Phosphoproteomics Uncovers Tumor Heterogeneity and Therapeutic Opportunities in Early-Stage Hepatocellular Carcinoma

Liangliang Ren[1,4], Jiahan Jiang[1,2,4], Weiran Lin[1], Lihui Qi[1], Chaoying Li[1, *] and Ying Jiang[1, 2, 3, *]

[1]State Key Laboratory of Medical Proteomics, National Center for Protein Sciences (Beijing), Beijing Proteome Research Center, Beijing Institute of Lifeomics, Beijing, China
[2]Hebei University, Hebei, China
[3]Anhui Medical University, Hefei, China
[4]These authors contributed equally to this work.
[*]Correspondence: Chaoying Li (lichaoying@ncpsb.org.cn), Ying Jiang (jiangying@ncpsb.org.cn)

February 11, 2025

Abstract

Hepatocellular carcinoma (HCC) is the third deadliest cancer in the world, however, the mechanisms of genesis, development and recurrence of HCC are not fully clarified yet. Previous studies on phosphoproteome in large-scale of HCC merely investigated the specific regulation on protein phosphorylation levels and identified subgroups consistent with proteomic subgroups. Here, we systematically interrogated the phosphoproteome of early-stage HCC and uncovered the dysregulated signaling pathways and kinases. Unlike previous studies on mRNA and protein levels, we observed a significant decline in phosphorylation regulation of EGFR-mediated canonical ERK/MAPK signaling alongside elevation of atypical MAPK signaling. Classification based on the heterogeneity of the phosphoproteome in tumor identified three phosphoproteomic subtypes, each characterized by distinct phosphorylation signatures, clinical characteristics, cellular signaling pathways and kinase activities. Further validation revealed hyperphosphorylation of SRSF3, one signature in the most malignant subtype, promote proliferation and metastasis of HCC cells. The mechanism of which was further demonstrated to be coordinated by activation of SRPK1/CLKs and potential suppression of PP1 phosphatase via phosphorylation of PPP1R7. Furthermore, inhibitors targeting SRPK1 or CLKs effectively suppressed hyperphosphorylation of SRSF3 mediated migration and invasion in HCC. Our study uncover the dysregulated phosphorylation signatures, offer new insights into the tumor heterogeneity, and identify phosphorylation regulating mechanism and corresponding key kinases with therapeutic value in early-stage HCC, which also spotlight the importance of protein phosphorylation in cancer progression.

Introduction

Hepatocellular carcinoma (HCC) is the third deadliest cancer worldwide, with a 5-year survival rate of 18% [1]. Even among patients undergoing surgical therapies, up to 70% suffer HCC recurrence within 5 years [1], whereas the mechanisms determining carcinogenesis and recurrence of HCC are still not clearly clarified and need further investigation. Genomic and transcriptomic studies of HCC have uncovered proliferation class and non-proliferation class of HCC [2, 3]. Recently three proteomic subtypes with distinct proteomic and prognosis features were uncovered with studies by Jiang *et al.* [4] and Gao *et al.* [5], respectively.

Post-translational modifications (PTMs) of proteins on specific amino acids, especially the phosphorylation that coordinated by kinases and phosphatases, dramatically expand the diversity of proteins and regulate signaling and protein turnover. Previous studies focused on the phosphorylation signatures that varied in distinct proteomic subtypes[4, 6], or signaling pathways at collapsed protein phosphorylation level and revealed subgroups consistent with the proteomic subgroups [5]. However, the phosphoproteome in these studies were not extensively explored as only the collapsed phosphorylation level of a protein was used and had lost the information of diverse phosphorylation regulation on different amino acids of proteins. While the mRNA and protein expression has shown a moderate to low positive correlation in large-scales of cancer tissues [7, 8], the correlation of protein expression and phosphorylation level has seldom been explored in cancer, particularly in HCC. Therefore, interrogation the signaling determined by protein phosphorylation in early-stage HCC may uncover novel signaling pathways that contribute to initiation and progression of HCC.

Targeting cancer dependency is a critical strategy for cancer therapy. Treatment of HCC with tyrosine kinase inhibitors such as sorafenib or lenvatinib is still the most effective single-drug therapies [9, 10], and showed an objective response rate (ORR) with 10-15% and 24.1% [10, 11], respectively. The relative low ORR raise the concern of tumor heterogeneity and activation of other oncogenic signaling pathways in distinct patients, therefore, stratifying the patients on protein phosphorylation level provides an opportunity to uncover the tumor heterogeneity on signaling pathways and to nominate kinases that contribute to them, which in turn should expedite the clinical classification and targeted therapies of HCC patients.

In this study, we performed integrated phosphoproteomic analysis of early stage HCC, and found a significant decline of phosphorylation regulation in EGFR-mediated canonical ERK/MAPK signaling as well as an elevation of atypical MAPK signaling, and the corresponding kinases involved in HCC. Investigation on the heterogeneity of phosphoproteome presented three phosphoproteomic subtypes that characterized by distinct phosphosite signatures, signaling pathways and kinase activities. Further interrogation on the worst prognosis subtype yields that phosphorylation on Ser148 and Ser150 of SRSF3 promotes the malignancy of HCC cells, which is coordinated by activated SRPK1/CLKs with potential suppression of PP1 phosphatase via phosphorylation of PPP1R7. Our investigation on the phosphoproteome of early-stage HCC uncover the dysregulated phosphorylation regulation in HCC, promoting understanding of the signaling pathways and related kinases in distinct subtype of HCC patients, offering opportunities for personalized therapeutics with improved medication selection.

Results

Phosphoproteomic characterization of early-stage HCC

In 103 paired tumor and adjacent non-tumor tissues of clinical early-stage hepatocellular carcinoma related to hepatitis B virus infection, phosphoproteomics identified 27,651 class I phosphosites (localization score > 0.75) and 9,254 proteins [4], quality control was performed by ROBPCA and 6 paired samples were identified as outliers and excluded from further analysis (Figure S1 A). Among the remaining 97 pairs of samples, 88 had corresponding proteomic expression data (Supplementary Table 0). Principal component analysis revealed a clear separation between tumor and non-tumor samples at the phosphoproteome level, while also highlighting the heterogeneity in tumor samples (Figure 1 A). We then examined the correlation between the abundance of 27,651 phosphosites and their aggregated protein phosphorylation levels with the corresponding protein abundance in the 88 paired tumor and non-tumor samples, respectively. In non-tumor tissue samples, 7,593 of the 27,651 phosphosites and their corresponding protein pairs were analyzed (Figure 1 B). The overall correlation between the phosphoproteomic and proteomic data was relatively weak, with a median Spearman's r value of 0.076, where 66.5% of the phosphosite abundances were positively correlated with their respective protein expression levels across the 88 samples, but only 1.6% showed significant positive correlation (Benjamin-Hochberg *adj.p*-value < 0.05). In the tumor tissues, 8,912 of the 27,651 phosphosites and their corresponding protein pairs were analyzed, presenting a median Spearman's r value of 0.169 of the overall correlations, where 80.7% of the phosphosite abundances were positively correlated with their protein expression levels, and 9.7% show significant positive correlation (Benjamin-Hochberg *adj.p*-value < 0.05). Even when considering aggregated protein phosphorylation levels from these phosphosites, only 6.7% and 24.5% of the correlations showed significant positive correlation in non-tumor and tumor tissues, respectively (Figure S1 B). Gene set enrichment analysis (GSEA) revealed that proteins with high correlations between protein expression and phosphorylation were enriched in hallmarks of glycolysis and epithelial mesenchymal transition (EMT), while those with low correlations were enriched in cellular signaling transduction such as signaling by GPCR, indicating that these signaling could be modulated extensively via protein phosphorylation in HCC (Figure 1 C, Supplementary Table 1).

The dysregulated phosphorylation signatures and signaling pathways involved in carcinogenesis and progression of early-stage HCC

Differential expressed phosphorylation sites (DES) analysis on the phosphoproteome revealed 1,950 upregulated and 1,660 downregulated phosphosites in early-stage HCC (Supplementary Table 2), and differential expressed proteins (DEP) analysis on the proteome showed 2,949 upregulated proteins and 679 downregulated proteins (Supplementary Table 5). Enrichment analysis of biological processes for the proteins with DES revealed that mRNA processing, chromatin remodeling, RNA catabolic process etc. are regulated with elevated phosphorylation in HCC, while metabolic processes were regulated with decreased phosphorylation (Figure S1 C, Supplementary Table 3). In terms of signaling pathways and hallmarks, mRNA splicing, spliceosome, Rho GTPase cycle, RNA polymerase II-mediated transcription, and transport of mature transcript to cytoplasm etc. were regulated with elevated phosphorylation in HCC (Figure S1D, Supplementary Table 4). Correspondingly, insulin

signaling pathway and metabolic pathways such as metabolism of glucose, carbohydrates, amino acid and derivatives showed decreased phosphorylation regulation in HCC (Figure S1 D, Supplementary Table 4), indicating the reduced physiological functions related to liver

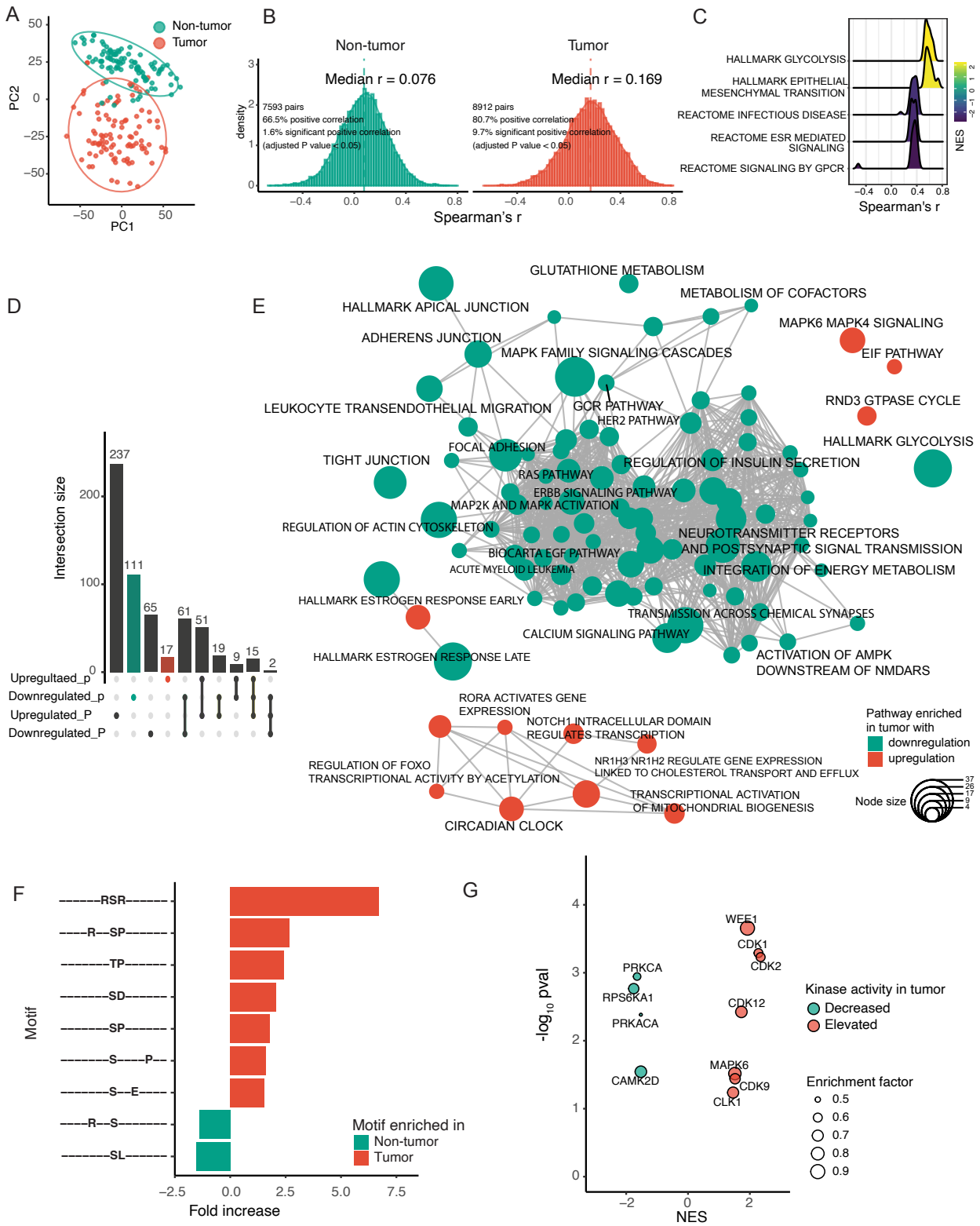


Figure 1: (Caption next page.)

Figure 1: (previous page.) The dysregulated protein phosphorylation in early-stage HBV-related HCC cohort (Jiang et al.'s cohort). (A) Principal component analysis of the phosphoproteome in the tumor and non-tumor tissues. (B) Correlation of levels between phosphorylation at phosphosite and protein expression in the non-tumor and tumor tissues (Benjamin-Hochberg adjusted p-value < 0.05). (C) GSEA enrichment analysis of pathways and hallmarks for proteins with the high or low correlations of protein expression and phosphorylation level in tumor tissues. (D) UpSet plot showing the number of pathways and hallmarks enriched from regulated proteins and phosphoproteins. Pathways and hallmarks that were upregulated or downregulated in the phosphoproteome are denoted as "Upregulated_p" or "Downregulated_p", respectively, while those in the proteome are denoted as "Upregulated_P" or "Downregulated_P". Pathways and hallmarks observed in one or multiple conditions are connected by black dots. The bar plot of pathways and hallmarks only observed with downregulation on phosphoproteins is highlighted with dark green, while that only observed with upregulation is highlighted with dark red. (E) Visualization of pathways and hallmarks significantly enriched and only observed with dysregulated protein phosphorylation in the tumor tissues. The similarity of the pathway enrichment is calculated by the Jaccard correlation coefficient method, and the node size indicates the number of overlapping proteins between gene sets. (F) Bar plot showing motif fold changes for the upregulated and downregulated phosphosites (compared to all observed phosphosites). The upregulated phosphosites were enriched with basophilic motifs (XRSRX, RXXSP), proline-directed motifs (XXTP and XXSP), and acidophilic motifs (SXXE, XXSD). Those downregulated were enriched with motifs of RXXS and XXSL. (G) Kinase-substrate enrichment analysis (KSEA) showed decreased and elevated kinase activities in tumor, with corresponding normalized enrichment score (NES). The size of the kinase indicates the enrichment factor.

metabolism.

By comparing the signaling pathways and hallmarks that are significantly regulated at the protein expression level and the phosphorylation level in early-stage HCC (Figure 1 D, see also Supplementary Table 8, 6 and 7), 51 hallmarks and pathways were observed with both elevated protein expression and phosphorylation, including the pathways involved in chromatin modifying enzymes, mRNA processing, cell cycle, E2F and MYC targets etc. (Supplementary table 7) . In addition, 61 hallmarks and pathways were observed with both decreased protein expression and phosphorylation (Figure 1 D, see also Supplementary table 7), including pathways involves lipid metabolism such as fatty acid metabolism, and other metabolism pathways like synthesis of bile acids and bile salts, metabolism of carbohydrate, xenobiotics, amino acid and derivatives, which has recently been found playing a protumor role in HCC [12], highlighting the loss of metabolism functions of the liver in HCC. Alongside the unaltered signaling on protein expression, 17 and 111 pathways or hallmarks were observed to be solely regulated with elevation or reduction at the phosphorylation level (Figure 1 D). To identify functional modules of pathways that were only observed on phosphoproteome, we created an enrichment map based on the UpSet plot of pathway enrichment result (Figure 1 E, Supplementary Data 1). One cluster and several pathways were identified in the pathways with elevation only on phosphorylation regulation, these include the circadian clock, transcriptional activation of mitochondrial biogenesis, and gene expression regulated by RORA, FOXO, NR1H3/NR1H2, EIF pathway, RND3 (RHOE) GTPase cycle, MAPK6 and MAPK4 signaling (Figure 1 E, see also Supplementary Table 8) etc. Whereas, a network with two clusters involving in metabolic and signaling pathways was observed with only decrease on phosphorylation regulation, one of which consisted of metabolism of cofactors integration of energy metabolism, AMPK downstream signaling, neurotransmitter receptors, transmission across chemical synapses, calcium signaling pathway, and regulation of insulin secretion etc. Another cluster consisted of modules on cell adhesion, migration and cancer-related signaling pathways, including hallmark apical junction, adherens junction, focal adhesion, regulation of actin cytoskeleton, as well as leukocyte transendothelial migration, signaling of MAP2K, MAPK family cascades, ERBB, HER2, EGF, and hallmark of estrogen response (Figure 1 E, Supplementary Table 8) etc. Overall, the dysregulated hallmarks and pathways indicate the post-translational regulation by phosphorylation in HCC.

To further explore the kinases that may determine the dysregulated signaling in HCC, by analyzing with the Motif-X algorithm on the amino acid sequences surrounding the 1,950

elevated phosphorylation sites and the 1,660 decreased phosphorylation sites [13], several basophilic kinase motifs, such as XRSRX (SRPKs [14]), RXXSP (SRPKs, CLKs [14]), and proline-directed motifs XXTP and XXSP (MAPK family kinases [15] and CDK family kinases [15]), as well as acidophilic kinase motifs, such as SXXE and XXSD (CK1 [15], CK2 [15, 16]) were significantly enriched in the elevated phosphorylation sites in early-stage HCC (Figure 1 F, Supplementary Table 9), suggesting that the activities of SRPKs and CLKs, members of MAPKs, CDKs, CK1 and CK2 family kinases were likely elevated in HCC. In contrast, the phosphosites with decreased levels in HCC were enriched in motifs of RXXS (CaM kinase II [17], PKA [17], PKC family kinases [18]) and XXSL, suggesting that the activities of members of PKC family kinases and PKA kinases were likely decreased in early-stage HCC (Figure 1 F, Supplementary Table 9).

We further examined the changes in kinase activities in early-stage HCC, applying another approach by utilizing the database with curated kinase-substrate relationships in Omnipath [19], and performed kinase-substrate enrichment analysis (KSEA) to interpret the regulation of kinase activities in HCC. We focused on the kinases with significant phosphorylation regulation in early-stage HCC, and the KSEA analysis further revealed 7 kinases with significantly elevated activities in HCC (Figure 1 G, Supplementary Table 10), including several members of cyclin-dependent kinases (CDK1, CDK2, CDK7, CDK9, and CDK12) and cell cycle regulator WEE1, mitogen-activated protein kinase 6 (MAPK6), and dual specificity protein kinase CLK1. In contrast, 4 kinases showed decreased activities in early-stage HCC, including protein kinase C alpha (PRKCA), cAMP-dependent protein kinase catalytic subunit alpha (PRKACA), calcium/calmodulin-dependent protein kinase type II subunit delta (CAMK2D) and ribosomal protein S6 kinase alpha-1 (RPS6KA1) (Figure 1 G). As indicated by the result of motif and KSEA analysis, further verification of the phosphorylation levels on activation sites of these kinases confirmed that the activity elevation of CDK1, CDK2, CDK7, CDK9, CK1D, CK1E, SRPK1, and CLK1-4, as well as the activity decrease of PRKACA, PRKCA, and CAMK2D in Jiang et al.'s cohort [4] and Gao et al.'s cohort [5] (Figure S2). In summary, by integrating the motif enrichment and KSEA analysis, we confirmed the elevated or decreased activities of these kinases, which likely play a major role in driving the dysregulated protein phosphorylation in HCC.

The EGFR-mediated canonical ERK/MAPK and atypical MAPK signaling cascades in early stage HCC

The enrichment map of pathways that dysregulated only on protein phosphorylation showed that several pathways including the EGF pathway and MAPK family cascades were observed with solely decrease on phosphorylation levels (Figure 1 E). To our surprise, even though the majority of EGFR-mediated canonical ERK/MAPK signaling molecules, especially the MAPK1 and MAPK3, showed significantly elevated protein expression in early-stage HCC, phosphorylation regulation on the activation sites of the proteins in EGFR-mediated canonical ERK/MAPK signaling was significantly decreased in HCC (Figure 2 A). In contrast, the phosphorylation regulation on the activation sites of RPS6KA3, together with MAPK6 (an atypical MAPK), was significantly elevated in HCC in both Jiang et al.'s cohort and Gao et al.'s cohort (Figure 2 A, Figure S3 A and B).

Phosphorylation of the sites Thr693, Ser1039, Ser1064, Ser1166 and Tyr1197 on EGFR underwent increased regulation upon EGF stimulation [20]. The canonical ERK/MAPK signaling acts as one downstream effector of several tyrosine kinases including EGFR, and

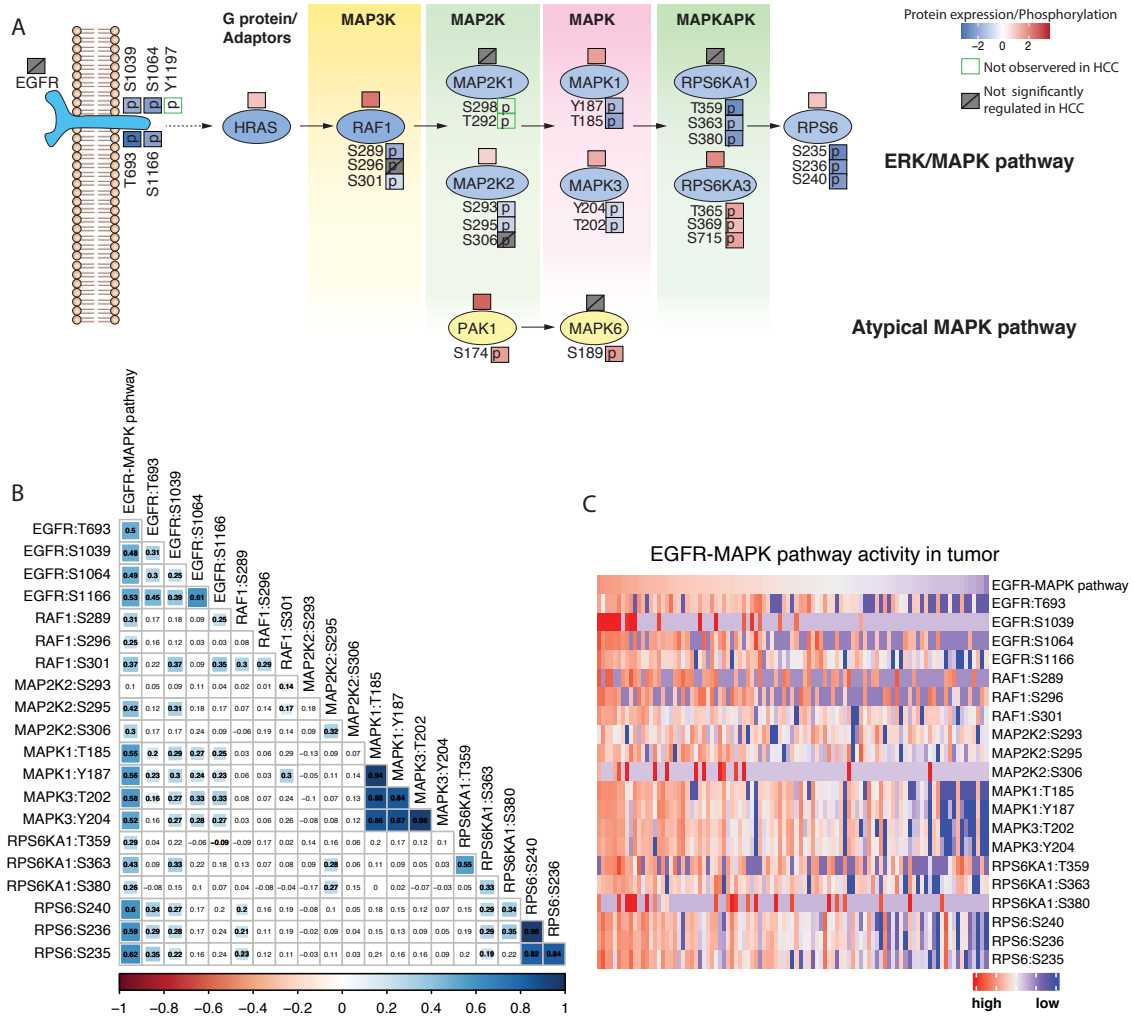


Figure 2: Expression and phosphorylation level of the EGFR-mediated canonical ERK/MAPK and atypical MAPK signaling cascades in HCC in Jiang et al.'s cohort. (A) The diagram of expression and functional phosphosite of proteins in EGFR-mediated ERK/MAPK pathway and atypical MAPK pathway in HCC. The upper box on the protein diagram represents the expression level of the protein, while the lower box indicates the phosphorylation status of the specific site, marked with a 'p' inside the box. Phosphosites not detected were indicated by a blank green box, while proteins and phosphosites with no significant regulation in HCC were marked with a grey box containing a forward slash. (B) The correlation of the phosphorylation level of activation sites of proteins in EGFR-mediated canonical ERK/MAPK signaling cascades in HCC. The phosphorylation levels of activation sites in proteins within the EGFR-mediated canonical ERK/MAPK signaling cascade in HCC were analyzed for correlations. Correlation coefficients are shown in squares, with significant ones highlighted in color. The median value of their activation sites was taken as a proxy of EGFR-mediated canonical ERK/MAPK pathway and (C) patients were ranked from high to low by that and the heatmap of their phosphorylation levels was shown (see also Figure S3).

RPS6KA1 (p90RSK) lies at the terminus of the canonical ERK/MAPK pathway [21, 22]. Once activated, RPS6KA1 dissociates from MAPK1/MAPK3 and translocate into the nucleus [21, 23], thereby phosphorylating the downstream substrate RPS6 [24]. Additionally, phosphorylation of Ser235, Ser236 and Ser240 on RPS6 also underwent increased regulation upon EGF stimulation [20, 24]. Therefore, using the median of z-score-scaled phosphorylation levels of these phosphosites as a proxy of EGFR-mediated canonical ERK/MAPK pathway activity, which showed significantly positively correlated with the phosphorylation levels of the activation sites on majority of molecules in this signaling (Figure 2 B, Figure S3 C). Furthermore, ranking patients from high to low based on the EGFR-mediated canonical ERK/MAPK pathway activity showed a consistent regulation on this signaling cascades in HCC (Figure 2 C, Figure S3 D).

Phosphoproteomic subtypes resolve the heterogeneity of early-stage HCC

The PCA analysis showed the complex heterogeneity of the phosphoproteome in HCC (Figure 1 A), and the dysregulation in signaling pathways mediated by phosphorylation also revealed distinct regulation on phosphoproteome (Figure 1 E, Figure 2), which prompted us to explore the molecular subtypes in the phosphoproteome of HCC. NMF factorization and consensus clustering of the phosphoproteome presented three subtypes from Jiang et al.'s cohort: pS1 subtype (n=35), pS2 subtype (n=28) and pS3 subtype (n=34) (Figure S4 A-C, Figure 3 A). Alignment of the proteomic subtypes [4] and clinical characteristics showed no significant difference on serum AFP status (>= 200 ng ml^{-1}) and patient age among the three phosphoproteomic subtypes (Figure 3 C and D). Interestingly, all female patients were distributed in pS1 and pS2 subtypes, whereas pS3 subtype contained only male patients. The pS1 subtype comprised the most portion of patients in the proteomic SI subtype, low degree of microvascular invasion (MVI^+) status and small tumor size (Figure 3 C and D). The pS2 subtype was the smallest subtype and characterized by larger tumor size than the pS1 subtype (Figure 3 D). Patients in pS3 subtype had a higher degree of MVI^+ status (P-value = 0.04) and the worst prognosis in overall survival (Figure 3 B and C).

We identified 1,816 phosphosites that significantly regulated among the phosphoproteomic subtypes and were categorized into four clusters (Figure 3 A, Supplementary Table 11): pC1-pC4, with which a clear separation of the three phosphoproteomic subtypes was presented (Figure S4 D). Assignment of the phosphoproteomic subtypes onto the proteome identified 566 proteins that differently expressed among the phosphoproteomic subtypes, which were also divided into four clusters: C1-C4 (Figure S6 A). The pS1 subtype is characterized with higher abundance of phosphosites on proteins involved in metabolism of carbohydrates (ChREBP pathway), fatty acid, organic acid, vitamins, and branched chain amino acids catabolism (Figure S5 B and C, Supplementary Table 12) in the tumor tissues. Meanwhile, the higher abundance of metabolism related pathways was also observed on proteome in pS1 (Figure S6 B and C, Supplementary Table 14). In addition, signaling transduction with higher phosphorylation regulation in pS1 subtype included mTOR, BRAF and RAF1 fusion, ERBB, IGF1R, activation of AMPK downstream of NMDARs, CDC42 GTPase cycle and CREB signaling etc. (Figure S5 D, Supplementary Table 12), among these included the cascades on EGFR-mediated canonical ERK/MAPK signaling, ie. Ser1064 and Ser1166 on EGFR, Thr202 and Tyr204 on MAPK3, and Thr185 and Tyr187 on MAPK1 (Figure S5 A).

The pS2 subtype presented significantly low abundance of phosphosites on proteins in-

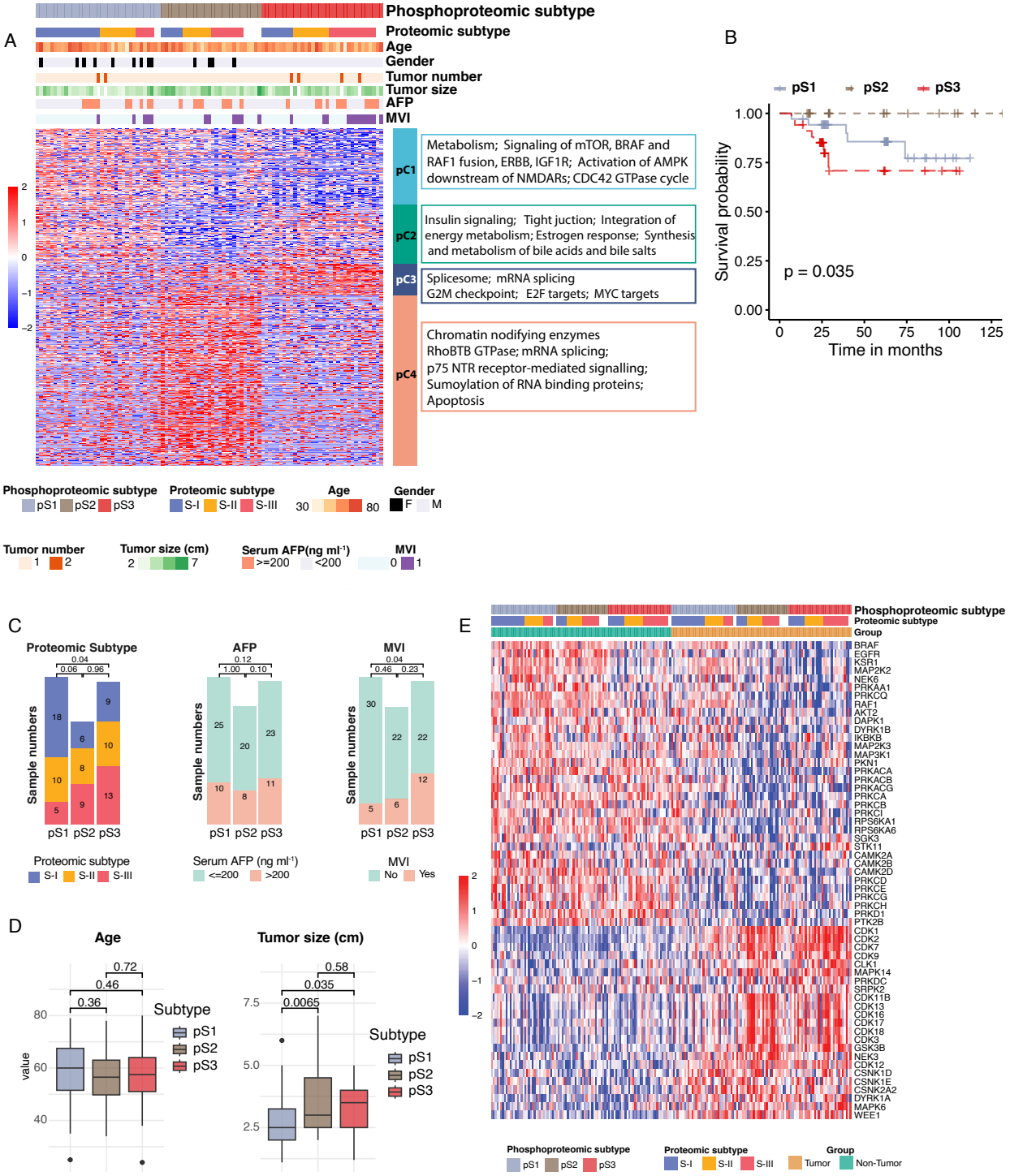


Figure 3: (Caption next page.)

volved in insulin signaling pathway, tight junction, PPARA and keratinocyte pathway, estrogen response, integration of energy metabolism, EIF4 pathway, synthesis and metabolism of bile acids and bile salts etc, a correlation of dysregulation of which and HBV-related HCC had been recently reported [5] (Figure 3 A , Figure S5 B and D). Besides, lower protein expression on metabolism of carbohydrates, amino acids and derivatives, fatty acids etc. was observed in pS2 subtype (Figure S6 B and C). On other side, pS2 subtype showed distinct

Figure 3: (previous page.) Phosphoproteomic subtypes identified in HCC tissues and their correlation with clinical characteristics and kinase activities. (A) Nonnegative matrix factorization and consensus-clustering analysis of the phosphoproteome in HCC tissues stratified patients into three phosphoproteomic subtypes. The proteomic subtypes and clinical characteristics from Jiang et al.'s cohort were annotated in the column annotation. A total of 1,816 phosphosites showed significantly differential levels among the three subtypes and were divided into four clusters. The relative abundance of phosphosites (log2-transformed) were shown in the heatmap. Specifically, pC1 cluster: phosphosites with the highest abundance in the pS1 subtype. pC2 cluster: phosphosites exhibiting the lowest abundance in the pS2 subtype. pC3 cluster: phosphosites showing the highest abundance in the pS3 subtype. pC4 cluster: phosphosites with the highest abundance in the pS2 subtype. (B) Kaplan–Meier curves for overall survival (OS), based on the phosphoproteomic subtypes (log-rank test). (C) Correlation of phosphoproteomic subtypes with proteomic subtypes and clinical features, including serum AFP levels, MVI, (D) patient age and tumor size. (E) Kinase inferred with significantly different activities in the phosphoproteomic subtypes in HCC, shown together with the corresponding non-tumor samples.

high abundance of phosphosites on proteins (pC4 cluster) participating in biological processes such as RNA splicing, chromatin remodeling, histone modification, DNA repair and nuclear transport (Figure S5 C, Supplementary Table 13). Additionally, signaling pathways related to RhoBTB GTPase, mRNA splicing, p75 NTR receptor-mediated signalling, apoptosis, chromatin-modifying enzymes, sumoylation of RNA binding proteins were presented (Figure S5 B and D), suggesting regulation on transcription and epigenetic changes may have occurred in pS2 subtype. As expected, the proteins with higher abundance in pS2 subtype enriched on epigenetic regulation of gene expression, basal transcription factors, rRNA processing and modification, transcription regulation by polymerase I, II, III and hallmark of MYC targets etc. (Figure S6, Supplementary Table 14 and 15). The pS3 subtype is characterized with lower abundance of phosphosites on proteins that participate in metabolism, especially liver function-related pathways. Additionally, pS3 subtype showed high abundance of phosphosites on proteins involved in mRNA processing and spliceosome, base excision repair and cellular component disassembly, mRNA splicing, G2M checkpoint, and hallmarks of E2F targets and MYC targets (Figure 3 A, Figure S5 B and D).

Moreover, the pS2 and pS3 subtypes presented regulation on distinct phosphosites of proteins in same the pathways. For instance, phosphorylation of Ser40 on MCM2, a marker of DNA replication during cell cycle [25] was hyperphosphorylated in the pS2 subtype (Figure S5 A) . Whereas Ser27 on MCM2, phosphorylated by CDKs, showed higher phosphorylation level in both pS2 and pS3 subtypes (Figure S5 A), suggesting proliferative characteristics mediated by distinct signaling in pS2 and pS3 subtypes. Interestingly, proteins with high abundance in pS3 subtype were involved in extracellular matrix organization (ECM), collagen formation, hallmarks of epithelial mesenchymal transition (EMT) and angiogenesis (Figure S6 B and C). Notably, the pS3 subtype showed high MVI status and poor prognosis. EMT progression is associated with angiogenesis and HCC cells acquire a more invasive and aggressive phenotype when they undergo this characteristic [26], implying a role of ECM and EMT in progression of pS3 subtype.

Blocking aberrant kinase activities is one of the strategies for anticancer therapeutics, therefore, we inferred the kinase activities by single-sample kinase-substrate relation set enrichment analysis (ssKSEA). We found activities of 57 kinases were significantly regulated in the phosphoproteomic subtypes (Figure 3 E, Supplementary Table 16). Notably, EGFR, canonical ERK/MAPK signaling included MAP2K2 and RAF1, showed low activities in pS3 subtype, consistent with our previous analysis. In addition, members of the protein kinase A family and the protein kinase C family showed declined activities in pS3, with even greater declines in pS2. Meanwhile, CDKs, MAPK14, GSK3B showed not only elevated activities in tumors but even greater increases in pS2 and pS3 subtypes. We noticed that CLK1 and SRPK2, two kinases contributing to phosphorylation of splicing factors, presented higher

activities the pS3 subtype. Taken together, the integrated analysis on phosphoproteome and proteome uncovered distinct phosphoproteomic subtypes in early-stage HCC, characterized by unique molecular signatures, clinical features, signaling pathways and kinase activities among the HCC patients.

SRSF3 contributes to proliferation, migration and invasion of HCC cells

We then focused on pS3 subtype, that was characterized with the worst prognosis and hyperphosphorylation regulation on proteins involved in spliceosome, mRNA processing (Figure 3 B). Among the phosphosites in pC3 that showed high level in pS3 subtype, these involved in phosphosites on spliceosome component such as SRRM2, PRPF3 and splicing factor SRSF3. Overexpression of SRSF3 has been shown to contribute the tumor progression of glioblastoma (GBM), head and neck squamous cell carcinoma (HNSCC) [27], and promotes gemcitabine resistance in pancreatic cancer [28]. SRSF3 is crucial for functional differentiation of hepatocytes [29], however, decreased levels of SRSF3 had been reported in HCC, and its loss in hepatocytes promoted development of metastatic HCC in mice [30]. In contrast, mRNA level of SRSF3 was reported to be upregulated in HCC [31]. We found SRSF3 was significantly upregulated in tumor tissues from HCC patients in Jiang et al.'s cohort and Gao et al.'s cohort (Figure 4 A), and this was further confirmed by immunohistochemistry (IHC) staining of non-tumor and tumor tissues from HCC patients (Figure 4 B and C). Therefore, we hypothesized that elevation of SRSF3 may play a crucial role in carcinogenesis and development of HCC. Stable SRSF3 knockdown using short hairpin RNA (shRNA) in MHCC-97H and Huh-7 cell lines significantly inhibited cell proliferation (Figure 4 D and E), cell monoclonal formation (Figure 4 F), and reduced the ability of migration and invasion of HCC cells (Figure 4 G). Conversely, stable overexpression of SRSF3 in PLC/PRF/5 and MHCC-97H cell lines significantly promoted cell proliferation (Figure S7 A and B), cell monoclonal formation (Figure S7 C), and increased capabilities of cell migration and invasion (Figure S7 D) in both the two cell lines. These above results indicated that SRSF3 contributes to proliferation, migration and invasion of HCC cells.

Phosphorylation on Ser148 and Ser150 of SRSF3 promotes migration, invasion and proliferation of HCC cells

The phosphorylation of serine/arginine-rich splicing factor 3 (SRSF3) on Ser148, Ser150 was significantly upregulated in the pS3 subtype, with no corresponding elevation in SRSF3 protein expression (Figure 5 A). These residues localize in the domain rich in arginines and serines (RS domain) at the C-terminus of SRSF3, which is intrinsically unstructured, and reversible phosphorylation on the RS domain of SRSF family was found to regulate their activities in spliceosome maturation, splicing catalysis and mRNA export [32, 33]. Therefore, we further interrogate the function of phosphorylation of SRSF3 in HCC in details.

We hypothesized that elevated phosphorylation levels at Ser148 and Ser150 on SRSF3 in the pS3 subtype of HCC may be an important factor affecting the prognosis of HCC patients. Therefore, we further constructed stable cell lines expressing SRSF3 with mutations simulating dephosphorylated form (S148A, S150A), phosphorylated form (S148D, S150D), and wide-type form (Figure S8 A). The construction of the cell lines was verified by

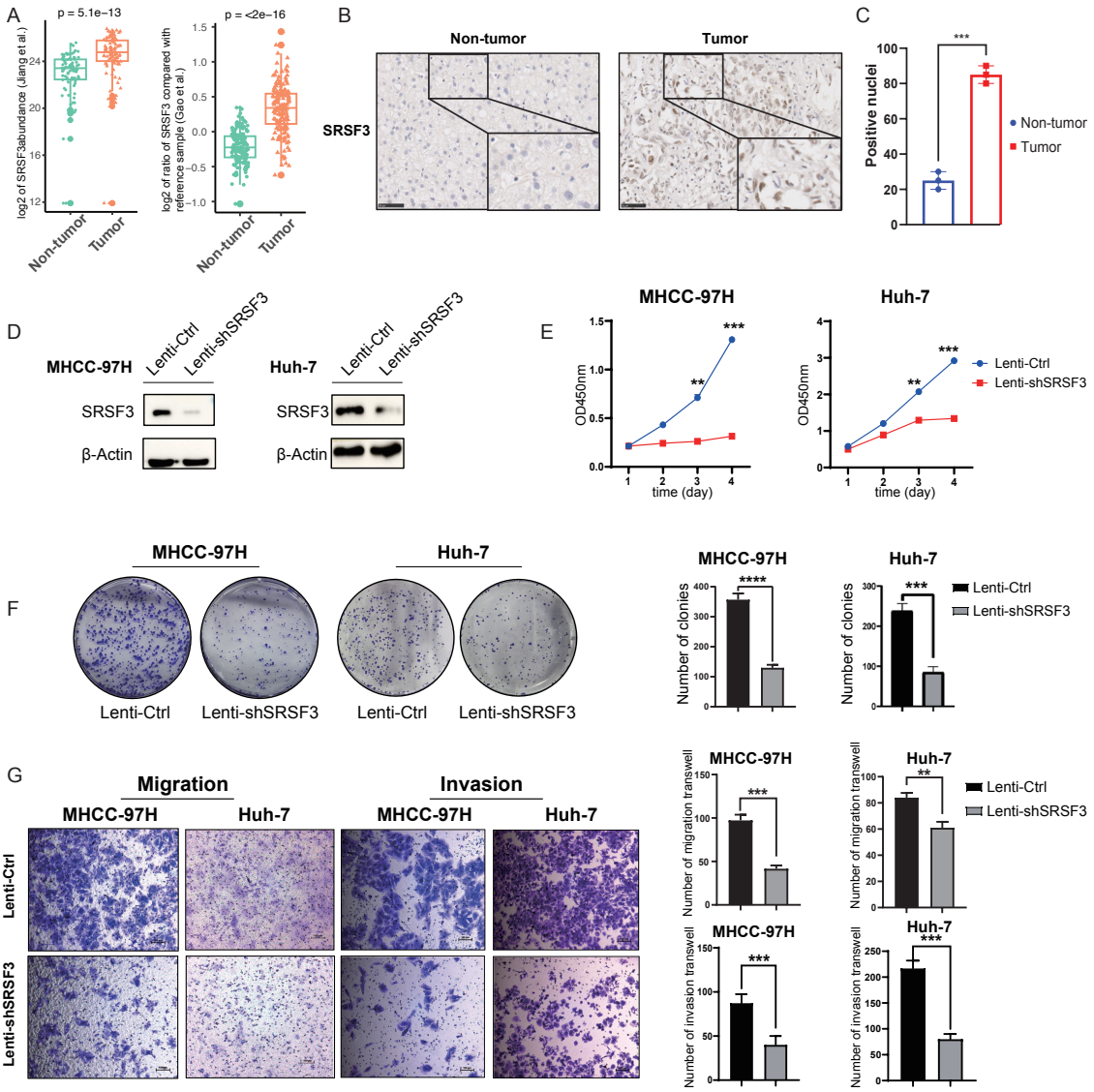


Figure 4: SRSF3 was upregulated in HCC and stable SRSF3 knockdown in HCC cell lines significantly decreased the malignancy of HCC cells. (A) SRSF3 was significantly upregulated in tumor tissues of HCC patients in Jiang et al.'s cohort and Gao et al.'s cohort. (B) Digital scans of representative IHC staining of SRSF3 and (C) statistical analysis of the positive nuclei in the non-tumor and tumor tissues of HCC patients. (D) Stable SRSF3 knockdown in MHCC-97H and Huh-7 HCC cell lines significantly reduced the capabilities of (E) cell proliferation, (F) cell clone formation, (G)migration and invasion. For all graphs, asterisks show statistical significance by two-sided Wilcoxon rank-sum test, $*p < 0.05$; $**p < 0.01$; $***p < 0.001$.

western blot analysis (Figure S8 B). As expected, overexpression of phosphorylated mutant (S148D/S150D) and the SRSF3-WT in PLC/PRF/5 and MHCC-97H cell lines showed significantly increased migration and invasion capabilities than the dephosphorylated mutant (S148A/S150A) in HCC cells (Figure 5 B), increased N-cadherin and vimentin expression and decreased E-cadherin expression, suggesting promotion of epithelial-mesenchymal transition (EMT) in HCC cells, while the dephosphorylated mutant (S148A/S150A) had few effect on promotion of EMT (Figure 5 C). In addition, the phosphorylated mutant (S148D/S150D) and the SRSF3-WT form both showed significantly increased abilities of cell monoclonal formation and cell proliferation than the dephosphorylated mutant (S148A/S150A) (Figure S8 C and D). Meanwhile, overexpression of SRSF3-WT resulted in a significant increase in tumor weight compared with the control group, whereas the SRSF3-S148D/S150D mutant

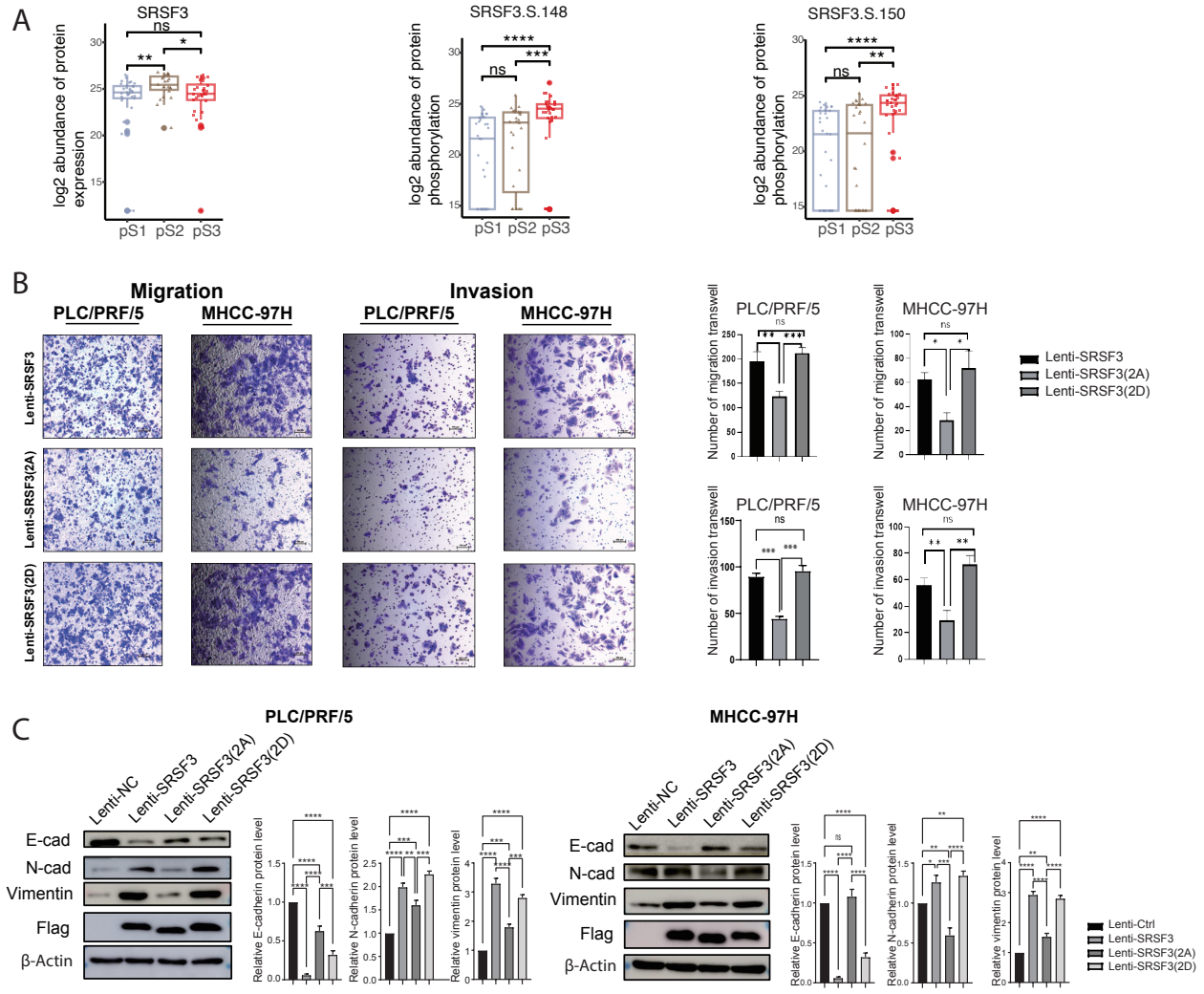


Figure 5: Phosphorylation on Ser148 and Ser150 of SRSF3 promotes migration and invasion in HCC cells. (A) The phosphorylation levels on Ser148 and Ser150 of SRSF3 were elevated in the phosphoproteomic pS3 subtype, while no elevation of the protein expression of SRSF3 observed. (B) Overexpression of the phosphorylated mutant (S148D/S150D) and the wild-type SRSF3 (SRSF3-WT) significantly increased migration and invasion capabilities compared to the dephosphorylated mutant (S148A/S150A) in HCC cells. (C) Overexpression of phosphorylated mutant (S148D/S150D) and the SRSF3-WT promoted epithelial-mesenchymal transition (EMT) of HCC cells, while the dephosphorylated mutant (S148A/S150A) showed limited effect on promotion of EMT than the phosphorylated mutant.

resulted in a significant increase in tumor weight compared to the SRSF3-S148A/S150A mutant in PLC/PRF/5 cell line xenograft model (Figure S8 E and F). There was not significant difference on the ability of migration, invasion, colony formation between the SRSF3-S148D/S150D mutant and the SRSF3-WT (Figure 5 B and Figure S8 C). Interestingly, the SRSF3 S148D/S150D mutant showed a significant increase in xenograft growth than the SRSF3-WT form (Figure S8 G). Ki-67, a prominent proliferation marker used in pathology [34], not only showed significantly elevated expression in the SRSF3-S148D/S150D mutant xenograft than in the SRSF3-S148A/S150A mutant xenograft, but also showed higher expression compared to the SRSF-WT xenograft (Figure S8 H and I), suggesting a higher proliferative ability of HCC cells mediated by phosphorylation regulation of Ser148 and Ser150 on SRSF3. Taken together, these results indicate that the higher phosphorylation levels of Ser148 and Ser150 SRSF3 in HCC cells promote the growth, proliferation, and metastatic invasion .

Activated SRPK1 and CLKs mediate phosphorylation of Ser148 and Ser150 on SRSF3 and inhibitors targeting SRPK1 or CLKs significantly suppress hyperphosphorylation of SRSF3-mediated migration and invasion in HCC

SRPK1, SRPK2, and members of the CLK family act as upstream kinases of SRSF family [35, 36], cooperating tightly to regulate the function of SRSF proteins [37]. The protein expression of SRPK1 and SRPK2 were significantly increased in HCC, while CLK1-4 were not detected at the protein level (Supplementary Table 5). Phosphorylation of SRPK1 on Ser309 and Ser311 was significantly elevated in HCC (Figure S2 A and B), with Ser309 serving as a self-activation site for the kinase [38], indicating that SRPK1 was activated in HCC. Although SRPK2 kinase protein was significantly upregulated in HCC, the phosphorylation levels of its kinase activation sites, Ser494 and Ser497 [39], the former was significantly downregulated in HCC whereas the latter showed inconsistent trends between Jiang et al's cohort and Gao et al's cohort (Figure S2 A and B). Despite that CLK1-4 kinases (CLKs) were not detected at the protein level, phosphorylation levels of activation site on CLKs [40] were all significantly increased in HCC (Figure S2). Given the elevated kinase activities of SRPK1 and CLKs in HCC (Figure S2 A and B), we proposed that they may be the upstream kinases that phosphorylate SRSF3 in HCC. We then treated the PLC/PRF/5 and MHCC-97H cells with SRPK1 inhibitor SRPIN340 [41] and CLKs inhibitor TG003 [42], in each cell line respectively. Both SRPIN340 and TG003 effectively reduced the phosphorylation levels of Ser148 and Ser150 on SRSF3 in PLC/PRF/5 and MHCC-97H cell models (Figure 6 A and B), confirming that the phosphorylation of Ser148 and Ser150 on SRSF3 is regulated by SRPK1 and CLKs in HCC. We then treated the PLC/PRF/5 and MHCC-97H cells that stably overexpress SRSF3-WT with individual inhibitors, and the migration and invasion ability of SRSF3-WT cells significantly declined to a similar level with the dephosphorylated mutant SRSF3 (S148A/S150A) (Figure 6 C). Moreover, under treatment with SRPIN340 or TG003, the SRSF3-WT showed significantly decreased abilities in migration and invasion in comparison with the phosphorylated mutant SRSF3 (S148D/S150D) (Figure S9 A). Taken together, these findings indicate that SRPK1 and CLKs phosphorylate SRSF3 on Ser148 and Ser150, thereby promoting the migration and invasion of HCC cells. Consequently, targeting these kinases could offer a promising strategy for the development of targeted therapies for HCC patients.

Hyperphosphorylation of SRSF3 in pS3 subtype is potentially mediated through suppression of PP1 phosphatase activity via phosphorylation on PPP1R7

Given that phosphorylation on Ser148 and Ser150 on SRSF3 is significantly elevated in pS3 subtype (Figure 5 A), as well as SRPK1 and CLKs contribute to phosphorylation of Ser148 and Ser150 on SRSF3, we speculate that phosphorylation of activation sites of these kinases would have been significantly elevated in pS3 subtype of early-stage HCC. To our surprise, neither the protein expression nor the phosphorylation of active sites of these kinases was observed with elevation in pS3 subtype (Figure S9 B and C), suggesting that dephosphorylation regulation of SRSF3 by phosphatases may be inhibited in pS3 subtype. Protein phosphatase 1 (PP1) mediates the dephosphorylation in the RS domain of SRSF1 [43], PPP1R7 (Sds22) is

a conserved PP1 regulator and a PP1 trap [44], phosphorylation of the N-terminal of PPP1R7 inhibits PP1 activity [45]. We found that the abundance of phosphorylation levels at Ser12, Ser24 and Ser27 of PPP1R7 showed high abundance in pS3 subtype (Figure S9 C), therefore, it is possible that the activity of PP1 phosphatase is suppressed in the pS3 subtype, which led to elevation of SRSF3 phosphorylation in pS3 subtype. As expected, treatment of HCC cells with Tautomycetin, the PP1 phosphatase inhibitor, significantly increased the phosphorylation levels of Ser148 and Ser150 on SRSF3 (Figure S9 E), confirming that PP1 regulates the dephosphorylation of SRSF3 in HCC. Taken together, we have demonstrated that the activation of SRPK1 and CLK1-4, potentially coordinated with suppression of PP1 phosphatase activity which is mediated by phosphorylation of PPP1R7, determined the elevation of phosphorylation of SRSF3 in pS3 subtype, thereby contributing to the worst prognosis of HCC patients in the pS3 subtype.

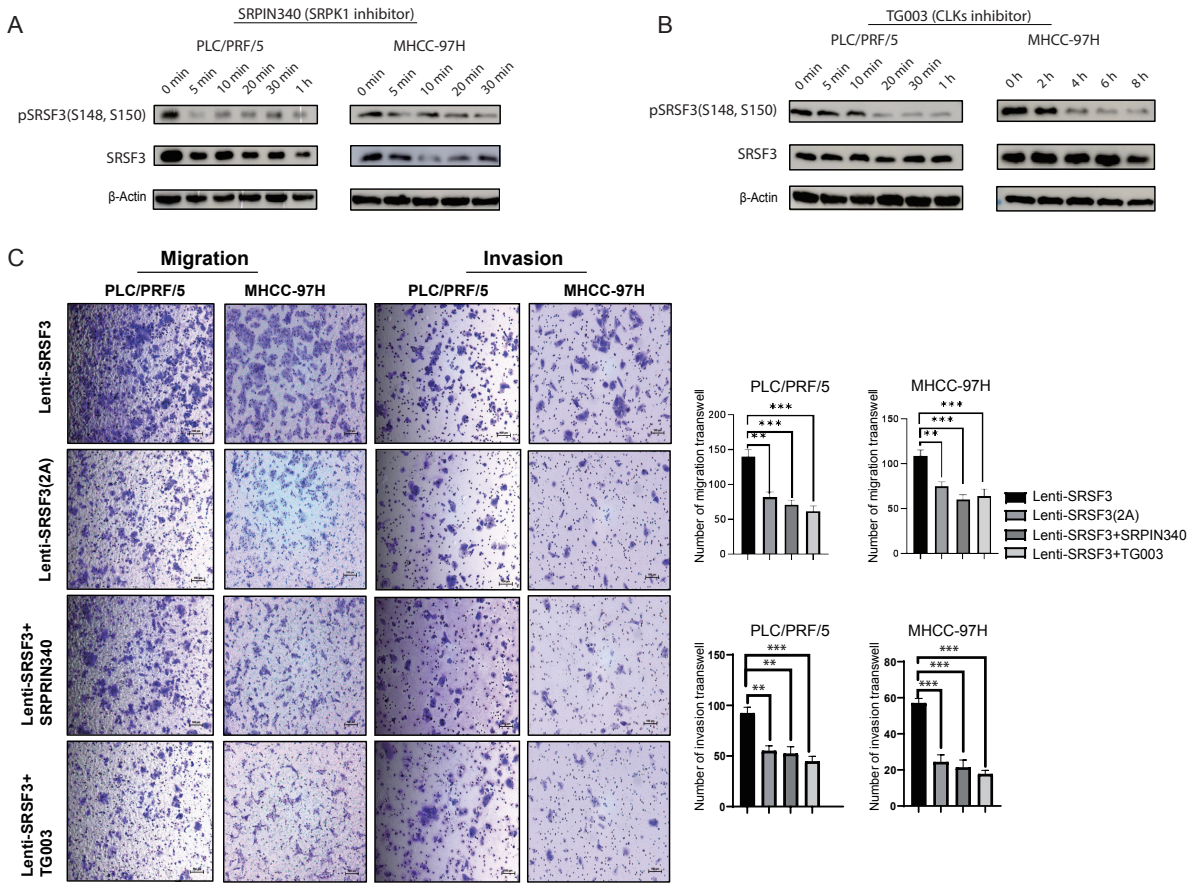


Figure 6: SRPK1/CLKs phosphorylate Ser148 and Ser150 on SRSF3 and inhibitors targeting SRPK1 or CLKs significantly suppress migration and invasion of HCC cells. (A) Both SRPIN340 and TG003 were applied at a concentration of 50 μM to treat the cells (see also Figure S9D), addition of the SRPK1 kinase inhibitor SRPIN340 or (B) the CLKs kinase inhibitor TG003 to PLC/PRF/5 and MHCC-97H cells significantly reduced the abundance of phosphorylation at Ser148 and Ser150 on SRSF3. (C) Treatment with SRPIN340 or TG003 to SRSF3-WT overexpressing PLC/PRF/5 and MHCC-97H cells significantly reduced the migration and invasion capabilities of HCC cells. The dephosphorylated mutant SRSF3(S148A/S150A), without treatment with SRPIN340 or TG003, also showed decreased migration and invasion capabilities compared to the SRSF-WT. For all graphs, asterisks show statistical significance by two-sided Wilcoxon rank-sum test, $^{*}P < 0.05$; $^{**}P < 0.01$; $^{***}P < 0.001$.

Discussion

HCC is one of the most challenging cancers characterized by high recurrence and mortality, and the nearly identical incidence and mortality underscore the dismal prognosis of this disease [1, 46, 47], which calls urgent need to systematically understand its molecular landscape. PTMs dramatically increase the diversity of protein functions; however, previous studies have seldom investigated the specific regulation of protein phosphorylation at a large-scale in HCC, and the information on dynamic regulation of phosphosites was lost when considering them with collapsed phosphoproteins [5, 6]. In this study, protein phosphorylation and expression levels showed weak positive correlation in early-stage HCC, suggesting the role of post-translational regulation by phosphorylation in cellular signaling (Figure 1 B and C). The phosphoproteome provided us a different perspective to investigate the regulation on signaling pathways and kinase activities in early-stage HCC. Integration of phosphoproteome with the proteome in early-stage HCC revealed unique regulation observed only at the level of protein phosphorylation, such as the transcription regulated by FOXO, NOTCH1, RORA (Figure 1 E), which may involve in carcinogenesis and development of HCC.

Persistent activation of Ras and the downstream ERK in canonical ERK/MAPK signaling cascades were previously reported in HCC [48]. Conversely, in this study, phosphorylation levels (rather than the protein expression) of EGFR-mediated canonical ERK/MAPK signaling cascades, together with PRKCA, PRKACA, CAMK2D are significantly downregulated in HCC. Meanwhile, the activities of MAPK6, RPS6KA3 and members of CDKs, CLKs and SRPK1 etc. are significantly upregulated in HCC (Figure 2 B and S3 A), indicating the signaling mediated by other kinases may contribute to the progress of HCC. The canonical ERK/MAPK signaling cascades act as downstream for several tyrosine kinases, including the drug targets of sorafenib and lenvatinib [3]. Sorafenib and lenvatinib achieved an ORR with 10-15% and 24.1% respectively, in targeted therapies of HCC [10, 11]; however, most patients don't respond to sorafenib or lenvatinib and finally succumb to HCC. Since the canonical ERK/MAPK signaling is downstream of their targets, their significant downregulation in HCC may explain the low ORR of sorafenib and lenvatinib in HCC patients. On the other hand, patients with high EGFR-mediated canonical ERK/MAPK activity may benefit from therapies targeting EGFR in combination with sorafenib or lenvatinib. Indeed, activation of EGFR and its downstream signaling contributes to resistance of HCC cells against sorafenib [49] and limits the response of HCC cells to lenvatinib [50]. . EGFR-mediated reactivation of canonical ERK/MAPK signaling contributes to resistance against targeted therapies in colorectal cancer and head and neck squamous cell carcinoma (HNSCC)[51, 52]. Therefore, targeting EGFR in combination with targeting canonical ERK/MAPK cascades seems promising for treating cancer patients characterized with high EGFR-mediated canonical ERK/MAPK signaling activity. Meanwhile, MAPK6, an atypical MAPK, is significantly elevated in HCC. This highlights the role of atypical MAPK signaling in HCC and need for further exploration.

Integration of heterogeneity of the phosphoproteome uncovered three phosphoproteomic subtypes of HCC, and each was characterized by distinct phosphosite signatures, kinase activities and signaling pathways. For instance, several CDKs, WEE1, MAPK6, CK1 and CK2 etc. showed significantly higher activities in all three phosphoproteomic subtypes compared to the non-tumor tissues, which can be targeted and could expand the selection of treatment for HCC patients. Regarding the signaling pathways, the pS1 subtype presented abundant phosphorylation regulation in liver-related pathways, including metabolism-related pathways,

signaling of mTOR, canonical ERK/MAPKs, and AMPK (Figure 3 B) etc. The pS2 subtype showed abundant phosphorylation regulation involved in transcription and epigenetic changes. Meanwhile, the pS3 subtype was characterized by less abundant phosphorylation regulation on liver function related pathways, but exhibited abundant phosphorylation regulation in spliceosome, MYC targets etc. Several signaling pathways, such as ChREBP pathway, showed distinct patterns in pS2 and pS3 subtypes. Pathways including spliceosome, G2M checkpoint etc. presented active phosphorylation regulation both in pS2 and pS3 subtypes. Notably, the pS3 subtype also is associated with abundant proteins involved EMT, a process further confirmed to be mediated by hyperphosphorylation of Ser148 and Ser150 on SRSF3.

The pS3 subtype is characterized with poorest prognosis and hyperphosphorylation of Ser148 and Ser150 on SRSF3. SRSF3 shuttles between the nucleus and cytoplasm and requires phosphorylation regulation on the RS domain [53]. We then performed immunofluorescence experiments to examine whether the different phosphorylation mutants of SRSF3 influenced its subcellular localization. The results showed that all of SRSF3-WT, SRSF3-S148A/S150A and SRSF3-S148D/S150D localize in the nuclear speckles and nucleoplasm of PLC/PRF/5 cells (Figure S9 F). We have demonstrated that hyperphosphorylation of Ser148 and Ser150 on SRSF3 promotes EMT, suggesting that it may be mediated other functions of SRSF3 such as mRNA export, mRNA stability, or regulation on translation [54], which deserves further investigation. In summary, we have spotlighted the importance of phosphorylation regulation in early-stage HCC, uncovered the dysregulated phosphorylation signatures, kinases, and pathways in early-stage HCC and across the phosphoproteomic subtypes. Moreover, we further demonstrated that the phosphorylation of SRSF3 mediates mechanism for development and metastasis of HCC, which should facilitate the development of proteomics-driven precision medicine for HCC patients.

Materials and methods

Normalization of phosphoproteomic data and quality assessment

The phosphoproteomic dataset with 27,651 phosphosites (with site localization scores > 0.75), and the proteomic dataset with 9,254 proteins from 103 tumor and non-tumor paired samples in Jiang et al.'s cohort [4] were processed in R environment (version 4.2.1). The intensities of phosphosites were log_2 transformed and normalized by *normalizeQuantiles* function from the R package limma. To detect outliers in the dataset, we applied robust principal component analysis (ROBPCA) to yield a robust estimate on the data [55]. As a result, 6 paired samples were examined as outliers and excluded from further analysis (Figure S1 A). The phosphoproteomic dataset with 26,418 phosphosites, and proteomic dataset with 6,478 proteins from 159 tumor and non-tumor paired samples in Gao et al.'s cohort [5] were used to verify the findings in this study.

Principal component analysis

We performed PCA on the 97 tumor and non-tumor paired samples to elucidate the global phosphoproteomic differences between tumor and non-tumor samples. The analysis was restricted to 13,303 phosphosites with quantification coverage of at least 25% in either tumor or

non-tumor samples individually. The samples were projected onto the first two components, and the 95% confidence intervals for each group were visualized using colored eclipse.

Correlations of protein expression and phosphorylation level

The 88 tumor and non-tumor paired samples were analyzed, as they contained both the proteomic and phosphoproteomic data. The correlations between protein expressions and phosphorylation levels were calculated separately for tumor and non-tumor tissues. Only proteins or phosphosites quantified in at least 25% of the samples were included for the analysis. The protein IDs were mapped into the gene names. The Spearman's correlation coefficients and P-values were calculated between phosphosite levels and protein expression levels. The P-values were further adjusted by using the Benjamin-Hochberg procedure, and pairs with an $adj.p$-value < 0.05 was determined as significant. For aggregated phosphoprotein analysis, the intensities of multiple phosphosites were combined for each phosphoprotein, and correlations were calculated as described above.

To investigate pathways and hallmarks associated with high or low correlations between protein expression and phosphorylation levels, gene sets from the R package msigdbr (version 7.5.1) were used. These included 186 Kyoto Encyclopaedia of Genes and Genomes (KEGG) gene sets, 292 Biocarta gene sets, 1615 Reactome gene sets, 196 PID gene sets, 10 NABA gene sets and 50 cancer hallmark gene sets. Proteins were ranked based on their Spearman's correlation coefficients (from high to low), and the *GSEA* function in R package clusterProfiler (version 4.8.3) was applied for the enrichment analysis. The minimum gene set size (minGSSize) was set to 10, and $p-$values were adjusted by Benjamin-Hochberg procedure, and a $adj.p-$value cut-off of 0.05 was used as the significance threshold.

Differential expressed phosphosites (DES) and proteins (DEP) between tumor and non-tumor

The protein expression/phosphorylation values were log_2 transformed and normalized using the *normalizeQuantiles* function in the R package limma. After filtering with a coverage cutoff $\geq$ 50% in at least one group of tumor or non-tumor tissues, missing values were imputed by the global minimum in the entire dataset. To identify the significantly regulated phosphosites/proteins between early stage HCC and paired non-tumor samples, a linear model was fitted using *limma*, accounting for within-patient correlations estimated by the *duplicateCorrelation* function. *Moderated t-statistics* were implemented, and p-values were adjusted by Benjamini-Hochberg procedure to control the false discovery rate. Following differential analysis, a total of 3,610 DESs and 3,628 DEPs were observed with an $adj.p$-value < 0.05 and a $|log_2$ (fold change)$| > 1$.

Motif enrichment and kinase-substrate enrichment analysis (KSEA)

The significantly elevated and decreased DESs were used for motif enrichment and kinase activity prediction. For motif enrichment, the Motif-X algorithm [13] was used with a p-value cutoff of 1e-5, minimum sequence length (min.seq) =20 for the amino acid 'S', 'T' and 'Y', and the sequences of all phosphosites observed in the dataset were used as the background sequences, respectively. For KSEA, the enzsub database was downloaded from Omnipath[19]. The DESs together with the phosphosites not significantly regulated (n =

7,575) were used and ranked by `-log10(p-value) * sign(coef(x))` and *GSEA* function with `minGSSize=5, nPerm=10,000` was applied to infer the kinase activities. A q-value less than 0.25 was considered as the cut-off. Only kinases detected with regulated phosphosites in HCC were considered as kinases with elevated or decreased activities in HCC.

Biological process and pathway alteration on protein expression and phosphorylation level

The *enrichGO* function in the R package clusterProfiler was used for Gene Ontology (GO) enrichment analysis. For pathway and hallmark enrichment analysis, the KEGG, Biocarta, Reactome, PID, NABA gene sets and cancer hallmark gene sets in migsdb database in R package msigdbr combined with function *enricher* in R package clusterProfiler were applied. For GO, pathway and hallmark enrichment results, p-values were adjusted by Benjamin-Hochberg procedure, with a cut-off of 0.05. The similarity in the pathway/hallmarks enrichment was calculated by the Jaccard correlation coefficient (JC) method, and the enrichment map was visualized by function *emapplot* in clusterProfiler.

UpSet plots of pathways and hallmarks

To visualize the overlap between pathways and hallmarks, UpSet plots were created using the R package UpSetR.

Identification of phosphoproteomic subtypes

Nonnegative matrix factorization (NMF) is widely used for dimension reduction and has demonstrated its ability to extract biological meaningful information from high-dimensional data[4, 56]. To uncover the subtypes in early-stage HCC, we first filtered out phosphosites quantified in less than 20% of the tumor samples. The median absolute deviation (MAD) was calculated for each phosphosite across the samples. Subsequently, the bottom 20% of phosphosites with low MAD were removed, leaving 11,083 phosphosites for NMF factorization. The *nsNMF* algorithm in R package NMF (v.0.25) was used for subtyping the phosphoproteomic data. We evaluated rank from 2 to 7, using 100 iterations for rank survey and 500 iterations for unsupervised consensus clustering. Based on cophenetic scores and consensus membership heatmaps, we selected rank=3 as the optimal subtype solution for the phosphoproteomic data (Figure S4).

Correlations between phosphoproteomic subtypes and clinical characteristics

The correlations were evaluated by Fisher's exact test for categorical data and Wilcoxon rank-sum test for continuous data, respectively. Log-rank test and Kaplan–Meier survival curves were applied to compare the overall survival among the phosphoproteomic subtypes, and a significance level of 0.05 was used as the cut-off threshold.

Signature phosphosites for phosphoproteomic subtypes and pathways/hallmarks, biological processes, kinase activities across the subtypes

To identify the signature phosphosites of phosphoproteomic subtypes, we compared the abundance of phosphosites in each phosphoproteomic subtype with other subtypes and adjacent non-tumor groups. The coverage of phosphosite with a quantification cut-off of 50% in at least one phosphoproteomic subtype or the adjacent non-tumor groups was required, and the missing values were imputed with global minimum value. A linear model was fitted in *limma* with considering the within-patient correlation estimated by *duplicateCorrelation* function, and *moderated t-statistics* was implemented and *p*-values were adjusted by Benjamin-Hochberg procedure. To identify signature phosphosites/proteins of the phosphoproteomic subtypes, the following criteria were required: all *adj.p*-values < 0.05 and at least two fold change (`lfc=1`) when comparing to other subtypes and adjacent non-tumor samples. 1,816 phosphosites and 566 proteins were identified as the signature phosphosites/proteins, which were further divided into four clusters (pC1-pC4 for phosphosites and C1-C4 for proteins): the pC1/C1 cluster comprises those with the highest abundance in the pS1 subtype, the pC2/C2 cluster exhibits the lowest abundance in the pS2 subtype, the pC3/C3 cluster shows the highest abundance in the pS3 subtype, and the pC4/C4 cluster shows the highest abundance in the pS2 subtype.

For the enrichment of biological processes, pathway/hallmark of the clusters, the protein IDs (of phosphosites) in each cluster were used for analysis. The *enrichGO* function in R package clusterProfiler was used for biological processes enrichment analysis. The KEGG, Biocarta, Reactome, PID, NABA gene sets and cancer hallmark gene sets in migsdb database in R package msigdbr combined with function *enricher* in R package clusterProfiler were used for the pathway and hallmark enrichment analysis. For biological processes, pathway and hallmark enrichment results, the *p*-values were adjusted by Benjamin-Hochberg procedure and with a cut-off of 0.05. The similarity in the pathways/hallmarks enrichment was calculated by the Jaccard correlation coefficient (JC) method and the enrichment map was visualized by function *emapplot* in clusterProfiler. And for dotplot of the enrichment results, the enrichment factor was calculated by the ratio of two proportions, ie. `Enrichment=log2(ProteinRatio(k/n)/BgRatio(M/N))`, in which `ProteinRatio(k/n)` represents ratio of input proteins that are annotated in a term, and `BgRatio (M/N)` represents ratio of all proteins that are annotated in this term.

For the kinase activities in the subtypes, the enzsub database from Omnipath[19] was used as a resource for kinase-substrate relation sets to infer kinase activities. We applied single-sample kinase-substrate relation set enrichment analysis (ssKSEA) to infer the kinase activities in each sample, including the tumor and non-tumor tissues. This analysis utilized the *gsva* algorithm with the *ssgsea* method from R package GSVA [57]. The kinase scores reported by the *gsva* algorithm were used as a proxy for kinase activities, and only those observed in the phosphoproteomic data were retained for inferring kinase activities in the phosphoproteomic subtypes. By applying the same procedure with signature phosphosites in the subtypes, as described above, and including the kinases that elevated or decreased in the tumor compared to non-tumor tissues across all three phosphoproteomic subtypes, we identified 57 kinases whose activities were significantly regulated.

Cell culture

The cell lines PLC/PRF/5, MHCC-97H, and Huh7 were cultured in DMEM/MEM medium containing 10% fetal bovine serum (FBS). They were maintained at 37°C with 5% CO2 in a cell culture incubator.

Lentivirus infected cells

When the cell confluence reached 60%-70%, the medium was changed to serum-free DMEM or MEM, then lentivirus and Polybrene were added to each well. The cells were then incubated. After 24 hours of infection, the medium was replaced with complete medium. For Huh-7 and MHCC-97H cells knockdown, puromycin (working concentration 2 μg/mL, Shanghai Taosen Biochemical Technology) was used for screening over 72-96 hours. For MHCC-97H and PLC/PRF/5 cells overexpressing SRSF3 and phosphomutants, Geneticin (G418) (working concentration 800 μg/mL, Beyotime Biotechnology) was used for screening for at least one week.

Design and synthesis of SRSF3 protein phosphomutants and interference sequences

The nucleotide sequence of SRSF3 gene was obtained from NCBI databases. Lentivirus vectors LV-EFS-Kozak-Flag-SRSF3 (Ser148, Ser150), LV-EFS-Kozak-Flag-SRSF3 (S148A,S150A), and LV-EFS-Kozak-Flag-SRSF3 (S148D,S150D), as well as control group LV-EFS-Kozak-Flag-NC were constructed by Sigma (Suzhou) Biotechnology Co., Ltd. The S148A (TCC to GCC), S150A (TCC to GCC) mutations and the S148D (TCC to GAC), S150D (TCC to GAC) mutations were introduced. The knockdown vectors, LV2-SRSF3-shRNA-PURO and control group LV2-PURO-NC, were constructed by Suzhou Gima Company. The interfering sequence of SRSF3 was 5'-CAGGTCCCTTTCTAGAGATAG-3'.

Extraction of total cellular protein and western blot

An appropriate volume of 4% SDS lysis buffer was added to the cell pellet followed by incubation on ice for 20 minutes. Ultrasonic cell disruption was performed, and after centrifugation at 12,000 rpm, 4 °C for 20 minutes, the supernatant was collected for protein concentration determination. For western blot, samples were denatured at 95 °C for 5 min, then resolved on a 12% SDS-PAGE gel, and transferred to a nitrocellulose membrane. To block the membrane, 5% skim milk was used, followed by incubation with the following antibodies: anti-beta-actin (Beijing Zhongshan Jinqiao Biotechnology Co., Ltd.), anti-SRSF3 (Invitrogen, USA), anti-Ki67 (Abcam Company), anti-SRSF3 (S148, S150) phosphorylated antibody (prepared by Wuhan Dion Biotechnology Co., Ltd.), Flag antibody (Proteintech Company), E-cadherin antibody (CST Company USA), anti-N-cadherin, anti-Vimentin (Proteintech Company), and goat anti-mouse/anti-rabbit secondary antibody (Beijing Kangwei Century Co., Ltd.).

Colony formation assay

Cells were seeded in 6-well plates at a density of 800-1,000 cells per well, and the cell culture plates were cultured in triplicate at 37 °C for 8 to 10 days. Cells were then fixed in 4%

paraformaldehyde for 15 min and stained with crystal violet for 15 min at room temperature. After washing with PBS, the colonies were counted to see if there was any significant differences.

Cell Proliferation

A CCK-8 assay was used to detect cell proliferation. The cells were plated in a 96-well plate at a density of 6×10^3 cells per well (100 μL total volume). After being cultured in an incubator for 5 hours to allow adhesion, the 96-well plate marked 0h was take out, the medium was discarded, and prepared 10% CCK-8 medium was added. The plate was then incubated for one hour in the cell culture box. A microplate reader (Biotek USA) was used to measure the absorbance value at OD 450 nm. Absorbance values were recorded at 24h, 48h, 72h, and 96h, and the data were saved.

Cell migration and invasion

Cells were starved cells overnight. Subsequently, 600 μL of complete medium was added to the lower compartment of a 24-well plate, and approximately 8×10^4 to 10×10^4 cells/100 μL of pre-calculated cell suspension were added to the upper compartment of each well. There were three replicates for each control and experimental group. The cells were incubated in a cell culture incubator for 24 or 48 hours. After removal from the chamber, the cells were fixed, stained, and photographed.

Immunohistochemistry IHC

The required tissue sections were retrieved from the oven and immersed in xylene twice for 20 minutes each. They were then transferred sequentially through 100%, 100%, 95%, 95%, 90%, and 80% ethanol for 5 minutes each, followed by washing with 1×TBS for 5 minutes. Antigen retrieval was performed by citrate acid treatment, followed by blocking with 5% BSA and incubation with the primary antibody overnight. After adding enhancement solution and secondary antibody, the sections were washed three times with TBS, stained with DAB, counterstained with hematoxylin, and finally dehydrated, mounted, and scanned. The sections were differentiated with hydrochloric alcohol and blued under running water.

Confocal Laser Scanning Microscopy (CLSM)

Cells were fixed with 4% paraformaldehyde for 15 minutes and washed with PBS. Blocking was performed using 10% normal goat serum for 40 minutes. Incubation with primary antibody diluted in 5% goat serum was done overnight at 4°C. The next day, the cells were incubated with secondary antibody, stained for nuclei, mounted, and imaged.

In vivo xenograft studies

The animal care and experiments were approved by the Institutional Animal Care and Use Committee (IACUC) of the National Center for Protein Sciences, Beijing, China (NCPSB). Male NOD/SCID mice (Beijing Vital River Laboratory Animal Technology Co., Ltd., Beijing, China), aged 5 weeks, were housed in a strictly specified pathogen-free (SPF) animal facility.

Stable NC, SRSF3-WT, SRSF3-S148A/S150A, SRSF3-S148D/S150D of PLC/PRF/5 cells (5×10^6) were injected subcutaneously with the right flank of each NOD/SCID mouse. Tumor dimensions were measured every 4 days using a caliper after tumor formation. The tumor volume was calculated using the formula: $0.5 \times$ length (L) $\times$ width $(W)^2$. The mice were euthanized at the end of the trials, and solid tumors were dissected and weighed.

Data Collection and Statistical Analysis

The data provided in this study were statistically analyzed in R environment or using GraphPad Prism 9 software and presented as mean $\pm$ standard deviation (mean $\pm$ SD). Two-sample comparisons were performed using two-sided Wilcoxon rank-sum test, with p-values <0.05 was considered as significant. $^*p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$; $^{****}p < 0.0001$.

Author contributions

L.L.R., C.Y.L. and Y.J. designed and conceived the studies; L.L.R. performed the computational analysis; J.H.J., W.R.L. and L.H.Q. performed the functional assays; C.Y.L. and Y.J. supervised the work; L.L.R., J.H.J., C.Y.L., and Y.J. wrote the manuscript; all authors proofread the manuscript. The authors declare no competing interests.

Acknowledgements

This work was supported by National Key R&D Program of China (Nos. 2024ZD0520500, 2021YFA1301601, 2020YFE0202200), National Natural Science Foundation of China (Nos. 82090051, 82273243, 92168207, 32400596), Open Project Program of the State Key Laboratory of Proteomics (SKLP-X202402).

References

1. Ishizawa, T. *et al.* Neither Multiple Tumors nor Portal Hypertension Are Surgical Contraindications for Hepatocellular Carcinoma. *Gastroenterology* **134,** 1908–1916 (2008).
2. Villanueva, A. Hepatocellular Carcinoma. *New England Journal of Medicine* **380** (ed Longo, D. L.) 1450–1462 (2019).
3. Llovet, J. M. *et al.* Hepatocellular Carcinoma. *Nature Reviews Disease Primers* **7,** 6 (2021).
4. Jiang, Y. *et al.* Proteomics Identifies New Therapeutic Targets of Early-Stage Hepatocellular Carcinoma. *Nature* **567,** 257–261 (2019).
5. Gao, Q. *et al.* Integrated Proteogenomic Characterization of HBV-Related Hepatocellular Carcinoma. *Cell* **179,** 561–577.e22 (2019).
6. Xing, X. *et al.* Integrated Omics Landscape of Hepatocellular Carcinoma Suggests Proteomic Subtypes for Precision Therapy. *Cell Reports Medicine* **4,** 101315 (2023).
7. Chen, Y.-J. *et al.* Proteogenomics of Non-smoking Lung Cancer in East Asia Delineates Molecular Signatures of Pathogenesis and Progression. *Cell* **182,** 226–244.e17 (2020).

8. Wang, Z. *et al.* Proteomic and Phosphoproteomic Landscape of Localized Prostate Cancer Unveils Distinct Molecular Subtypes and Insights into Precision Therapeutics. *Proceedings of the National Academy of Sciences* **121,** e2402741121 (2024).

9. Llovet, J. M. *et al.* Sorafenib in Advanced Hepatocellular Carcinoma. *New England Journal of Medicine* **359,** 378–390 (2008).

10. Kudo, M. *et al.* Lenvatinib versus Sorafenib in First-Line Treatment of Patients with Unresectable Hepatocellular Carcinoma: A Randomised Phase 3 Non-Inferiority Trial. *The Lancet* **391,** 1163–1173 (2018).

11. Montal, R. *et al.* Reply to: "mRECIST for Systemic Therapies: More Evidence Is Required before Recommendations Could Be Made". *Journal of Hepatology* **67,** 196–197 (2017).

12. Ericksen, R. E. *et al.* Loss of BCAA Catabolism during Carcinogenesis Enhances mTORC1 Activity and Promotes Tumor Development and Progression. *Cell Metabolism* **29,** 1151–1165.e6 (2019).

13. Schwartz, D. & Gygi, S. P. An Iterative Statistical Approach to the Identification of Protein Phosphorylation Motifs from Large-Scale Data Sets. *Nature Biotechnology* **23,** 1391–1398 (2005).

14. Lipp, J. J. *et al.* SR Protein Kinases Promote Splicing of Nonconsensus Introns. *Nature Structural & Molecular Biology* **22,** 611–617 (2015).

15. Songyang, Z *et al.* A Structural Basis for Substrate Specificities of Protein Ser/Thr Kinases: Primary Sequence Preference of Casein Kinases I and II, NIMA, Phosphorylase Kinase, Calmodulin-Dependent Kinase II, CDK5, and Erk1. *Molecular and Cellular Biology* **16,** 6486–6493 (1996).

16. Rusin, S. F. *et al.* Identification of Candidate Casein Kinase 2 Substrates in Mitosis by Quantitative Phosphoproteomics. *Frontiers in Cell and Developmental Biology* **5,** 97 (2017).

17. Kennelly, P. J. & Krebs, E. G. Consensus Sequences as Substrate Specificity Determinants for Protein Kinases and Protein Phosphatases. *The Journal of Biological Chemistry* **266,** 15555–15558 (1991).

18. Nishikawa, K. *et al.* Determination of the Specific Substrate Sequence Motifs of Protein Kinase C Isozymes. *Journal of Biological Chemistry* **272,** 952–960 (1997).

19. Türei, D. *et al.* OmniPath: Guidelines and Gateway for Literature-Curated Signaling Pathway Resources. *Nature Methods* **13,** 966–967 (2016).

20. Zhang, X. *et al.* Identifying Novel Targets of Oncogenic EGF Receptor Signaling in Lung Cancer through Global Phosphoproteomics. *PROTEOMICS* **15,** 340–355 (2015).

21. Roux, P. P. *et al.* Phosphorylation of P90 Ribosomal S6 Kinase (RSK) Regulates Extracellular Signal-Regulated Kinase Docking and RSK Activity. *Molecular and Cellular Biology* **23,** 4796–4804 (2003).

22. Frödin, M. & Gammeltoft, S. Role and Regulation of 90 kDa Ribosomal S6 Kinase (RSK) in Signal Transduction. *Molecular and Cellular Endocrinology* **151,** 65–77 (1999).

23. Casalvieri, K. A. *et al.* Selective Targeting of RSK Isoforms in Cancer. *Trends in Cancer* **3,** 302–312 (2017).

24. Pende, M. *et al.* S6K1(-/-)/S6K2(-/-) Mice Exhibit Perinatal Lethality and Rapamycin-Sensitive 5'-Terminal Oligopyrimidine mRNA Translation and Reveal a Mitogen-Activated Protein Kinase-Dependent S6 Kinase Pathway. *Molecular and Cellular Biology* **24,** 3112–3124 (2004).

25. Montagnoli, A. *et al.* Identification of Mcm2 Phosphorylation Sites by S-phase-regulating Kinases. *Journal of Biological Chemistry* **281,** 10281–10290 (2006).

26. Giannelli, G. *et al.* Role of Epithelial to Mesenchymal Transition in Hepatocellular Carcinoma. *Journal of Hepatology* **65,** 798–808 (2016).

27. Song, X. *et al.* SRSF3-Regulated RNA Alternative Splicing Promotes Glioblastoma Tumorigenicity by Affecting Multiple Cellular Processes. *Cancer Research* **79,** 5288–5301 (2019).

28. Wang, Z.-W. *et al.* SRSF3-mediated Regulation of N6-methyladenosine Modification-Related lncRNA ANRIL Splicing Promotes Resistance of Pancreatic Cancer to Gemcitabine. *Cell Reports* **39,** 110813 (2022).

29. Sen, S. *et al.* Splicing Factor SRSF3 Is Crucial for Hepatocyte Differentiation and Metabolic Function. *Nature Communications* **4,** 1336 (2013).

30. Sen, S. *et al.* Deletion of Serine/Arginine-Rich Splicing Factor 3 in Hepatocytes Predisposes to Hepatocellular Carcinoma in Mice. *Hepatology* **61,** 171–183 (2015).

31. Wang, H. *et al.* Alteration of Splicing Factors' Expression during Liver Disease Progression: Impact on Hepatocellular Carcinoma Outcome. *Hepatology International* **13,** 454–467 (2019).

32. Slišković, I. *et al.* Exploring the Multifunctionality of SR Proteins. *Biochemical Society Transactions* **50,** 187–198 (2022).

33. Huang, Y. *et al.* A Molecular Link between SR Protein Dephosphorylation and mRNA Export. *Proceedings of the National Academy of Sciences* **101,** 9666–9670 (2004).

34. Gerdes, J. *et al.* Production of a Mouse Monoclonal Antibody Reactive with a Human Nuclear Antigen Associated with Cell Proliferation. *International Journal of Cancer* **31,** 13–20 (1983).

35. Gui, J. F. *et al.* A Serine Kinase Regulates Intracellular Localization of Splicing Factors in the Cell Cycle. *Nature* **369,** 678–682 (1994).

36. Duncan, P. I. *et al.* The Clk2 and Clk3 Dual-Specificity Protein Kinases Regulate the Intranuclear Distribution of SR Proteins and Influence Pre-mRNA Splicing. *Experimental Cell Research* **241,** 300–308 (1998).

37. Aubol, B. E. *et al.* Release of SR Proteins from CLK1 by SRPK1: A Symbiotic Kinase System for Phosphorylation Control of Pre-mRNA Splicing. *Molecular Cell* **63,** 218–228 (2016).

38. Zhou, Z. *et al.* The Akt-SRPK-SR Axis Constitutes a Major Pathway in Transducing EGF Signaling to Regulate Alternative Splicing in the Nucleus. *Molecular Cell* **47,** 422–433 (2012).

39. Lee, G. *et al.* Post-Transcriptional Regulation of De Novo Lipogenesis by mTORC1-S6K1-SRPK2 Signaling. *Cell* **171,** 1545–1558.e18 (2017).

40. Xu, Q. *et al.* Identifying Three-Dimensional Structures of Autophosphorylation Complexes in Crystals of Protein Kinases. *Science Signaling* **8** (2015).

41. Fukuhara, T. *et al.* Utilization of Host SR Protein Kinases and RNA-splicing Machinery during Viral Replication. *Proceedings of the National Academy of Sciences* **103,** 11329–11333 (2006).

42. Muraki, M. *et al.* Manipulation of Alternative Splicing by a Newly Developed Inhibitor of Clks. *Journal of Biological Chemistry* **279,** 24246–24254 (2004).

43. Ma, C.-T. *et al.* Mechanism of Dephosphorylation of the SR Protein ASF/SF2 by Protein Phosphatase 1. *Journal of Molecular Biology* **403,** 386–404 (2010).

44. Choy, M. S. *et al.* SDS22 Selectively Recognizes and Traps Metal-Deficient Inactive PP1. *Proceedings of the National Academy of Sciences* **116,** 20472–20481 (2019).

45. Duan, H. *et al.* Phosphorylation of PP1 Regulator Sds22 by PLK1 Ensures Accurate Chromosome Segregation. *Journal of Biological Chemistry* **291,** 21123–21136 (2016).

46. Vogel, A. *et al.* Hepatocellular Carcinoma. *The Lancet* **400,** 1345–1362 (2022).

47. Siegel, R. L. *et al.* Cancer Statistics, 2022. *CA: A Cancer Journal for Clinicians* **72,** 7–33 (2022).

48. Calvisi, D. F. *et al.* Ubiquitous Activation of Ras and Jak/Stat Pathways in Human HCC. *Gastroenterology* **130,** 1117–1128 (2006).

49. Ezzoukhry, Z. *et al.* EGFR Activation Is a Potential Determinant of Primary Resistance of Hepatocellular Carcinoma Cells to Sorafenib. *International Journal of Cancer* **131,** 2961–2969 (2012).

50. Jin, H. *et al.* EGFR Activation Limits the Response of Liver Cancer to Lenvatinib. *Nature* **595,** 730–734 (2021).

51. Corcoran, R. B. *et al.* EGFR-Mediated Reactivation of MAPK Signaling Contributes to Insensitivity of *BRAF* -Mutant Colorectal Cancers to RAF Inhibition with Vemurafenib. *Cancer Discovery* **2,** 227–235 (2012).

52. Rong, C. *et al.* Adaptive ERK Signalling Activation in Response to Therapy and in Silico Prognostic Evaluation of EGFR-MAPK in HNSCC. *British Journal of Cancer* **123,** 288–297 (2020).

53. Cáceres, J. F. *et al.* A Specific Subset of SR Proteins Shuttles Continuously between the Nucleus and the Cytoplasm. *Genes & Development* **12,** 55–66 (1998).

54. Jia, R. & Zheng, Z.-M. Oncogenic SRSF3 in Health and Diseases. *International Journal of Biological Sciences* **19,** 3057–3076 (2023).

55. Hubert, M. *et al.* ROBPCA: A New Approach to Robust Principal Component Analysis. *Technometrics* **47,** 64–79 (2005).

56. Alexandrov, L. B. *et al.* The Repertoire of Mutational Signatures in Human Cancer. *Nature* **578,** 94–101 (2020).

57. Hänzelmann, S. *et al.* GSVA: Gene Set Variation Analysis for Microarray and RNA-Seq Data. *BMC Bioinformatics* **14,** 7 (2013).

Supplementary Information

Supplementary Information includes:
Figure S1. The outliers detection, phosphoproteome and proteome relationship, and dysregulated Gene Ontology terms and pathways in early stage HCC.
Figure S2. Regulated kinase activities in HCC.
Figure S3 Phosphorylation of the EGFR-mediated canonical ERK/MAPK and atypical MAPK signaling cascades in Jiang et al.'s cohort and Gao et al.'s cohort.
Figure S4. Nonnegative matrix factorization and consensus-clustering identified three phosphoproteomic subtypes in HCC tissues.
Figure S5. Characterization of representative phosphosite signature, pathway, hallmark, Gene Ontology across the three phosphoproteomic subtypes in HCC tissues.
Figure S6. Characterization of pathway, hallmark of the proteome across the three phosphoproteomic subtypes in HCC tissues.
Figure S7. SRSF3 contributes to proliferation, clone formation, migration and invasion of HCC cells.
Figure S8. Phosphorylation on Ser148 and Ser150 of SRSF3 promotes proliferation of HCC cells.
Figure S9. Phosphorylation on Ser148/Ser150 of SRSF3 promotes migration and invasion of HCC cells, which is regulated by SRPKs/CLKs and PP1.
Supplementary Table 0. Phosphoproteomic data QC status of 103 HCC cases in Jiang et al.'s cohort.
Supplementary Table 1. GSEA enrichment analysis of pathways and hallmarks for proteins with the high or low correlations of protein expression and phosphorylation level in tumor tissues.
Supplementary Table 2. Differential expressed phosphorylation sites (DES) analysis on the phosphoproteome in early-stage HCC.
Supplementary Table 3. Enrichment analysis of biological processes of Gene Ontology for the proteins with differential expressed phosphorylation sites (DES) in early-stage HCC.
Supplementary Table 4. Enrichment analysis of pathways and hallmarks for the proteins with differential expressed phosphorylation sites (DES) in early-stage HCC.
Supplementary Table 5. Differential expressed proteins (DEP) analysis on the proteome in early-stage HCC.
Supplementary Table 6. Enrichment analysis of biological processes of Gene Ontology for the differential expressed proteins (DEP) in early-stage HCC.
Supplementary Table 7. Enrichment analysis of pathways and hallmarks for the differential expressed proteins (DEP) in early-stage HCC.
Supplementary Table 8. Pathways and hallmarks enriched from regulated proteins and phosphoproteins in early-stage HCC.
Supplementary Table 9. Motifs enriched from the upregulated and downregulated phosphosites in early-stage HCC.
Supplementary Table 10. Kinase activities inffered by KSEA in early-stage HCC.
Supplementary Table 11. Signature phosphosites for different phosphoproteomic subtypes in early-stage HCC.
Supplementary Table 12. Pathways and hallmarks enriched from phosphoproeins with signature phosphosites in different cluster in tumor tissues.
Supplementary Table 13. GO terms enriched from phosphoproeins with signature phospho-

sites in different cluster in tumor tissues.

Supplementary Table 14. Pathways and hallmarks enriched from signature proteins in different cluster in tumor tissues.

Supplementary Table 15. GO terms enriched from signature proteins in different cluster in tumor tissues.

Supplementary Table 16. Kinase activities inferred by ssKSEA that significantly regulated in the phosphoproteomic subtypes.

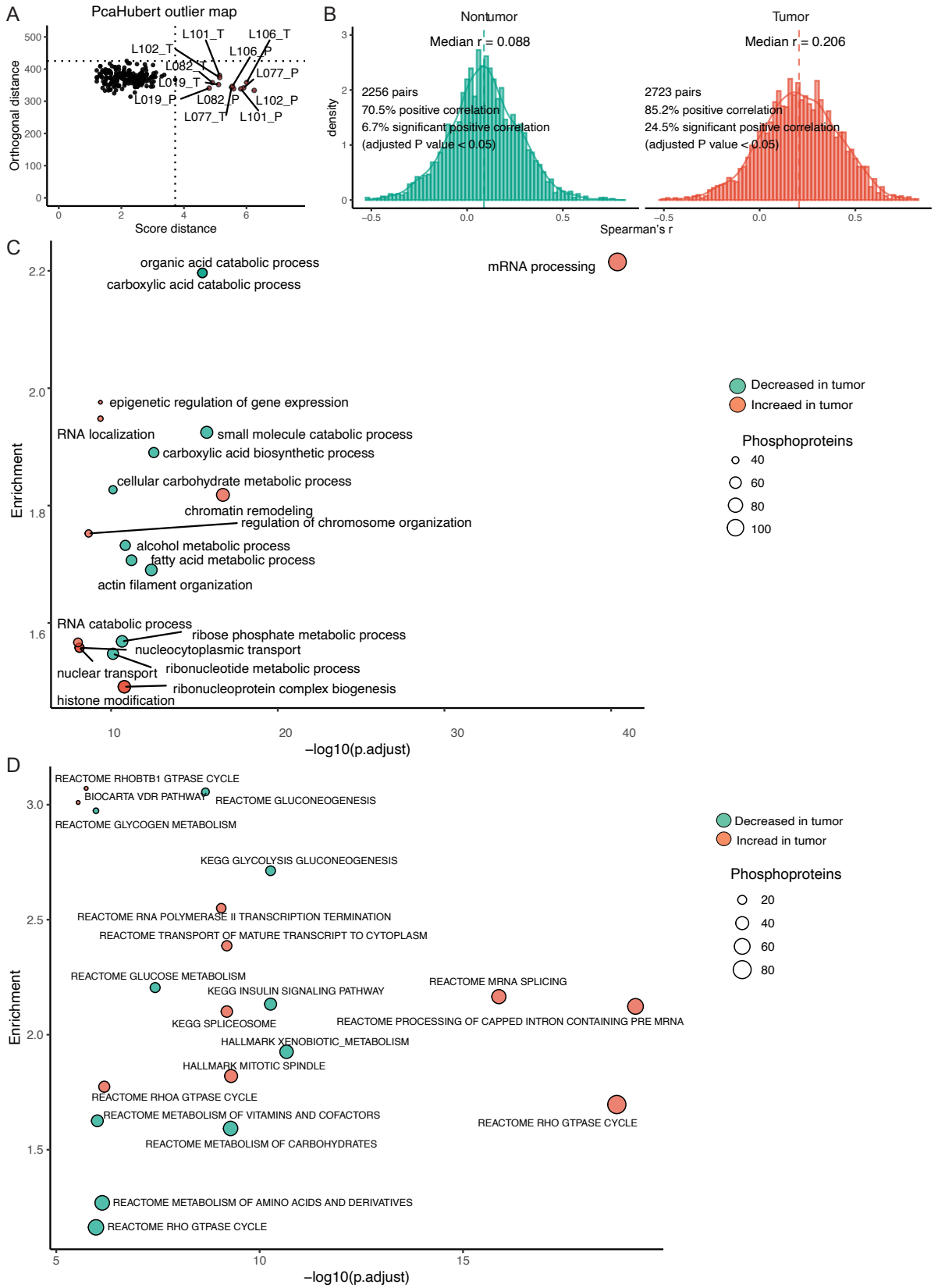
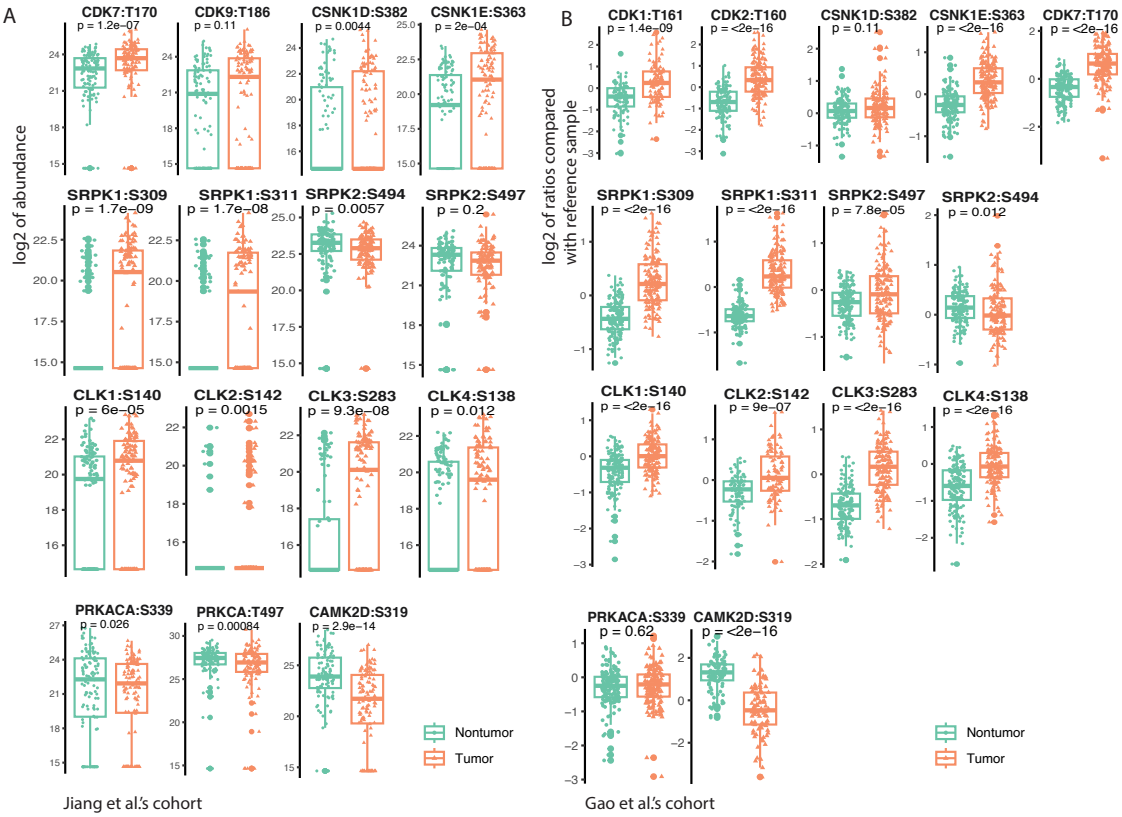


Figure S1: Outlier detection, phosphoproteome and proteome relationship, and dysregulated Gene Ontology terms and pathways in early stage HCC. (A) RobPCA identified 12 outliers (6 paired samples) in the 103 paired non-tumor and tumor samples in Jiang et al.'s cohort. (B) Correlation between protein expression levels and collapsed protein phosphorylation levels in the non-tumor and tumor tissues. (C) The top 10 biological processes of Gene Ontology, (D) pathways and hallmarks with increased and decreased protein phosphorylation levels in tumor compared with non-tumor tissues (Benjamin-Hochberg $adj.p$-value < 0.05).



C

```
Sequence alignment                                                    Activation site

sp|P49759|CLK1_HUMAN     CEPGHRQRDHESRYQNHSSKSSGRSGRSSYKSKHRIHHSTSHRRSHGKSHRRKRTRSVED  143
sp|P49760|CLK2_HUMAN     RGDAYYDTDYRHSYEYQRENSSYRSQRSSRRKHRRRRRRSRTFSRSSSQHSSRRAKSVED  145
sp|P49761|CLK3_HUMAN     FGEDYYGPSRSRHRRRSRERGPYRTRKHAHHCHKRRTRSCSSASSRSQQSSKRSSRSVED  286
sp|Q9HAZ1|CLK4_HUMAN     YVPRHYHRDIESGYRIHCSKSSVRSRRSSPKRKRNRHC--SSHQSRSKSHRRKRSRSIED  141
```

Figure S2: Regulated kinase activities in HCC. (A) Kinase observed with increased and decreased phosphorylation on activation sites in tumor tissues compared to the non-tumor tissues in Jiang et al.'s cohort and (B) Gao et al.'s cohort. The gene name of kinases and the amino acid sequence on their activation sites was labeled in each boxplot, the p-values are calculated by two-sided Wilcoxon rank-sum test. (C) Sequence aligment of CLK1-4 kinases showed the conserved activation sites, which are highlighted in red.

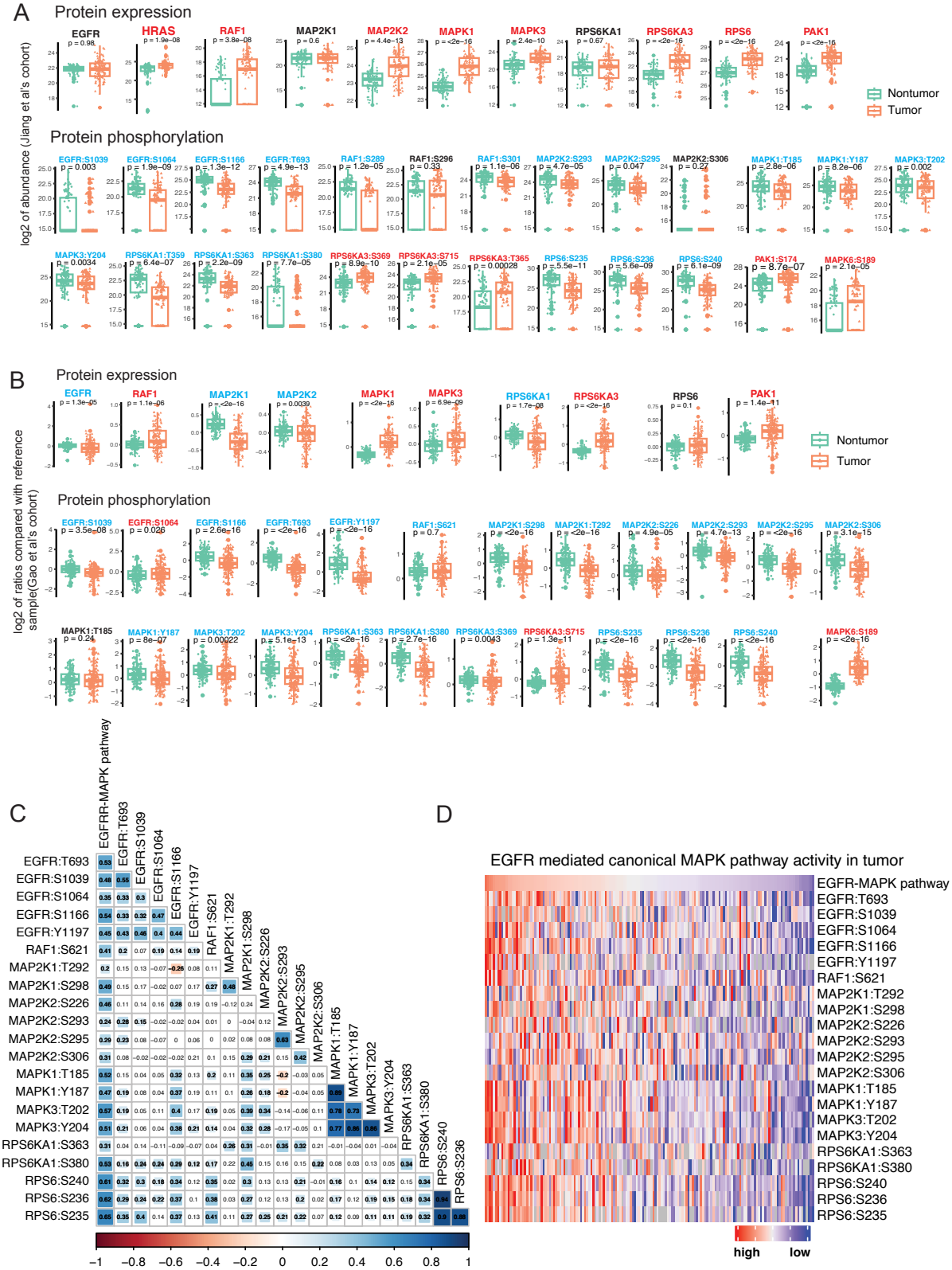


Figure S3: Phosphorylation of the EGFR-mediated canonical ERK/MAPK and atypical MAPK signaling cascades in Jiang et al.'s cohort and Gao et al.'s cohort. (A) The protein expression and phosphorylation levels of key signaling molecules in EGFR-mediated canonical ERK/MAPK and atypical MAPK signaling in HCC and non-tumor tissues in Jiang et al.'s cohort and (B) Gao et al.'s cohort are shown. The expression of proteins and functional phosphosites in non-tumor and tumor tissues are shown in boxplots. Proteins/phosphosites highlighted in red indicate significantly upregulated levels in HCC, while those in blue indicate significantly downregulted levels in HCC. Those in black indicate no significant regulation in HCC. (C) The correlation of the phosphorylation levels of activation sites of proteins in EGFR-mediated canonical ERK/MAPK signaling cascades in Gao et al.'s cohort. The median value of their activation sites was taken as a proxy for the EGFR-mediated canonical ERK/MAPK pathway and (D) patients were ranked from high to low by that and the heatmap of their phosphorylation levels is displayed.

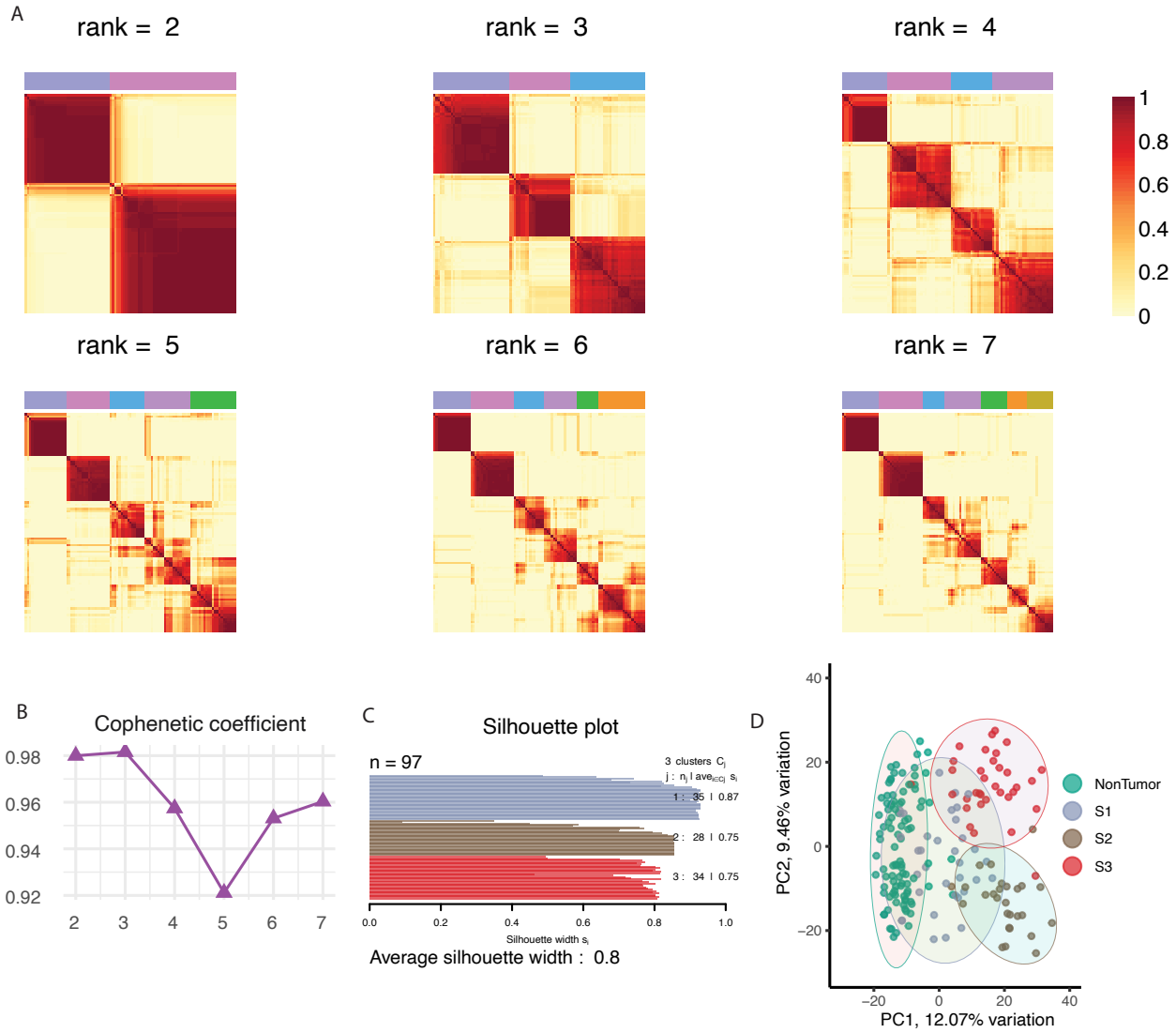


Figure S4: Nonnegative matrix factorization and consensus-clustering identified three phosphoproteomic subtypes in early-stage HCC tissues. (A) NMF and consensus-clustering analysis were applied on 97 samples with rank from 2 to 7 on phosphoproteomic data in Jiang et al.'s cohort [4], and the consensus matrix is shown in the heatmap. (B) Rank = 3 was determined as the final solution since it yieled the largest cophenetic coefficient with (C) average silhouette width of 0.8. (D) Annotation of the three phosphoproteomic subtypes on the scatter plot of principal component analysis of 97 paired samples by the 1,816 phosphoisites, the ellipse presents the 0.95 confidence intervals for each group.

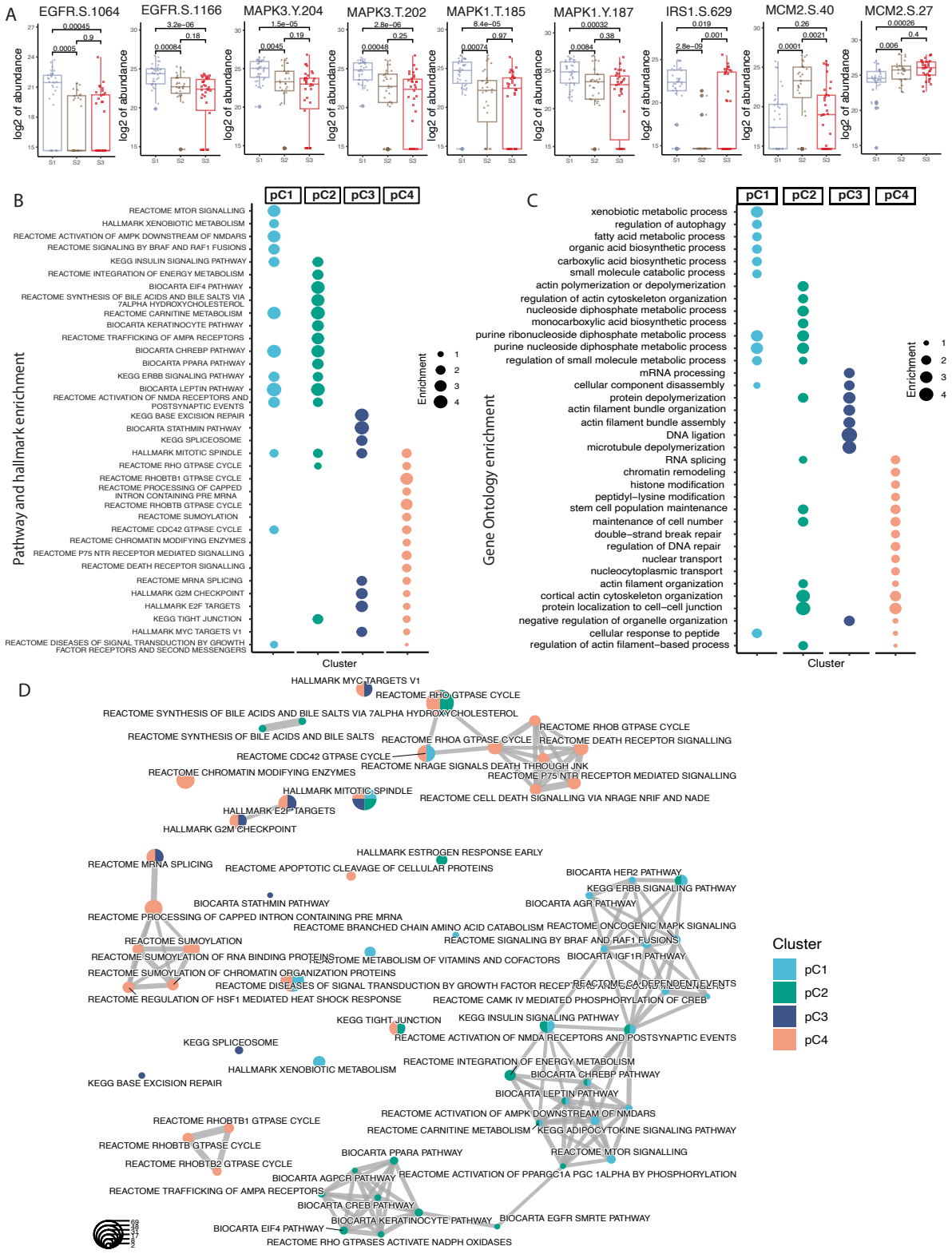


Figure S5: Characterization of representative phosphosite signatures, pathways and hallmarks, biological processes across the three phosphoproteomic subtypes in HCC tissues. (A) Representative phosphosites relating to EGFR-mediated canonical ERK/MAPK signaling, insulin signaling pathway, and DNA replication during cell cycle across the phosphoproteomic subtypes. (B) Clusters of significantly differently expressed phosphosites among the phosphoproteomic subtypes with top 10 pathways and hallmarks, (C) biological processed (Benjamin-Hochberg *adj.p*-value < 0.05) for each cluster. (D) The emmaplot of top 10 pathways and hallmarks from proteins in four clusters with significantly differently expressed phosphosites across three phosphoproteomic subtypes.

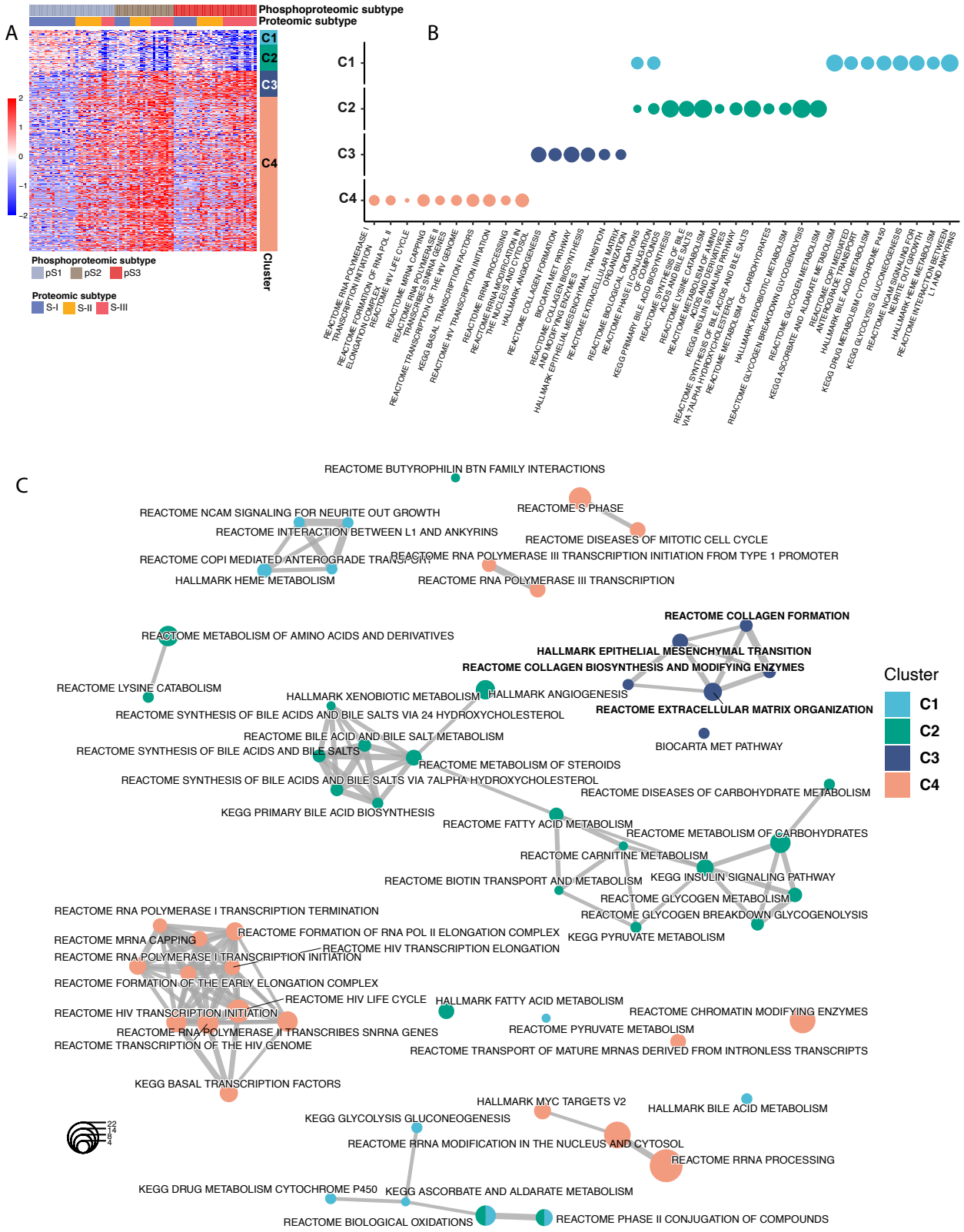


Figure S6: Characterization of pathways and hallmarks of the proteome across the three phosphoproteomic subtypes in HCC tissues. (A) 566 proteins showed significantly different expression levels across the three phosphoproteomic subtypes and were divided into four clusters. The relative abundance of proteins (log2transformed) were shown in the heatmap. (B) Clusters of significantly differently expressed proteins across the phosphoproteomic subtypes with top 10 pathways and hallmarks (Benjamin-Hochberg *adj.p*-value < 0.05) for each cluster. (C) The emmaplot of top 10 pathways and hallmarks from proteins in four clusters with significantly different expression levels across three phosphoproteomic subtypes. The pathways related to collagen formation, EMT, and ECM were highlighted in bold.

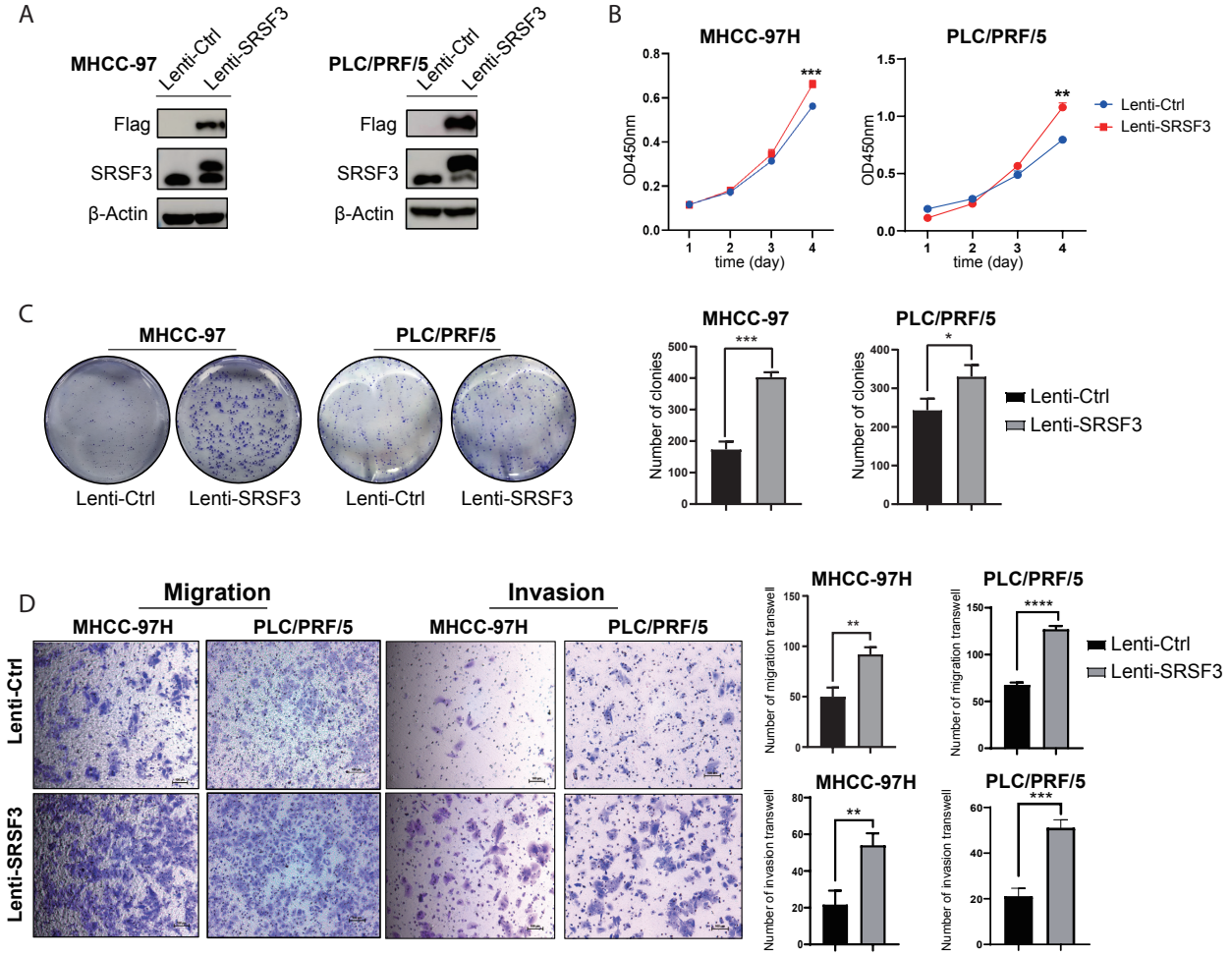


Figure S7: SRSF3 contributes to proliferation, clone formation, migration and invasion of HCC cells. (A) Stable SRSF3 overexpression in MHCC-97H and PLC/PRF/5 cell lines were constructed and significantly enhanced the abilities of (B) proliferation, (C) cell clone formation, (D) migration and invasion of HCC cells. For all graphs, asterisks show statistical significance by two-sided Wilcoxon rank-sum test, $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

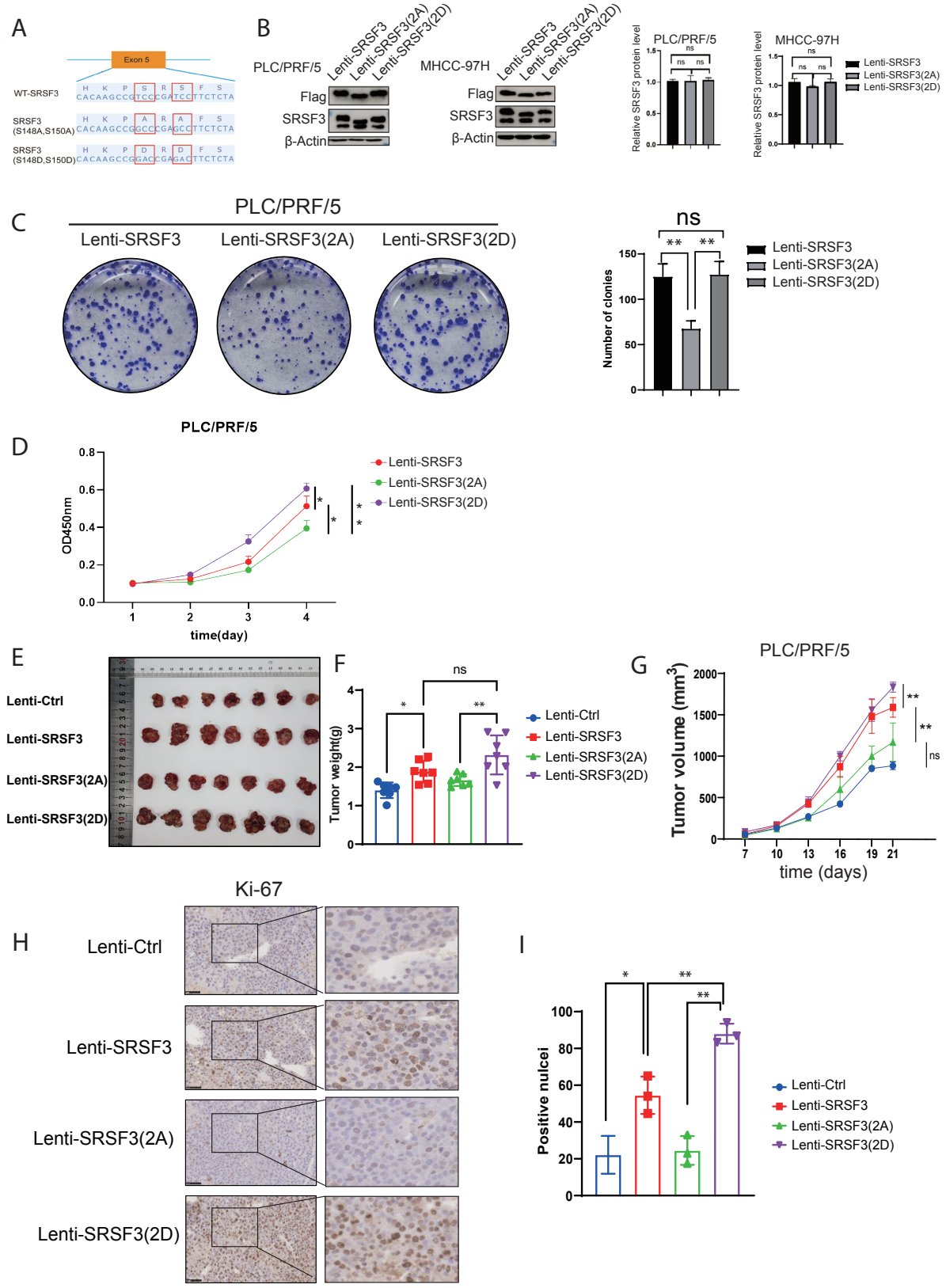


Figure S8: Phosphorylation on Ser148 and Ser150 of SRSF3 promotes proliferation of HCC cells. (A) Stable cell lines were constructed expressing wide-type form, mutant simulating dephosphorylated form (S148A, S150A), mutant simulating phosphorylated form (S148D, S150D) and (B) were confirmed by Western-Blot analysis. (C) Colony formation assays showed that SRSF3 (S148A/S150A) mutant significantly inhibited cell clone formation and (D) proliferation of HCC cells. (E) Images of the gross tumor, (F) the plot of tumor weight and (G) the growth curve from xenograft mouse model in different groups. (H) Immunohistochemistry (IHC) using the Ki-67 antibody highlighted proliferative activity in tumor tissues from the xenograft mouse model in groups with different SRSF3 mutants. (I) Statistical analysis of Ki-67-positive nuclei proportions in groups with different SRSF3 mutants. For all graphs, asterisks show statistical significance by two-sided Wilcoxon rank-sum test , $*p < 0.05$; $**p < 0.01$; $***p < 0.001$.

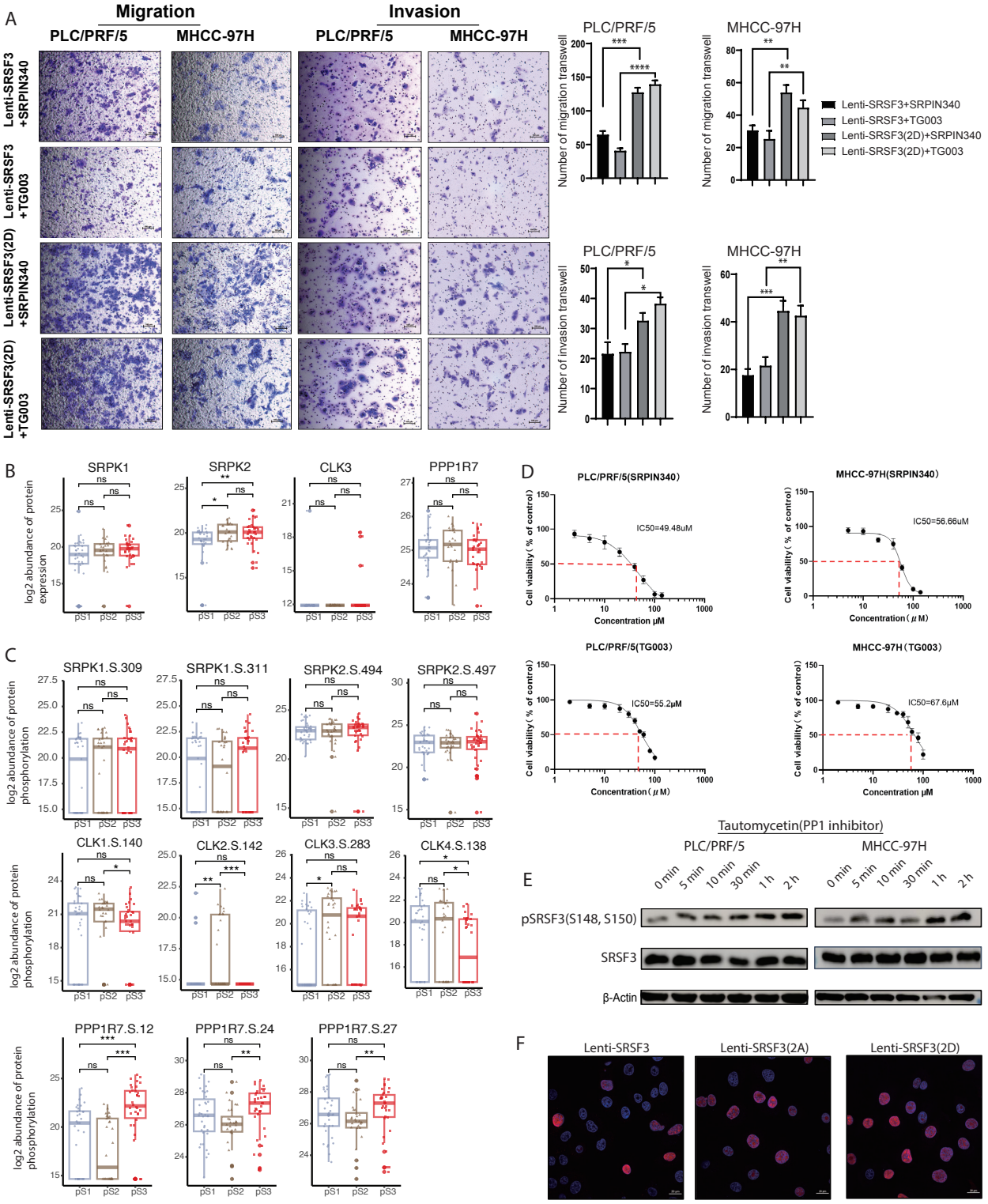


Figure S9: Phosphorylation on Ser148/Ser150 of SRSF3 promotes migration and invasion of HCC cells, regulated by SRPKs/CLKs and PP1. (A) Treatment with SRPIN340 or TG003 was applied to SRSF3-WT overexpressing or phosphorylated mutant SRSF3(S148D/S150D) overexpressing PLC/PRF/5 and MHCC-97H HCC cell lines, respectively. The treatment significantly reduced the migration and invasion abilities of HCC cells in the SRSF3-WT cell lines compared to the phosphorylated mutant. (B) The protein expressions of CLK3, SRPK1, SRPK2 and PPP1R7, and (C) protein phosphorylation level of functional regulating phosphosites on CLKs, SRPKs and PPP1R7 across the three phosphoproteomic subtypes were analyzed. (D) The IC50 of SRPIN340 and TG003 inhibitors in PLC/PRF/5 and MHCC-97H cell lines were determinated. (E) The addition of tautomycetin (PP1 inhibitor) to PLC/PRF/5 and MHCC-97H cell lines significantly increased the abundance of phosphorylation of Ser148 and Ser150 on SRSF3. (F) No significant differences were observed in the subcellular localization of SRSF3 across the SRSF3 mutants. For all graphs, asterisks show statistical significance by two-sided Wilcoxon rank-sum test , $*p < 0.05$; $**p < 0.01$; $***p < 0.001$.