smads, and hypoxia/ autophagy-associated pathways were also activated in

380 iHCC3 (Figure 5A, Table S5).

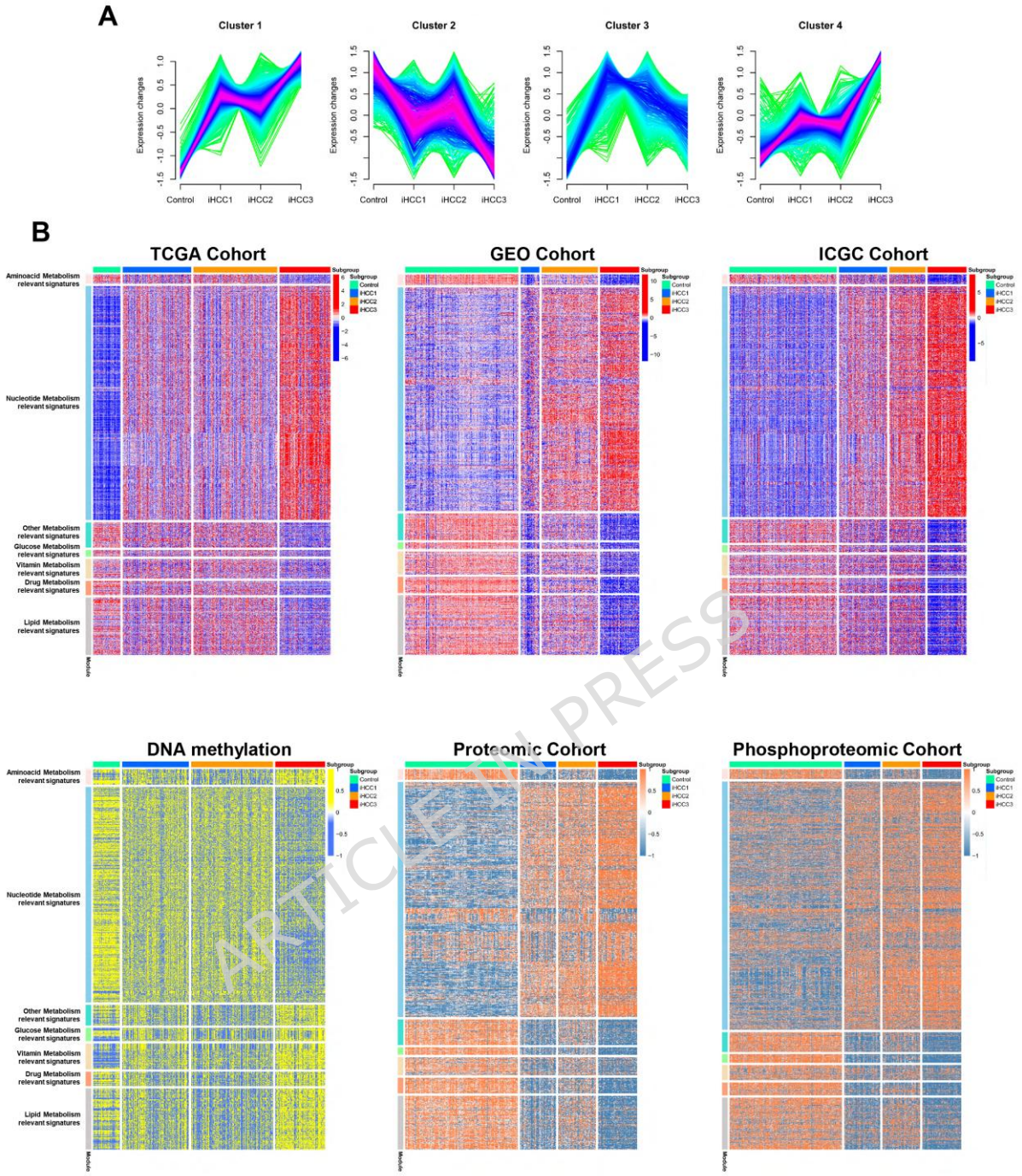


381

Figure 5. Integrated analysis of altered pathways at transcriptome and proteome levels among the 3 HCC subgroups. (**A**) Heatmap of alteration pathways between 3 HCC subtypes in the TCGA, GEO, ICGC, and Proteogenomic cohorts. The color of each cell represents the average ssGSEA enrichment scores for that subtype; red denotes activation and blue denotes inhibition. (**B**) A boxplot of the signature score for 3 HCC progression-associated signatures distinguished by different subtypes. P values are derived from the Kruskal-Wallis test. Boxplots show the median (central line), and the 25-75% IQR (box limits).

3.5 Fuzzy C-means clustering reveals a dynamic expression landscape of subtype-specific signatures

Four groups (Control group, iHCC1, iHCC2, iHCC3) of genes were clustered according to their distinct temporal expressing patterns using the fuzz C-means clustering algorithm. Some DEGs were specifically up-regulated or down-regulated between controls and iHCC subtypes (Figure 6A). The analysis of metabolic gene mapping to their respective pathways reveals that genes associated with nucleotide metabolism exhibited elevated expression levels in the iHCC3 subtype, while showing a significant decrease in expression in the control group. Genes related to other metabolic pathways were highly expressed in the control group and lowly expressed in the iHCC3 subtype. Consistently, similar trend was observed in GEO and ICGC cohorts (Figure 6B, top). Moreover, aberrant hypomethylation of promoters associated nucleotide metabolism and increased levels of proteins and phosphorylation were detected in iHCC3. Conversely, in normal liver tissues, genes involved in metabolic pathways other than nucleotide metabolism were hypermethylated and decreased protein abundance was accompanied by a low phosphorylation level (Figure 6B, bottom).



404

405 **Figure 6.** Analysis of metabolic gene expression patterns in various subgroups. (**A**) Fuzzy c-means
406 clustering identified two distinct temporal patterns of protein expression. The x-axis represents four
407 developmental clusters, while the y-axis represents log2-transformed, normalized intensity ratios in
408 each subgroup. (**B**) A global heatmap depicting the relative abundance of metabolism-related
409 signatures in HCC. Biological functions related to these signatures are denoted by the color panel on

the left and annotated according to the pathways based on the enrichment of MSigDB ontologies. Top: RNA-seq data in TCGA, GEO and ICGA; Bottom: Methylation, proteomic, and phosphoproteomic data.

3.6 Global landscape of metabolic heterogeneity at single-cell resolution

In addition, we endeavored to evaluate the metabolic heterogeneity of HCC at the single-cell level. UMAP dimensionality reduction identified 37 major clusters subdivided into eight major cell populations according to ElbowPlot and clustering tree in PCA algorithm, following the correction of batch effects (five CD8+ T clusters (C0, C1, C20, C27, C36), six Myeloid clusters (C2, C4, C5, C11, C29, C30), twelve Hepatocyte clusters (C3, C7, C9, C12, C14, C16, C18, C21, C25, C26, C31, C34), two Endothelial clusters (C6, C33), six CD4+ T clusters (C8, C19, C22, C24, C28, C32), one Fibroblast cluster (C10), one Cholangiocyte cluster (C13), and 4 B cell cluster (C15, C17, C23, C35)) (Figure S8). Canonical lineage markers were specifically expressed among different major cell populations (Figure S9). To further investigate the metabolic heterogeneity of HCC at the single-cell level, we initially analyzed the panoramic view of metabolic gene expression in these 37 primary clusters using UMAP based on 120 metabolic and HCC-related pathways. GSVA analysis revealed the cell-cluster-specific differences between 37 cells clusters in different pathways (Figure 7A). After mapping these cell clusters to the major cell populations, hepatocytes and cholangiocyte cells exhibited significant metabolic characteristics in amino acid, glucose, lipids, vitamin metabolisms, and drug related metabolic pathways. Conversely, CD4+ T cells displayed heightened nucleotide metabolic activity (Figure 7B). Then, 12 hepatocyte clusters were re-clustered into 15 major clusters, based on 3214 metabolic genes, instead of HVGs (Figure 7C). It was further identified that hepatocytes derived from normal tissues (NTL) were grouped into one cell population within the

designated region (a part of the dotted circle). Conversely, hepatocytes derived from tumor tissues (PT, MNL, and PVTT) were clustered into independent cell populations and separated from their tissue of origin, indicating that cells derived from HCC tumors also exhibited metabolic heterogeneity between individual patients (Figure 7D). Additionally, Figure 7D suggested that malignant hepatocyte clusters derived from the same patient (HCC7T and HCC7P, HCC08T and HCC08P, HCC10T and HCC10L) do not completely overlap, suggesting the presence of metabolic heterogeneity among them. The application of clustering analysis following UMAP dimensionality reduction demonstrated that the malignant hepatocyte cells exhibited discernible clusters, which aligned with the primary origin of their respective tumors in HCC (i.e. from which tumor site, stage or viral infection status the cells were derived) (Figure S10). Since hepatocyte cells derived from normal tissues were clustered and exhibited no observable interpatient heterogeneity within the same cell types, suggesting that metabolic gene expression in malignant cells was predominantly determined by patient-specific factors. Single-cell GSVA analysis depicted a striking difference in verious metabolic and HCC-related pathways between tumor cells derived from 15 hepatocyte clusters (Figure 7E). Furthermore, we tried to characterize the unique global metabolic landscape of different Patients, Samples, Tissues, Sites, Viral infection status and Stage, ultimately revealing the presence of location-specific metabolic heterogeneity (Figure S11A-E). The Pearson correlation also suggested three major subsets of hepatocyte clusters (Figure S11F) and the NMF algorithm also divided malignant hepatocytes into three major clusters (Figure S11G). Figure S12 showed that patients or tumor tissues with similar metabolic status are divided into the same NMF cluster (Figure S12A), whereas metabolic genes exhibited different global structures among different clusters (Figure S12B). To further determine major contributors to intra-tumoral metabolic heterogeneity of HCC (i.e. variation of metabolic characteristics among malignant cells from different HCC patients), CD4+ T, CD8+ T, B cell, and Myeloid cell clusters were then re-clustered based on 3214 metabolic genes. The UMAP clustering analysis revealed that there was no discernible heterogeneity among

different patients in terms of CD4+ T, CD8+ T, and B cell populations (Figure S13A-C). Figure S13D illustrated that the re-clustered Myeloid cells exhibited a higher degree of distinctiveness in their clustering patterns across different patients. Consequently, these cells were subsequently redefined into eight prominent cell clusters (Figure S13E), and GSVA analysis revealed high metabolic activity in myeloid-2, 3, and 6 (Figure S13F). We subsequently endeavored to identify the specific variation of metabolic pathway among the different myeloid cell clusters. Following this, we conducted enrichment scoring utilizing established markers for recognized cell types. Through this approach, we identified 11 major Myeloid populations (including both cytotoxic lymphocyte cells (CTL), Proliferative CD4 T cells, regulatory T cells (Treg), Mature B cells, MARCO+ macrophage, Dendritic cells (DCs), mucosal-associated invariant T cells (MAIT), resident memory T cells (Trm), Monocyte-derived macrophage, Plasma B cells, and MMP9+ macrophage) (Figure S14A-B). GSVA analysis demonstrated that Proliferative CD4 T cells showed higher metabolic activity compared to other major cell populations, followed by MMP9+ macrophage, MARCO+ macrophage, and Monocyte derived macrophage (Figure S14C). Stacked barplots showed significant depletion of CTL cells and enrichment of MARCO+ macrophage, Treg, and Proliferative CD4 T cells in tumor tissues (PT, MNL, and PVTT) as compared to normal tissues (NTL) (Figure S14D). Enrichment analysis revealed that multiple amino acid and vitamin-related metabolic pathways were activated in the CTL cell population (Figure S14E). We conducted GSVA analyses on each macrophage cell cluster and observed that metabolic heterogeneity in all macrophage cell types was primarily driven by variations in metabolic activities. More specifically, lipid-related metabolic pathways were enriched in the MARCO+ macrophage cluster, while the MMP9+ macrophage cluster demonstrated a predominant enrichment of pathways associated with nucleotide metabolism (Figure S14F). Collectively, these findings indicated that hepatocyte and Myeloid cells derived from HCC tumor exhibited higher metabolic plasticity, potentially resulting in patient-specific metabolic reprogramming of malignant cells as opposed to other supporting cells in the TME. Our comprehensive analysis determined

significant metabolic heterogeneity in HCC, which may be related to metabolic reprogramming during HCC occurrence and development.

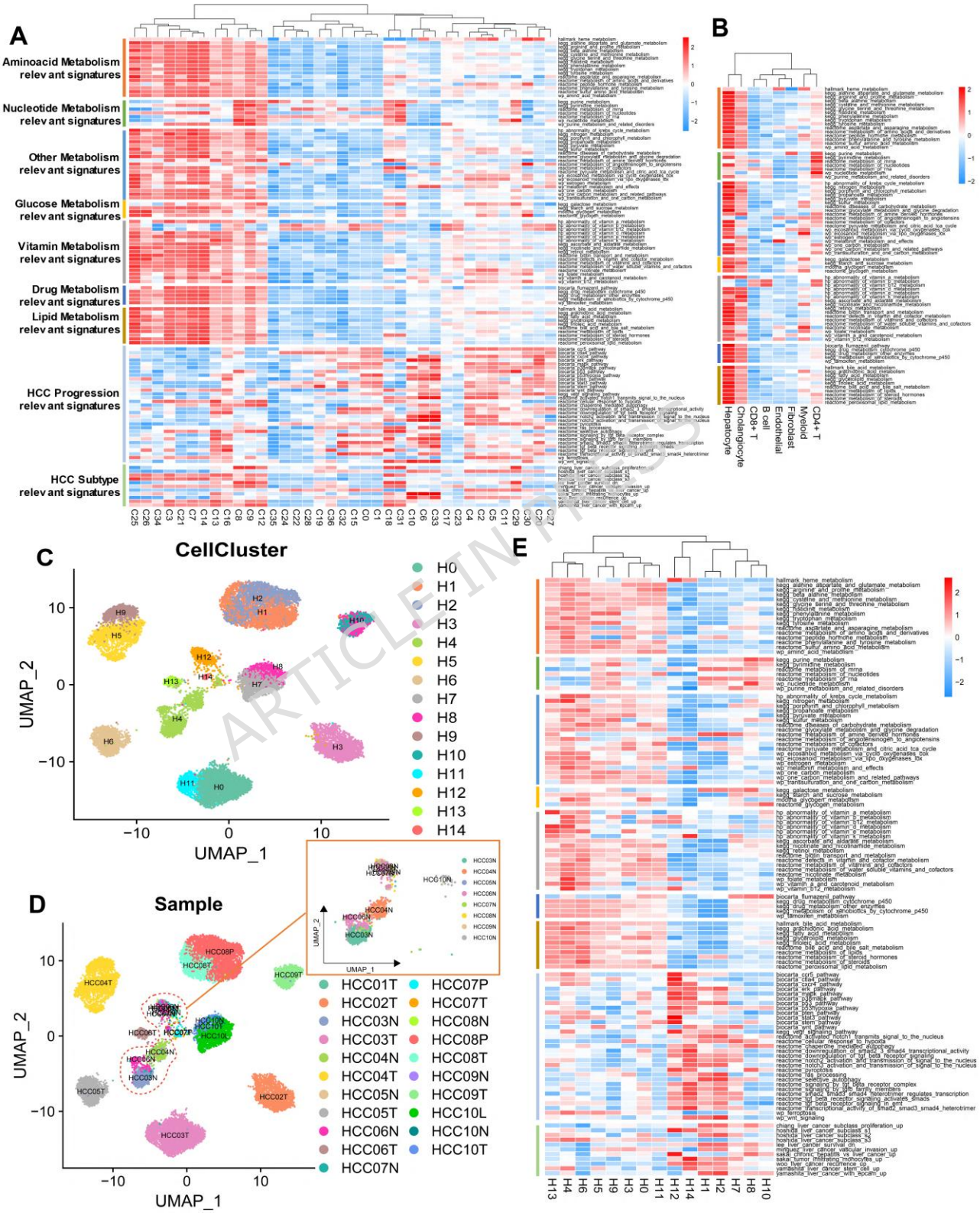


Figure 7 Unsupervised clustering revealed the metabolic heterogeneity of HCC at the single-cell level. (**A**) Heatmap of GSVA results for 120 pathways altered in 37 primary clusters; upregulated pathways are colored in red, and downregulated pathways are colored in blue. (**B**) The activity of metabolic pathway in 8 major cell populations. (**C**) Unsupervised clustering reclassified 12 hepatocyte cell clusters into 15 new clusters based 3214 metabolic genes. (**D**) UMAP plot of single-cell clusters from the multiple HCC samples. The color of each dot indicates the tumor which the cell comes from, and most tumor cells derived from a specific HCC patient are exclusive to a specific cell cluster. (**E**) Metabolic pathway activities in different hepatocyte clusters, metabolic activity varied significantly between clusters.

3.7 Immune microenvironment heterogeneity among the three HCC subtypes

Since iHCC3 held the worst prognosis, we further characterized their immune landscape across the 22 human hematopoietic cell phenotypes with the CIBERSORT algorithm from 3 HCC cohorts (Figure 8A-C). Specifically, CD8+ T cells and M0 Macrophages exhibited lower abundance in iHCC3, while a higher abundance of naive CD4+ T cells and M2 Macrophages in the TCGA cohort ($P < 0.05$). Consistently, this conclusion was also applicable in the GEO and ICGC cohorts (Table S6). We conducted additional analysis to examine the characteristics of immune checkpoint genes that may be targeted in these three subtypes. Figure 8D indicated that iHCC3 had significantly higher expression of 14 immune checkpoint signatures (excluding IL-6) compared to iHCC1 and iHCC2. This implied that different metabolic subtypes may yield disparate outcomes in forthcoming immune checkpoint inhibitor therapy, which needed further investigation. Our study compared the gene expression patterns between different subtypes of HCC and other patients undergoing various immunotherapies, including anti-CTLA4, anti-PD1, Pembrolizumab, anti-PD-L1, and anti-MAGE-

A3 therapies. The comparison was conducted using the TCGA and GEO cohorts, employing the Submap algorithm. Bonferroni-corrected P-values indicated iHCC3 was more promising to respond to anti-PD-L1 therapy, whereas iHCC2 was more favorable to the Pembrolizumab treatment (Figure 8E-F). Based on the Genomics of Cancer Drug Sensitivity (GDSC) database, we predicted the sensitivity of each subtype to 198 drugs using the R package oncoPredict (version 0.2). Figure 8G highly responded to the effectiveness of anti-pyrimidine drugs in iHCC3. Notably, PI3K/AKT, growth factor, GSK/ ERK/ TKI, and Wnt/ TGF-β pathway inhibitors may be applicable in different metabolic subtypes (Figure S15). These results emphasized the potential role of specific TME in tumor metabolic heterogeneity.

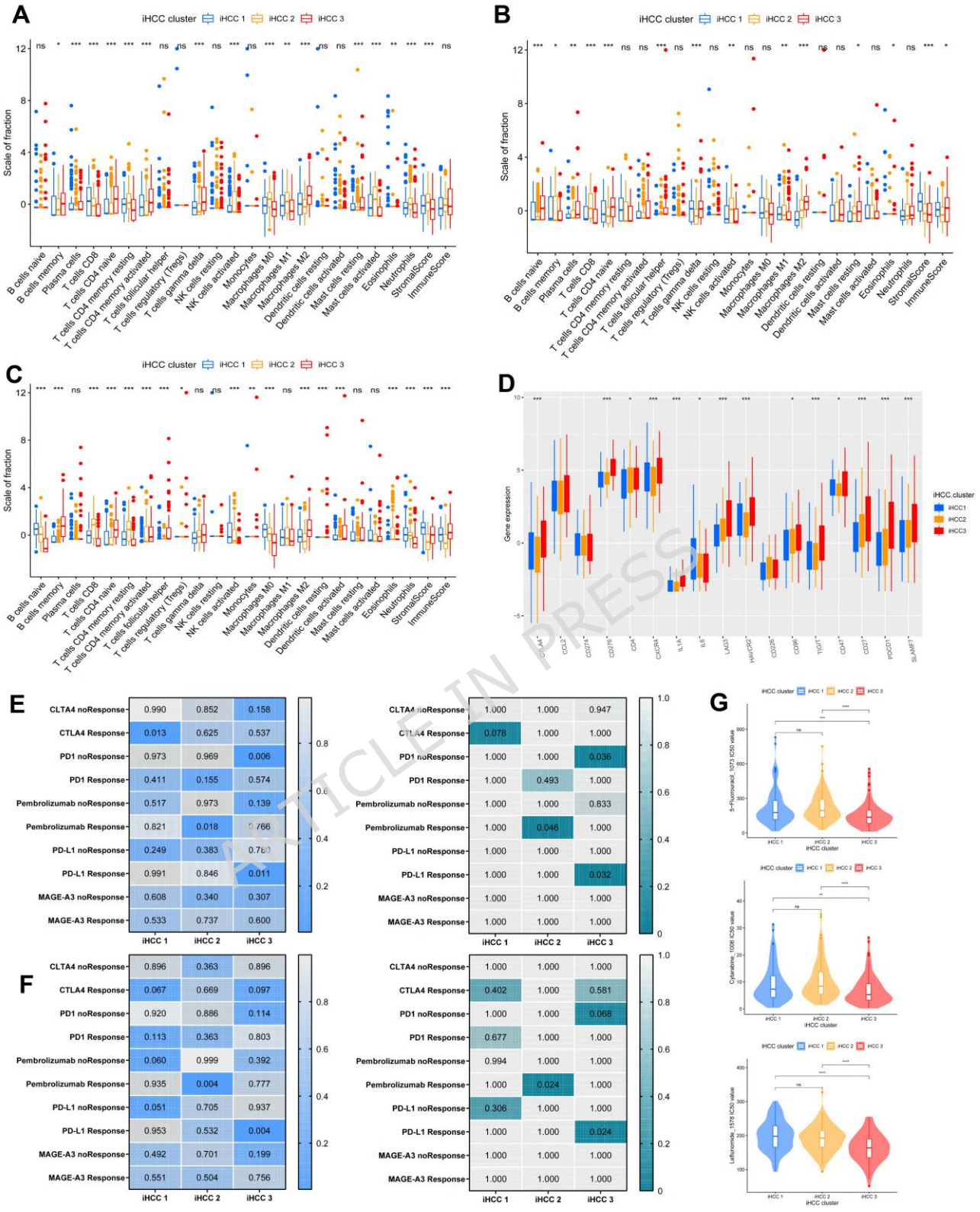


Figure 8 Immune characteristics of 3 HCC subtype. (**A-C**) Boxplot of the abundance of immune and stromal cell populations distinguished by different subclasses in the TCGA, GEO, and ICGC cohort.

(**D)** Expression levels (normalized count) of 17 immune checkpoint genes in three HCC subclasses. (**E**, **F**) Associations of the efficacy of conventional therapy with the HCC subtypes. The similarity of gene expression profiles between HCC subtypes and patients receiving anti-CTLA4, anti-PD1, Pembrolizumab, anti-PD-L1 or anti-MAGE-A3 therapies was shown in TCGA and GEO cohorts. (G) Sensitivity analysis of different subtypes to 5-fluorouracil, Cytarabine, and Leflunomide.

3.8 Mutation landscape of metabolic subtypes in HCC

Metabolic reprogramming and epigenetic modifications, which were inextricably linked and reciprocally regulate each other[26,64]. We analyzed the global mutational landscape in four independent cohorts. TP53 (22%-58%), TTN (21%-37%), MUC16 (12%-23%), and CTNNB1 (18%-36%) were frequently observed as missense mutation in HCC (Figure S16A). Further analysis revealed that iHCC3 had a higher frequency of TP53 mutations (iHCC1: 31/94 (24.8%), iHCC2: 37/116 (24.2%), iHCC3: 42/51 (45.2%), $P < 0.001$), while CTNNB1 mutations were more frequently observed in iHCC1 (iHCC1: 49/76 (39.2%), iHCC2: 20/133 (13.1%), and iHCC3: 22/71 (23.7%), $P < 0.001$) (Table S7). Stratified survival curves demonstrated that patients with TP53 mutations had significantly shorter OS ($P = 0.005$) and RFS ($P = 0.012$). Conversely, patients with CTNNB1 mutations exhibited a statistically significantly favorable prognosis than CTNNB1-wild-type patients in each HCC subtypes (OS: $P = 0.01$, RFS: $P = 0.002$), respectively (Figure S16B). Proteins enriched in CTNNB1-mutated and TP53-mutated HCC tissues have been associated with several metabolic pathways incorporating lipids, glycolysis/ gluconeogenesis, amino acid metabolism, and drug-related metabolism [26,65,66]. We performed a GSEA analysis among CTNNB1-mutated and CTNNB1-wild-type samples to investigate the molecular changes associated with CTNNB1 mutation. Amino acid, fatty acid, and drug-related metabolism pathways were enriched in CTNNB1-mutant tissues (Table

S8). In contrast, TP53 mutations were primarily associated with the activation of pyrimidine, polyamine, and glucose metabolism (Figure S16C, FDR q-value < 0.05) (Table S8).

3.9 Machine learning models for metabolic subtypes classification and performance evaluation

An attempt was made to identify subtype-specific signatures in the three HCC subtypes, based on AdaBoost, XgBoost, Random Forest, and Boruta algorithms. A total of 39 genes were selected for further analysis on the basis of the intersection of the four feature selection algorithms (Figure S17A). Utilizing these features, we built a multi-gene classifier based on eight machine-learning algorithms. The XgBoost model demonstrated superior performance in terms of balanced accuracy, overall accuracy, and Kappa score compared to the other seven machine learning models, resulting in improved classification performance on the test set (96.9~98.9% balanced accuracy, 0.973 overall accuracies with 95.8% Kappa score) (Table S9). These findings indicated that the multi-gene classifier was promising in identifying various metabolic subtypes in HCC, it needed to be validated in real-world datasets. Figure S17B and S17C demonstrated that these features are demarcated and independent of each other between 3 HCC subtypes and normal liver tissues, implying that subtype-specific genes may drive different metabolic subtypes. These results may reveal the potential role of subtype-specific gene signatures in driving the metabolic landscape of different HCC patients. Furthermore, we identified the expression of ALDOB, GNMT, HAAO, HMGCS2, IMPDH2, MTHFD1L, PYCR2, SLC10A1, and TRMT61A were significantly negatively correlated with CpG methylation (Figure S17D). Moreover, their expression levels and CpG site methylation levels significantly differed among normal liver tissues and metabolic subtypes. We hypothesized that this differential expression might be driven by DNA methyltransferases between different metabolic subtypes (Figure S17E), which needs to be further validated at the histopathological and cytological levels. Both bulk RNA-seq and single cell RNA-seq analyses confirmed the dysregulation of these nine genes in HCC (Figure S17F, Figure S18).

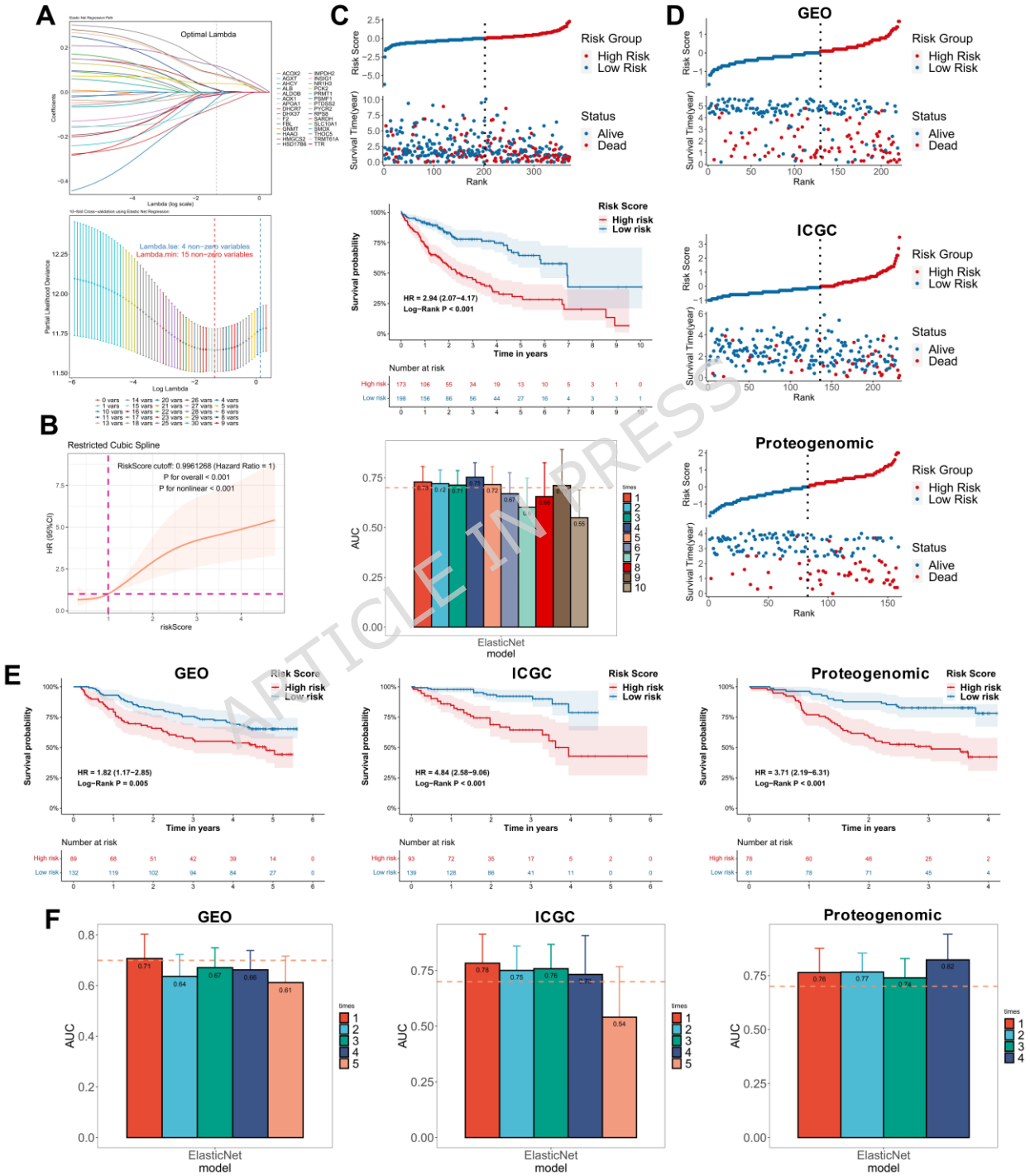
3.10 Development and validation of a machine learning model as a novel stratification strategy for HCC patients

To further expand the clinical availability of metabolic subtypes, 39 subtype-specific signatures were employed as an input matrix to construct a prognostic model. A total of 30 genes reached the threshold of $P < 0.2$ in univariate cox regression, then 15 non-zero variables were further identified through Elastic Net regression analysis (Figure 9A). Finally, DHX37, FBL, HMGCS2, ALDOB, NR1H3, and TRMT61A were identified by stepwise multivariate regression analysis in the final model. The model exported the tumor metabolic index (TMI) of each patient by the formula below: (0.602*DHX37) + (0.040*FBL) + (-0.0571 * HMGCS2) + (- 0.383 * ALDOB) + (-0.018 * NR1H3) + (0.15 * TRMT61A). The RCS curve effectively determined the optimal threshold value at which the hazard ratio (HR) equaled 1 (Figure 9B). 173 HCC patients were assigned into the high-risk group, while the other 198 were classified into the low-risk group. The KM plot revealed a 194% increased risk of death in the high-risk group compared with the low-risk group (Log-Rank $P < 0.001$). The ROC curve (AUC) showed that the TMI model provided better prognostic prediction efficiency at every temporal node (Figure 9C). Three independent validation cohorts confirmed the robustness of the TMI model, revealing that high TMI had a higher risk of mortality (GEO cohort: HR = 1.82, P = 0.005; ICGC cohort: HR = 4.82, $P < 0.001$; Proteogenomic cohort: HR = 3.71, $P < 0.001$) (Figure 9D and 9E), and AUC indicated the prediction efficiency of the model at different time points (Figure 9F). Furthermore, in four independent cohorts, both univariate and multivariate Cox regression analyses identified TMI as a significant independent prognostic factor ($P < 0.05$) (Table S10). To provide better clinical applicability, we developed a multivariate nomogram based on risk score, and various clinical features (Figure 10A) to intuitively understand the 1-, 2-, 3- and 5-year survival probabilities of each HCC patient. A good degree of consistency between the calibration curve and the ideal curve was observed (Figure 10B). Time-dependent AUC demonstrated

592 better accuracy compared to clinical models, TMI, and other single features at different time nodes

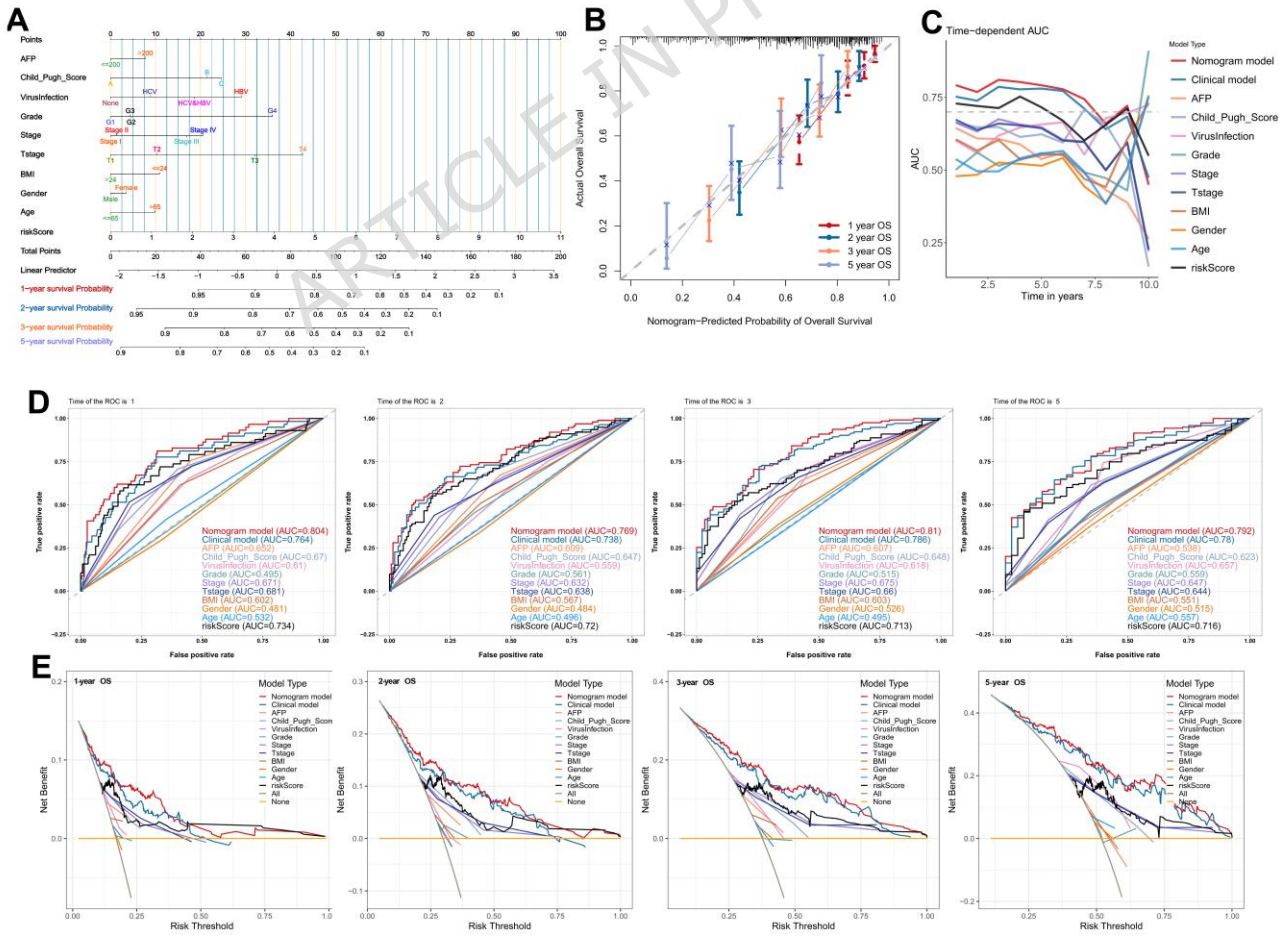
593 (Figure 10C-D). Furthermore, the application of decision curve analysis (DCA) demonstrated that the

594 Nomogram model exhibited a superior net benefit across various hierarchical strategies (Figure 10E).



595

Figure 9. Construction and validation of the TMI in TCGA cohort. (**A**) Elastic Net regression revealed the partial likelihood deviance in the 10-fold cross-validation. The red and blue vertical dotted lines were drawn at the minimum and 1-SE criteria value. (**B**) The restricted cubic spline determined the best cutoff for the risk score when HR was 1. (**C**) 371 HCC patients were divided into high- and low-risk groups based on the RCS cutoff value and represent survival difference between high- and low-group (top); KM survival curve of OS between two subgroups (medium); ROC analyses at different time nodes (bottom). The orange dashed line indicated the threshold of 0.70. (**D**) HCC patients in three independent cohorts were divided into high- and low-risk group based on the optimal cutoff in the TCGA cohort. (**E**) High risk scores were associated with poor overall survival in three validation cohorts. (**F**) Time-dependent ROC curves of prognostic prediction on the 1-, 2-, 3-, and 5-year survival rates for three validation cohorts.



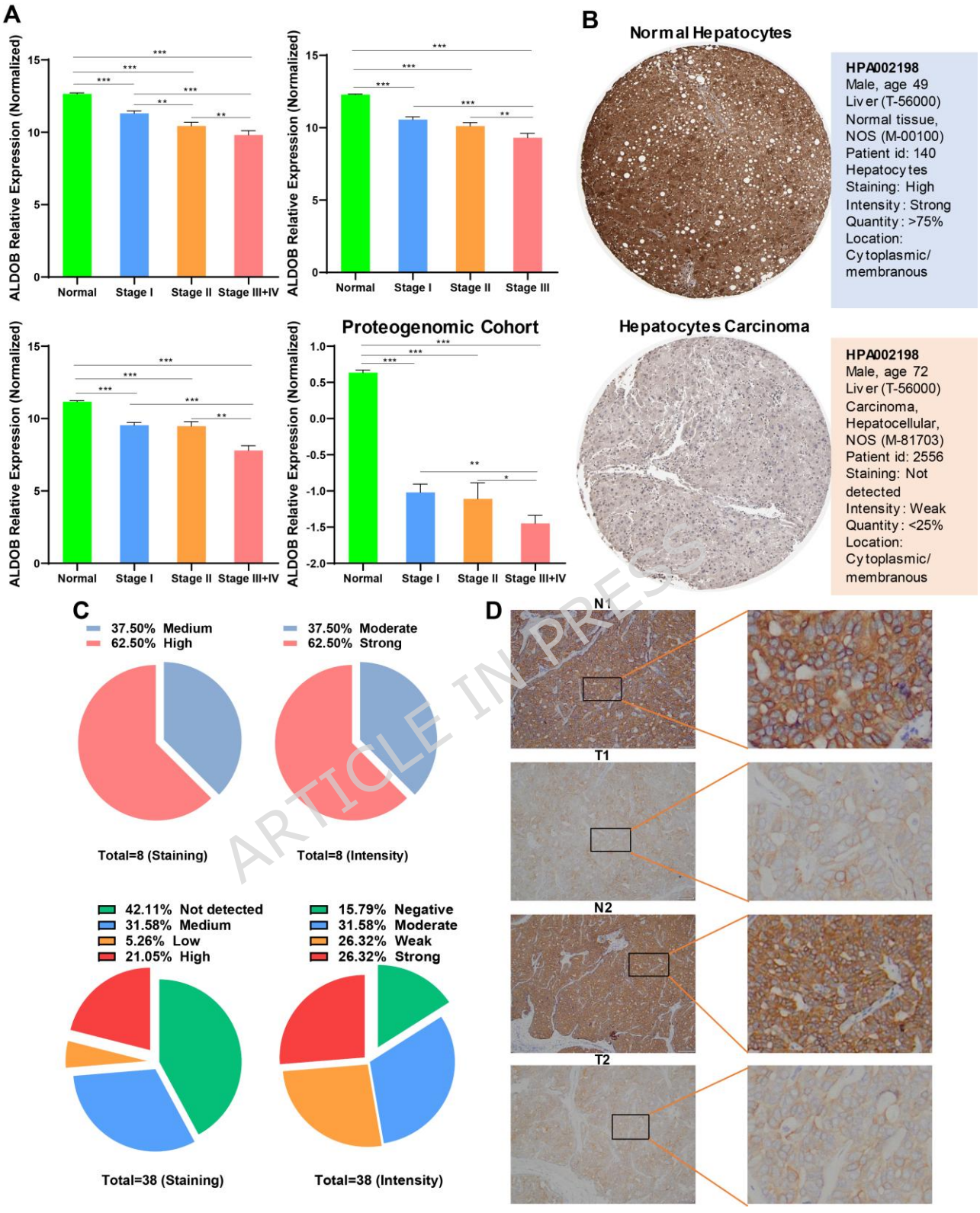
608 **Figure 10.** Nomogram developed and evaluated for predicting the probability of 1-, 2-, 3- and 5-year
609 OS in TCGA cohort. (**A**) The nomogram was constructed, with the TMI and clinical features
610 incorporated. (**B**) Calibration plot of the nomogram in terms of agreement between the predicted and
611 observed. (**C**, **D**) Time-dependent area under the ROC curve for the comparison of TMI with other
612 model. (**E**) Decision curve analysis of the nomogram for 1-, 2-, 3- and 5-year risk.

613

614 **3.11 ALDOB represents a potential prognostic biomarker and therapeutic target in HCC**

615 Previous results have demonstrated a significant down-regulation of ALDOB in HCC, particularly in
616 iHCC3. Furthermore, the KM plot analysis revealed a correlation between decreased ALDOB
617 expression and reduced OS and RFS (Figure S19A). Further analysis also demonstrated dramatically
618 decreased phosphorylation of ALDOB at eight sites (Figure S19B). We also observed that the
619 methylation level of multiple CG sites on ALDOB was negatively correlated with gene expression
620 (Figure S19C), and there were numerous CpG islands in the promoter region, suggesting that DNA
621 methylation may mediate the deficiency of ALDOB (Figure S19D). DNA methyltransferases
622 (DNMTs) are a family of "writer" enzymes responsible for DNA methylation. As the three active
623 enzymes that maintain DNA methylation, our results showed that they were significantly negatively
624 correlated with ALDOB expression and had opposite expression pattern to ALDOB in normal liver
625 tissues and three HCC subtypes (Figure S19E-F). Further, we observed a continuous decrease in
626 ALDOB with HCC progression in four cohorts (Figure 11A). In HPA database, ALDOB showed
627 weak or moderate protein levels in HCC tissues compared to normal liver tissues (Figure 11B-C). We
628 verified the low protein levels of ALDOB in HCC by immunohistochemical staining (IHC) in real
629 world cohort (Abcam, ab153828, 1:500) (Figure 11D). In previous results, we observed significant
630 upregulation of the Notch, PI3K/ AKT, mTOR, and Wnt pathways, which were closely associated

631 with lipid metabolism (Figure S20A). Furthermore, GSEA analysis showed that high levels of
632 DNMTs and low levels of ALDOB were closely associated with activation of Notch, PI3K/ AKT,
633 mTOR and Wnt pathways and dysregulation of lipid metabolism pathways (Figure S20B). The above
634 results suggest that the DNMTs/ALDOB/Notch/PI3K-AKT/mTOR metabolic axis may play a role in
635 the occurrence and development of HCC, which needs further verification.



636

637 **Figure 11.** The expression level of ALDOB in normal liver, and HCC tissues. (**A**) Pooled analysis

638 showing ALDOB mRNA level in different pathology grade. (**B**, **C**) IHC results indicating deficiency

of ALDOB in HCC tissues from the Human Protein Atlas database. (**D**) The representative immunohistochemical staining images showing ALDOB protein expression in HCC and normal liver tissues. Scale bar: 100 μm.

4. DISCUSSION

HCC is one of the most common categories of liver cancer, and effective treatment often varies significantly among patients owing to inherent tumor heterogeneity. A comprehensive and hierarchical approach to managing HCC patients is anticipated, aiming to elucidate the molecular basis of heterogeneity and identify potential therapeutic targets. Early-stage diagnosis and management of HCC can be effectively achieved through a hierarchical framework based on pathological and histological classification. Recognizing high-risk HCC patients to reduce mortality and customize individualized treatment before the emergence of severe symptoms and complications remains a critical task. In this study, we stratified and characterized each HCC subtype by molecular signatures using an unsupervised algorithm (NMF) based on the expression profiles of metabolic genes and/ or proteins. Integrated multi-omics analysis identified 3 distinct HCC subtypes, each with significant differences in metabolic and HCC-related pathways projecting at the genomic, transcriptomic, proteomic, and phosphoproteomic levels. These subtypes differ in clinical outcomes due to altered metabolic activities, Wnt, Notch/PI3K-AKT/mTOR, TGF-β/ Smads, and hypoxia/ autophagy-associated pathways, etc.

Compared to previous studies, patients classified under the iHCC3 subtype, which is associated with the poorest survival and disease progression, exhibited several indicators of reduced survival and an increased likelihood of disease recurrence. In contrast, in the high- and moderate-survival groups, iHCC1 and iHCC2 were associated with signatures of low recurrence, high survival, and

tumor cell differentiation. Furthermore, the highest levels of AFP and GGT were observed in iHCC3, accompanied by a higher pathological stage (Stage III/IV) or grade (G3/G4). Regarding prognosis and pathway activities, iHCC1 exhibited similarities to iHCC2 overall, while still maintaining subtype-specific characteristics. Notably, with the exception of nucleotide metabolism, nearly all metabolic pathways showed significant down-regulation in iHCC3 compared to iHCC1 and iHCC2. Subsequent analysis suggested that this phenomenon may be linked to hypermethylation and reduced protein phosphorylation of the relevant metabolic signatures. Previous studies have also reported the down-regulation of multiple metabolic enzymes to be associated with a poor prognosis in HCC[67-70]. Similarly, phosphorylation of proteins associated with metabolic enzymes played a crucial part in the metabolic reprogramming of HCC, including phosphoproteomic alterations in glycolysis, fatty acid-lipid metabolism, bile acid metabolism, TCA cycle, and xenobiotic metabolism[26,71]. We also conducted a pan-cancer GSVA investigation. HCC demonstrated the most distinct metabolic pattern separated from other 14 common cancer types (Figure S21), highlighting the significance of remarkable metabolic alterations in HCC.

Previously reported HCC molecular subclasses, including Yamshita's classification ($EpCAM^+$, $EpCAM^-$), Chiang's classification (CTNNB1, Proliferation, IFN-related, polysomy 7, and an unannotated class), Lee's classification, Minguez's classification, Sakai's classification, Woo's classification, and Hoshida's classification (termed S1, S2, and S3) have all been used to compare with our subtype classification. Generally speaking, our classification validates previously reported HCC subclasses while retaining their characteristics. Specifically, iHCC3 corresponded to the $EpCAM^+$ subtype, accompanied by hepatic progenitor cell features and high AFP levels[72]. iHCC1 corresponds to Chiang's CTNNB1 subclass, whereas iHCC3 corresponds to the proliferation subclass, with IFN-related and polysomy 7 subclasses showing a stronger match with iHCC2[73]. The high frequency of CTNNB1 mutations in the iHCC1 subtype suggested a potential benefit from

686 inhibitors targeting Wnt signaling[26,74]. Hoshida's S1 and S2 subtypes are more associated with
687 iHCC3, characterized by high proliferation, low differentiation and activation of MYC, AKT, and
688 Wnt[75]. In addition, iHCC3 has been associated with vascular invasion[76], intrahepatic metastasis[77],
689 higher mononuclear cell infiltration [78], and a disappointing outcome [79,80]. The metabolic subtypes in
690 our study were consistent with the previously identified molecular subtypes.

691 Metabolic reprogramming is driven not only by a series of oncogenic alterations but also
692 through complex interactions with the tumor ecosystem, influencing tumorigenesis and progression.
693 It is also worth noting that the three metabolic subtypes differed significantly in the immune
694 microenvironment. Prognostic features include the abundance and characteristics of immune cells in
695 the TME and response to immunotherapy. Notably, the iHCC3 subtype is characterized by low levels
696 of CD8 T cells and high infiltration of M2 macrophages, which have been associated with poorer
697 outcomes in previous studies[45,81-83]. In addition, the differential expression of immune checkpoint
698 signatures predicted the response of different metabolic subtypes to immunotherapy[84,85]. Moreover,
699 the findings of the present study demonstrating the reciprocal interactions between the anti-tumor
700 microenvironment and metabolic heterogeneity in HCC may provide mechanistic insights into the
701 benefits of immunotherapy. However, the robustness of the predictive capability of the metabolic
702 phenotyping assay needs to be further validated prospectively in a larger patient cohort before
703 translating to the clinic, including responses to immunotherapy and small molecule inhibitors.

704 The present study investigated a range of subtype-specific gene signatures among 3 iHCC
705 subtypes. We attempted to determine the optimal combinations of features to improve outcome
706 prediction and identify reproducible metabolic subtypes in HCC. Our findings indicate that the
707 XgBoost model provides better discrimination and can recognize different metabolic subtypes with
708 high sensitivity and specificity using a reduced set of 39 signatures. These results provided a
709 potential combination of diagnostic biomarkers for metabolic subtype classification (Table S5).

These signatures represent different metabolic proclivities that may contribute to metabolic heterogeneity in HCC. Furthermore, significant alterations caused by hypermethylation or hypomethylation of genes may lead to the reprogramming of HCC (ALDOB [86,87], GNMT [88], HAAO [89], HMGCS2 [90], IMPDH2 [91], MTHFD1L [92], PYCR2 [26,93], SLC10A1 [26,94], and TRMT61A [95]) were observed, suggesting the potential role of DNA methylation in the metabolic reprogramming of tumors[96-98]. Finally, we identified the DNMTs/ ALDOB/ Notch/ PI3K-AKT/ mTOR metabolic axis as a potential marker and therapeutic target for HCC patients. Meanwhile the internal validation was performed in an unforeseeable test set. Further validation in external cohorts would be beneficial before clinical translation. We aimed to capture multi-omics-integrated data, including radiomics, metabolomics, pathomics and even hematological tests. These multi-omics data should be prospectively extracted from new samples drawn from a larger population and used for independent validation. This would facilitate the ML model more robust and generally applicable for a multi-centres disposition[99,100]. The refined ML model can be used as a webserver-available classifier in primary care to identify HCC patients. Those identified as high risk will be referred to a medical center with appropriate expertise for an individualized appraisal, if necessary. Overall, this presents as an extremely promising pilot study which needs further validation in a larger cohort.

Regarding intra-tumor metabolic heterogeneity, our findings suggest that the major hypometabolic subtype (iHCC3) has significant prognostic significance in HCC, regardless of the presence or absence of other subordinate subtypes. Further investigation into the metabolic subtypes of HCC in advanced cases through the analysis of tumor biopsies or biofluid-based metabolomics is necessary to understand intra- and inter-tumoral metabolic dysregulation that may aid in early detection and treatment of HCC. Therefore, we conducted extensive GSVA analysis on CCLE cell lines and an additional 81 HCC cell lines based on metabolic and HCC-related pathways[101]. Our PCA and t-SNE results show that cell lines derived from HCC exhibit distinctly different metabolic

profiles from other cell lines (hepatocyte clusters are independent of other tumor cell clusters of different origins) (Figure S22A and 22B). NMF revealed that 81 HCC cell lines replicated the typical three metabolic subtypes (Figure S22C). NMF revealed that the 81 HCC cell lines replicated the typical three metabolic subtypes, with the iHCC3 subtype showing good similarity in specific pathway activity to an independent bulk-seq RNA panel. This suggests that a subtype with high nucleic acid metabolic activity may consistently exist in many HCC cell lines, regardless of the specific cell line or HCC cohort. These results suggest that cell lines such as iHCC3 (JHH2, CLC5, SNU761) may serve as representative cell lines for subtype 3 in subsequent in vitro experiments (Figure S22D). Furthermore, we validated the expression of nine characteristic genes in different cell lines, which also showed good consistency with other HCC cohorts (Figure S22E). Future, according to the results of metabolomics, we analyzed the distribution levels of different metabolites in the three metabolic subtypes through NMF in TIGER-LC cohort, and the results showed that the distribution of different metabolites in the three metabolic subtypes was different (Figure S23A).[21] In addition, subtype-specific metabolites that were different in at least two HCC subtypes were considered as subtype-specific metabolites. The enrichment degree of these subtype-specific metabolites in these HCC subtypes was analyzed, and some specific primary and secondary metabolites were found to be more abundant in the iHCC subtype (Figure S23B-D). In addition, we also analyzed the effects of different metabolites on the prognosis of HCC patients, and the abundance of metabolites including N-acetyltryptophan and taurocholate had a significant impact on the prognosis of HCC patients (Figure S24). These results further confirmed the impact of metabolic heterogeneity on the risk stratification of HCC patients, which will greatly promote our understanding of the specific mechanisms of metabolic heterogeneity in future studies.

Bulk RNA-seq evaluates the average gene expression levels across a mixture of various cell types, thereby obscuring the distinctions between cell types within the same sample. This observation

was further supported by an extensive analysis of pathway activity distributions in single cells and bulk samples, which demonstrated greater variability in pathway activities among different types of single cells compared to the variability between bulk tumors and normal tissues. Single-cell sequencing revealed that the metabolic activity of various pathways differs in each HCC patient compared to normal hepatocytes. The above results suggest that metabolic heterogeneity may not be completely the same in different subtypes or even within the same subtype, and each HCC patient has its own unique characteristics. This will encourage us to take full account of each independent individual when conducting heterogeneity studies. Tumor heterogeneity plays a crucial role in cancer progression and clinical treatment, and is an important factor in tumor treatment resistance and progression. A comprehensive exploration of tumor heterogeneity at the single-cell level is anticipated to greatly enhance overall tumor survival rates. Similarly, the integration of single-cell multi-omics analysis, akin to bulk multi-omics analyses, is expected to offer substantial insights into unraveling the complexities of cellular heterogeneity.

Due to the high metabolic adaptability of tumor cells, when any metabolic pathway encounters obstacles, tumor cells will automatically switch or enable other pathways, so as to avoid stress damage. Consequently, therapeutic strategies for tumor metabolic regulation should be combined to block or regulate multiple metabolic pathways simultaneously. Both tumor cells and stromal cells undergo rapid metabolic adaptations during tumor proliferation and metastasis. Tumor cells and their "neighbors" maintain their biosynthetic and energy metabolic needs through metabolic coordination or competition, while evading immune surveillance or therapeutic intervention. There is a dual mechanism in tumor evolution. The balance between mutation level and selective metabolic pressure determines the trajectory of tumor evolution, leading to different temporal sequences of accumulation of different somatic cell aberrations. In addition to genomic evolution, cancer cells also evolve through epigenetic and metabolic adaptations and may exhibit a more complex adaptive landscape

than dual mechanisms. Elucidation of the interplay between the different types of genetic and non-genetic changes occurring in cancer cells will help reveal the weaknesses of tumors and improve clinical treatment outcomes. The causes of metabolic heterogeneity are multifaceted. Metabolic dysregulation in HCC encompasses systemic perturbations influenced by a multitude of factors, including age, gender, dietary patterns, circadian rhythms, and others. The complex interplay of genetic, epigenetic, microbiome, lifestyle, dietary, and environmental factors contributes to metabolic heterogeneity[102,103]. In future studies, metabolic heterogeneity will encourage in-depth exploration from a more comprehensive perspective.

In summary, the integration of tumor metabolism and multi-omics holds immense promise for advancing risk stratification management, treatment, and prognostic monitoring of HCC patients. However, several challenges and deficiencies remain to be addressed. Specifically, the acquisition and analysis of certain omics data are constrained by factors such as sample quality, data accuracy, and reproducibility. Although multi-omics data offers more comprehensive information, its accuracy in prognostic monitoring still requires enhancement. To overcome these limitations, our future research will prioritize the optimization of metabolic therapy, ensuring a more precise targeting of tumor cells while minimizing toxicity to normal tissues. Additionally, there will be a focus on deepening the integration of multi-omics data, leveraging advanced algorithms and tools to improve data processing accuracy and efficiency. By establishing comprehensive databases containing large samples and diverse omics data, we aim to provide richer resources for tumor stratification, treatment, and prognostic monitoring. In conclusion, the integration of tumor metabolism and multi-omics represents a significant step forward in the management of HCC patients. However, continued research and innovation are essential to overcome current challenges and deficiencies, ensuring that we can provide the most precise and effective treatment for these patients.

5. CONCLUSIONS

On the whole, these observations above highlight distinct discrepancies in metabolic and HCC-related signaling pathways across three HCC subtypes. These differences were thought to result from high tumor heterogeneity among tumors and appeared to be a factor influencing factor in determining the prognosis. In a comprehensive multi-omics analysis, we identified several potential subtype-specific therapeutic targets involved in regulating various metabolic networks. Revealing the mechanistic differences between HCC subtypes and identifying subtype-specific drug targets should accelerate efficient treatment strategies and precision medicine.

HIGHLIGHTS

- Our study revealed subtype-specific signatures connecting metabolite alterations to clinical prognosis, epigenetic modifications and TME.
- Our classification strategy validated the previously reported HCC subclasses while retaining their characteristics.
- This study identified subtype-specific metabolic preferences, proposed new insights into the heterogeneity of HCC from a metabolic perspective.

ABBREVIATIONS

HCC: hepatocellular carcinoma; NMF: Non-negative Matrix Factorization; GSVA: gene set variation analysis; NAFLD: nonalcoholic fatty liver disease; T2DM: type 2 diabetes mellitus; MsigDB: Molecular Signatures Database; TCGA: the Cancer Genome Atlas; ICGC: the International Cancer Genome Consortium; TPM: transcripts per kilobase million; FDR: false discovery rate; UMAP: Uniform Manifold Approximation and Projection; PCA: principal component analysis; ML: machine

learning; DT: Decision Tree; RF: Random Forest; SVM: support vector machine; KNN: k-nearest neighbor; LDA: linear discriminant analysis; AdaBoost: Adaptive Boosting; XgBoost: extreme gradient boosting; AFP: alpha-fetoprotein; GGT: γ-Glutamyl Transferase; FCM: Fuzzy C-Means Clustering

DATA AVAILABILITY STATEMENT

The data sets (TCGA, GEO, and ICGC) generated and/or analyzed during the current study period can be found in the MATERIALS AND METHODS section of the original manuscript file. The microarray data of GSE14520 were downloaded from the GEO database. Proteome and phosphoproteome data of 159 HCC patients were approved and provided by the researchers in supplementary data. The DNA methylation expression profile was derived from TCGA-LIHC. We have described in detail the processes including data set selection, filtering and/or standardization in materials and methods. Any further data or information are available from the corresponding author upon reasonable request.

AUTHOR CONTRIBUTIONS

All authors read and approved the final version of the manuscript. L.Y.Z. and Y.L. conceived and designed the study. Y.L., B.Y.Z., and J.L.Z. developed and implemented ML pipeline under the supervision of L.Y.Z. and J.L.Z. L.Y.Z., J.L.Z., and W.X.Z. developed the methodology for multi-omics data generation and phenotypic data. L.Y.Z. and Y.L. performed formal analysis and investigation on omics data. L.Y.Z., Y.L., B.Y.Z., W.Y.Z. and Z.Z.L, were under the responsibility of the production of Figures and Tables, and manuscript writing. Z.H.L performed additional dataset retrieval and analysis in the revised manuscript.

FUNDING

This work has been supported by grants from the National Natural Science Foundation of China (No. 82303218), the Natural Science Foundation of Anhui Province (No.2308085QH285), the Key Program of Anhui Provincial Department of Education (No.2022AH051438) and the Natural Science Key Project of Bengbu Medical College (No.2021BYZD238).

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

Ethics approval for the study was granted by the clinical trial ethics committee (No. 2023023) at the First Affiliated Hospital of Bengbu Medical University, and all patients provided written informed consent. Ethic approval was exempted for public resource by the First Affiliated Hospital of Bengbu Medical University according to the description of national legislation guidelines, such as item 1 and 2 of Article 32 of the Measures for Ethical Review of Life Science and Medical Research Involving Human Subjects dated February 18, 2023, China, because the patients involved in the database have obtained ethical approval.

ACKNOWLEDGMENTS

We would like to thank all participants for their understanding and support of our research.

CONFLICT OF INTEREST

All the authors declare that they have no conflict of interest.

REFERENCES:

1. Llovet, J. M. et al. Hepatocellular Carcinoma. *Nat. Rev. Dis. Primers*. **7**, 7 (2021).
2. Collaboration, G. B. O. D. The Burden of Primary Liver Cancer and Underlying Etiologies From 1990 to 2015 at the Global, Regional, and National Level: Results From the Global Burden of Disease Study 2015. *Jama Oncol.* **3**, 1683-1691 (2017).

3. Martínez-Reyes, I. & Chandel, N. S. Cancer Metabolism: Looking Forward. *Nat. Rev. Cancer*. **21**, 669-680 (2021).

4. Gomes, A. P. et al. Altered Propionate Metabolism Contributes to Tumour Progression and Aggressiveness. *Nat. Metab.* **4**, 435-443 (2022).

5. Gong, Y. et al. Metabolic-Pathway-Based Subtyping of Triple-Negative Breast Cancer Reveals Potential Therapeutic Targets. *Cell Metab.* **33**, 51-64 (2021).

6. Chen, D. et al. Identification and Characterization of Robust Hepatocellular Carcinoma Prognostic Subtypes Based On an Integrative Metabolite-Protein Interaction Network. *Adv. Sci.* **8**, 2100311 (2021).

7. Faubert, B., Solmonson, A. & DeBerardinis, R. J. Metabolic Reprogramming and Cancer Progression. *Science*. **368**, eaaw5473 (2020).

8. Sun, J. et al. Decreased Propionyl-Coa Metabolism Facilitates Metabolic Reprogramming and Promotes Hepatocellular Carcinoma. *J. Hepatol.* **3**, 627-642 (2022).

9. Li, Y. et al. Got2 Silencing Promotes Reprogramming of Glutamine Metabolism and Sensitizes Hepatocellular Carcinoma to Glutaminase Inhibitors. *Cancer Research (Chicago, Ill.)*. **82**, 3223-3235 (2022).

10. Fang, C. et al. Metabolic Reprogramming and Risk Stratification of Hepatocellular Carcinoma Studied by Using Gas Chromatography–Mass Spectrometry-Based Metabolomics. *Cancers*, 2022.

11. Xue, R. et al. Liver Tumour Immune Microenvironment Subtypes and Neutrophil Heterogeneity. *Nature*. **612**, 141-147 (2022).

12. Zhang, Q. et al. Integrated Multiomic Analysis Reveals Comprehensive Tumour Heterogeneity and Novel Immunophenotypic Classification in Hepatocellular Carcinomas. *Gut*. **68**, 2019-2031 (2019).

13. Losic, B. et al. Intratumoral Heterogeneity and Clonal Evolution in Liver Cancer. *Nat. Commun.*

11, 291-305 (2020).

14. Zhai, W. et al. Dynamic Phenotypic Heterogeneity and the Evolution of Multiple Rna Subtypes in Hepatocellular Carcinoma: The Planet Study. *Natl. Sci. Rev.* **9**, nwab192-nwab206 (2022).
15. Liu, J. et al. Integrative Metabolomic Characterisation Identifies Altered Portal Vein Serum Metabolome Contributing to Human Hepatocellular Carcinoma. *Gut*. **71**, 1203-1213 (2022).
16. Qiu, X. et al. Spatial Single-Cell Protein Landscape Reveals Vimentinhigh Macrophages as Immune-Suppressive in the Microenvironment of Hepatocellular Carcinoma. *Nat. Cancer*. **5**, 1557-1578 (2024).
17. Wang, N. et al. Piwil1 Governs the Crosstalk of Cancer Cell Metabolism and Immunosuppressive Microenvironment in Hepatocellular Carcinoma. *Signal Transduct. Target. Ther.* **6**, 86 (2021).
18. Donne, R. & Lujambio, A. The Liver Cancer Immune Microenvironment: Therapeutic Implications for Hepatocellular Carcinoma. *Hepatology (Baltimore, Md.)*. **77**, 1773-1796 (2023).
19. Kurebayashi, Y. et al. Immunovascular Classification of Hcc Reflects Reciprocal Interaction Between Immune and Angiogenic Tumor Microenvironments. *Hepatology*. **75**, 1139-1153 (2022).
20. Gong, J. et al. Hcc Subtypes Based On the Activity Changes of Immunologic and Hallmark Gene Sets in Tumor and Nontumor Tissues. *Brief. Bioinform.* **22**, bbaa427 (2021).
21. Chaisaingmongkol, J. et al. Common Molecular Subtypes Among Asian Hepatocellular Carcinoma and Cholangiocarcinoma. *Cancer Cell*. **32**, 57-70 (2017).
22. Bidkhori, G. et al. Metabolic Network-Based Stratification of Hepatocellular Carcinoma Reveals Three Distinct Tumor Subtypes. *Proceedings of the National Academy of Sciences*. **115**, E11874-E11883 (2018).
23. Yang, C., Huang, X., Liu, Z., Qin, W. & Wang, C. Metabolism-Associated Molecular Classification of Hepatocellular Carcinoma. *Mol. Oncol.* **14**, 896-913 (2020).

24. Roessler, S. et al. A Unique Metastasis Gene Signature Enables Prediction of Tumor Relapse in Early-Stage Hepatocellular Carcinoma Patients. *Cancer Res.* **70**, 10202-10212 (2010).

25. Chen, C. et al. Removing Batch Effects in Analysis of Expression Microarray Data: An Evaluation of Six Batch Adjustment Methods. *Plos One*. **6**, e17238 (2011).

26. Gao, Q. et al. Integrated Proteogenomic Characterization of Hbv-Related Hepatocellular Carcinoma. *Cell*. **179**, 561-577 (2019).

27. Luo, Y. Evaluating the State of the Art in Missing Data Imputation for Clinical Data. *Brief. Bioinform.* **23**, bbab489 (2022).

28. Blazek, K., van Zwieten, A., Saglimbene, V. & Teixeira-Pinto, A. A Practical Guide to Multiple Imputation of Missing Data in Nephrology. *Kidney Int.* **99**, 68-74 (2021).

29. Xi, Y. et al. Multi-Omic Characterization of Genome-Wide Abnormal Dna Methylation Reveals Diagnostic and Prognostic Markers for Esophageal Squamous-Cell Carcinoma. *Signal Transduct. Target. Ther.* **7**, 53-65 (2022).

30. Jammula, S. et al. Identification of Subtypes of Barrett's Esophagus and Esophageal Adenocarcinoma Based On Dna Methylation Profiles and Integration of Transcriptome and Genome Data. *Gastroenterology*. **6**, 1682-1697 (2020).

31. Aryee, M. J. et al. Minfi: A Flexible and Comprehensive Bioconductor Package for the Analysis of Infinium Dna Methylation Microarrays. *Bioinformatics*. **30**, 1363-1369 (2014).

32. Tian, Y. et al. Champ: Updated Methylation Analysis Pipeline for Illumina Beadchips. *Bioinformatics*. **33**, 3982-3984 (2017).

33. Subramanian, A. et al. Gene Set Enrichment Analysis: A Knowledge-Based Approach for Interpreting Genome-Wide Expression Profiles. *Proceedings of the National Academy of Sciences*. **102**, 15545-15550 (2005).

34. Hänzelmann, S., Castelo, R. & Guinney, J. Gsva: Gene Set Variation Analysis for Microarray and Rna-Seq Data. *Bmc Bioinformatics*. **14**, 7 (2013).

945 35. Rosario, S. R. et al. Pan-Cancer Analysis of Transcriptional Metabolic Dysregulation Using the
946 Cancer Genome Atlas. *Nat. Commun.* **9**, 5330-5346 (2018).
947 36. Désert, R. et al. Human Hepatocellular Carcinomas with a Periportal Phenotype Have the Lowest
948 Potential for Early Recurrence After Curative Resection. *Hepatology*. **66**, 1502-1518 (2017).
949 37. Newman, A. M. et al. Robust Enumeration of Cell Subsets From Tissue Expression Profiles. *Nat.*
950 *Methods*. **12**, 453-457 (2015).
951 38. Tian, J. et al. Cancerimmunityqtl: A Database to Systematically Evaluate the Impact of Genetic
952 Variants On Immune Infiltration in Human Cancer. *Nucleic Acids Res.* **49**, D1065-D1073
953 (2021).
954 39. Hoshida, Y., Brunet, J. P., Tamayo, P., Golub, T. R. & Mesirov, J. P. Subclass Mapping:
955 Identifying Common Subtypes in Independent Disease Data Sets. *Plos One*. **2**, e1195 (2007).
956 40. Ulloa-Montoya, F. et al. Predictive Gene Signature in Mage-a3 Antigen-Specific Cancer
957 Immunotherapy. *J. Clin. Oncol.* **31**, 2388-2395 (2013).
958 41. Roh, W. et al. Integrated Molecular Analysis of Tumor Biopsies On Sequential Ctla-4 and Pd-1
959 Blockade Reveals Markers of Response and Resistance. *Sci. Transl. Med.* **9**, eaah3560-eaah3572
960 (2017).
961 42. Liu, Z. et al. Machine Learning-Based Integration Develops an Immune-Derived Lncrna Signature
962 for Improving Outcomes in Colorectal Cancer. *Nat. Commun.* **13**, 816 (2022).
963 43. Kumar, L. & E Futschik, M. Mfuzz: A Software Package for Soft Clustering of Microarray Data.
964 *Bioinformation*. **1**, 5-7 (2007).
965 44. Shen, B. et al. Proteomic and Metabolomic Characterization of Covid-19 Patient Sera. *Cell*. **182**,
966 59-72 (2020).
967 45. Lu, Y. et al. A Single-Cell Atlas of the Multicellular Ecosystem of Primary and Metastatic
968 Hepatocellular Carcinoma. *Nat. Commun.* **13**, 4594-4611 (2022).
969 46. Tran, H. T. N. et al. A Benchmark of Batch-Effect Correction Methods for Single-Cell Rna

970 Sequencing Data. *Genome Biol.* **21**, 12-43 (2020).
971 47. Korsunsky, I. et al. Fast, Sensitive and Accurate Integration of Single-Cell Data with Harmony.
972 *Nat. Methods*. **16**, 1289-1296 (2019).
973 48. Ruan, H. et al. Single-Cell Rna Sequencing Reveals the Characteristics of Cerebrospinal Fluid
974 Tumour Environment in Breast Cancer and Lung Cancer Leptomeningeal Metastases. *Clin.*
975 *Transl. Med.* **12**, e885 (2022).
976 49. Sharma, A. et al. Onco-Fetal Reprogramming of Endothelial Cells Drives Immunosuppressive
977 Macrophages in Hepatocellular Carcinoma. *Cell*. **183**, 377-394 (2020).
978 50. Becht, E. et al. Dimensionality Reduction for Visualizing Single-Cell Data Using Umap. *Nat.*
979 *Biotechnol.* **37**, 38-44 (2019).
980 51. Zhang, X. et al. Cellmarker: A Manually Curated Resource of Cell Markers in Human and Mouse.
981 *Nucleic Acids Res.* **47**, D721-D728 (2019).
982 52. Ahn, S. et al. Genomic Portrait of Resectable Hepatocellular Carcinomas: Implications of Rb1 and
983 Fgf19 Aberrations for Patient Stratification. *Hepatology*. **60**, 1972-1982 (2014).
984 53. Schulze, K. et al. Exome Sequencing of Hepatocellular Carcinomas Identifies New Mutational
985 Signatures and Potential Therapeutic Targets. *Nature Genet.* **47**, 505-511 (2015).
986 54. Mayakonda, A., Lin, D., Assenov, Y., Plass, C. & Koeffler, H. P. Maftools: Efficient and
987 Comprehensive Analysis of Somatic Variants in Cancer. *Genome Res.* **28**, 1747-1756 (2018).
988 55. Tsiliki, G. et al. Rregrs: An R Package for Computer-Aided Model Selection with Multiple
989 Regression Models. *J. Cheminformatics*. **7**, 46 (2015).
990 56. Kuhn, M. Building Predictive Models in R Using the Caret Package. *J. Stat. Softw.* **28**, 1-26 (2008).
991 57. Mukherjee, S. et al. Radiomics-Based Machine-Learning Models Can Detect Pancreatic Cancer
992 On Prediagnostic Computed Tomography Scans at a Substantial Lead Time Before Clinical
993 Diagnosis. *Gastroenterology*. **163**, 1435-1446 (2022).
994 58. Meißner, A. et al. Radiomics for the Noninvasive Prediction of the Braf Mutation Status in Patients

with Melanoma Brain Metastases. *Neuro-Oncology*. **24**, 1331-1340 (2022).

59. Shu, T. et al. Plasma Proteomics Identify Biomarkers and Pathogenesis of Covid-19. *Immunity*. **53**, 1108-1122 (2020).

60. Liu, D. et al. Integrated Analysis of Plasma and Urine Reveals Unique Metabolomic Profiles in Idiopathic Inflammatory Myopathies Subtypes. *Journal of Cachexia, Sarcopenia and Muscle*. **13**, 2456-2472 (2022).

61. Reel, P. S. et al. Machine Learning for Classification of Hypertension Subtypes Using Multi-Omics: A Multi-Centre, Retrospective, Data-Driven Study. *Ebiomedicine*. **84**, (2022).

62. Sajjadian, M. et al. Machine Learning in the Prediction of Depression Treatment Outcomes: A Systematic Review and Meta-Analysis. *Psychol. Med.* **51**, 2742-2751 (2021).

63. Bi, J. & Zhang, C. An Empirical Comparison On State-of-the-Art Multi-Class Imbalance Learning Algorithms and a New Diversified Ensemble Learning Scheme. *Knowledge-Based Syst.* **158**, 81-93 (2018).

64. Sun, L., Zhang, H. & Gao, P. Metabolic Reprogramming and Epigenetic Modifications On the Path to Cancer. *Protein Cell*. **13**, 877-919 (2022).

65. Ng, C. K. Y. et al. Integrative Proteogenomic Characterization of Hepatocellular Carcinoma Across Etiologies and Stages. *Nat. Commun.* **13**, 2436-2453 (2022).

66. Chen, X. et al. Molecular Profiling Identifies Distinct Subtypes Across Tp53 Mutant Tumors. *Jci Insight*. **7**, e156485 (2022).

67. Betto, R. M. et al. Metabolic Control of Dna Methylation in Naive Pluripotent Cells. *Nature Genet.* **53**, 215-229 (2021).

68. Luo, X. et al. Dnmt1 Mediates Metabolic Reprogramming Induced by Epstein–Barr Virus Latent Membrane Protein 1 and Reversed by Grifolin in Nasopharyngeal Carcinoma. *Cell Death Dis.* **9**, 619-633 (2018).

69. F. C. Lopes, A. Mitochondrial Metabolism and Dna Methylation: A Review of the Interaction

Between Two Genomes. *Clin. Epigenetics*. **12**, 182-194 (2020).

70. Madsen, A. et al. An Important Role for Dnmt3a-Mediated Dna Methylation in Cardiomyocyte Metabolism and Contractility. *Circulation*. **142**, 1562-1578 (2020).

71. Jiang, Y. et al. Proteomics Identifies New Therapeutic Targets of Early-Stage Hepatocellular Carcinoma. *Nature*. **567**, 257-261 (2019).

72. Yamashita, T. et al. Epcam and A-Fetoprotein Expression Defines Novel Prognostic Subtypes of Hepatocellular Carcinoma. *Cancer Res.* **68**, 1451-1461 (2008).

73. Chiang, D. Y. et al. Focal Gains of Vegfa and Molecular Classification of Hepatocellular Carcinoma. *Cancer Res.* **68**, 6779-6788 (2008).

74. Buechel, D. et al. Parsing B-Catenin'S Cell Adhesion and Wnt Signaling Functions in Malignant Mammary Tumor Progression. *Proceedings of the National Academy of Sciences*. **118**, e2020227118 (2021).

75. Hoshida, Y. et al. Integrative Transcriptome Analysis Reveals Common Molecular Subclasses of Human Hepatocellular Carcinoma. *Cancer Res.* **69**, 7385-7392 (2009).

76. Villanueva, A. et al. Combining Clinical, Pathology, and Gene Expression Data to Predict Recurrence of Hepatocellular Carcinoma. *Gastroenterology*. **140**, 1501-1512 (2011).

77. Liao, Y. L. et al. Identification of Sox4 Target Genes Using Phylogenetic Footprinting-Based Prediction From Expression Microarrays Suggests that Overexpression of Sox4 Potentiates Metastasis in Hepatocellular Carcinoma. *Oncogene*. **27**, 5578-5589 (2008).

78. Sakai, Y. et al. Common Transcriptional Signature of Tumor-Infiltrating Mononuclear Inflammatory Cells and Peripheral Blood Mononuclear Cells in Hepatocellular Carcinoma Patients. *Cancer Res.* **68**, 10267-10279 (2008).

79. Lee, J. et al. Classification and Prediction of Survival in Hepatocellular Carcinoma by Gene Expression Profiling. *Hepatology*. **40**, 667-676 (2004).

80. Woo, H. G. et al. Gene Expression–Based Recurrence Prediction of Hepatitis B Virus–Related

Human Hepatocellular Carcinoma. *Clin. Cancer Res.* **14**, 2056-2064 (2008).

81. Zhang, L. et al. Glycolysis-Related Gene Expression Profiling Serves as a Novel Prognosis Risk Predictor for Human Hepatocellular Carcinoma. *Sci Rep*. **11**, 18875-18895 (2021).

82. Cheng, Y. et al. Non-Terminally Exhausted Tumor-Resident Memory Hbv-Specific T Cell Responses Correlate with Relapse-Free Survival in Hepatocellular Carcinoma. *Immunity*. **54**, 1825-1840 (2021).

83. Patel, S. A. et al. Molecular Mechanisms and Future Implications of Vegf/Vegfr in Cancer Therapy. *Clin. Cancer Res.* **29**, 30-39 (2023).

84. He, W. et al. Hepatocellular Carcinoma-Infiltrating Γδ T Cells are Functionally Defected and Allogenic Vδ2+ Γδ T Cell Can be a Promising Complement. *Clin. Transl. Med.* **12**, e800 (2022).

85. Wang, T. et al. Integrating Bulk and Single-Cell Rna Sequencing Reveals Cellular Heterogeneity and Immune Infiltration in Hepatocellular Carcinoma. *Mol. Oncol.* **16**, 2195-2213 (2022).

86. Li, M. et al. Aldolase B Suppresses Hepatocellular Carcinogenesis by Inhibiting G6Pd and Pentose Phosphate Pathways. *Nat. Cancer*. **1**, 735-747 (2020).

87. Liu, G. et al. Fructose-1,6-Bisphosphate Aldolase B Depletion Promotes Hepatocellular Carcinogenesis through Activating Insulin Receptor Signaling and Lipogenesis. *Hepatology*. **74**, 3037-3055 (2021).

88. Rome, F. I. & Hughey, C. C. Disrupted Liver Oxidative Metabolism in Glycine N-Methyltransferase-Deficient Mice is Mitigated by Dietary Methionine Restriction. *Mol. Metab.* **58**, 101452-101467 (2022).

89. Shi, Z., Gan, G., Gao, X., Chen, F. & Mi, J. Kynurenine Catabolic Enzyme Kmo Regulates Hcc Growth. *Clin. Transl. Med.* **12**, e697 (2022).

90. Wang, Y., Suk, F. & Liao, Y. Loss of Hmgcs2 Enhances Lipogenesis and Attenuates the Protective Effect of the Ketogenic Diet in Liver Cancer., 2020:1797-1812.

91. Hayward, D. et al. Ankrd9 is a Metabolically-Controlled Regulator of Impdh2 Abundance and

Macro-Assembly. *J. Biol. Chem.* **294**, 14454-14466 (2019).

92. Lee, D. et al. Folate Cycle Enzyme Mthfd1L Confers Metabolic Advantages in Hepatocellular Carcinoma. *J. Clin. Invest.* **127**, 1856-1872 (2017).

93. Sun, C. et al. Spatially Resolved Metabolomics to Discover Tumor-Associated Metabolic Alterations. *Proceedings of the National Academy of Sciences*. **116**, 52-57 (2019).

94. Wang, J. et al. Metabolomics Study of the Metabolic Changes in Hepatoblastoma Cells in Response to Ntcp/Slc10a1 Overexpression. *Int. J. Biochem. Cell Biol.* **125**, 105773-105781 (2020).

95. Wang, Y. et al. N1-Methyladenosine Methylation in Trna Drives Liver Tumourigenesis by Regulating Cholesterol Metabolism. *Nat. Commun.* **12**, 6314-6332 (2021).

96. Zhang, M. et al. Pancreatic Cancer Cells Render Tumor-Associated Macrophages Metabolically Reprogrammed by a Garp and Dna Methylation-Mediated Mechanism. *Signal Transduct. Target. Ther.* **6**, 366-383 (2021).

97. Park, J. et al. Targeted Erasure of Dna Methylation by Tet3 Drives Adipogenic Reprogramming and Differentiation. *Nat. Metab.* **4**, 918-931 (2022).

98. Gomez-Alonso, M. D. C. et al. Dna Methylation and Lipid Metabolism: An Ewas of 226 Metabolic Measures. *Clin. Epigenetics*. **13**, 7-28 (2021).

99. Ayton, S. G., Pavlicova, M., Robles-Espinoza, C. D., Tamez Peña, J. G. & Treviño, V. Multiomics Subtyping for Clinically Prognostic Cancer Subtypes and Personalized Therapy: A Systematic Review and Meta-Analysis. *Genet. Med.* **24**, 15-25 (2022).

100. Wang, F. et al. Identification of blood-derived exosomal tumor RNA signatures as noninvasive diagnostic biomarkers for multi-cancer: a multi-phase, multi-center study. Mol Cancer. **24**, 60-80 (2025).

101. Qiu, Z. et al. A Pharmacogenomic Landscape in Human Liver Cancers. *Cancer Cell*. **36**, 179-193 (2019).

102. Mukherjee, S. et al. Familiarity Breeds Strategy: In Silico Untangling of the Molecular Complexity On Course of Autoimmune Liver Disease-to-Hepatocellular Carcinoma Transition Predicts Novel Transcriptional Signatures. *Cells*. **10**, 1917 (2021).

103. Xu, X. Q. et al. Stage-Specific Transcriptome Landscape of Hepatocellular Carcinoma: Insights From Super and Poor Survivors with Prognostic Signature Identification. *J. Dig. Dis.* **26**, 371-384 (2025).

TABLE AND FIGURES LEGENDS

Figure 1. (**A**) Flow chart of Gene Set Enrichment Analysis (GSEA). (**B**) GSEA analysis showed that amino acid, glucose, lipids, nucleotide, vitamin metabolism and other related metabolic pathways were significantly dysregulated in HCC tissues.

Figure 2. Identification of HCC subtypes using NMF consensus clustering in the TCGA cohort. (**A**) Schematic of the NMF algorithm for unsupervised clustering of RNA-seq expression profiles. (i) Strategy for selecting the optimal rank number (ii) Consensus matrix heatmap when NMF rank was 3 (iii) Silhouette plot displaying the composition (n = number of samples) and stability (average width) of clustering (**B**) NMF clustering using 3124 metabolism-associated genes. The cophenetic correlation coefficient for k =2-10 is shown (left); The consensus map (k = 3) of NMF classification. Three HCC subtypes (iHCC 1, iHCC 2, iHCC 3) were identified in the TCGA cohort (middle); the Silhouette Coefficient confirmed the robust subtype Classification (left). (**C**) OS and RFS of 3 subclasses (iHCC1, iHCC2, and iHCC3) in the TCGA cohort, with iHCC 3 having the worst prognosis. (**D**) PCA and t-SNE analyses supported the stratification into three HCC subclasses.

Figure 3. Consensus clustering for Proteogenomic data. (**A**) NMF clustering for k = 2-10 (left); Consensus map for k = 3 (middle); the Silhouette Coefficient supported the robust 3 subtype Classification (left). (**B**) OS and RFS curves based on 3 NMF clusters in the Proteogenomic cohort. (**C**) PCA and t-SNE projection of 3 HCC subtypes based on Proteogenomic data.

Figure 4. Clinical characteristics of 3 HCC subtypes. (**A**) Correlation of our classification (iHCC1, iHCC2, and iHCC3) with clinical characteristics in the TCGA, GEO, ICGC, and Proteogenomic cohorts. (**B**) AFP distribution among different HCC subtypes, the iHCC 3 cluster has a higher level of AFP in the TCGA, GEO, and Proteogenomic cohort. (**C**) Kaplan-Meier curves show distinct prognoses among two AFP levels. Higher AFP levels usually predicted shorter OS or RFS.

Figure 5. Integrated analysis of altered pathways at transcriptome and proteome levels among the 3 HCC subgroups. (**A**) Heatmap of alteration pathways between 3 HCC subtypes in the TCGA, GEO, ICGC, and Proteogenomic cohorts. The color of each cell represents the average ssGSEA enrichment scores for that subtype; red denotes activation and blue denotes inhibition. (**B**) A boxplot of the signature score for 3 HCC progression-associated signatures distinguished by different subtypes. P values are derived from the Kruskal-Wallis test. Boxplots show the median (central line), and the 25-75% IQR (box limits).

Figure 6. Analysis of metabolic gene expression patterns in various subgroups. (**A**) Fuzzy c-means clustering identified two distinct temporal patterns of protein expression. The x-axis represents four developmental clusters, while the y-axis represents log2-transformed, normalized intensity ratios in each subgroup. (**B**) A global heatmap depicting the relative abundance of metabolism-related signatures in HCC. Biological functions related to these signatures are denoted by the color panel on the left and annotated according to the pathways based on the enrichment of MSigDB ontologies. Top: RNA-seq data in TCGA, GEO and ICGA; Bottom: Methylation, proteomic, and phosphoproteomic data.

Figure 7. Unsupervised clustering revealed the metabolic heterogeneity of HCC at the single-cell level.

(**A**) Heatmap of GSVA results for 120 pathways altered in 37 primary clusters; upregulated pathways are colored in red, and downregulated pathways are colored in blue. (**B**) The activity of metabolic pathway in 8 major cell populations. (**C**) Unsupervised clustering reclassified 12 hepatocyte cell clusters into 15 new clusters based 3214 metabolic genes. (**D**) UMAP plot of single-cell clusters from the multiple HCC samples. The color of each dot indicates the tumor which the cell comes from, and most tumor cells derived from a specific HCC patient are exclusive to a specific cell cluster. (**E**) Metabolic pathway activities in different hepatocyte clusters, metabolic activity varied significantly between clusters.

Figure 8. Immune characteristics of 3 HCC subtype. (**A**-**C**) Boxplot of the abundance of immune and stromal cell populations distinguished by different subclasses in the TCGA, GEO, and ICGC cohort. (**D**) Expression levels (normalized count) of 17 immune checkpoint genes in three HCC subclasses. (**E**, **F**) Associations of the efficacy of conventional therapy with the HCC subtypes. The similarity of gene expression profiles between HCC subtypes and patients receiving anti-CTLA4, anti-PD1, Pembrolizumab, anti-PD-L1 or anti-MAGE-A3 therapies was shown in TCGA and GEO cohorts. (G) Sensitivity analysis of different subtypes to 5-fluorouracil, Cytarabine, and Leflunomide.

Figure 9. Construction and validation of the TMI in TCGA cohort. (**A**) Elastic Net regression revealed the partial likelihood deviance in the 10-fold cross-validation. The red and blue vertical dotted lines were drawn at the minimum and 1-SE criteria value. (**B**) The restricted cubic spline determined the best cutoff for the risk score when HR was 1. (**C**) 371 HCC patients were divided into high- and low-risk groups based on the RCS cutoff value and represent survival difference between high- and low-group (top); KM survival curve of OS between two subgroups (medium); ROC analyses at different time nodes (bottom). The orange dashed line indicated the threshold of 0.70. (**D**) HCC patients in three independent cohorts were divided into high- and low-risk group based on the optimal cutoff in the TCGA cohort. (**E**) High risk scores were associated with poor overall survival in three validation cohorts. (**F**) Time-dependent ROC curves of prognostic prediction on the 1-, 2-, 3-, and 5-year survival rates for three validation cohorts.

Figure 10. Nomogram developed and evaluated for predicting the probability of 1-, 2-, 3- and 5-year OS in TCGA cohort. (**A**) The nomogram was constructed, with the TMI and clinical features incorporated. (**B**) Calibration plot of the nomogram in terms of agreement between the predicted and observed. (**C**, **D**) Time-dependent area under the ROC curve for the comparison of TMI with other model. (**E**) Decision curve analysis of the nomogram for 1-, 2-, 3- and 5-year risk.

Figure 11. The expression level of ALDOB in normal liver, and HCC tissues. (**A**) Pooled analysis showing ALDOB mRNA level in different pathology grade. (**B**, **C**) IHC results indicating deficiency

of ALDOB in HCC tissues from the Human Protein Atlas database. (**D**) The representative immunohistochemical staining images showing ALDOB protein expression in HCC and normal liver tissues. Scale bar: 100 μm.