Cell Reports Medicine, Volume *4*

Supplemental information

Integrated omics landscape of hepatocellular carcinoma suggests proteomic subtypes for precision therapy

Xiaohua Xing, En Hu, Jiahe Ouyang, Xianyu Zhong, Fei Wang, Kaixin Liu, Linsheng Cai, Yang Zhou, Yingchao Wang, Geng Chen, Zhenli Li, Liming Wu, and Xiaolong Liu

SUPPLEMENTAL FIGURES

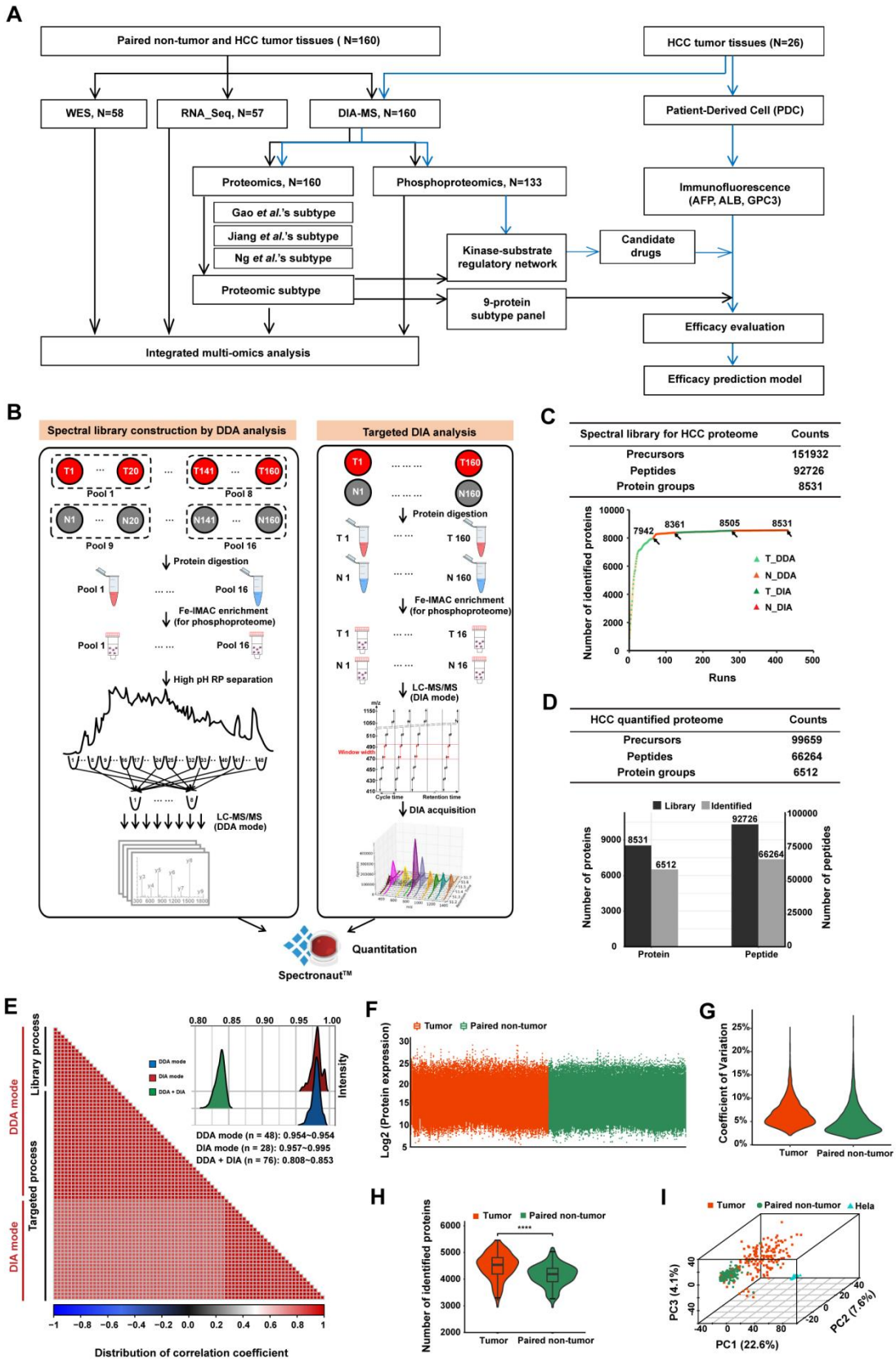

Figure S1. Study design and overview of proteomic landscape of HCC. Related to Figure 1.

(A) Overview of multi-omics landscape of HCC. 160 paired tumor and non-tumor HCC tissues were subjected to multi-omics analysis. All tissues were performed with proteomic analysis to verify the proteomic subtypes of HCC and construct prediction model, of which 132 paired tissues were selected for phosphoproteomic analysis to further screen for drug targets, and of which 58 paired tissues were selected for whole-exome sequencing (WES), and of which 57 paired tissues were selected for total transcriptome sequencing (RNA_Seq) for integrated multi-omics analysis. For constructing the subtype-based therapeutic effect prediction model for candidate drugs, we also performed proteome and phosphoproteome profiling on 26 paired tumor and non-tumor HCC tissues before PDC culture, respectively.
(B) Overview of the proteomics workflow. To construct the spectral library, the HCC tumor and paired non-tumor tissues were divided into 16 pool samples, and each pool sample created by pooling 20 samples with equal contribution. The pool samples were then digested, fractionated and subjected to LC-MS/MS with DDA mode. For individual samples, the digestion and LC-MS/MS analysis with DIA mode were performed individually. The proteins were detected and quantified using software Spectronaut.
(C) Overview of the spectral library HCC tissues. The upper table shows the information of the spectral library, including precursors, peptides and protein groups, fractions and the addition of DIA data significantly increased the coverage of reference spectral library. The lower panel was the protein number accumulation curve distinguishing the sample type and the data acquisition mode.
(D) Summary of the DIA proteome of HCC tissues. The upper table shows the information of the DIA proteome, including precursors, peptides and protein groups. The lower figure shows the proportion of identified proteins and peptides in the reference library.
(E) Robust and precise proteomic platforms. The bottom-left half of the panel represents the pairwise Pearson's correlation coefficients of the Hela cell samples through library process and targeted process including DDA mode (technical replicate n = 48) and DIA mode (technical replicate n = 28), and the top-right half of the panel depicts the distribution of Pearson's correlation of Hela samples for DDA mode, DIA mode and DDA+DIA mode.
(F) Distribution of protein abundance identified in HCC tumor (biological replicate n = 152) and paired non-tumor tissues (biological replicate n = 152). Red presents tumor samples, Green denotes paired non-tumor samples. In the box plots, the middle bar represents the median, and the box represents the interquartile range; bars extend to 2 × the interquartile range.
(G) Distribution of coefficient of variation of HCC tumor and paired non-tumor samples.
(H) The protein number shows significant difference between HCC tumors and paired non-tumors (two-tailed Wilcoxon test). Boxplots show median (central line), upper and lower quartiles (box limits), 1.5 × interquartile range (whiskers).
(I) Principal component analysis. The tumor samples exhibit higher heterogeneity than the paired non-tumor samples.

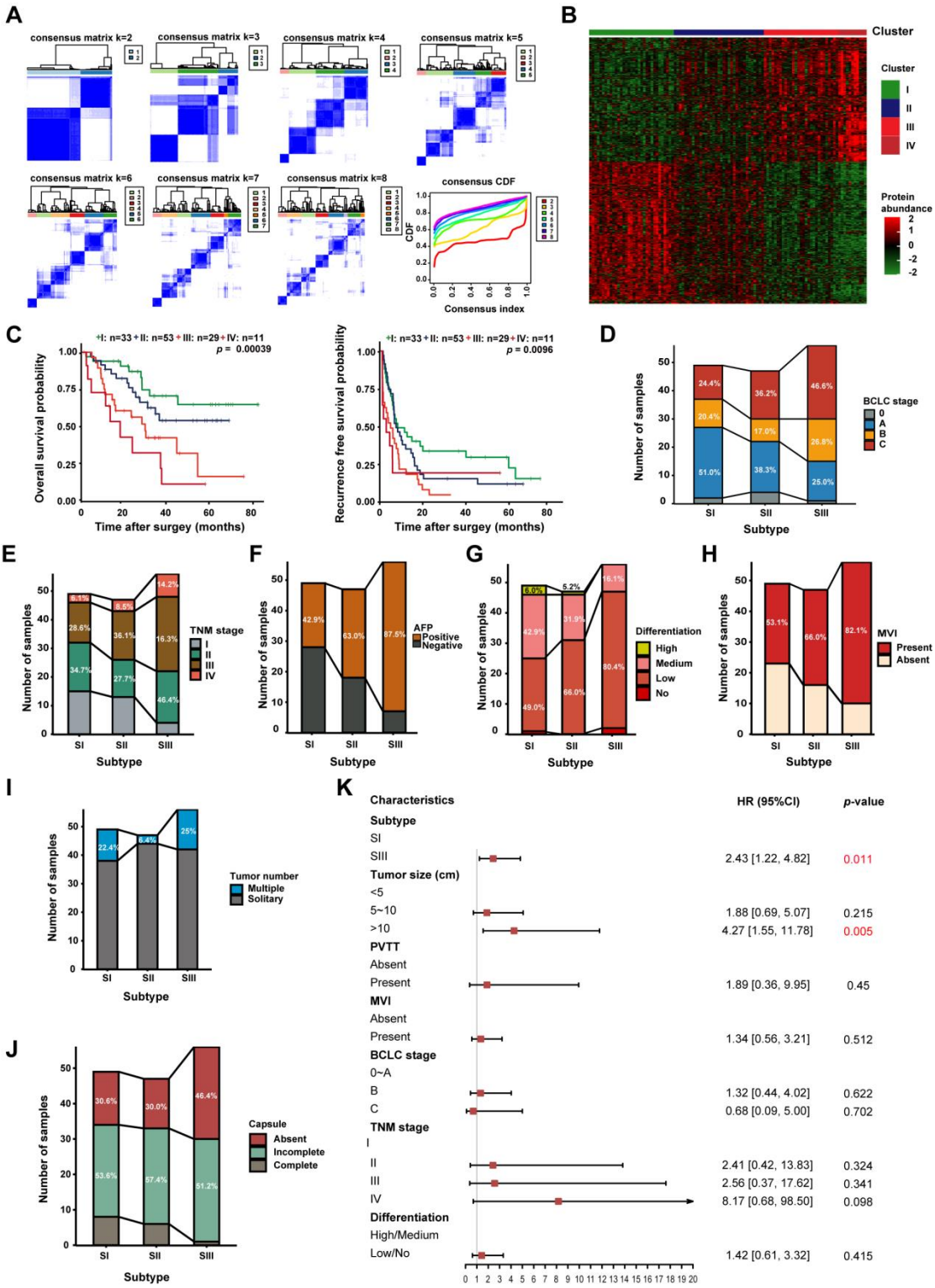

Figure S2. The proteomic subtypes of HCC. Related to Figure 1.

(A) Consensus clustering of HCC tumors based on the relative abundance of most variant proteins.

(B) The heatmap of the relative abundance of signature proteins (log_2-transformed) in four clusters (cluster I = 33, cluster II = 53, cluster III = 29, cluster IV = 11).

(C) Kaplan-Meier curves of OS and RFS for each cluster. The *p* values were calculated by log-rank test. Due to the small sample size of the fourth cluster and its similar protein expression and prognosis to the third cluster, it was merged with the third cluster as an integrated subtype.

(D-J) Association of BCLC stage (D), TNM stage (E), serum AFP levels (F), tumor differentiation (G), MVI (H), tumor number (I) tumor capsule (J) with proteomic subtypes.
(K) Multivariable Cox analysis of the proteomic subtypes with known clinical and pathologic risk factors for progression of HCC (log-rank test).

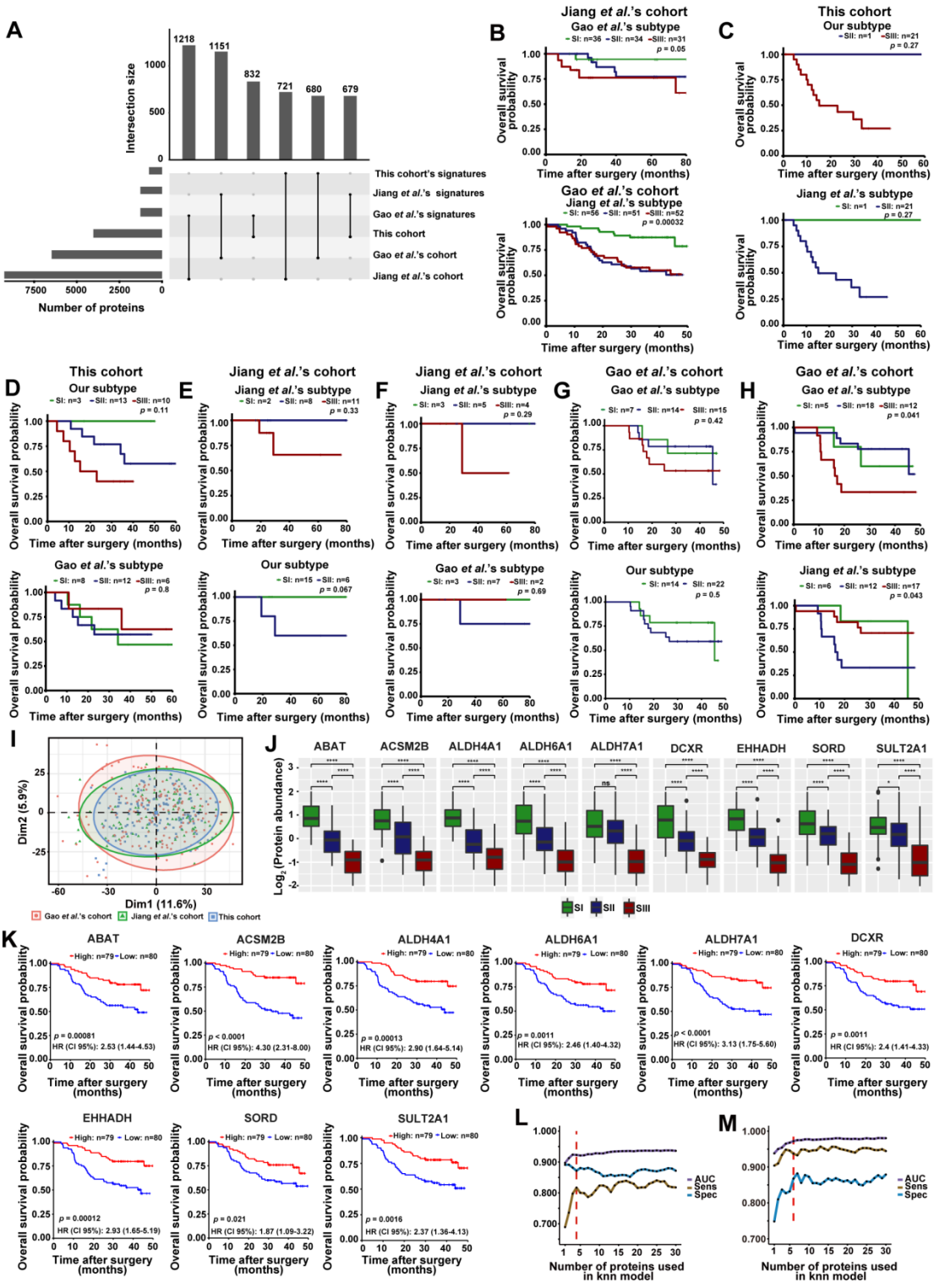

Figure S3. Cross validation of proteomic subtypes in 3 cohorts and the simplified panel for distinguishing HCC proteomic subtypes. Related to Figure 2.

(A) The upset diagram shows three subtype-specific signatures in three cohorts (Gao *et al.'s* cohort: N = 159, Jiang *et al.'s* cohort: N = 101, This cohort: N = 152).

(B) The validation of Jiang *et al.*'s and Gao *et al.*'s subtype-specific signatures in the cohort of each other. The Kaplan-Meier curves of OS were shown. The *p* values were calculated by log-rank test.

(C-H) Prognostic difference of the discordant patients based on Jiang *et al.*'s subtypes in our cohort

(C), Gao *et al*.'s in our cohort (D), our subtypes in Jiang *et al*.'s cohort (E), Gao *et al*.'s subtypes in Jiang *et al*.'s cohort (F), our subtypes in Gao *et al*.'s cohort (G), Jiang *et al*.'s subtypes in Gao *et al*.'s cohort (H). The *p* values were calculated by log-rank test.

(I) The PCA plot among 3 cohorts after removing the batch effect.

(J) The abundance of 9 proteins altered among the 3 subtypes. The *p* values were calculated with two-tailed Wilcoxon test with *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$; ****, $p < 0.001$. Boxplots show median (central line), upper and lower quartiles (box limits), 1.5 × interquartile range (whiskers).

(K) The Kaplan-Meier curves of OS and RFS for 9 proteins. *p* values were calculated by log-rank test.

(L-M) The ROC accuracy, sensitivity and specificity for SI (L) and SIII (M) distinguishing in the training data set.

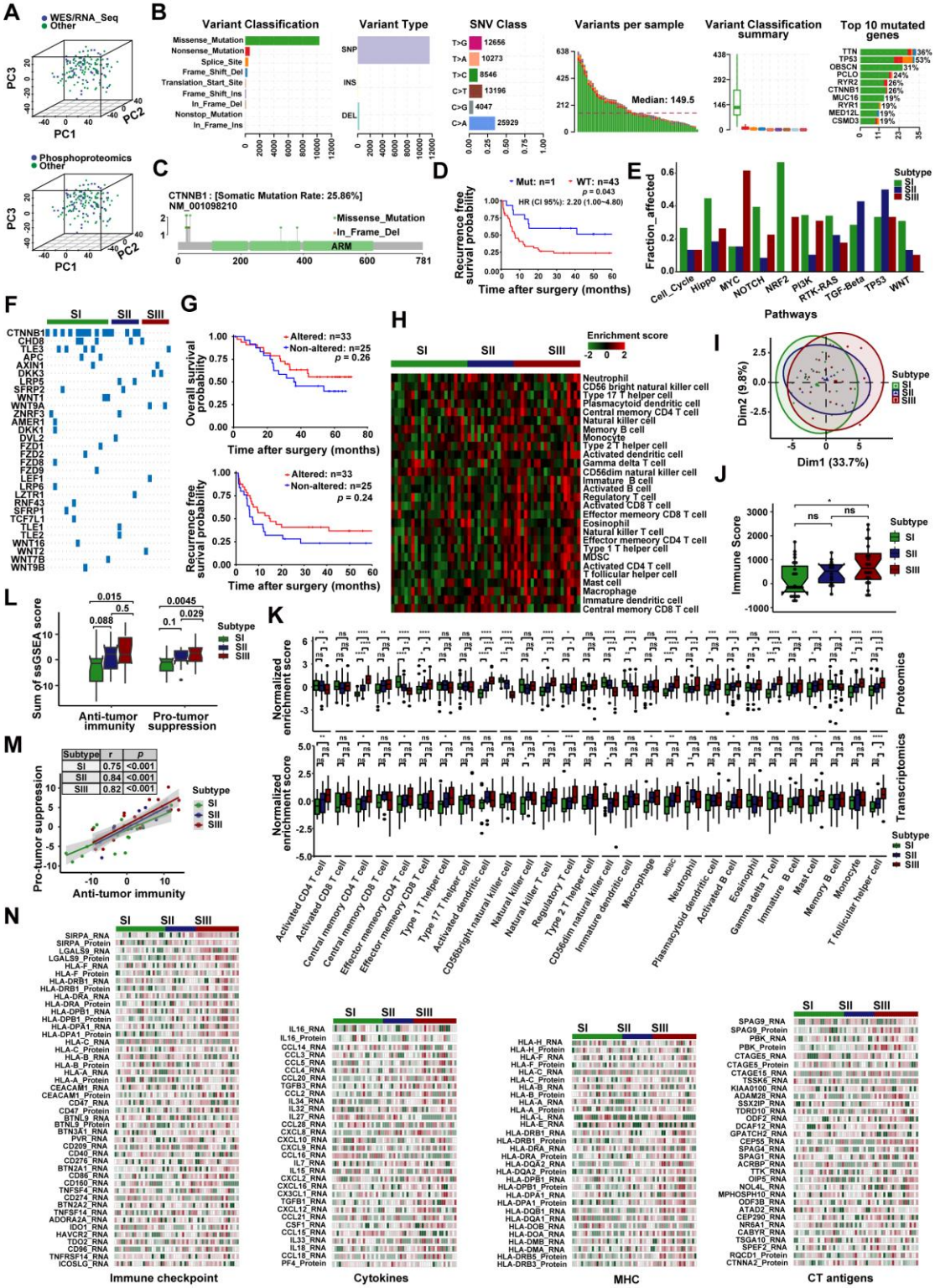

Figure S4. Mutation and immune landscape of 3 HCC proteomic subtypes. Related to Figure 3.

(A) The PCA plot between individual omics cohort and proteomics cohort. The upper panel was WES/RNA_Seq cohort, and lower panel was phosphoproteomics cohort.

(B) Summary of the mutation landscape.

(C) Lollipop plot of CTNNB1 alterations with ARM domain annotation. Mutations was annotated with gray lines, green circles were missense mutation and red circles were in-frame deletion.

(D) Kaplan-Meier curves for RFS of patients with CTNNB1 mutation or wild-type (log-rank test).

(E) Mutations-based pathways enriched in 3 proteomic subtypes.
(F) Mutation frequency of the genes involved in the Wnt pathway.
(G) Kaplan-Meier curve of WNT pathway alterations and OS/RFS (log-rank test).
(H) Heatmap shows the immune cell populations of 3 proteomic subtypes in transcriptome.
(I) The principal component analysis plot of immune scores of immune cell populations based on transcriptomic data in 3 proteomic subtypes.
(J) Boxplot showing proteomic- and transcriptomic-based immune cell abundance stratified by 3 proteomic subtypes. Significance was evaluated by two-tailed Wilcoxon test with *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$; ****, $p < 0.001$. The box portion is defined by two lines at the 75th percentile and the 25th percentile of the values. The middle line indicates 50th percentile (median).
(K) Transcriptome-based immune scores in 3 proteomic subtypes. Significance was evaluated by a two-tailed Wilcoxon test. Boxplots show median (central line), upper and lower quartiles (box limits), 1.5 × interquartile range (whiskers).
(L) Transcriptomic-based immune scores of immune activation and immunosuppression in 3 proteomic subtypes (two-tailed Wilcoxon test). Boxplots show median (central line), upper and lower quartiles (box limits), 1.5 × interquartile range (whiskers).
(M) The correlation between immune activation (anti-tumor immunity) and immunosuppression (pro-tumor suppression) based on transcriptome in 3 proteomic subtypes. Pearson's correlation coefficient (r) and *p* values are present in the table. The *p* values were calculated using the Pearson's correlation method.
(N) The expression of HLA molecule, checkpoints, CT antigens and cytokines in three proteomic subtypes.

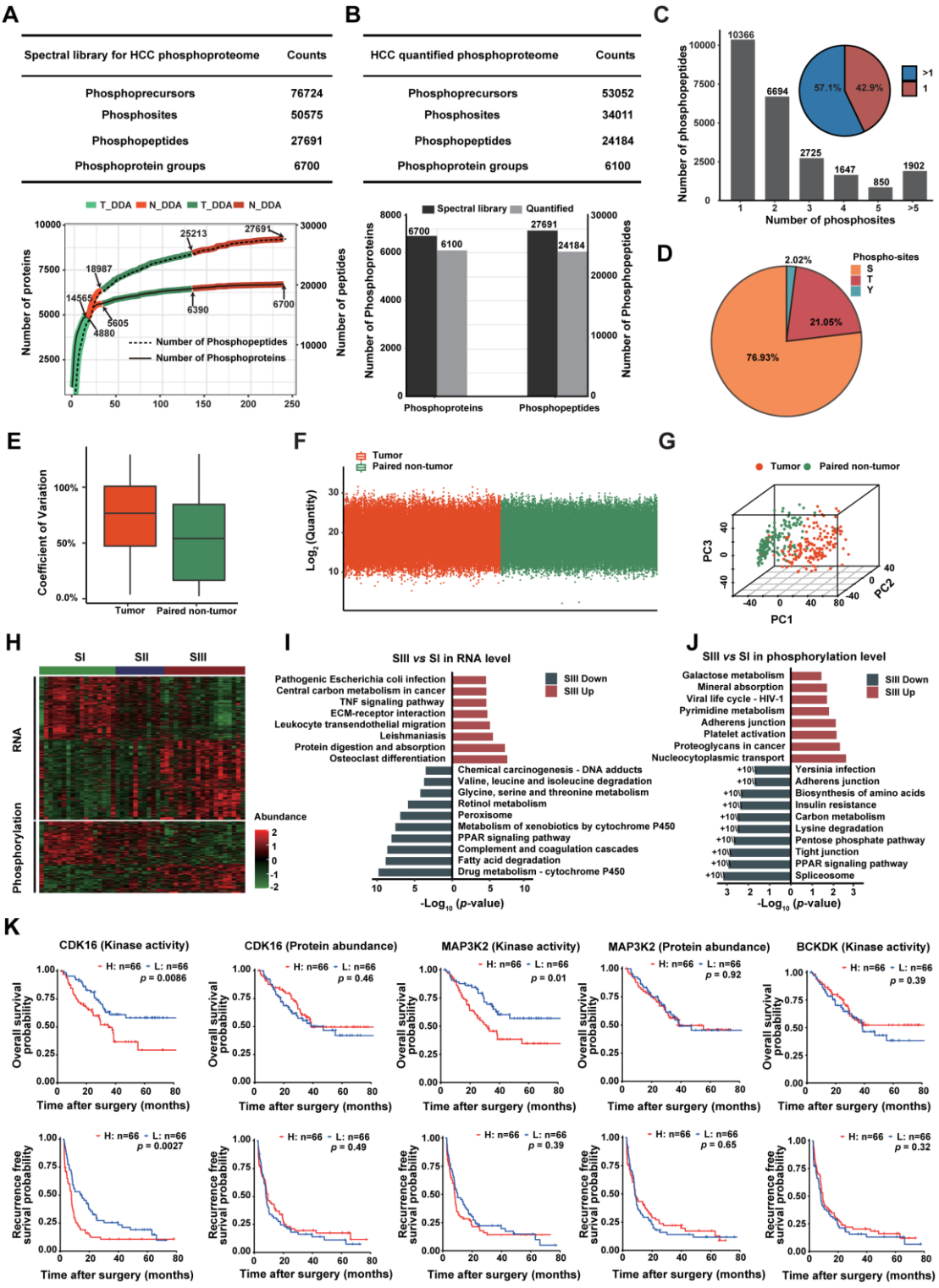

Figure S5. Phosphoproteomic and kinase profile of 3 proteomic subtypes of HCC. Related to Figure 4.

(A) Overview of the spectral library of HCC tissues for phosphoproteomics. The upper table shows the information of the spectral library, including phosphoprecursors, phosphosites, phosphopeptides and phosphoprotein groups. The lower panel was the phosphoprotein number accumulation curve distinguishing the sample type and the data acquisition mode.

(B) Summary of the DIA proteome of HCC tissues. The upper table shows the information of the DIA

phosphoproteome, including phosphoprecursors, phosphosites, phosphopeptides and phosphoprotein groups. The lower figure shows the proportion of identified phosphoproteins and phosphopeptides in the reference library.
(C) Distribution of phosphopeptides depending on their number of p-sites.
(D) Distribution of phosphorylation serine (S), phosphorylation threonine (T) and phosphorylation tyrosine (Y) sites.
(E) Distribution of coefficient of variation of HCC tumor and paired non-tumor samples. Boxplots show median (central line), upper and lower quartiles (box limits), 1.5 × interquartile range (whiskers).
(F) Distribution of proteins abundance identified in HCC tumor (n = 132) and paired non-tumor tissues (n = 132). Red presents tumor samples, Green denotes paired non-tumor samples. In the box plots, the middle bar represents the median, and the box represents the interquartile range; bars extend to 2 × the interquartile range.
(G) Principal component analysis. The tumor samples exhibit higher heterogeneity than the paired non-tumor samples.
(H) The abundance of RNA and phosphopeptides with the highest variation among 3 proteomic subtypes (SI = 17, SII = 11, SIII = 18).
(I-J) Pathway alterations in SIII versus SI at RNA level (I) and phosphorylation level (J).
(K) Kaplan-Meier curves of OS and RFS for kinase activity and kinase abundance in HCC. *p* values were calculated by log-rank test.

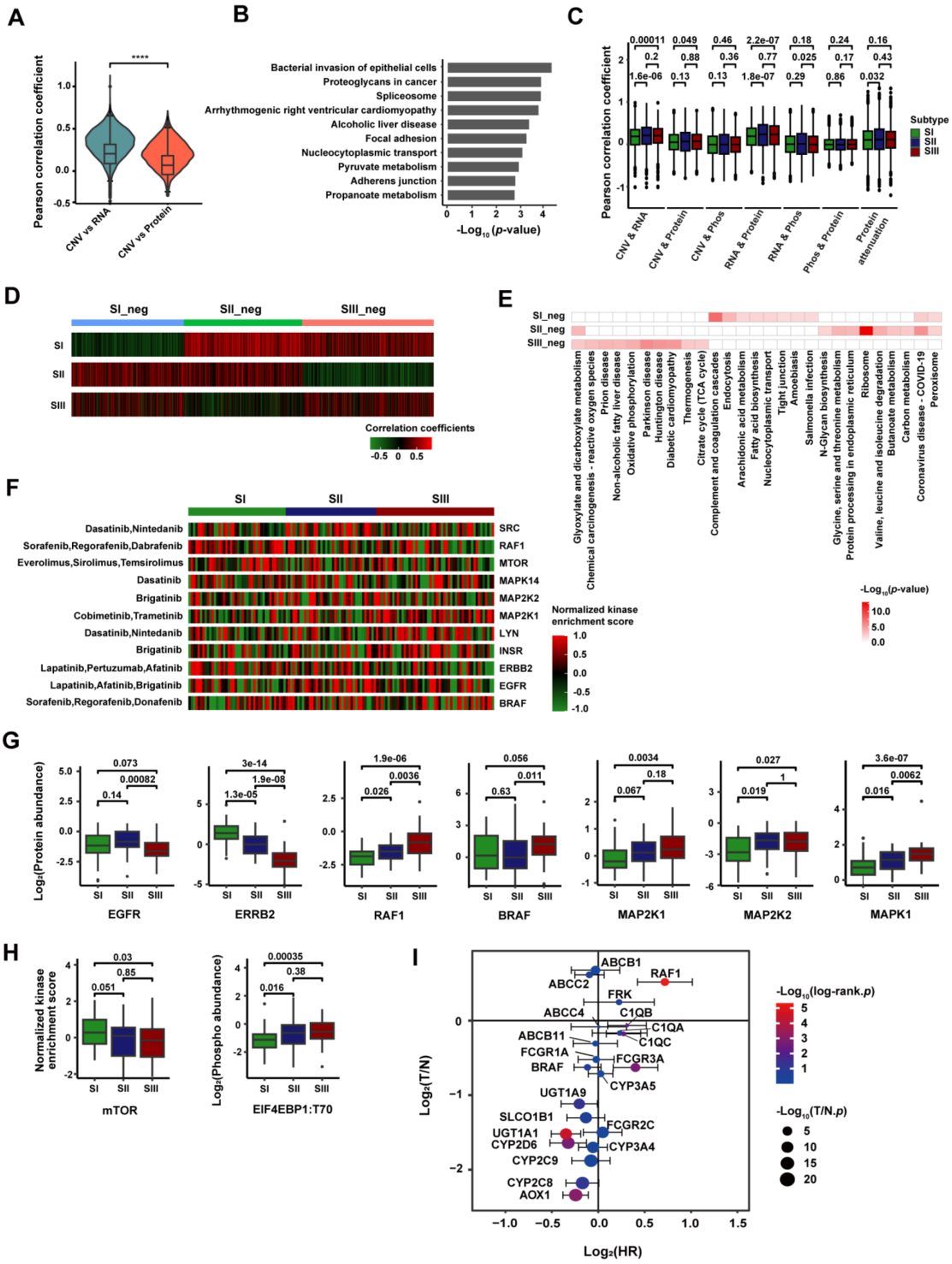

Figure S6. Integrated multi-omics analysis and key drug target screening for 3 proteomic subtypes of HCC. Related to Figure 5 and Figure 6.

(A) Comparisons of correlations between CNV *vs* RNA and CNV *vs* protein (two-tailed Wilcoxon test with *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$; ****, $p < 0.001$).

(B) The pathways enriched using negatively correlated RNA-proteins.

(C) Comparisons of correlations between every two individual omics. The *p* values were calculated with two-tailed Wilcoxon test.

(D) Hierarchical clustering analysis map of significantly changed RNA-protein correlations among 3

proteomic subtypes. Pearson's correlation coefficients of 3 proteomic subtypes between matched RNA abundances and protein abundances were calculated.

(E) Functional enrichment for significant RNA-protein correlations in each cluster.

(F) The kinase activity of FDA-approved drug targets in 3 proteomic subtypes.

(G) The kinase abundance of RAF-MEK-ERK signaling pathway related proteins in 3 proteomic subtypes (two-tailed Wilcoxon test). Boxplots show median (central line), upper and lower quartiles (box limits), 1.5 × interquartile range (whiskers).

(H) The kinase activity of mTOR and its substrate EIF4EBP1 phosphorylation in 3 proteomic subtypes. The *p* values were calculated with two-tailed Wilcoxon test. Boxplots show median (central line), upper and lower quartiles (box limits), 1.5 × interquartile range (whiskers).

(I) The recurrence risk scores of each target from FDA-approved HCC clinical drugs. The x-axis indicates $\log_2$-transformed hazard ratio for each target (log-rank test); y-axis indicate $\log_2$-transformed T/N fold change for each target (two-tailed Wilcoxon test).

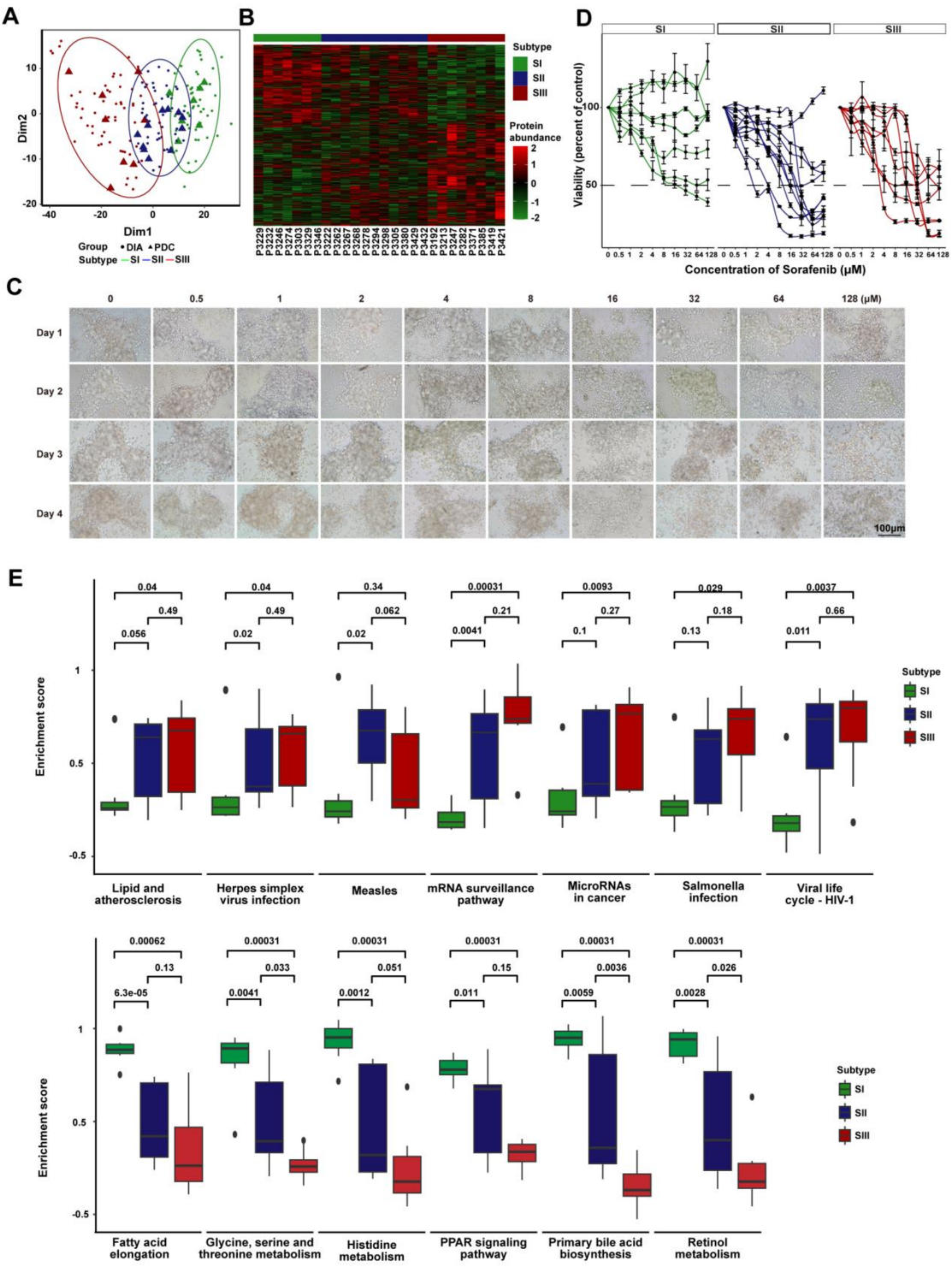

Figure S7. Subtype-specific drug sensitivities based on PDC models. Related to Figure 7.

(A) The principal component analysis plot of PDC samples and DIA samples from discovery cohort based on proteomic data.

(B) The proteomic subtypes of HCC patients for PDCs. The heatmap was shown (SI = 7, SII = 11, SIII = 8).

(C) A representative image of PDC cells at different treatment times and concentrations under microscopic examination. Scale bar, 100 μm.

(D) Dose-response curves of PDC cells to Sorafenib treatment for 3 proteomic subtypes, with an

endpoint measurement at 96 h (median ± SD, n = 3 biological repeats).
(E) The enrichment of pathways associated with Sorafenib sensitivity in 3 proteomic subtypes (two-tailed Wilcoxon test). Boxplots show median (central line), upper and lower quartiles (box limits), 1.5 × interquartile range (whiskers).

SUPPLEMENTAL TABLES

Table S7. Prognosis of 22 drug targets related to HCC. Related to Figure 6.

		T/N	OS		RFS	
Gene symbol	**Log2 (FC)**	***p* value**	***p* value**	**HR [CI 95%]**	***p* value**	**HR [CI 95%]**
ABCB1	0.68	6.81E-10	4.32E-01	0.91 [0.71,1.16]	8.31E-01	0.98 [0.82,1.18]
ABCB11	-0.31	2.61E-03	7.12E-01	0.96 [0.78,1.19]	7.99E-01	0.98 [0.83,1.15]
ABCC2	0.61	1.07E-04	9.64E-01	1.00 [0.86,1.16]	2.43E-01	0.94 [0.84,1.05]
ABCC4	-0.08	3.46E-01	2.86E-01	1.17 [0.88,1.57]	9.65E-01	1.01 [0.82,1.24]
AOX1	-2.34	2.46E-23	1.80E-03	0.83 [0.74,0.93]	5.37E-04	0.85 [0.77,0.93]
BRAF	-0.63	1.67E-04	3.84E-01	0.94 [0.81,1.08]	1.21E-01	0.92 [0.84,1.02]
C1QA	-0.17	1.38E-02	4.00E-04	1.60 [1.23,2.07]	1.26E-01	1.17 [0.96,1.44]
C1QB	-0.07	2.89E-01	0.00E+00	1.50 [1.25,1.80]	3.38E-03	1.24 [1.07,1.43]
C1QC	-0.18	5.07E-02	4.00E-04	1.34 [1.14,1.57]	3.69E-03	1.21 [1.06,1.37]
CYP2C8	-2.18	2.46E-23	5.84E-01	0.96 [0.81,1.13]	5.85E-02	0.89 [0.79,1.00]
CYP2C9	-1.88	1.47E-23	8.78E-02	0.84 [0.69,1.03]	4.51E-01	0.95 [0.82,1.09]
CYP2D6	-1.64	5.53E-21	2.30E-03	0.77 [0.65,0.91]	1.32E-03	0.80 [0.70,0.92]

CYP3A4	-1.70	5.58E-19	1.84E-01	0.91 [0.78,1.05]	4.76E-01	0.96 [0.86,1.07]
CYP3A5	-0.71	4.45E-05	7.30E-01	0.98 [0.86,1.11]	7.05E-01	1.02 [0.93,1.12]
FCGR1A	-0.52	3.51E-05	1.96E-01	0.89 [0.76,1.06]	8.23E-01	0.99 [0.86,1.13]
FCGR2C	-1.50	4.72E-19	1.13E-01	0.85 [0.69,1.04]	6.47E-01	1.03 [0.90,1.19]
FCGR3A	-0.63	4.37E-10	3.20E-03	1.41 [1.12,1.78]	1.02E-03	1.32 [1.12,1.56]
FRK	0.25	1.30E-03	2.30E-02	1.53 [1.06,2.22]	2.46E-01	1.17 [0.90,1.52]
RAF1	0.52	4.25E-07	0.00E+00	1.87 [1.43,2.44]	1.97E-06	1.65 [1.34,2.02]
SLCO1B1	-1.30	2.57E-18	8.40E-02	0.86 [0.72,1.02]	1.99E-01	0.91 [0.79,1.05]
UGT1A1	-1.52	5.32E-18	5.00E-04	0.78 [0.67,0.90]	1.58E-05	0.79 [0.70,0.88]
UGT1A9	-1.12	1.00E-15	4.43E-02	0.84 [0.71,1.00]	4.08E-02	0.87 [0.76,0.99]