Am J Cancer Res 2023;13(7):2814-2840
www.ajcr.us /ISSN:2156-6976/ajcr0144936

Original Article

Construction of molecular typing in LIHC microenvironment based on lipid metabolism-related genes

Shan Cong[1*], Shanshan Bai[2*], Yanfang Bi[3*], Yu Wang[1], Shi Jin[1], Hui He[1]

[1]Department of Laparoscopic Surgery, The First Affiliated Hospital of Dalian Medical University, Dalian 116000, Liaoning, China; [2]Department of Ultrasound, The First Affiliated Hospital of Dalian Medical University, Dalian 116000, Liaoning, China; [3]Department of Nursing, The First Affiliated Hospital of Dalian Medical University, Dalian 116000, Liaoning, China. [*]Equal contributors.

Received June 24, 2022; Accepted April 11, 2023; Epub July 15, 2023; Published July 30, 2023

Abstract: Consensus on the stage of liver hepatocellular carcinoma (LIHC) in patients is difficult, which restricts the diagnosis and treatment of liver cancer. Molecular typing based on genes related to the lipid metabolism pathways can reflect deeper characteristics of liver cancer and complement the deficiency of the clinical staging system. In this study, we constructed and verified two cell subtypes: C1 and C2 in LIHC, based on six lipid metabolic pathway-associated genes identified in two independent external validation cohorts comprising single-cell RNA-sequencing technology (scRNA-Seq) data and bulk RNA-seq data downloaded from Gene Expression Omnibus (GEO) database and The Cancer Genome Atlas (TCGA) database. The C2 subtype showed poorer prognosis, higher immune scores, and greater correlation with pathways associated with tumor progression as compared to the C1 subtype. Moreover, the sensitivity of many tested targeted drugs in C1 was relative to C2. Furthermore, Gene Set Enrichment Analysis (GSEA) revealed several significantly enriched oncological signatures and metabolic processes, which might help elucidate the underlying molecular mechanisms. At the same time, we identified there were significantly different metabolites in C1 and C2 subtypes using 11 LIHC tissue samples. In conclusion, we constructed two molecular subtypes based on the lipid metabolism-associated genes, which may provide valuable information to further study the pathogenesis and devise clinical management strategies for LIHC.

Keywords: Liver hepatocellular carcinoma, lipid metabolism, data analysis, scRNA-Seq, heterogeneity, prognosis

Introduction

Liver cancer is the most frequent fatal malignancy, originating from epithelial or mesenchymal tissues of the liver. Liver hepatocellular carcinoma (LIHC) is a common histologic type of primary liver cancer, accounting for nearly 90% of all cases of primary liver cancer [1]. Over 0.5 million new LIHC cases are reported annually and the associated morbidity is on the rise [2, 3]. LIHC is now the second leading cause of cancer-related deaths worldwide [4], and is mainly caused by a viral infection and liver fibrosis [5]. At present, the main therapeutic strategies for LIHC include hepatic resection, liver transplantation, radiotherapy, chemotherapy, and molecular targeted therapy, which are mainly decided by the clinical staging system. Despite the development of new staging systems for LIHC worldwide, there is no global consensus on a staging system that allows for the comparison of existing management protocols among heterogeneous populations [6], and personalized care for patients is lacking. Moreover, LIHC has a poor prognosis with a five-year recurrence rate reaching up to 70% post-resection and post-ablation with a five-year survival rate of 30-40% [7, 8].

The formation of tumors is the result of the accumulation of mutations in the genetic material in cells. Tumor cells with different genomic alterations can develop into different subclones owing to cell phenotypic heterogeneity induced by genomic heterogeneity, leading to differential responses to therapy and metastatic poten-

tial of tumors [9-11]. Following the development of the gene chip and high-throughput sequencing technologies, network biology has emerged as an integrative and systems-level approach to analyze tumor-related genes and their underlying regulatory mechanisms, particularly using the Gene Expression Omnibus (GEO) and The Cancer Genome Atlas (TCGA) databases, which have become essential in the field of cancer genomics [12]. Traditional methods of tumor gene expression analyses are based on large mixed cell samples, and the information on gene expression represents the average, ignoring the important but small information [13]. In 2009, Tang et al. proposed scRNA-Seq for the first time [14]. Using scRNA-Seq to investigate genomic heterogeneity and colon structures in intratumor cells of liver cancer not only provides a theoretical basis for accurate clinical classification of liver cancer, but also facilitates devising potential precise treatment plans based on molecular characteristics and discovery of new biomarkers related to diagnosis and prognosis. As a difficult tumor to treat, liver cancer has been studied extensively using the scRNA-Seq approach. Aizarani et al. established a human liver cell atlas by performing scRNA-Seq analysis of cells from normal liver tissues of 9 human donors [15]. Their findings may advance the understanding of the pathogenesis of liver cancer at the molecular level and help improve diagnostic and treatment strategies.

Abnormal metabolism exhibited by tumor cells with uncontrolled growth even in the presence of nutrient deficiencies is a hallmark of cancer, and is strongly associated with the occurrence, development, recurrence, metastasis, and prognosis of tumor [16-18]. On the one hand, carcinogenic factors disrupt the metabolic balance in the body, inducing metabolic reorganization and development of cancerous cells; on the other hand, reprogramming the metabolic system induces several biological behaviors involved in the proliferation, invasion, and metastasis of cancer cells [19, 20]. Metabolic reprogramming is a hallmark of cancer, among which, alterations in lipid metabolism are crucial. Fat is the main energy storage material in the human body, and it not only participates in energy metabolism but is required for the growth and proliferation of cells. Lipid molecules constitute the basic structure of cell membranes, which provide energy and function as signaling molecules. Merino Salvador et al. showed high alterations in lipid metabolism in tumor cells. Unlike normal cells that rely primarily on the uptake of exogenous fatty acids (FA), tumor cells show enhanced nascent adipogenesis, crucial for membrane biosynthesis and signaling molecules [21]. Fatty acid synthase (FAS, FASN) is a metabolic oncogene that supports tumor cell growth and survival and is highly expressed in many cancers [22]. Accumulating evidence shows that dysregulation of lipid metabolism is a significant metabolic alteration in cancer that not only affects primary tumor growth but also mediates its progression and metastasis [23]. Research on the molecular pathological characteristics of liver cancer cell metabolism is underway [24]. JIANG et al. found higher expression of cholesteryl ester in 254 liver cancer tissues as compared to the adjacent normal tissues through tissue microarray chip analysis, suggesting that lipid metabolism may be involved in the occurrence and development of liver cancer [25]. Relative to non-malignant liver tissues, genes involved in FA biosynthesis are generally up-regulated in most liver cancer tissues [26]. However, the relationship between lipid metabolism and related genes and the prognosis of liver cancer is unclear.

Notably, population changes with respect to liver cancer in the past several years have been detected. Liver cancer rates associated with nonalcoholic fatty liver disease (NAFLD) and nonalcoholic steatohepatitis (NASH), mainly caused by metabolic syndromes and obesity, are increasing gradually, even in a large proportion without underlying cirrhosis [27, 28]. NASH is the fastest growing indication for liver transplantation (LT) in the USA [29], however, studies on long-term outcomes and molecular pathological data are scarce, relative to liver cancer associated with other traditional liver diseases [30, 31]. Moreover, with the development of tumor screening procedures for patients with liver cirrhosis in recent years, the proportion of patients with early-stage liver cancer in developed countries has shown a gradual increase [32], posing new challenges for prognostic risk assessment, as these patients often lack previously identified risk factors associated with a poor prognosis. Thus, molecular typing based on genomic features to

improve the clinical staging system for these specific populations of liver cancer patients is crucial [33-35].

In this study, we constructed a classifier for liver cancer based on six pathways related to lipid metabolism and analyzed the heterogeneity using bulk RNA-seq and scRNA-seq datasets. Subsequently, we evaluated the relationship between the two molecular subtypes and their prognostic and clinical features. We also identified different subtypes using non-negative matrix factorization (NMF), and distinct tumor microenvironment and immune characteristics were detected. In conclusion, we established a molecular classification model for liver cancer based on lipid metabolism characteristics, and further constructed the characteristic index of each subtype, in an attempt to resolve the insufficiency of the clinical staging system. Our findings will provide research ideas and a theoretical basis for prognostic prediction and designing individualized treatment strategies for LIHC patients.

Materials and methods

Data sources

The LIHC scRNA-seq dataset, GSE149614, was downloaded from the GEO (https://www.ncbi.nlm.nih.gov/) database and included 10 patients with a primary tumor (PT), 2 with portal vein tumor thrombus (PVTT), 1 with metastatic lymph node (MLN), and 8 with normal liver tissue (NLT). Publicly available clinical data and information on gene expression were retrieved from TCGA database (https://portal.gdc.cancer.gov/). Furthermore, GSE14520 was acquired from the GEO database to construct the molecular subtypes and as an independent cohort for the validation of molecular subtypes. GSE10143 dataset played as an external cohort for the validation of molecular subtypes. Well-annotated gene sets, representing the universal set of biological processes, are crucial for the meaningful and insightful interpretation of large-scale genomic data [36]. The Molecular Signatures Database (MSigDB) is a collection of annotated gene sets and can be used to perform GSEA (https://www.gsea-msigdb.org/gsea/msigdb/index.jsp) [37]. We screened six lipid metabolic pathway-related gene sets. These included biosynthesis of unsaturated FAs, FA metabolism, steroid biosynthesis, steroid hormone biosynthesis, alpha-Linolenic acid metabolism, and arachidonic acid metabolism. These gene sets are shown in Supplementary Table 1.

Data acquisition and processing

TCGA-LIHC data were processed as follows: (1) Samples without information on clinical follow-up were removed. (2) Ensemble ID was converted to Gene Symbol. (3) The average expression value was considered in the case of multiple Gene Symbols.

The GSE14520 microarray dataset was processed as follows: (1) Samples without information on clinical follow-up were removed. (2) Probes were converted to Gene Symbols. (3) Probes corresponding to multiple genes were removed. (4) The average expression value was considered in the presence of multiple Gene Symbols.

After preprocessing, a total of 360 samples in TCGA-LIHC, and 242 samples and 10501 genes in GSE14520 were obtained and used for further analysis. The flowchart of the study design is shown in **Figure 1**.

Single-sample gene set enrichment analysis (ssGSEA)

ssGSEA classifies marker gene sets with common biological functions, chromosomal localization, and physiological regulation in a single sample [38]. Each ssGSEA enrichment score represents the absolute degree of enrichment of genes of a particular gene set in the corresponding sample. Gene expression for a given sample is rank-normalized and enrichment scores are generated using the empirical cumulative distribution function (ECDF) of the genes in the signature and the remainder of the genes. ssGSEA was performed using the GSVA [39] package to obtain a hallmark gene set score and the hallmark gene set was obtained from MSigDB.

Identification of molecular subtypes using the consensus cluster plus algorithm

We calculated the sample scores of six lipid metabolism pathways in TCGA and GSE14520 datasets by ssGSEA using the GSVA package, separately shown as shown in Supplementary Tables 2 and 3. Next, scores of these six meta-

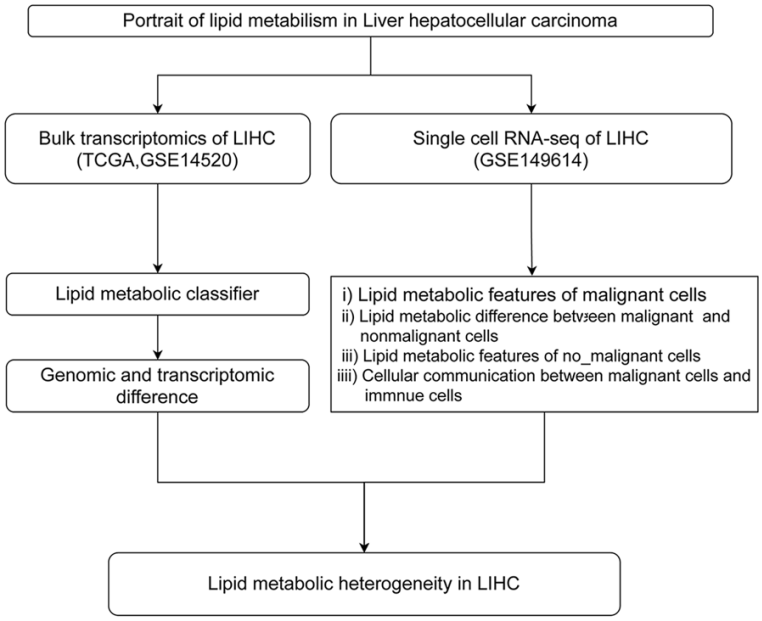


Figure 1. Technical road map.

bolic pathways were analyzed using the Consensus Cluster Plus package, using the hc algorithm and "canberra" as the metric distance. Consensus Cluster Plus verifies the rationality of clustering through resampling methods, which can disrupt the original dataset such that clustering analysis is performed on each resampled sample. Finally, the results of multiple clustering analyses are comprehensively evaluated to give a consensus of evaluation. A total of 500 bootstraps were performed and each bootstrap process included 80% of patients in the training set. The number of clusters was set from 2-10, and the optimal classification was determined by calculating the consistency matrix and the consistency cumulative distribution function (CDF). A total of 360 samples in TCGA_LIHC were clustered by Consensus Cluster Plus, and the optimal number of clusters was determined according to the CDF and CDF Delta area curves, which were used to construct metabolic subtypes for liver cancer.

Clinical association and prognostic analysis

Wilcoxon rank sum test and t-test were utilized to evaluate the clinical indexes (gender, stage, grade, age, and events) between the two molecular subtypes. The Kaplan-Meier (KM) survival curve combined with a log-rank test was used to compare the differences in survival between the two molecular subtypes using the R package, "survival".

Differentially expressed genes (DEGs) and their functional enrichment analysis

DEGs were screened in the TCGA cohort using the "limma" package in R software, with |log2FC| > 1.0 and false discovery rate (FDR) < 0.05 as the cut-off values. The Volcano plot was generated to visualize the distribution of the identified DEGs. Subsequently, the Kyoto Encyclopedia of Genes and Genomes (KEGG) and Gene Ontology (GO) analyses were performed to investigate the most significantly enriched pathways and biological processes related to the DEGs using the "clusterProfiler" package in R software.

Status of immune functions between the molecular subtypes

ssGSEA (Rooney et al., 2015) was employed to estimate the infiltration score of immune cells and the activity of immune-related pathways using the R software packages, "GSVA" and "GSEA Base". The Wilcoxon rank-sum test was used to compare the statistical differences between high- and low-risk groups.

Biological samples

11 LIHC tissue samples were obtained from the First Affiliated Hospital of Dalian Medical University. All patients signed an informed consent for the use of their tissues in research studies.

RNA sequencing

Total RNA was extracted from the tissue using TRIzol® Reagent (Plant RNA Purification Reagent for plant tissue) according the manufacturer's instructions (Invitrogen) and genomic DNA was removed using DNase I (TaKara). Then RNA quality was determined by 2100

Bioanalyser (Agilent) and quantified using the ND-2000 (NanoDrop Technologies). Only high-quality RNA sample (OD260/280 = 1.8-2.2, OD260/230S2.0, RINN6.5, 28S: 18S^1.0, > 1 pg) was used to construct sequencing library. RNA-seq transcriptome library was prepared following TruSeqTM RNA sample preparation Kit from Illumina (San Diego, CA) using 1 pg of total RNA. Shortly, messenger RNA was isolated according to polyA selection method by oligo(dT) beads and then fragmented by fragmentation buffer firstly. Secondly double-stranded cDNA was synthesized using a SuperScript double-stranded cDNA synthesis kit (Invitrogen, CA) with random hexamer primers (Illumina). Then the synthesized cDNA was subjected to end-repair, phosphorylation and 'A' base addition according to Illumina's library construction protocol. Libraries were size selected for cDNA target fragments of 300 bp on 2% Low Range Ultra Agarose followed by PCR amplified using Phusion DNA polymerase (NEB) for 15 PCR cydes. After quantified by TBS380, paired-end RNA-seq sequencing library was sequenced with the Illumina HiSeq xten/NovaSeq 6000 sequencer (2 × 150 bp read length).

Transcriptomics data analysis

To identify DEGs between two different samples, the expression level of each transcript was calculated according to the transcripts per million reads (TPM) method. RSEM was used to quantify gene abundances. Essentially, differential expression analysis was performed using the DESeq2/DEGseq/EdgeR with Q value w 0.05, DEGs with |log2FC| > 1 and Q value ≤ 0.05 (DESeq2 or EdgeR)/Q value ≤ 0.001 (DEGseq) were considered to be significantly different expressed genes.

Liver metabolomics analysis

Metabolites were extracted as the following steps: 1. 100 uL plasma/serum + 300 uL IPA (-20°C pre-cooled), vortex shaking for 1 min, stand at -20°C overnight; 2. At 1,2000 rpm centrifuge 4°C for 20 min, remove the supernatant and take 100 uL to the sample vial; 3. 50 uL were taken to make quality control (QC) samples. In positive ion acquisition mode, the Masslynx software performs primary and secondary MS data acquisition based on the MSE function. Capillary voltage: 0.5 kV, cone hole voltage 40 V, ion source temperature 100°C, solvent removal gas flow rate 1000 L/h, cone hole gas flow rate 50 L/h, m/z 50-1200 Da in 18 min, 0.2 sec/cycle. In anion acquisition mode, Masslynx software performs primary and secondary MS data acquisition based on the MSE function. Capillary voltage: 0.8 kV, cone hole voltage 40 V, ion source temperature 100°C, solvent removal gas flow rate 1000 L/h, cone hole gas flow rate 50 L/h, m/z 100-1200 Da in 18 min, 0.2 sec/cycle. The databases used for this experiment were LipidMaps and HMDB, and metabolites with *P* value < 0.05 were selected as differentially expressed metabolites (DEMs). For the liquid phase conditions, Phase A: Water: acetonitrile = 6:4 + 0.1% formic acid + 10 mM ammonium formate; Phase B: Isopropanol: acetonitrile = 9:1 + 0.1% formic acid + 10 mM ammonium formate. The detail was showed in Supplementary Table 4.

Integrative metabolomics and transcriptomics analysis

To analyze the relationship between metabolomics and transcriptomics in 11 LIHC tissue samples, we made Network network using gephi software. Spearman correlation coefficients was calculated by WGCNA package in R and genes and metabolites with Spearman correlation coefficients > 0.8 and P < 0.1 were selected.

Statistical analysis

Statistical analysis was performed using SPSS 22.0. Normally distributed data were analyzed by one-way ANOVA or t-test, and non-normally distributed data were analyzed using the Kruskal-Wallis H test or Mann-Whitney U-test. Differential analysis for count data was performed using Pearson's chi-squared test (α = 0.05) or Fisher's exact test. A *P*-value < 0.05 was considered statistically significant.

Ethics approval and consent to participate

The Ethics Committee of the First Affiliated Hospital of Dalian Medical University approved the study design.

Results

Metabolism-related subtypes

Considering that consensus clustering results were relatively stable at k = 2 (**Figure 2A**, **2B**) and the CDF Delta area curve, we obtained two clusters, namely C1 and C2 (**Figure 2C**). There were significant differences in prognoses between C1 and C2 (**Figure 2D**). In general, the prognosis of patients in C1 was relatively good, whereas that in C2 was poor. The same analysis was performed and consistent results were obtained in the GSE14520 cohort (**Figure 2E**). To validate the metabolism-related subtypes, we performed K-M curves using the GSE10143 dataset. The patients in C2 showed shorter OS compared with those in C1 (**Figure 2F**) ($P < 0.05$). Furthermore, principal component analysis (PCA) was performed in the six lipid metabolic pathways (Supplementary Figure 1), which further supported heterogeneity in tumor cells and the reliability of dividing liver cancer patients into two clusters.

The differences in the expression of metabolic pathways between metabolism-related subtypes

Next, we compared the differences in genes associated with the six lipid metabolic pathways between the two clusters (**Figure 3A**) in TCGA database, which had significant differences in the scores for the six lipid metabolic pathways between the two clusters (**Figure 3C**). We performed the same analysis in the GSE31210 dataset and obtained similar results (**Figure 3B**, **3D**).

The differences in clinical characteristics between metabolism-related subtypes

To investigate whether metabolism-related subtypes influenced clinical characteristics, we compared the distribution of different clinical characteristics between the two clusters in TCGA database. The results showed significant differences in gender, T stage, stage, grade, and survival states between C1 and C2 (**Figure 4A**). What's more, significant differences in stage, age, and survival states between clusters in the GSE14520 dataset existed (**Figure 4B**).

Immune signatures between metabolism-related clusters

First, according to the previous study [40], we calculated the immune cell scores for each sample in TCGA by ssGSEA. The wilcox.test showed there were significant differences in abundances of T cells and B cells between the two clusters (**Figure 5A**). Next, ESTIMATE was performed to calculate the immune scores for each sample in TCGA. The wilcox.test showed significant differences between the two clusters and the score in C2 was higher than that in C1 (**Figure 5B**).

Based on the ten pathways associated with tumors obtained from a previous study [41], we calculated the enrichment scores for each sample in TCGA by ssGSEA. And we found significant differences in seven pathways: the scores of the cell cycle, NOTCH, RAS, TP53, and WNT were higher in C2 than C1, whereas the scores of NRF1 and PI3K were higher in C1 compared with C2 (**Figure 5C**). Based on the 31 genes associated with cell cycle progression (CCP) identified in a previous study [42], we calculated the enrichment scores for CCP in each sample in TCGA by ssGSEA. The enrichment scores for CCP in C2 were higher than in C1 according to the results of the wilcox.test (**Figure 5D**). As shown in **Figure 5A**, significant differences in macrophage abundance were obtained between the two clusters. Macrophages play an important role in immune regulation through antigen processing and presentation, Toll-like receptor signaling pathway, and natural killer cell-mediated cytotoxicity pathway which can kill tumor cells by mediating the antibody-dependent cell-mediated cytotoxicity (ADCC) effect. Thus, we analyzed the genes associated with these pathways in the MSigDB database by GSEA and calculated the scores of them for each sample by ssGSEA. Only the antigen processing and presentation score had significant significance between the two clusters and it was higher in C2 than C1 (**Figure 5E-G**).

Apart from macrophages, T cells including Type 2 T helper cells showed significant differences in abundance between the clusters. Thus, we extracted the genes associated with Th1 and Th2 cell differentiation and the IL-17 signaling pathway and calculated the scores for each sample by ssGSEA. The wilcox.test showed

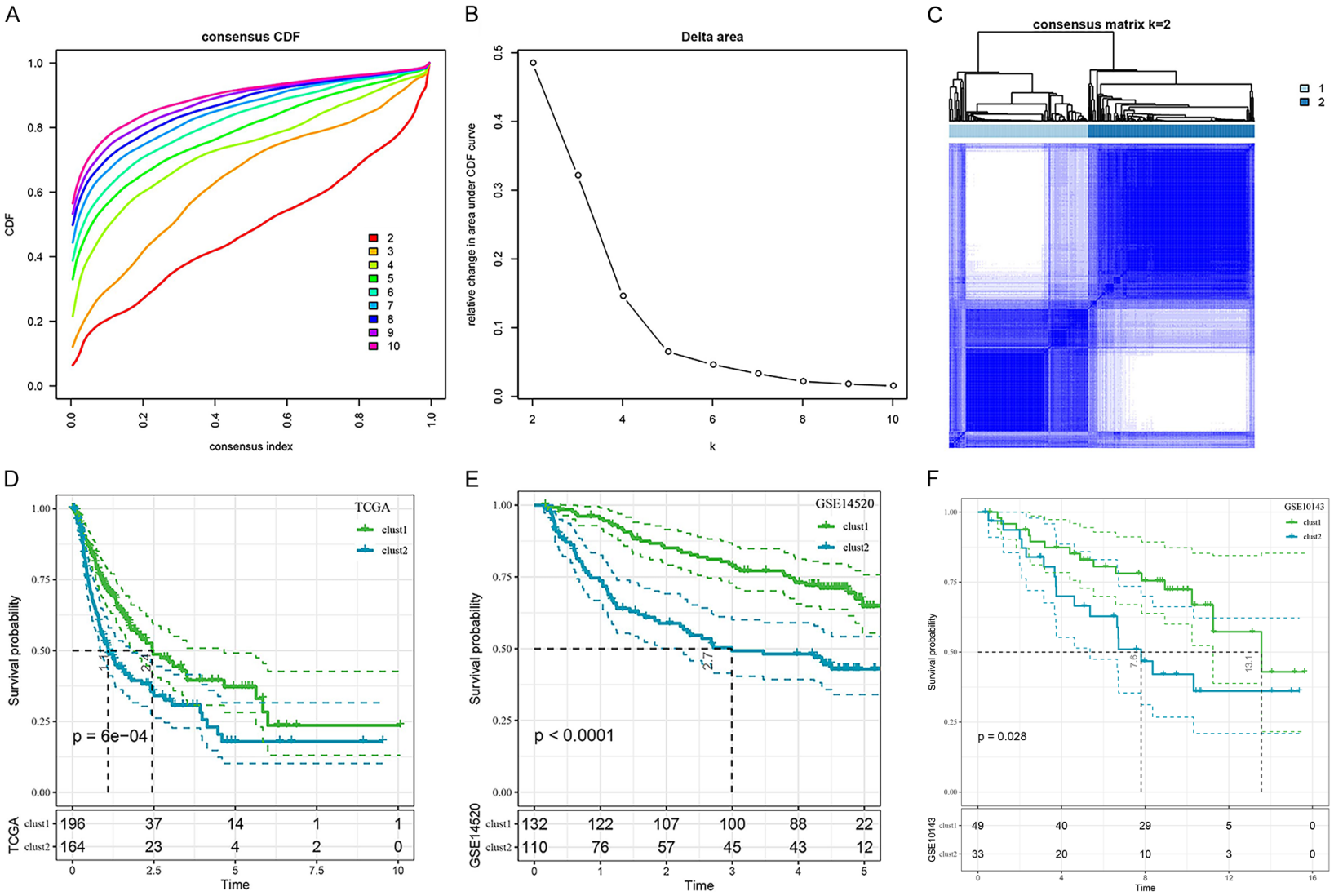


Figure 2. Metabolism-related subtypes. A: CDF curve for samples in TCGA-LIHC cohort; B: Cumulative distribution function (CDF) Delta area curve for samples in TCGA-LIHC cohort. The Delta area curve of consensus clustering indicates the relative change in the area under the CDF curve for each category number k as compared to k-1. The horizontal axis represents the category number, k, and the vertical axis represents the relative change in the area under the CDF curve; C: Sample cluster heatmap at consensus k = 2; D: KM curves for the prognosis of patients in TCGA belonging to the two subtypes; E: KM curves for the prognosis of patients in the GSE14520 cohort belonging to the two subtypes. F: KM curves for metabolism-related subtypes in GSE10143.

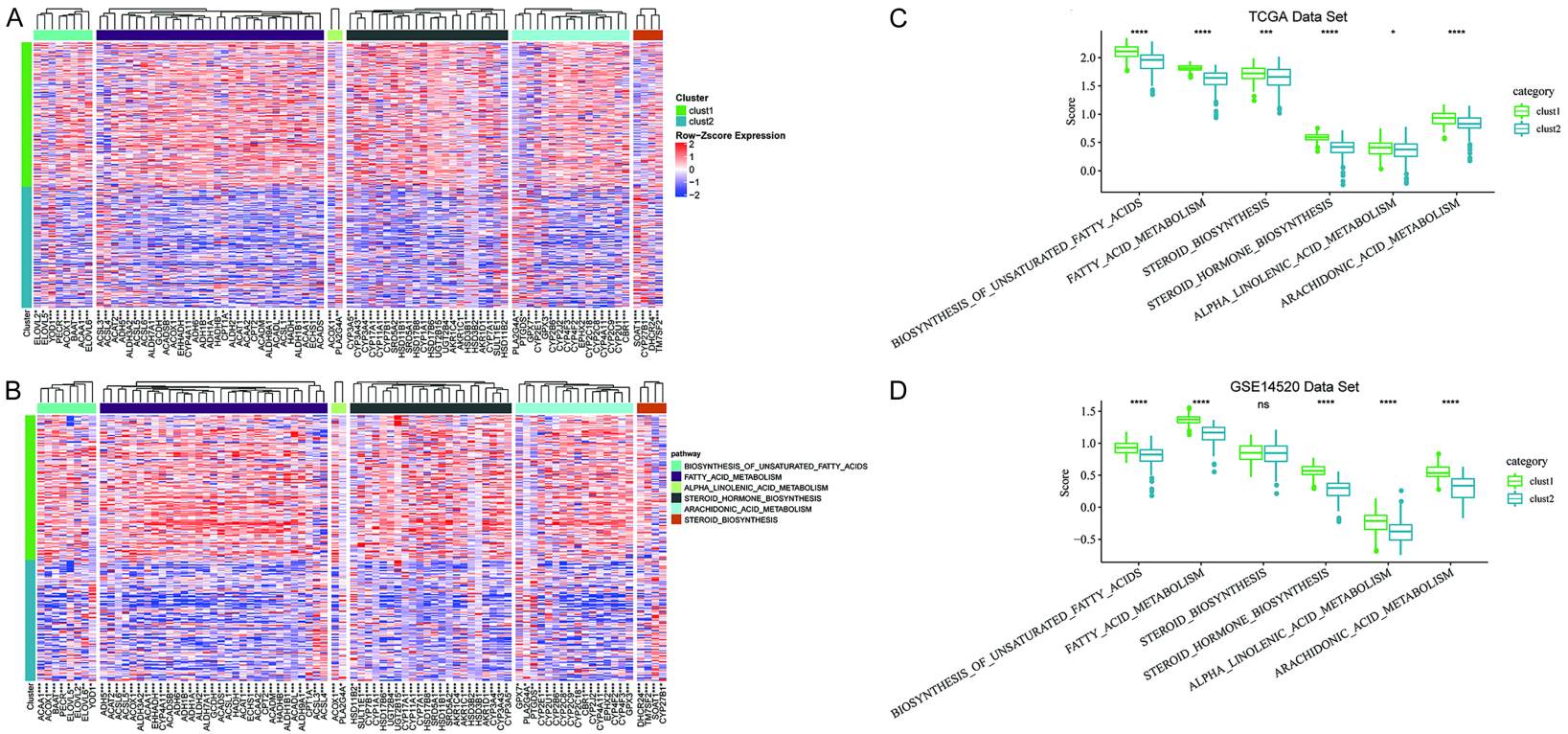


Figure 3. Differences in the expressions of metabolic pathways in metabolism-related subtypes. A: Heatmaps for two subtypes based on the expression of genes in the six lipid metabolism pathways in TCGA dataset; B: Heatmaps for two subtypes based on the expression of genes in the six lipid metabolism pathways in the GSE14520 cohort; C: Boxplots showing differences in six lipid metabolism pathway scores between the two subtypes in TCGA dataset; D: Boxplots showing differences in six lipid metabolism pathway scores between the two subtypes in the GSE14520 cohort.

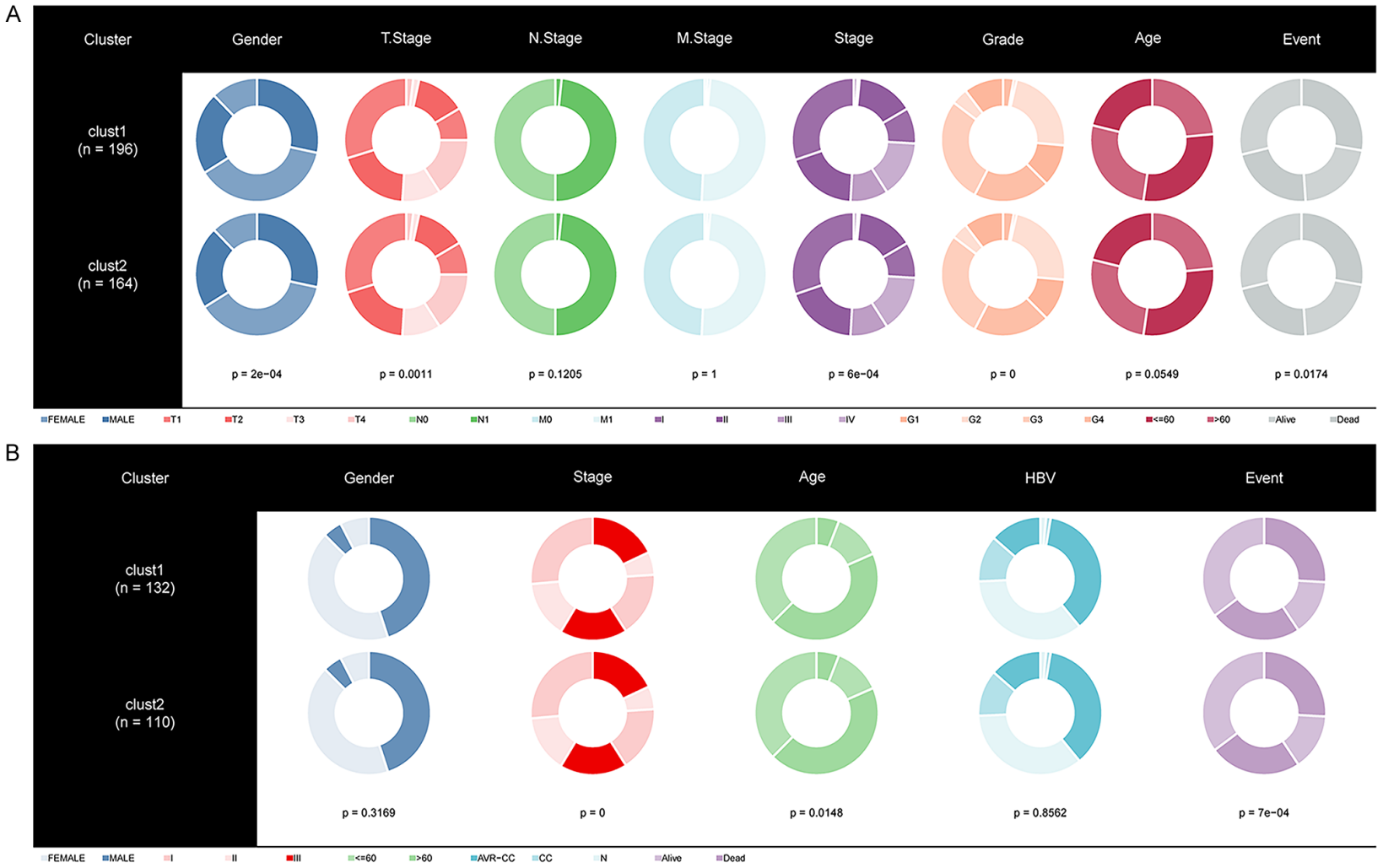


Figure 4. Differences in the clinical characteristics between metabolism-related subtypes. A: Comparison of the distribution of different clinical features between two molecular subtypes in TCGA dataset; B: Comparison of the distribution of different clinical features between two molecular subtypes in the GSE14520 cohort.

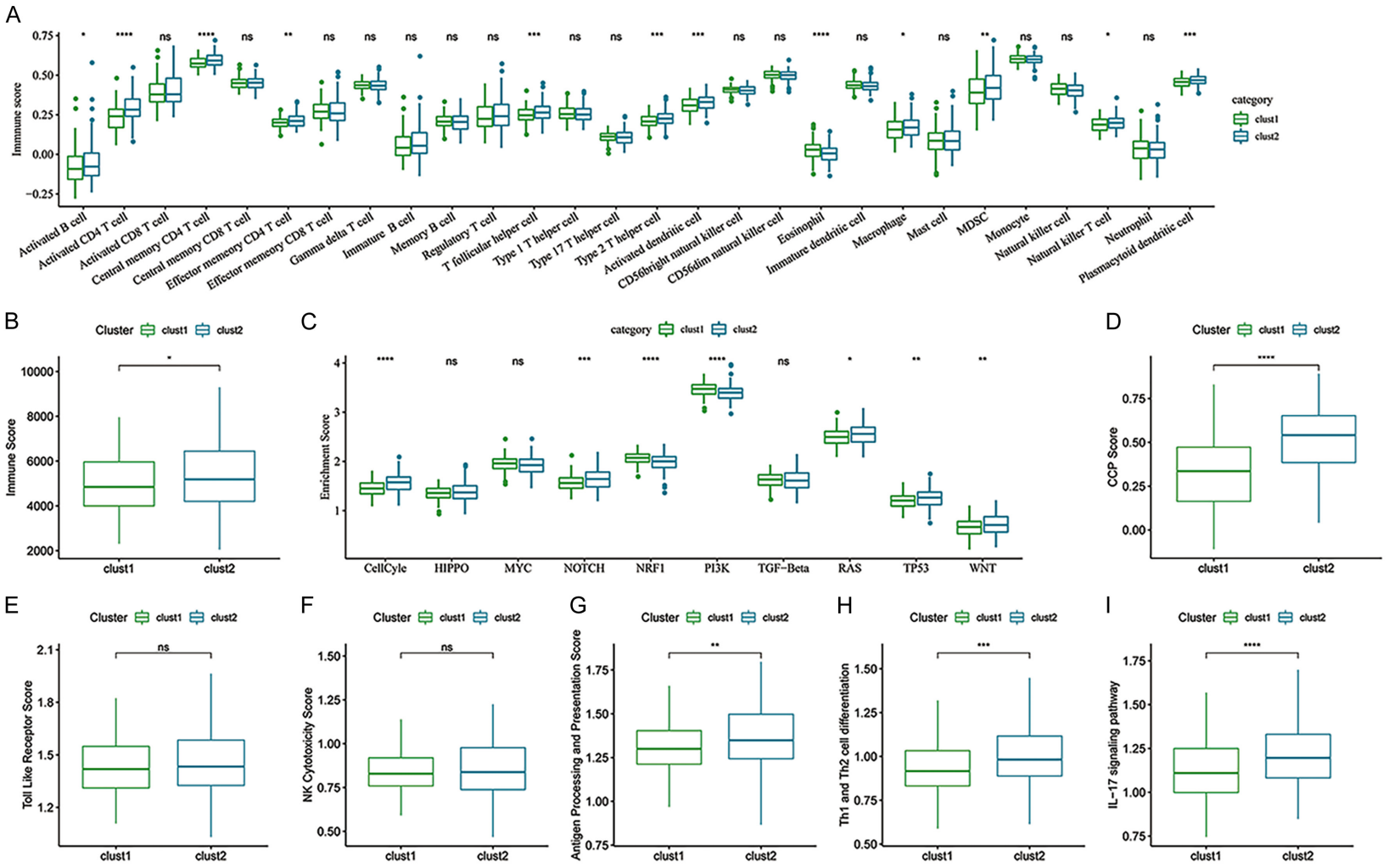


Figure 5. Immune signatures between metabolism-related clusters. A: The differences in immune cell scores between two subtypes in TCGA predicted by ssGSEA; B: The differences in immune cell scores between two subtypes in TCGA predicted by ESTIMATE; C: The differences in ten tumor-associated pathways among two subtypes; D: The differences in CCP scores between two subtypes in TCGA-LIHC dataset; E: The differences in scores for the Toll-like receptor signaling pathway between the two subtypes in TCGA dataset; F: The differences in scores for natural killer cell-mediated cytotoxicity between the two subtypes in TCGA dataset; G: The differences in scores for antigen processing and presentation between the two subtypes in TCGA dataset; H: The differences in scores for Th1 and Th2 cell differentiation between the two subtypes in TCGA dataset; I: The differences in scores for the IL-17 signaling pathway between the two subtypes in TCGA dataset.

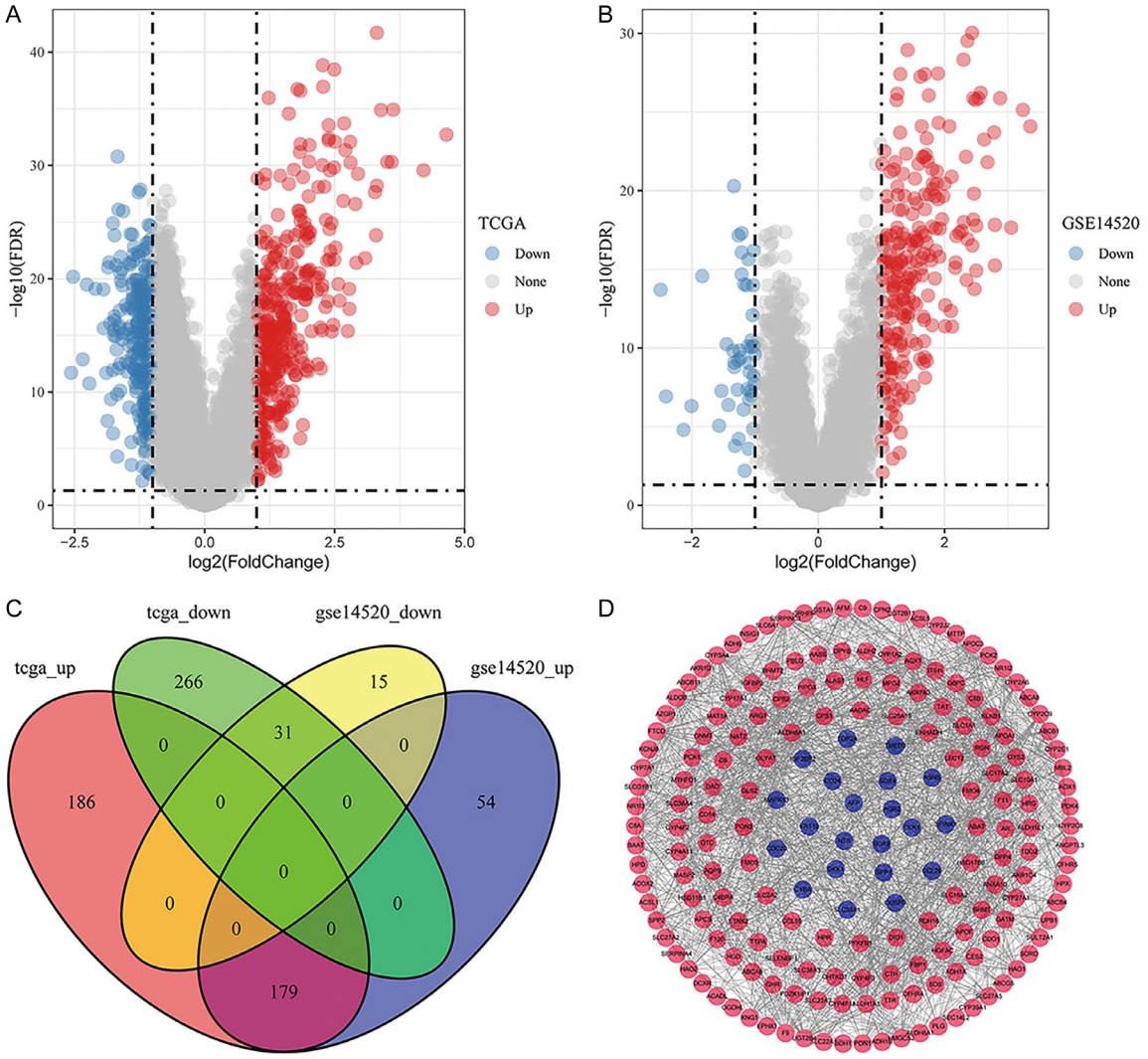


Figure 6. Differential gene expression analysis between metabolism-related subtypes. A: Differentially expressed genes between the two subtypes in TCGA dataset; B: Differentially expressed genes between the two subtypes in the GSE14520 dataset; C: The common differentially expressed genes between the two subtypes in TCGA and GSE14520 datasets; D: Close interactions among 190 genes.

that the scores in C2 were significantly higher than in C1 (**Figure 5H**, **5I**).

DEGs between metabolism-related subtypes

The results showed that there were 365 DEGs in TCGA dataset (**Figure 6A**), among which 233 genes were up-regulated while 46 were down-regulated genes in the GSE14520 dataset (**Figure 6B**). Overlapping analysis suggested that 179 up-regulated genes and 31 down-regulated genes were common between the datasets (**Figure 6C**). Subsequently, we screened protein-protein interaction through STRING (https://cn.string-db.org/) using a confidence score cut-off of ≥ 0.4. 190 genes showing significant interactions among the 210 genes were visualized in Cytoscape. The degree of the network was estimated using Analyze Network in Cytoscape. The higher the degree, the more central the gene in the network, which implied that the corresponding gene held greater importance (Supplementary Table 5). **Table 1** showed the genes with Degree ≥ 30.

Finally, GO and KEGG pathway enrichment analysis were performed to examine the functional characteristics of these 190 genes using Web Gestalt R. The results of GO analysis revealed that these genes were significantly enriched in 218 biological process (BP) terms, 61 molecular function (MF) functions and 22

Table 1. The degree information of crucial genes

Name	Degree
CYP3A4	44
FTCD	40
AGXT	40
CYP2E1	37
APOA1	37
SERPINC1	36
C8A	36
APOC3	35
CYP2C9	34
CYP7A1	33
HRG	33
SLC2A2	32
HGD	31
TAT	31
NR1I2	31
F9	30
F13B	30
CPB2	30

cellular component (CC) functions (FDR < 0.05). The top 10 most significant ones are listed (Supplementary Figure 2A-C). The results of KEGG analysis (FDR < 0.05) showed that these marker genes were significantly enriched in 26 pathways (Supplementary Figure 2D). The detailed information of KEGG and GO enrichment results are provided in Supplementary Table 6.

Unsupervised clustering and dimensionality reduction

First, cells that met the following criteria were filtered: (1) Each gene was expressed in at least 3 cells, and each cell expressed at least 250 genes. (2) Cell expressed 100 UMIs at least, more than 100 genes, and less than 6000 genes. (3) Less than 30% mitochondrial gene expression in UMI counts. PercentageFeatureSet was used to calculate the proportion of mitochondrial genes and rRNA. As shown in Supplementary Figure 3A, UMI and mRNA were significantly correlated with the content of mitochondrial genes, while UMI/mRNA was not. Supplementary Figure 3B, 3C showed the Violin diagrams before and after quality control, respectively.

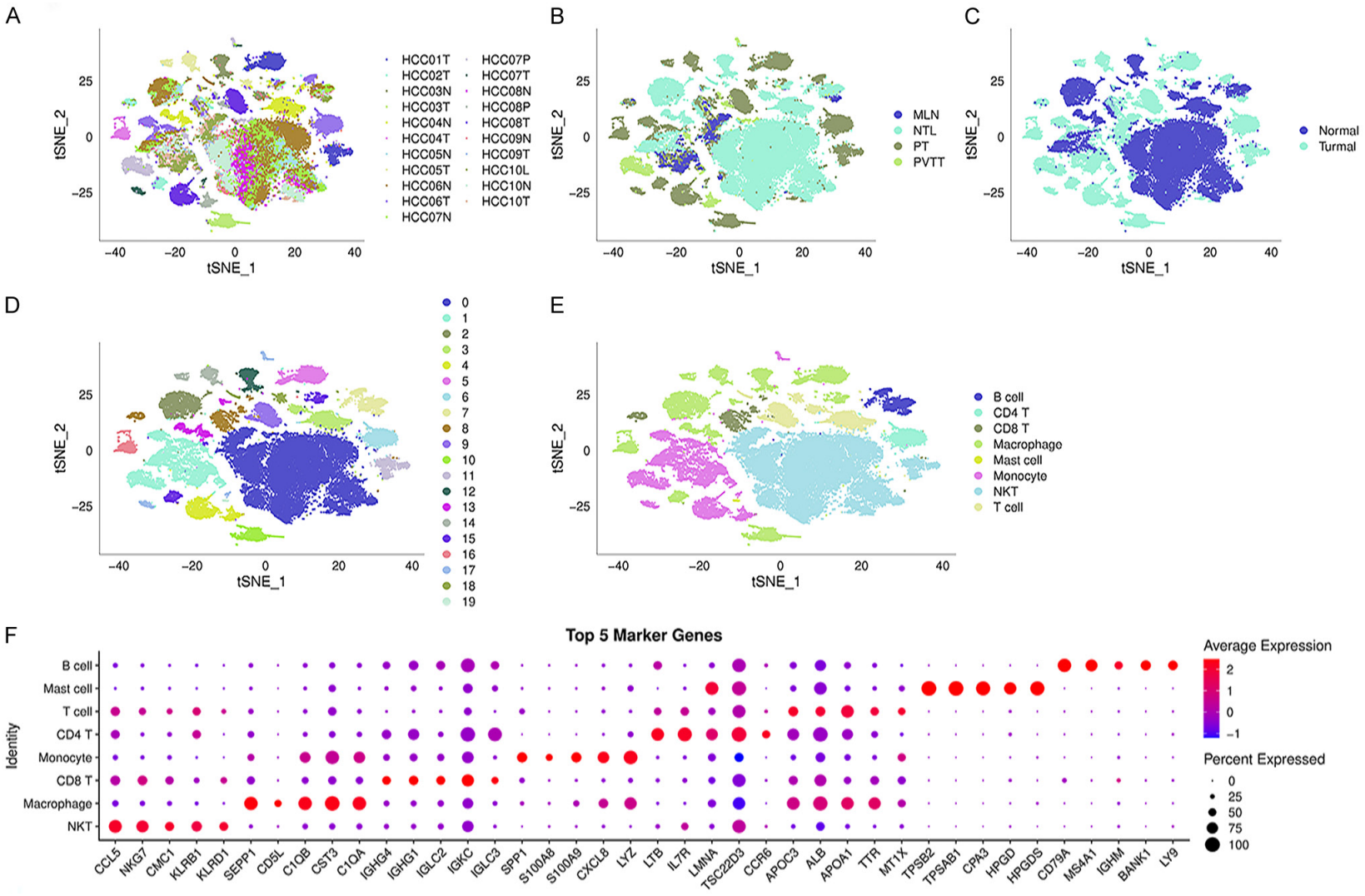
Further, we normalized the data of 4 samples separately using "Log Normalize". Variable genes were selected using FindVariableFeatures. The ScaleData function was used to scale all genes and perform PCA dimensionality reduction to identify anchor points. Dim = 35 was chosen. And then we used the FindNeighbors and FindClusters functions in Seurat to perform cell clustering (Resolution = 0.9). Finally, 51 cell clusters were obtained. Immune cell clusters were identified by PTPRC (CD45) expression in a total of 42,451 cells (Supplementary Figure 3D, 3E).

Next, we chose 42,451 immune cells, and variable genes were selected after normalization and identification of anchor points by PCA dimensionality reduction. Dim = 30 was chosen. We used the FindNeighbors and FindClusters functions in Seurat to perform cell clustering (Resolution = 0.1). Finally, 20 cell clusters were obtained. RunTSNE function was used to perform t-SNE dimensionality reduction in 42,451 immune cells. We also visualized the cell clusters using the RunTSNE function (**Figure 7A**) based on the expression of known immune marker genes (Supplementary Figure 4). **Figure 7B** showed different t-SNE diagrams. t-SNE plots showed the cell distribution between tumor and normal tissues, the distribution of 17 cell clusters and the distribution of cells after annotation (**Figure 7C-E**). **Table 2** listed the statistics of the number of cells in each sample before and after filtering. The FindAllMarkers function was used to screen marker genes in 7 subgroups using logfc = 0.5 (fold difference) and Minpct = 0.35 (the smallest expression ratio of differential genes), along with corrected P < 0.05. Here we only showed the expression of the top 5 marker genes with the most prominent contributions in each cluster (**Figure 7F**). Marker genes were listed in Supplementary Table 7.

Furthermore, we analyzed the proportions of tumor and para-tumor tissues in 8 cell clusters (**Figure 7G**). KEGG pathway enrichment analysis was performed using the cell markers of the 7 cell clusters using the clusterProfiler package (**Figure 7H**). No significantly enriched pathway in CD4 T cells was identified.

Abnormal metabolism at the single-cell level

Although tumor and normal tissues were selected when sampling, tumor tissues may contain some normal cells. To further demonstrate the conclusion obtained from bulk RNA-seq analysis, whereby activated metabolic pathways may be associated with poor prognosis based on the metabolic pathway enrichment score which


A
B
C
D
E
F
tSNE_1
tSNE_2
HCC01T
HCC02T
HCC03N
HCC03T
HCC04N
HCC04T
HCC05N
HCC05T
HCC06N
HCC06T
HCC07N
HCC07P
HCC07T
HCC08N
HCC08P
HCC08T
HCC09N
HCC09T
HCC10L
HCC10N
HCC10T
MLN
NTL
PT
PVTT
Normal
Turmal
B cell
CD4 T
CD8 T
Macrophage
Mast cell
Monocyte
NKT
T cell
Top 5 Marker Genes
Identity
Average Expression
Percent Expressed
CCL5 NKG7 CMC1 KLRB1 KLRD1 SEPP1 CD5L C1QB CST3 C1QA IGHG4 IGHG1 IGLC2 IGKC IGLC3 SPP1 S100A8 S100A9 CXCL8 LYZ LTB IL7R LMNA TSC22D3 CCR6 APOC3 ALB APOA1 TTR MT1X TPSB2 TPSAB1 CPA3 HPGD HPGDS CD79A MS4A1 IGHM BANK1 LY9

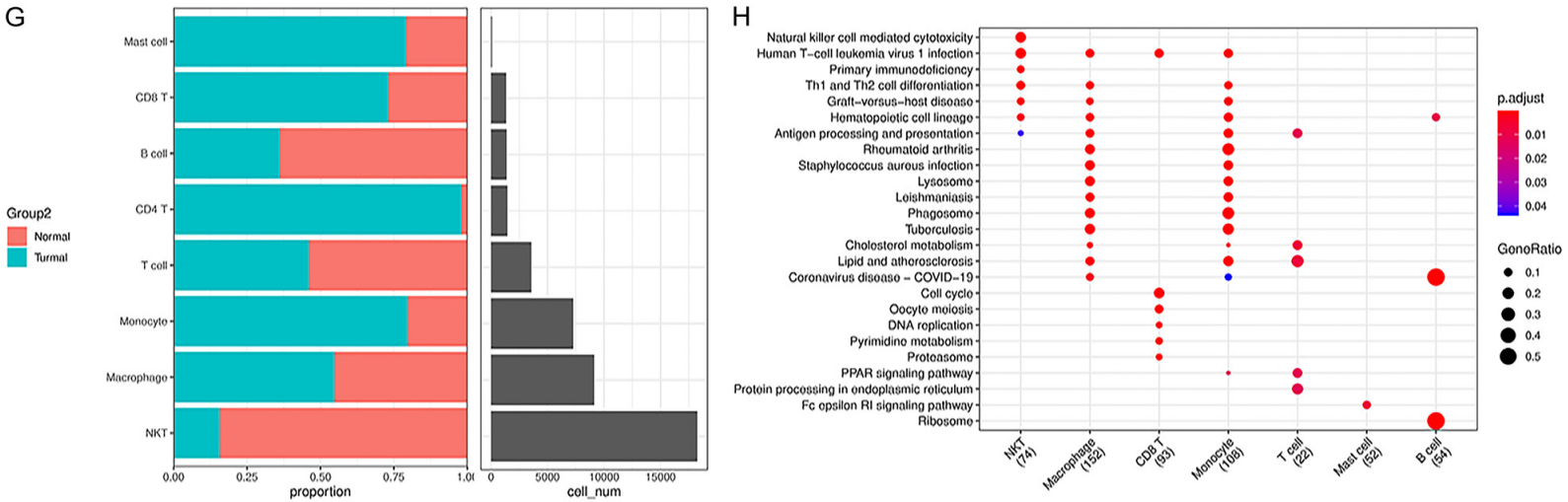


Figure 7. Unsupervised clustering and dimensionality reduction. A: t-SNE plot for 21 sample distributions; B: Different types of t-SNE plots; C: t-SNE plots of cell distributions between tumor and normal tissues; D: t-SNE plot of the distribution in 17 subtypes; E: t-SNE plots of cell clusters after cellular annotation; F: Dot plot for the top 5 marker genes in clusters after annotation; G: The proportions and cell numbers of annotated cell clusters between cancer and para-cancerous tissues; H: Dot plot for KEGG enrichment analysis in subpopulations annotated cell clusters.

Table 2. Statistics of cell counts before and after sample filtration

Sample	raw_count	clean_count	percent %
N1	4243	4052	95.5
N2	2466	2249	91.2
T1	2642	2345	88.76
T2	3203	2571	80.27

was higher in C2 as compared to C1, we distinguished malignant cells from normal cells in the sc-RNA seq immune cells data by comparing the copy number variations (CNV) using the copykat package. A total of 19435 malignant cells, 16398 normal cells, and 6618 unknown cells were identified. Next, we predicted the scores for lipid metabolism-related subtypes (**Figure 8**) based on the 179 co-up-regulated genes (marker genes of C1) and 31 co-down-regulated genes (marker genes of C2) screened by ssGSEA as shown in **Figure 6**. In the microenvironment, the scores for C1 in malignant cells were lower than normal cells, however, the scores of C2 in malignant cells were higher than in C1, consistent with the results of bulk RNA-seq analysis, which meant the prognosis of C1 was good but that of C2 was poor.

Interactions between malignant and immune cells

Cell-cell L/R pairs analysis was performed by cellphoned packbag in 35833 cells including 19435 malignant cells and 16398 normal cells. $P < 0.001$ was used for screening, and 38 L/R pairs were found to play an important role in the interaction between malignant cells and normal immune cells (**Figure 9A**). Furthermore, 11 L/R pairs played an important role in the interaction between malignant cells and normal immune cells (**Figure 9B**) and between normal immune cells and malignant cells (**Figure 9C**). The 11 L/R pairs were as followed: *CD8_LCK, KLRB1_CLEC2D, FAM3C_CLEC2D, CD2_CD58, CD160_TNFRSF14, CD55_ADGRE5, CLEC2B_KLRF1, ICAM1_AREG, TNFSF14_TNFRSF14, SELL_SELPLG,* and *LAMP1_FAM3C*.

The relationship between metabolism of malignant cells and the tumor immune microenvironment

Although malignant cells dominate the lipid metabolism patterns in tumors, we aimed to identify the role of TME and whether it influenced the metabolism of malignant cells. However, environmental factors cannot be directly incorporated into the metabolic analysis, thus, we chose two factors, hypoxia and angiogenesis, for our assessment. Genes associated with the hallmark-hypoxia pathway were used to calculate the hypoxia and angiogenesis scores in malignant cells based on the screened 24 genes using ssGSEA. ssGSEA was used to calculate the scores of the six lipid metabolic pathways in malignant cells. Pearson's chi-squared test was used to analyze the relationship between hypoxia scores, angiogenesis scores, and the six lipid metabolic pathways (**Figure 10A**). The results showed that hypoxia scores were negatively associated while angiogenesis scores were positively associated with alpha-linolenic acid metabolism.

Next, immune cells in 19435 malignant cells were assessed. As shown in **Table 3**, immune cells in malignant cells were almost MDM, and only one Mast cell and T cell were identified. We compared the scores of different immune cells between C1 and C2 (**Figure 10B**). The scores of C1 were significantly lower than those of C2 for all immune cells, indicating that poor prognosis in C2 was advantageous in the malignant cell microenvironment.

Finally, correlational analysis for C1, C2, hypoxia scores, angiogenesis scores, and the 6 lipid metabolic pathways was performed (**Figure 10C**). The result showed that the scores of C1 were significantly associated with steroid hormone biosynthesis and FA metabolism.

Analysis of drug sensitivity between the metabolism-related clusters

Based on the analysis of the two metabolism-related clusters, we identified significant differences in metabolic pathways and TME between them, which play an important role in tumor resistance. We further studied the influence of the targeted drugs in the two metabolism-related clusters using the pRRophetic package in R. The sensitivity to many targeted drugs, including cisplatin, rapamycin, and MG-132 in C1 was higher than in C2 (**Figure 11**).

Network analysis of differential metabolites and genes

According to the genes in **Table 1**, we divided 11 LIHC tissue samples into T and CK group

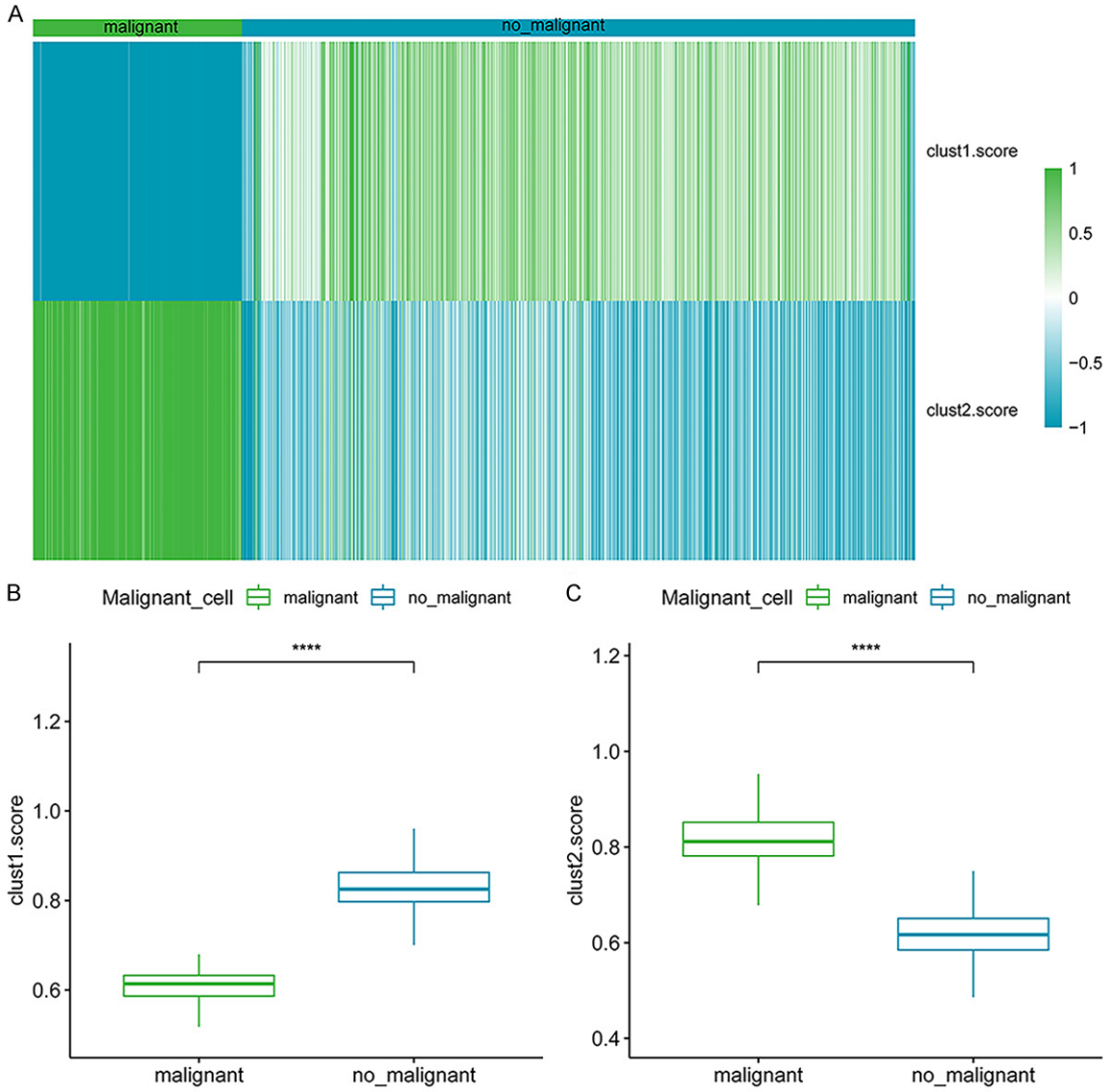


Figure 8. Abnormal metabolism at the single-cell level. A: Scores for malignant cells (cancer cells) and non-malignant cells (normal cells) in C1 and C2; B: The differences in scores between malignant cells (cancer cells) and non-malignant cells (normal cells) in C1; C: The differences in scores between malignant cells (cancer cells) and non-malignant cells (normal cells) in C2.

(**Figure 12A**). Then DEMs was got in both negative and positive ionization modes (**Figure 12B**, **12C**). To further analysis the relationship between genes and DEMs, we constructed a fully connected network of differentially metabolites and genes (**Figure 12D**). Next, genes and metabolites with Spearman correlation coefficients > 0.8 and P < 0.1 were selected. We got 7 genes: CPB2, CYP3A4, TAT, HGD, CYP2E1, NR1I2, CYP7A1 were significantly associated with 11 differentially metabolites (**Figure 12E**).

Discussion

At present, the treatment of liver cancer faces huge challenges of difficulty in prevention, late detection, and few available therapeutic targets. Targeted immunotherapy shows low overall response rates and varies greatly among patients, mainly due to the lack of knowledge of the molecular subtypes and factors driving the progression of liver cancer. At present, the clinical decisions for liver cancer treatment are often based on the TNM staging system, which

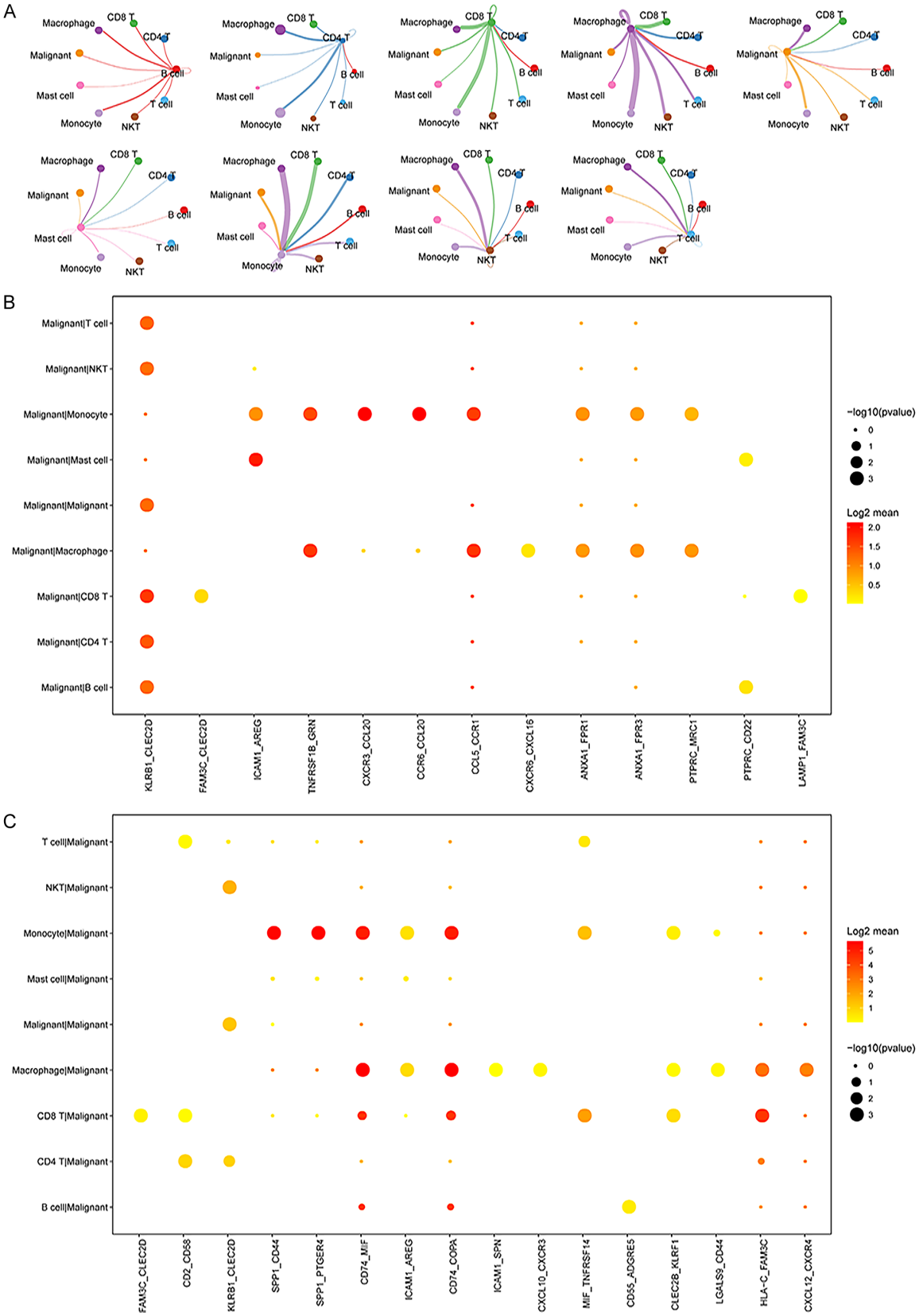
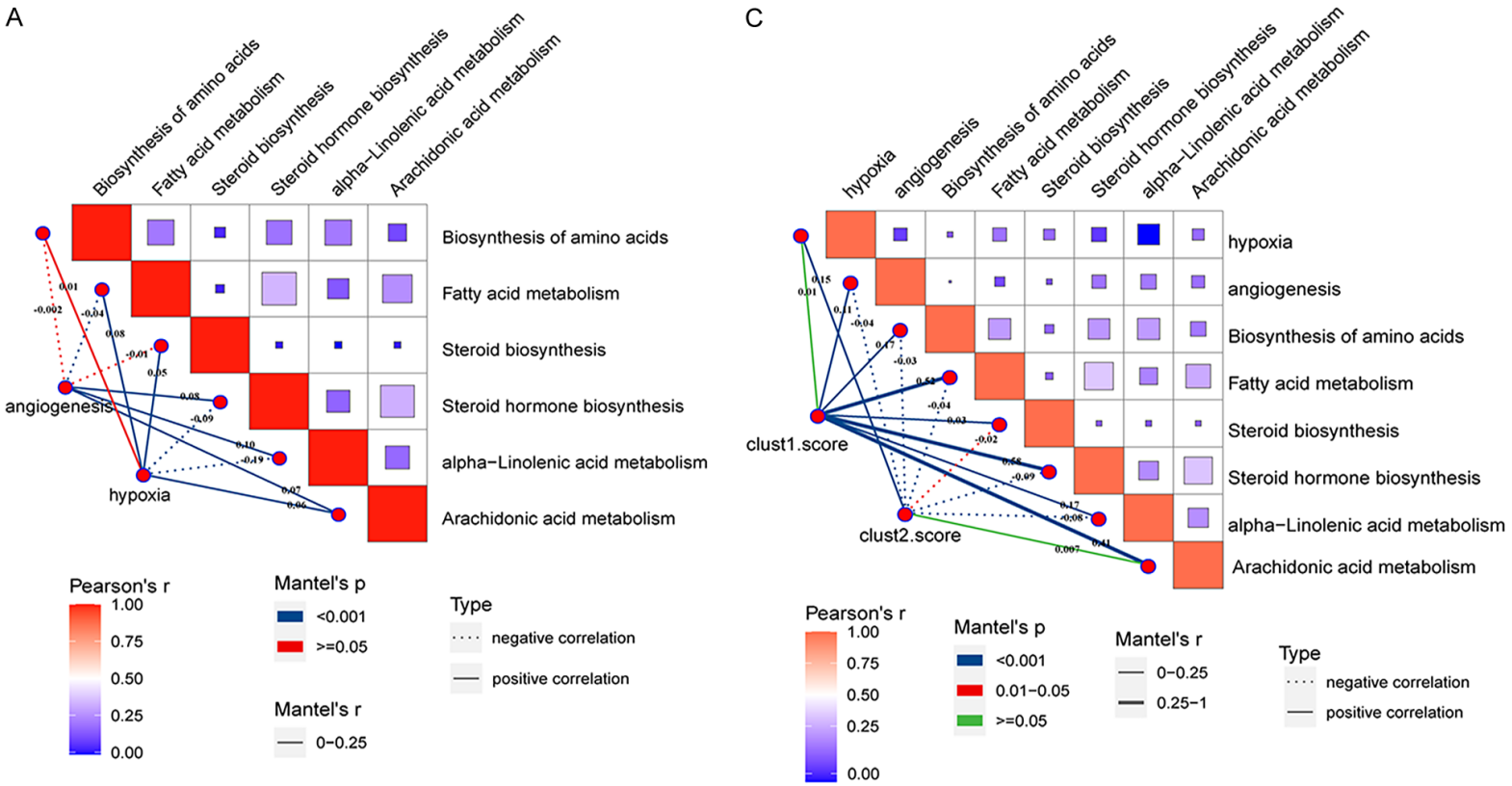


Figure 9. Interactions between malignant and immune cells. A: The network shows the number of significant interactions between different cell populations, with thicker lines representing higher numbers of interactions (more ligand receptors); B: Dot plot of ligand-receptor pairs between malignant cells and normal immune cells; C: Dot plot of ligand-receptor pairs between normal immune cells and malignant cells. The color represents the magnitude of the interaction, and the larger the point, the smaller the significance (*p*-value).


A
Biosynthesis of amino acids
Fatty acid metabolism
Steroid biosynthesis
Steroid hormone biosynthesis
alpha−Linolenic acid metabolism
Arachidonic acid metabolism
angiogenesis
hypoxia
Pearson's r
1.00
0.75
0.50
0.25
0.00
Mantel's p
<0.001
>=0.05
Mantel's r
0−0.25
Type
negative correlation
positive correlation
C
hypoxia
angiogenesis
Biosynthesis of amino acids
Fatty acid metabolism
Steroid biosynthesis
Steroid hormone biosynthesis
alpha−Linolenic acid metabolism
Arachidonic acid metabolism
clust1.score
clust2.score
Pearson's r
1.00
0.75
0.50
0.25
0.00
Mantel's p
<0.001
0.01−0.05
>=0.05
Mantel's r
0−0.25
0.25−1
Type
negative correlation
positive correlation

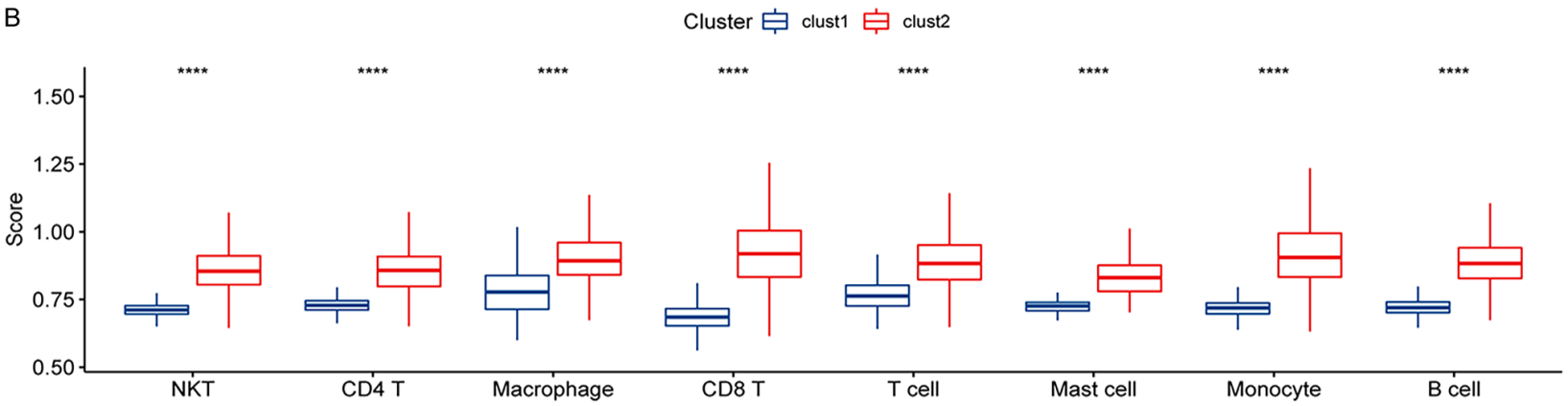

B
Cluster
clust1
clust2
Score
1.50
1.25
1.00
0.75
0.50
****
NKT
CD4 T
Macrophage
CD8 T
T cell
Mast cell
Monocyte
B cell

Figure 10. The relationship between metabolism of malignant cells and the tumor immune microenvironment. (A) Correlational analysis of hypoxic type scores and angiogenesis scores with six lipid metabolism pathways; (B) A differential analysis based on lipid metabolism-related subtypes; (C) The correlational analysis of the hypoxia-type, angiogenesis, and six lipid metabolism pathway scores with the subtypes; in (A-C), the thickness of the corresponding lines represent the absolute values of the correlation between angiogenesis scores, hypoxia scores, C1 scores, C2 scores, and lipid metabolism pathways. The type of line shows a positive or negative correlation, and the color indicates the magnitude of significance.

Table 3. The composition of malignant cells

	C0	C11	C1	C17	C4	C2	C5	C10	C12	C13	C14	C15	C16	C9	C18	C6	C7	C8	C19
NKT	12359	792	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Monocyte	0	0	457	63	36	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Macrophage	0	0	0	0	0	85	59	57	151	207	41	94	29	0	0	0	0	0	0
T cell	0	0	0	0	0	0	0	0	0	0	0	0	0	818	396	0	0	0	0
CD4 T	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1241	0	0	0
B cell	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	870	0	0
CD8 T	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	612	0
Mast cell	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	45

relies on dividing the patients into different stages according to clinical features associated with prognosis, especially pathological factors [43]. The clinical factors associated with the prognosis of liver cancer mainly reflect the extent of tumor spread including tumor size, the number of tumor nodules, vascular invasion, distant metastasis, and severity of liver damage including protein synthesis and detoxification, and liver decompensation symptoms [44]. Despite great progress made in guiding clinical practice for tangible therapeutic benefits for patients with liver cancer based on the classification methods of pathological diagnosis. However, it is difficult to reflect the biological nature of the tumor, especially essential differences at the molecular level, because tumor biological features can be defined only at the tissue level [45]. At the same time, it's noteworthy that despite in the same TNM stage, significant differences in survival time and responses to the therapy across patients are found owing to the high heterogeneity of LIHC.

With the advancement and popularization of high-throughput sequencing technology and parallel detection technology, related research and applications in the field of biomedicine have advanced. Traditional bulk RNA sequencing is based on tissue samples and reflects the average expression in cell populations. However, extensive heterogeneity is present between cells, which plays an important role in the response to targeted therapy in tumors. In recent years, scRNA-seq, which can reveal the expression of all genes in the whole genome at the single-cell level along with tumor heterogeneity, has been vigorously developed [46, 47].

The development of scRNA-seq has provided new methods for the investigation of liver cancer. Zhang et al. harvested 42 samples from eight LIHC patients and evaluated tumor heterogeneity by whole-exome sequencing, RNA sequencing, mass spectrometry-based proteomics and metabolomics, cytometry by time-of-flight, and single-cell analysis [48]. Hou et al. used scTrio-seq and simultaneously assessed the genome, methylome, and transcriptome at the single-cell level, to analyze 25 single cells derived from a LIHC tissue sample and found two subpopulations showing distinct DNA copy numbers, DNA methylation, RNA expression. Higher copy number variations and methylation levels were associated with a greater likelihood of metastasis [49]. These studies comprehensively reveal the mechanism and driving factors of liver cancer from multi-dimensional, multi-omics, and multi-system perspectives, facilitating the development of new and precise diagnostic and treatment strategies.

As the pivotal organ for carbohydrate, lipid, and protein metabolism, metabolic function is the most important feature of the liver tissue. Several factors lead to metabolism disorders of the liver and are implicated in liver cancer. In recent years, cases of liver cancer caused by NAFLD/NASH associated with metabolic syndromes and obesity are on the rise, highlight-

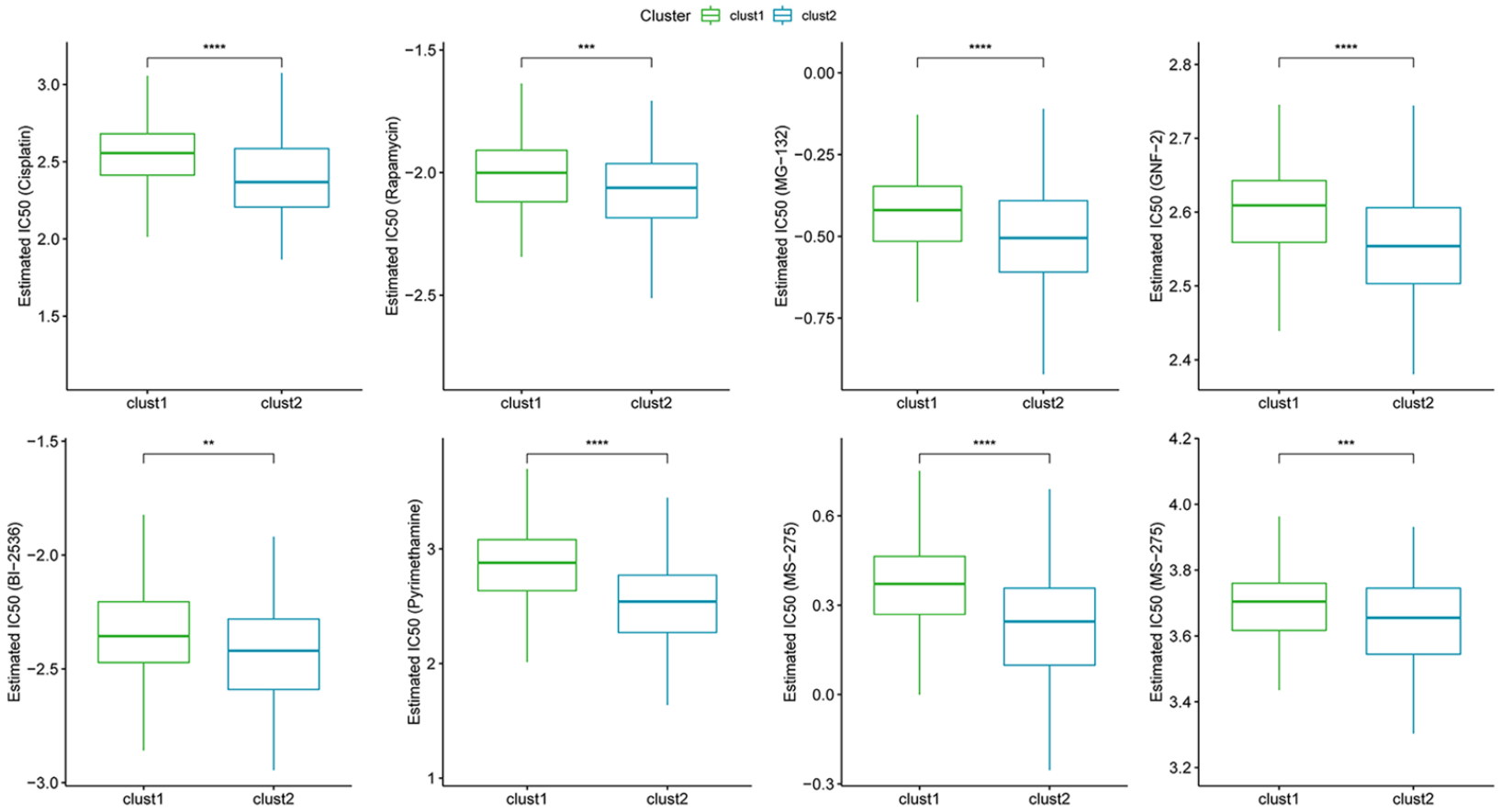
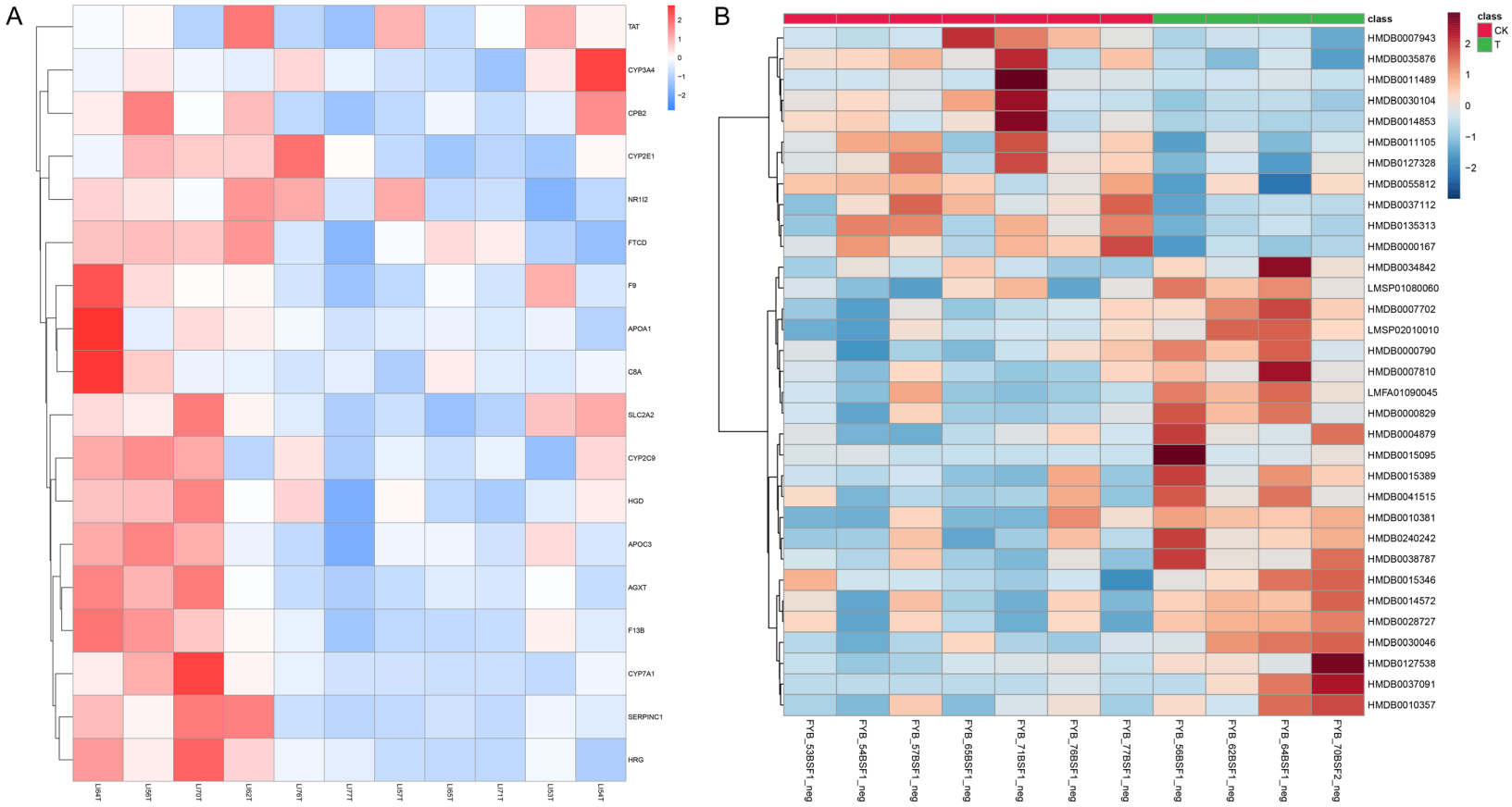


Figure 11. Analysis of drug sensitivity between the clusters (****P < 0.0001, ***P < 0.001, **P < 0.01).


A
TAT
CYP3A4
CPB2
CYP2E1
NR1I2
FTCD
F9
APOA1
C8A
SLC2A2
CYP2C9
HGD
APOC3
AGXT
F13B
CYP7A1
SERPINC1
HRG
2
1
0
−1
−2
B
class
HMDB0007943
HMDB0035876
HMDB0011489
HMDB0030104
HMDB0014853
HMDB0011105
HMDB0127328
HMDB0055812
HMDB0037112
HMDB0135313
HMDB0000167
HMDB0034842
LMSP01080060
HMDB0007702
LMSP02010010
HMDB0000790
HMDB0007810
LMFA01090045
HMDB0000829
HMDB0004879
HMDB0015095
HMDB0015389
HMDB0041515
HMDB0010381
HMDB0240242
HMDB0038787
HMDB0015346
HMDB0014572
HMDB0028727
HMDB0030046
HMDB0127538
HMDB0037091
HMDB0010357
FYB_53BSF1_neg
FYB_54BSF1_neg
FYB_57BSF1_neg
FYB_66BSF1_neg
FYB_71BSF1_neg
FYB_76BSF1_neg
FYB_77BSF1_neg
FYB_56BSF1_neg
FYB_62BSF1_neg
FYB_64BSF1_neg
FYB_70BSF2_neg
class
CK
T
2
1
0
−1
−2

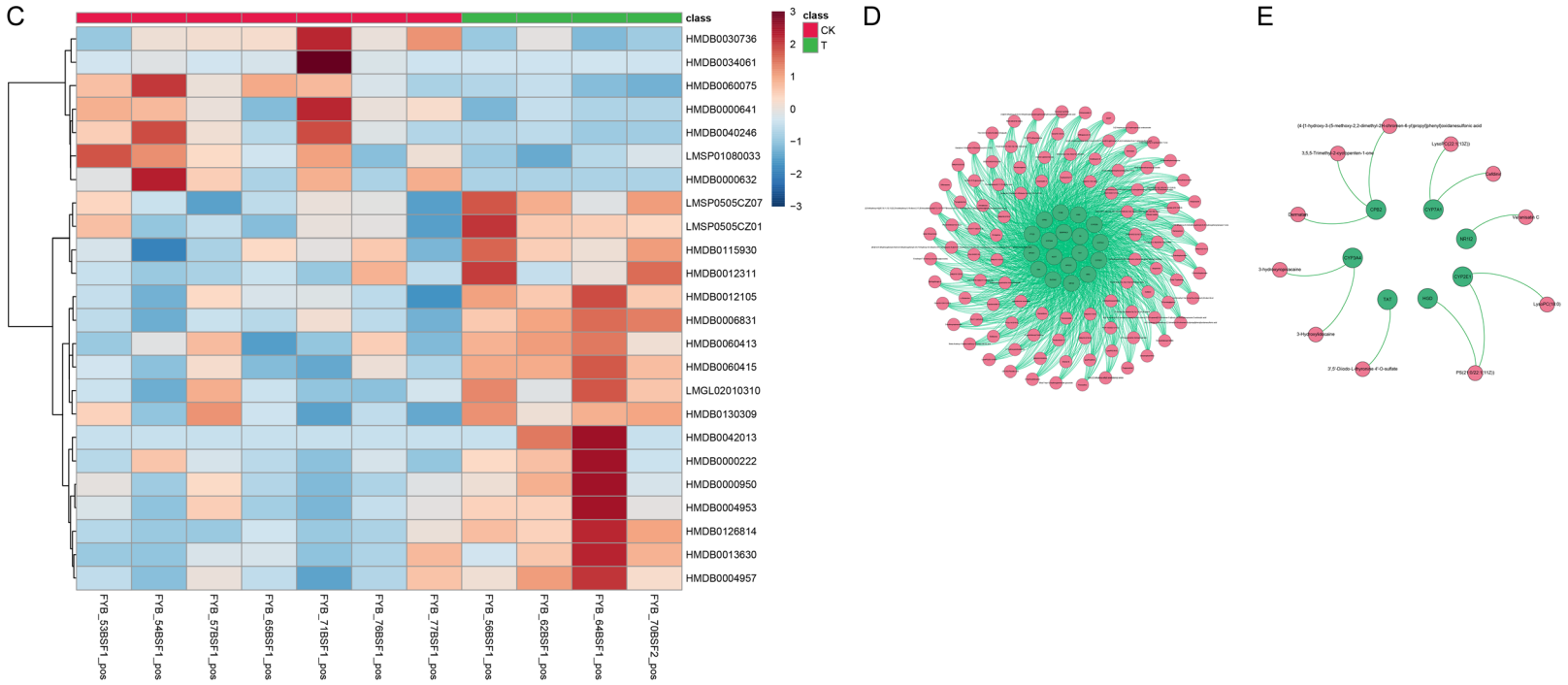


Figure 12. The relationship between DEGs and DEMs in 11 LIHC tissue samples. A: 11 LIHC tissue samples were devided into 2 groups according to the DEGs; B: The DEMs in negative and positive ionization modes; C: The DEMs in positive ionization modes; D: A fully connected network of differentially metabolites and genes; E: The network of significantly associated metabolites and genes.

ing the importance of lipid metabolism in hepatocarcinogenesis. Studies show that increased de novo synthesis of lipids is a common feature of many human cancers [50]. FA turnover is high in tumor cells to meet the energy and synthetic requirements for their [51]. Moreover, FAS and the key enzymes that catalyze FAS are overexpressed in human malignant tumors. Lipolytic enzymes that hydrolyze monoacylglycerols to release free FAs are overexpressed in aggressive tumor cells; inhibiting FAS can suppress the proliferative activity of several malignant tumor cell lines [52]. Luo et al. found that tumor-associated macrophages (TAMs) from tumor-bearing mice have higher lipid content with decreased phagocytic potency as compared to macrophages from tumor-free mice [53], suggesting that these tumors can evade the body's immune surveillance through mechanisms involving changes in tumor lipid pools and accumulation of lipids and FAs [54].

Taken together, this study is the first to analyze the heterogeneity of liver cancer from the perspective of lipid metabolism. Here we classified 360 LIHC samples from TCGA based on six lipid metabolic pathway-associated genes, and assigned them to two molecular subtypes (C1 and C2), exhibiting different clinical features and prognostic outcomes. Apart from metabolic activity, significant differences in clinical features and immune characteristics between the two subtypes were present and closely related to the prognosis. In general, patients in C2 showed poor prognosis, lower scores for lipid metabolic pathways, a higher proportion of deaths, higher T-stage, more advanced staging, and higher immune scores as compared to those in C1. To verify the effectiveness of this classifier in predicting prognosis, we further validated it in the GSE10143 and GSE149614 datasets, which confirmed significant differences in the survival time between the two subtypes. The sensitivity to many targeted drugs in the C1 group was significantly higher than that in C2, suggesting that patients in C1 could benefit substantially from targeted therapy.

ssGSEA was performed to further assess the two subtypes, and the results showed that different genes were significantly enriched based on tumor features and metabolism characteristics. The expression of several tumor-related pathways was significantly higher in the C2 subtype than in the C1 subtype, suggesting that tumors of the C2 category were more aggressive, consistent with the clinical characteristics, including advanced stage, high degree of differentiation, and high mortality. Compared with the C2 subtype, metabolism-related pathways were enriched in the C1 subtype and were mostly associated with the physiological and metabolic functions of hepatocytes such as FA metabolism, biosynthesis of unsaturated FAs, and alpha-linolenic acid metabolism, indicating that the hepatocyte function was more complete. Thus, the C1 subtype showed a better prognosis than the C2 subtype. The analysis of TME showed significant differences in B cell, T cell, and macrophage scores between the two subtypes. The results of bulk RNA-seq showed that the enriched scores of malignant cells in C2 with a poor prognosis were higher than those in C1. Furthermore, malignant cells dominated the lipid metabolism pattern of tumors and influenced their metabolism through the TME. Moreover, the two subtypes showed strong concordance with molecular subtypes obtained from previous studies in different LIHC populations, which supported the molecular typing based on the lipid metabolism function for liver cancer. These findings suggest that molecular typing can reflect the relationship between liver cancer at the molecular and pathological levels, and is worthy of further in-depth investigation.

High heterogeneity in liver cancer poses substantial barriers to improving patient outcomes. There are numerous published studies employing gene chips and next-generation sequencing for liver cancer [55-57]. Breuhahn et al. distinguished LIHC into group A (65%) and group B (35%) based on clustering using the most variable genes; cDNA microarray analyses provided subtyping for LIHC related to intratumor inflammation and tumor cell apoptosis [58]. Chiang et al. characterized copy number variant alterations and gene expression profiles and identified 5 classes based on the clustering of gene expression, further supported by the molecular data [59]. Dong et al. used Cox regression as well as SVM-RFE and FW-SVM algorithms to analyze DNA methylation data and constructed a model using three risk categories to predict the overall survival, showing satisfactory predictive power [60]. Hoshi-

da et al. performed a meta-analysis of gene expression profiles in a dataset comprising 603 patients, in an attempt to obtain universal molecular classifications for liver cancer. They found distinct molecular subclasses in several datasets [61], which supported the view that although the results of integrated genomic, transcriptomic, and epigenomic analyses across different studies vary, the high-frequency genes and the main pathways are consistent [62]. Regardless of the clinical heterogeneity across liver cancer patient populations worldwide, certain global features are shared at the molecular level.

Of course, several aspects need improvement and further investigation for utilizing molecular typing based on the metabolic functions in liver cancer. First, obtaining information more easily, given that analyses usually require complete gene data to assess the activity of pathways based on enrichment using scRNA-sq. The roles of specific metabolic and signaling pathways across metabolic subtypes of liver cancer and the differences need to be elucidated in the future. The effects of metabolic subtypes of liver cancer on the patient's systemic metabolic status and the crucial factors affecting the prognosis of these patients warrant further studies. In the future, we plan to perform more experiments to validate the results of this study and use more rational bioinformatic strategies to improve the model. Several changes have occurred in the population of liver cancer in the past several years. The proportions of early liver cancer cases and those caused by non-hepatitis virus are increasing [63, 64], the molecular characteristics and typing are thus essential in liver cancer research.

Conclusion

In summary, we constructed lipid metabolism subtypes using six lipid metabolism-related pathways, which may be used as independent prognostic factors for liver cancer. We further analyzed the characteristic differences in the tumor immune microenvironment and drug sensitivities between the two subtypes to assess the prognostic risk of liver cancer patients, which provides support for the clinical diagnosis based on the stage, individualized treatment strategies, and prognostic prediction for liver cancer.

Acknowledgements

This study was supported by the Natural Science Foundation of China (NSFC No. 81903964 to HH; NSFC No. 82174133 to SJ).

Disclosure of conflict of interest

None.

Address correspondence to: Shi Jin and Hui He, Department of Laparoscopic Surgery, The First Affiliated Hospital of Dalian Medical University, Lianhe Road 193#, Shahekou Distinct, Dalian 116000, Liaoning, China. Tel: +86-0411-83635963 Ext. 2066; E-mail: jinshi@dmu.edu.cn (SJ); hehui@dmu.edu.cn (HH)

References

[1] Petrick JL, Braunlin M, Laversanne M, Valery PC, Bray F and McGlynn KA. International trends in liver cancer incidence, overall and by histologic subtype, 1978-2007. Int J Cancer 2016; 139: 1534-1545.

[2] Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A and Bray F. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2021; 71: 209-249.

[3] Llovet JM, Kelley RK, Villanueva A, Singal AG, Pikarsky E, Roayaie S, Lencioni R, Koike K, Zucman-Rossi J and Finn RS. Hepatocellular carcinoma. Nat Rev Dis Primers 2021; 7: 6.

[4] Forner A, Reig M and Bruix J. Hepatocellular carcinoma. Lancet 2018; 391: 1301-1314.

[5] McGlynn KA, Petrick JL and London WT. Global epidemiology of hepatocellular carcinoma: an emphasis on demographic and regional variability. Clin Liver Dis 2015; 19: 223-238.

[6] Tellapuri S, Sutphin PD, Beg MS, Singal AG and Kalva SP. Staging systems of hepatocellular carcinoma: a review. Indian J Gastroenterol 2018; 37: 481-491.

[7] Cancer Genome Atlas Research Network. Electronic address: wheeler@bcm.edu; Cancer Genome Atlas Research Network. Comprehensive and integrative genomic characterization of hepatocellular carcinoma. Cell 2017; 169: 1327-1341, e23.

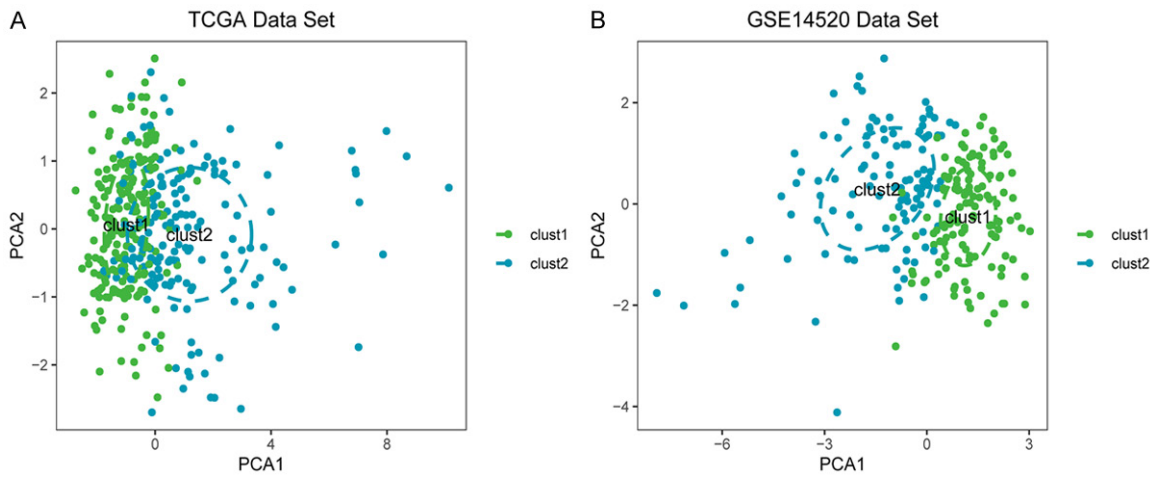
[8] Goh EL, Chidambaram S and Ma S. Laparoscopic vs open hepatectomy for hepatocellular carcinoma in patients with cirrhosis: a meta-analysis of the long-term survival outcomes. Int J Surg 2018; 50: 35-42.

[9] McGranahan N and Swanton C. Clonal heterogeneity and tumor evolution: past, present, and the future. Cell 2017; 168: 613-628.

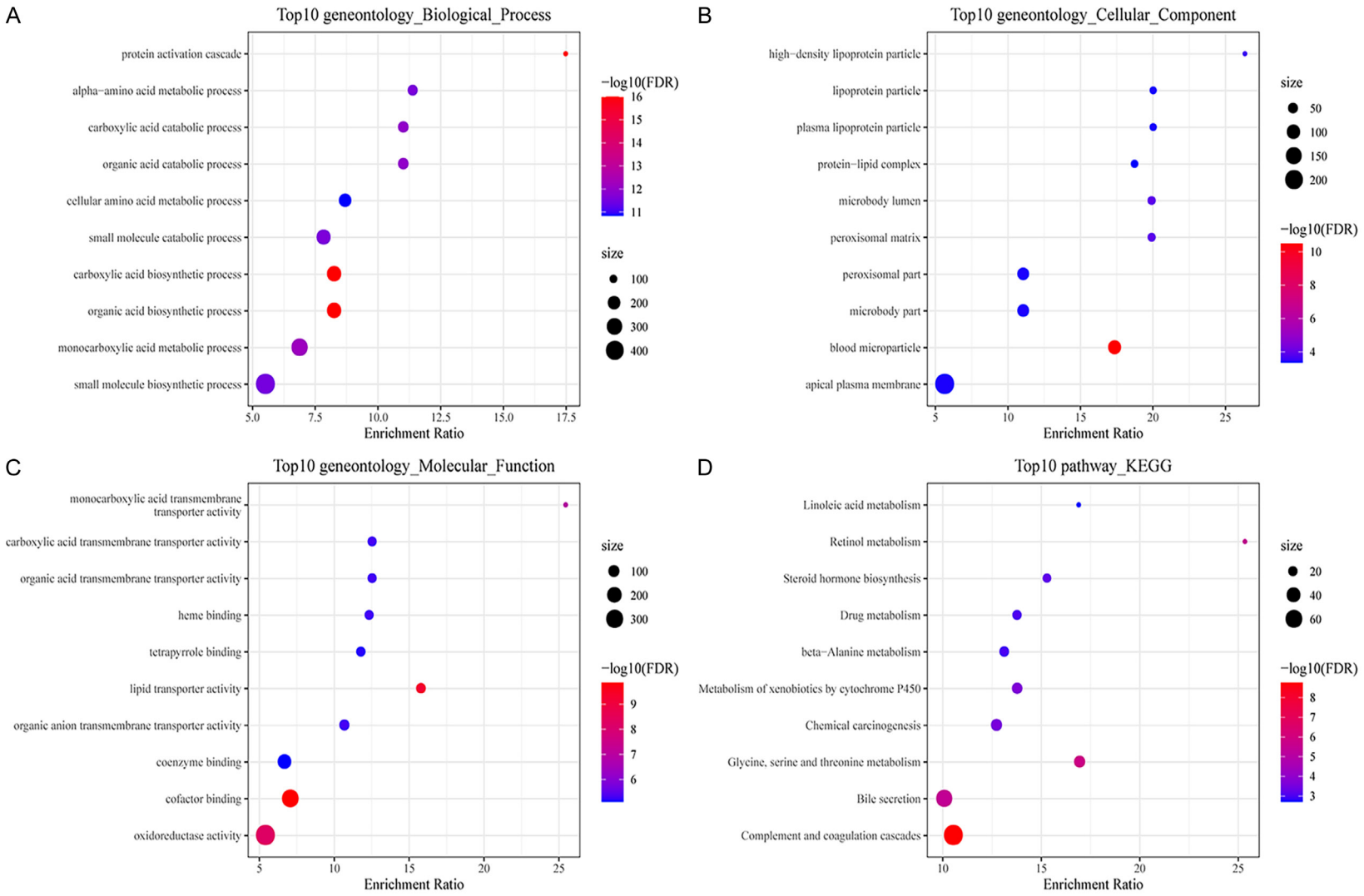
[10] Califano A and Alvarez MJ. The recurrent architecture of tumour initiation, progression and drug sensitivity. Nat Rev Cancer 2017; 17: 116-130.
[11] Stanta G, Jahn SW, Bonin S and Hoefler G. Tumour heterogeneity: principles and practical consequences. Virchows Arch 2016; 469: 371-384.
[12] Furlong LI. Human diseases through the lens of network biology. Trends Genet 2013; 29: 150-159.
[13] Gerlinger M, Rowan AJ, Horswell S, Math M, Larkin J, Endesfelder D, Gronroos E, Martinez P, Matthews N, Stewart A, Tarpey P, Varela I, Phillimore B, Begum S, McDonald NQ, Butler A, Jones D, Raine K, Latimer C, Santos CR, Nohadani M, Eklund AC, Spencer-Dene B, Clark G, Pickering L, Stamp G, Gore M, Szallasi Z, Downward J, Futreal PA and Swanton C. Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. N Engl J Med 2012; 366: 883-892.
[14] Tang F, Barbacioru C, Wang Y, Nordman E, Lee C, Xu N, Wang X, Bodeau J, Tuch BB, Siddiqui A, Lao K and Surani MA. mRNA-Seq whole-transcriptome analysis of a single cell. Nat Methods 2009; 6: 377-382.
[15] Aizarani N, Saviano A, Sagar, Mailly L, Durand S, Herman JS, Pessaux P, Baumert TF and Grun D. A human liver cell atlas reveals heterogeneity and epithelial progenitors. Nature 2019; 572: 199-204.
[16] Pavlova NN and Thompson CB. The emerging hallmarks of cancer metabolism. Cell Metab 2016; 23: 27-47.
[17] Tsoulfas G. Hepatocellular carcinoma and metabolic syndrome: the times are changing and so should we. World J Gastroenterol 2019; 25: 3842-3848.
[18] Guo JN, Chen D, Deng SH, Huang JR, Song JX, Li XY, Cui BB and Liu YL. Identification and quantification of immune infiltration landscape on therapy and prognosis in left- and right-sided colon cancer. Cancer Immunol Immunother 2022; 71: 1313-1330.
[19] DeBerardinis RJ, Lum JJ, Hatzivassiliou G and Thompson CB. The biology of cancer: metabolic reprogramming fuels cell growth and proliferation. Cell Metab 2008; 7: 11-20.
[20] Yang G, Wang Y, Feng J, Liu Y, Wang T, Zhao M, Ye L and Zhang X. Aspirin suppresses the abnormal lipid metabolism in liver cancer cells via disrupting an NFkappaB-ACSL1 signaling. Biochem Biophys Res Commun 2017; 486: 827-832.
[21] Merino Salvador M, Gomez de Cedron M, Moreno Rubio J, Falagan Martinez S, Sanchez Martinez R, Casado E, Ramirez de Molina A and Sereno M. Lipid metabolism and lung cancer. Crit Rev Oncol Hematol 2017; 112: 31-40.
[22] Zhao Y, Li H, Zhang Y, Li L, Fang R, Li Y, Liu Q, Zhang W, Qiu L, Liu F, Zhang X and Ye L. Oncoprotein HBXIP modulates abnormal lipid metabolism and growth of breast cancer cells by activating the LXRs/SREBP-1c/FAS signaling cascade. Cancer Res 2016; 76: 4696-4707.
[23] Fernandez LP, Gomez de Cedron M and Ramirez de Molina A. Alterations of lipid metabolism in cancer: implications in prognosis and treatment. Front Oncol 2020; 10: 577420.
[24] Park JH, Pyun WY and Park HW. Cancer metabolism: phenotype, signaling and therapeutic targets. Cells 2020; 9: 2308.
[25] Jiang Y, Sun A, Zhao Y, Ying W, Sun H, Yang X, Xing B, Sun W, Ren L, Hu B, Li C, Zhang L, Qin G, Zhang M, Chen N, Zhang M, Huang Y, Zhou J, Zhao Y, Liu M, Zhu X, Qiu Y, Sun Y, Huang C, Yan M, Wang M, Liu W, Tian F, Xu H, Zhou J, Wu Z, Shi T, Zhu W, Qin J, Xie L, Fan J, Qian X and He F; Chinese Human Proteome Project (CNHPP) Consortium. Proteomics identifies new therapeutic targets of early-stage hepatocellular carcinoma. Nature 2019; 567: 257-261.
[26] Nisthul A A, Retnakumari AP, A S, Anto RJ and Sadasivan C. In silico screening for identification of fatty acid synthase inhibitors and evaluation of their antiproliferative activity using human cancer cell lines. J Recept Signal Transduct Res 2018; 38: 335-341.
[27] Noureddin M and Rinella ME. Nonalcoholic fatty liver disease, diabetes, obesity, and hepatocellular carcinoma. Clin Liver Dis 2015; 19: 361-379.
[28] Dhamija E, Paul SB and Kedia S. Non-alcoholic fatty liver disease associated with hepatocellular carcinoma: an increasing concern. Indian J Med Res 2019; 149: 9-17.
[29] Wong RJ, Cheung R and Ahmed A. Nonalcoholic steatohepatitis is the most rapidly growing indication for liver transplantation in patients with hepatocellular carcinoma in the U.S. Hepatology 2014; 59: 2188-2195.
[30] Vigano L, Conci S, Cescon M, Fava C, Capelli P, D'Errico A, Torzilli G, Di Tommaso L, Giuliante F, Vecchio FM, Salizzoni M, David E, Pinna AD, Guglielmi A and Capussotti L. Liver resection for hepatocellular carcinoma in patients with metabolic syndrome: a multicenter matched analysis with HCV-related HCC. J Hepatol 2015; 63: 93-101.
[31] Weinmann A, Alt Y, Koch S, Nelles C, Duber C, Lang H, Otto G, Zimmermann T, Marquardt JU, Galle PR, Worns MA and Schattenberg JM. Treatment and survival of non-alcoholic steatohepatitis associated hepatocellular carcinoma. BMC Cancer 2015; 15: 210.

[32] Rich NE, Yopp AC and Singal AG. Medical management of hepatocellular carcinoma. J Oncol Pract 2017; 13: 356-364.
[33] Couri T and Pillai A. Goals and targets for personalized therapy for HCC. Hepatol Int 2019; 13: 125-137.
[34] Personeni N and Rimassa L. Hepatocellular carcinoma: a global disease in need of individualized treatment strategies. J Oncol Pract 2017; 13: 368-369.
[35] Huang A, Yang XR, Chung WY, Dennison AR and Zhou J. Targeted therapy for hepatocellular carcinoma. Signal Transduct Target Ther 2020; 5: 146.
[36] Liberzon A, Subramanian A, Pinchback R, Thorvaldsdóttir H, Tamayo P and Mesirov JP. Molecular signatures database (MSigDB) 3.0. Bioinformatics 2011; 27: 1739-1740.
[37] Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP and Tamayo P. The molecular signatures database (MSigDB) hallmark gene set collection. Cell Syst 2015; 1: 417-425.
[38] Kobayashi Y, Kushihara Y, Saito N, Yamaguchi S and Kakimi K. A novel scoring method based on RNA-Seq immunograms describing individual cancer-immunity interactions. Cancer Sci 2020; 111: 4031-4040.
[39] Ferreira MR, Santos GA, Biagi CA, Silva Junior WA and Zambuzzi WF. GSVA score reveals molecular signatures from transcriptomes for biomaterials comparison. J Biomed Mater Res A 2021; 109: 1004-1014.
[40] Zhuang W, Sun H, Zhang S, Zhou Y, Weng W, Wu B, Ye T, Huang W, Lin Z, Shi L and Shi K. An immunogenomic signature for molecular classification in hepatocellular carcinoma. Mol Ther Nucleic Acids 2021; 25: 105-115.
[41] Sanchez-Vega F, Mina M, Armenia J, Chatila WK, Luna A, La KC, Dimitriadoy S, Liu DL, Kantheti HS, Saghafinia S, Chakravarty D, Daian F, Gao Q, Bailey MH, Liang WW, Foltz SM, Shmulevich I, Ding L, Heins Z, Ochoa A, Gross B, Gao J, Zhang H, Kundra R, Kandoth C, Bahceci I, Dervishi L, Dogrusoz U, Zhou W, Shen H, Laird PW, Way GP, Greene CS, Liang H, Xiao Y, Wang C, Iavarone A, Berger AH, Bivona TG, Lazar AJ, Hammer GD, Giordano T, Kwong LN, McArthur G, Huang C, Tward AD, Frederick MJ, McCormick F, Meyerson M; Cancer Genome Atlas Research Network, Van Allen EM, Cherniack AD, Ciriello G, Sander C and Schultz N. Oncogenic signaling pathways in the cancer genome atlas. Cell 2018; 173: 321-337, e10.
[42] Cuzick J, Swanson GP, Fisher G, Brothman AR, Berney DM, Reid JE, Mesher D, Speights VO, Stankiewicz E, Foster CS, Møller H, Scardino P, Warren JD, Park J, Younus A, Flake DD 2nd, Wagner S, Gutin A, Lanchbury JS and Stone S; Transatlantic Prostate Group. Prognostic value of an RNA expression signature derived from cell cycle proliferation genes in patients with prostate cancer: a retrospective study. Lancet Oncol 2011; 12: 245-255.
[43] Gao TM, Bai DS, Qian JJ, Zhang C, Jin SJ and Jiang GQ. The growth rate of hepatocellular carcinoma is different with different TNM stages at diagnosis. Hepatobiliary Pancreat Dis Int 2021; 20: 330-336.
[44] Ingle PV, Samsudin SZ, Chan PQ, Ng MK, Heng LX, Yap SC, Chai AS and Wong AS. Development and novel therapeutics in hepatocellular carcinoma: a review. Ther Clin Risk Manag 2016; 12: 445-455.
[45] Chow AK, Yau SW and Ng L. Novel molecular targets in hepatocellular carcinoma. World J Clin Oncol 2020; 11: 589-605.
[46] Zhang Y, Wang D, Peng M, Tang L, Ouyang J, Xiong F, Guo C, Tang Y, Zhou Y, Liao Q, Wu X, Wang H, Yu J, Li Y, Li X, Li G, Zeng Z, Tan Y and Xiong W. Single-cell RNA sequencing in cancer research. J Exp Clin Cancer Res 2021; 40: 81.
[47] Li L, Xiong F, Wang Y, Zhang S, Gong Z, Li X, He Y, Shi L, Wang F, Liao Q, Xiang B, Zhou M, Li X, Li Y, Li G, Zeng Z, Xiong W and Guo C. What are the applications of single-cell RNA sequencing in cancer research: a systematic review. J Exp Clin Cancer Res 2021; 40: 163.
[48] Zhang Q, Lou Y, Yang J, Wang J, Feng J, Zhao Y, Wang L, Huang X, Fu Q, Ye M, Zhang X, Chen Y, Ma C, Ge H, Wang J, Wu J, Wei T, Chen Q, Wu J, Yu C, Xiao Y, Feng X, Guo G, Liang T and Bai X. Integrated multiomic analysis reveals comprehensive tumour heterogeneity and novel immunophenotypic classification in hepatocellular carcinomas. Gut 2019; 68: 2019-2031.
[49] Hou Y, Guo H, Cao C, Li X, Hu B, Zhu P, Wu X, Wen L, Tang F, Huang Y and Peng J. Single-cell triple omics sequencing reveals genetic, epigenetic, and transcriptomic heterogeneity in hepatocellular carcinomas. Cell Res 2016; 26: 304-319.
[50] Liao C, Li M, Li X, Li N, Zhao X, Wang X, Song Y, Quan J, Cheng C, Liu J, Bode AM, Cao Y and Luo X. Trichothecin inhibits invasion and metastasis of colon carcinoma associating with SCD-1-mediated metabolite alteration. Biochim Biophys Acta Mol Cell Biol Lipids 2020; 1865: 158540.
[51] Maan M, Peters JM, Dutta M and Patterson AD. Lipid metabolism and lipophagy in cancer. Biochem Biophys Res Commun 2018; 504: 582-589.
[52] Park EJ, Lee JH, Yu GY, He G, Ali SR, Holzer RG, Osterreicher CH, Takahashi H and Karin M. Dietary and genetic obesity promote liver inflammation and tumorigenesis by enhancing IL-6 and TNF expression. Cell 2010; 140: 197-208.

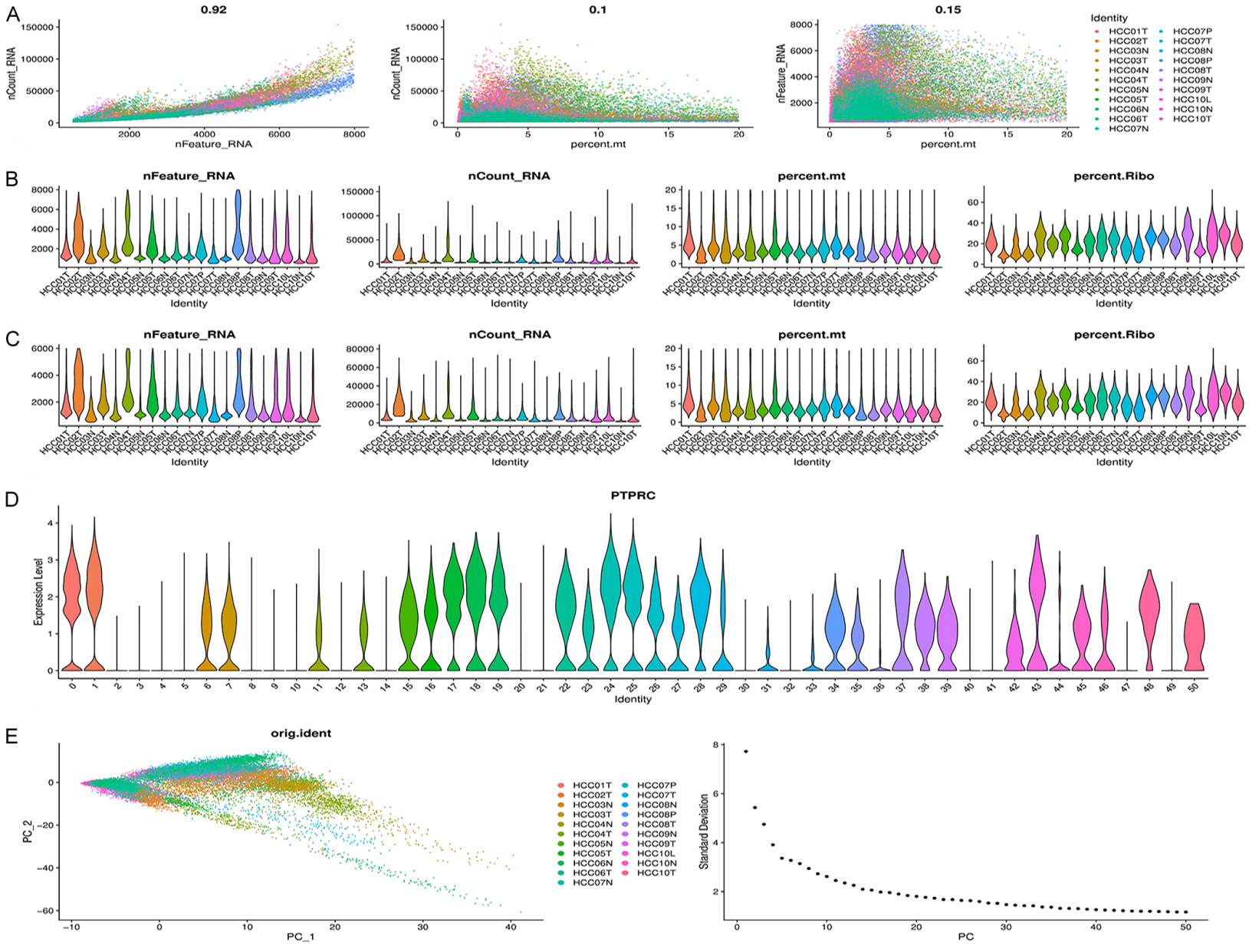
[53] Luo Q, Zheng N, Jiang L, Wang T, Zhang P, Liu Y, Zheng P, Wang W, Xie G, Chen L, Li D, Dong P, Yuan X and Shen L. Lipid accumulation in macrophages confers protumorigenic polarization and immunity in gastric cancer. Cancer Sci 2020; 111: 4000-4011.
[54] Tiwary S, Berzofsky JA and Terabe M. Altered lipid tumor environment and its potential effects on NKT cell function in tumor immunity. Front Immunol 2019; 10: 2187.
[55] Robison K. Application of second-generation sequencing to cancer genomics. Brief Bioinform 2010; 11: 524-534.
[56] Wu IC, Liu WC and Chang TT. Applications of next-generation sequencing analysis for the detection of hepatocellular carcinoma-associated hepatitis B virus mutations. J Biomed Sci 2018; 25: 51.
[57] Li S and Mao M. Next generation sequencing reveals genetic landscape of hepatocellular carcinomas. Cancer Lett 2013; 340: 247-253.
[58] Breuhahn K, Vreden S, Haddad R, Beckebaum S, Stippel D, Flemming P, Nussbaum T, Caselmann WH, Haab BB and Schirmacher P. Molecular profiling of human hepatocellular carcinoma defines mutually exclusive interferon regulation and insulin-like growth factor II overexpression. Cancer Res 2004; 64: 6058-6064.
[59] Chiang DY, Villanueva A, Hoshida Y, Peix J, Newell P, Minguez B, LeBlanc AC, Donovan DJ, Thung SN, Sole M, Tovar V, Alsinet C, Ramos AH, Barretina J, Roayaie S, Schwartz M, Waxman S, Bruix J, Mazzaferro V, Ligon AH, Najfeld V, Friedman SL, Sellers WR, Meyerson M and Llovet JM. Focal gains of VEGFA and molecular classification of hepatocellular carcinoma. Cancer Res 2008; 68: 6779-6788.
[60] Dong RZ, Yang X, Zhang XY, Gao PT, Ke AW, Sun HC, Zhou J, Fan J, Cai JB and Shi GM. Predicting overall survival of patients with hepatocellular carcinoma using a three-category method based on DNA methylation and machine learning. J Cell Mol Med 2019; 23: 3369-3374.
[61] Hoshida Y, Nijman SM, Kobayashi M, Chan JA, Brunet JP, Chiang DY, Villanueva A, Newell P, Ikeda K, Hashimoto M, Watanabe G, Gabriel S, Friedman SL, Kumada H, Llovet JM and Golub TR. Integrative transcriptome analysis reveals common molecular subclasses of human hepatocellular carcinoma. Cancer Res 2009; 69: 7385-7392.
[62] Villanueva A. Hepatocellular carcinoma. N Engl J Med 2019; 380: 1450-1462.
[63] Yang JD, Hainaut P, Gores GJ, Amadou A, Plymoth A and Roberts LR. A global view of hepatocellular carcinoma: trends, risk, prevention and management. Nat Rev Gastroenterol Hepatol 2019; 16: 589-604.
[64] Massoud O and Charlton M. Nonalcoholic fatty liver disease/nonalcoholic steatohepatitis and hepatocellular carcinoma. Clin Liver Dis 2018; 22: 201-211.



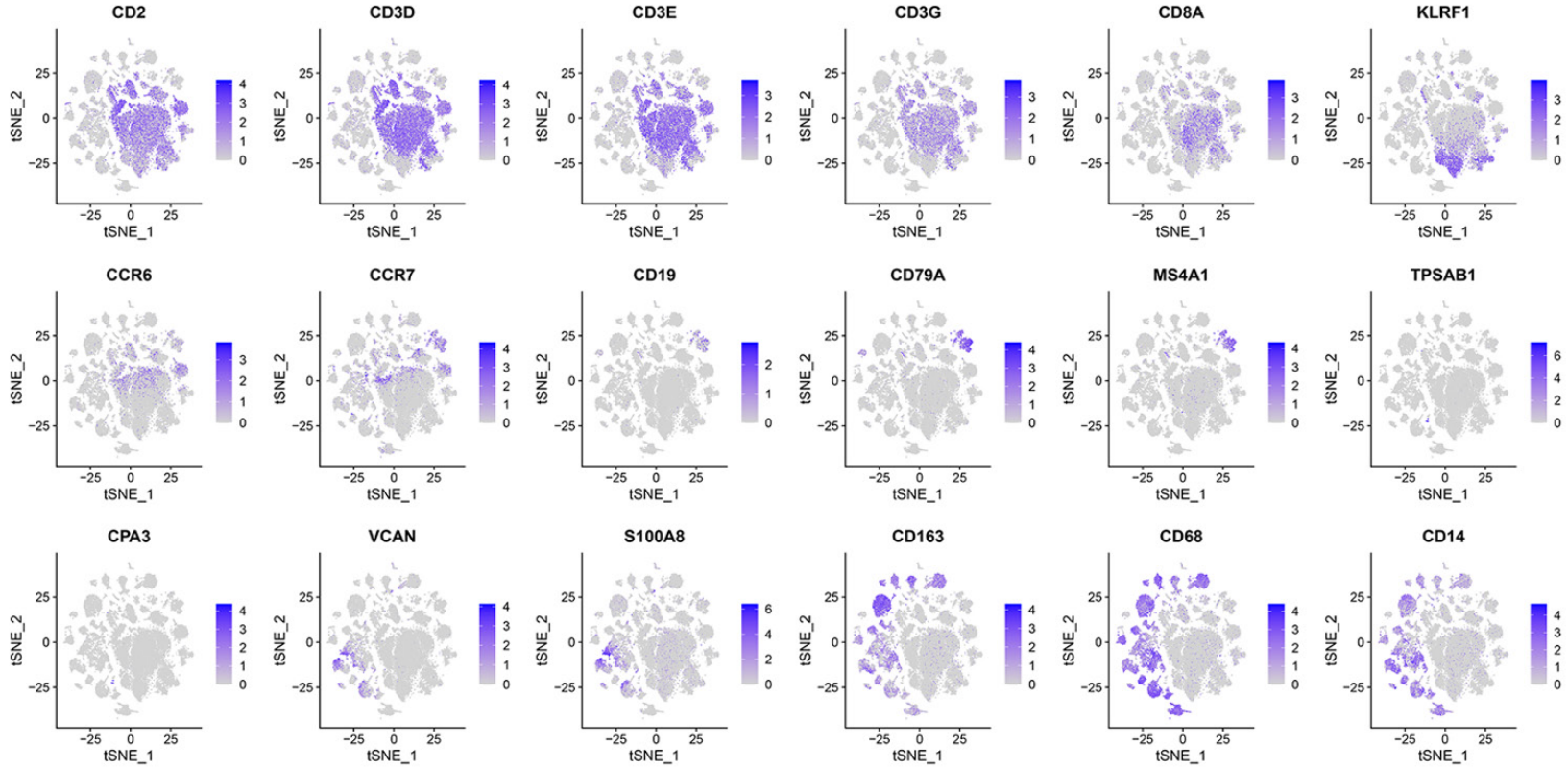
Supplementary Figure 1. PCA for two datasets based on six lipid metabolism pathways.



Supplementary Figure 2. The results of GO and KEGG enrichment analyses. A: BP annotation map for common genes; B: CC annotation map for common genes; C: MF annotation map for common genes; D: KEGG annotation map for common genes.


A
0.92
0.1
0.15
nCount_RNA
nFeature_RNA
percent.mt
Identity
HCC01T
HCC02T
HCC03N
HCC03T
HCC04N
HCC04T
HCC05N
HCC05T
HCC06N
HCC06T
HCC07N
HCC07P
HCC07T
HCC08N
HCC08P
HCC08T
HCC09N
HCC09T
HCC10L
HCC10N
HCC10T
B
nFeature_RNA
nCount_RNA
percent.mt
percent.Ribo
C
D
PTPRC
Expression Level
E
orig.ident
PC_1
PC_2
Standard Deviation
PC

Supplementary Figure 3. Preprocessing of sc-RNA-Seq data. A: The relationship between mitochondrial genes and UMI/mRNA quantity; the relationship between UMI and mRNA quantity; B: The relationship between mRNA/UMI/mitochondrial content/rRNA content in each sample before filtering; C: The relationship between mRNA/UMI/mitochondrial content/rRNA content in each sample after filtering; D: Violin plot for the gene expression of CD45; E: The sample distribution map for PCA dimensionality reduction and the anchor point map for PCA.



Supplementary Figure 4. TSNE plot showing the expression of marker genes.