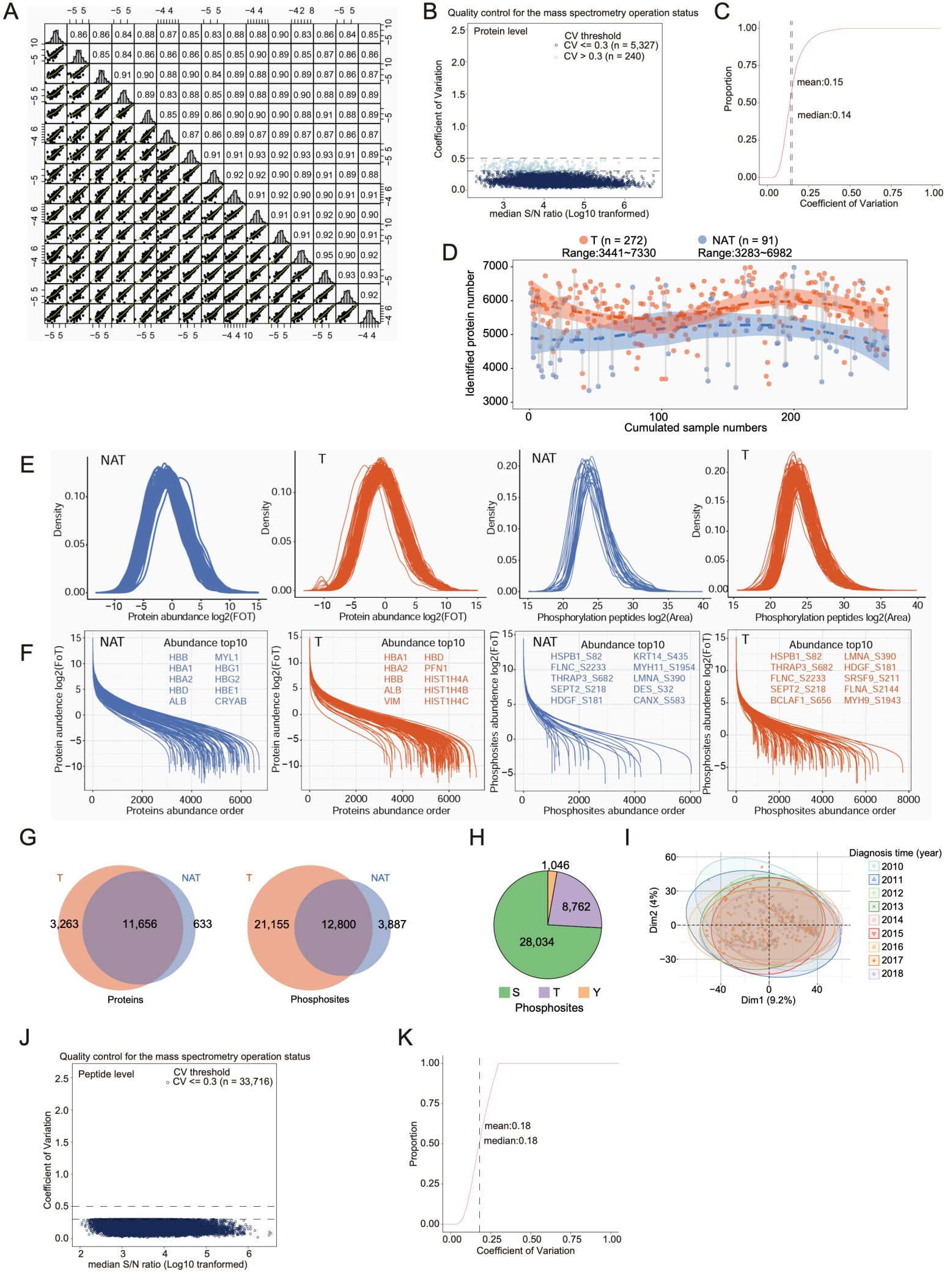
Supplementary Figure 1



Supplementary figure 1. Quality control of proteomic and phosphoproteomic data

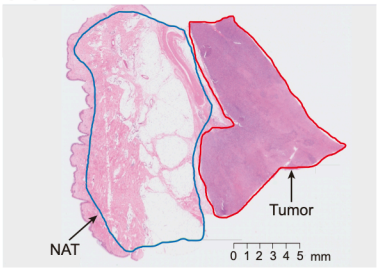
A. Longitudinal MS quality control using tryptic digest of HEK293T cells. Scatter plots and Spearman correlation coefficients are performed to evaluate the robustness of label-free quantification. Spearman correlation coefficients between repeat experiments with the same samples and the pairwise scatterplot comparison of the different samples, respectively.
B. The scatter plot illustrates the coefficient of variation (CV) and signal-to-noise ratio (S/N ratio) on the protein level among all HEK293 standard samples. The number of proteins with different CV level: less than 0.3 (n = 5,327) and more than 0.3 (n = 240).
C. Cumulative distribution curve illustrates the distribution of CVs presented on the supplementary figure 1B. Median CV: 0.14 and mean CV: 0.15.
D. Pairwise comparison of proteins is annotated in gray straight lines. The dashed curves fitted by less regression show the distribution of protein identifications in tumors (orange, n = 272) and NATs (blue, n = 91).
E. Density plot showing the distribution of protein and phosphosite abundance in tumors (orange) and NATs (blue). A unimodal distribution (dip test) is observed. All of the samples pass proteomic quality control.
F. Scatterplots showing the protein and phosphosite abundance dynamics in tumors (orange) and NATs (blue). Row data of proteins (area of peak value region) are transformed by fraction of total (FoT) and $\log_2$ transformed. The 10 highest abundance proteins are shown in the box.
G. Venn diagram shows the identified proteins and phosphosites in tumors and NATs, and their shared identification numbers.
H. Pie chart showing the proportion of three phosphorylated amino acid residues, serine, threonine, and tyrosine in all identified phosphosites.
I. Primary component analysis (PCA) of samples excised and saved in different years.
J. The scatter plot illustrates the CV and S/N ratio on peptide level among all HEK293 standard samples. The number of peptides with CV less than 0.3 is 33,716.
K. The cumulative distribution curve illustrates the distribution of CVs presented in supplementary figure J. Median CV: 0.18 and mean CV: 0.18.

Source data are provided as a Source Data file.

Supplementary Figure 2

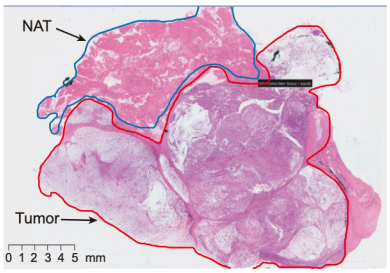
A

STS-120



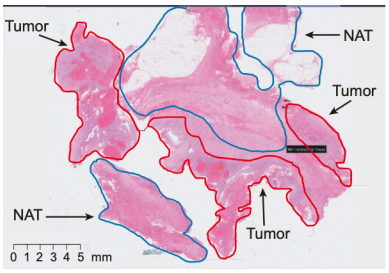
Tumor: UPS NAT : connective tissue

STS-193



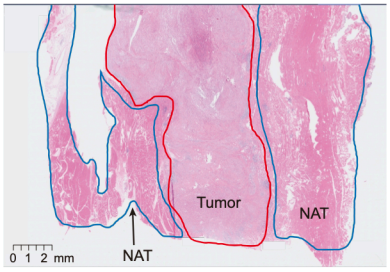
Tumor: MFS NAT : connective tissue + muscle

STS-296



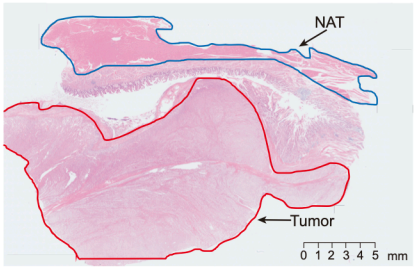
Tumor: AS NAT : connective tissue

STS-282



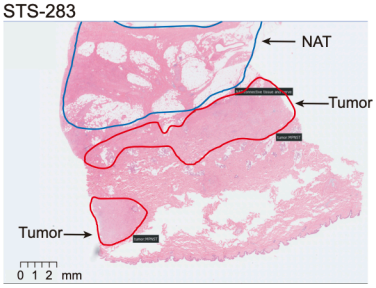
Tumor: RMS NAT : muscle

STS-28



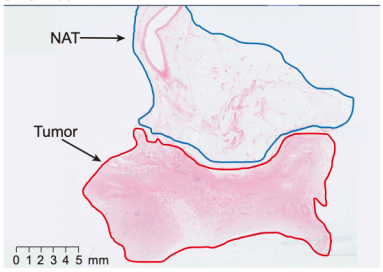
Tumor: LMS NAT : smooth muscle

STS-283



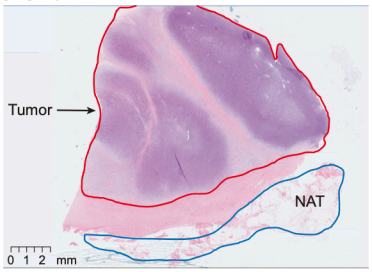
Tumor: MPNST NAT : connective tissue + nerve

STS-168



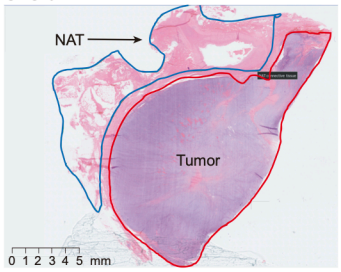
Tumor: DDLPS NAT : lipid tissue

STS-254



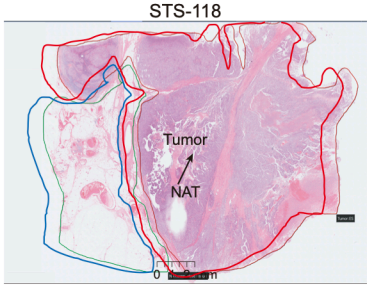
Tumor: MLPS NAT : lipid tissue

STS-81



Tumor: SS NAT : connective tissue

STS-118



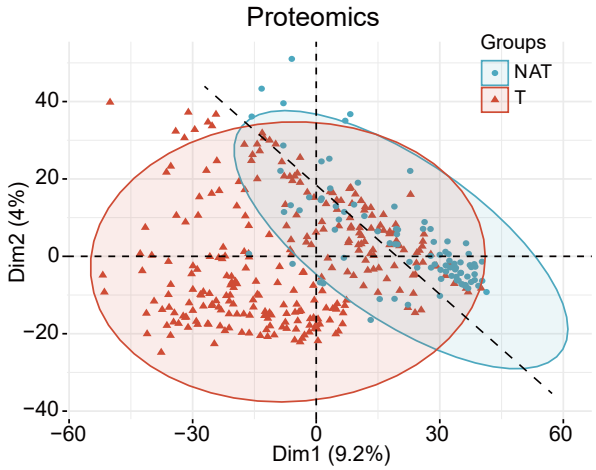
NAT : connective tissue

Supplementary figure 2. HE samples to represent the selection of NAT and tumor tissues

A. H&E-stained slices presents the regions of tumor and paired NATs. Different histological subtypes have distinguished tissue types of NATs.

Supplementary Figure 3

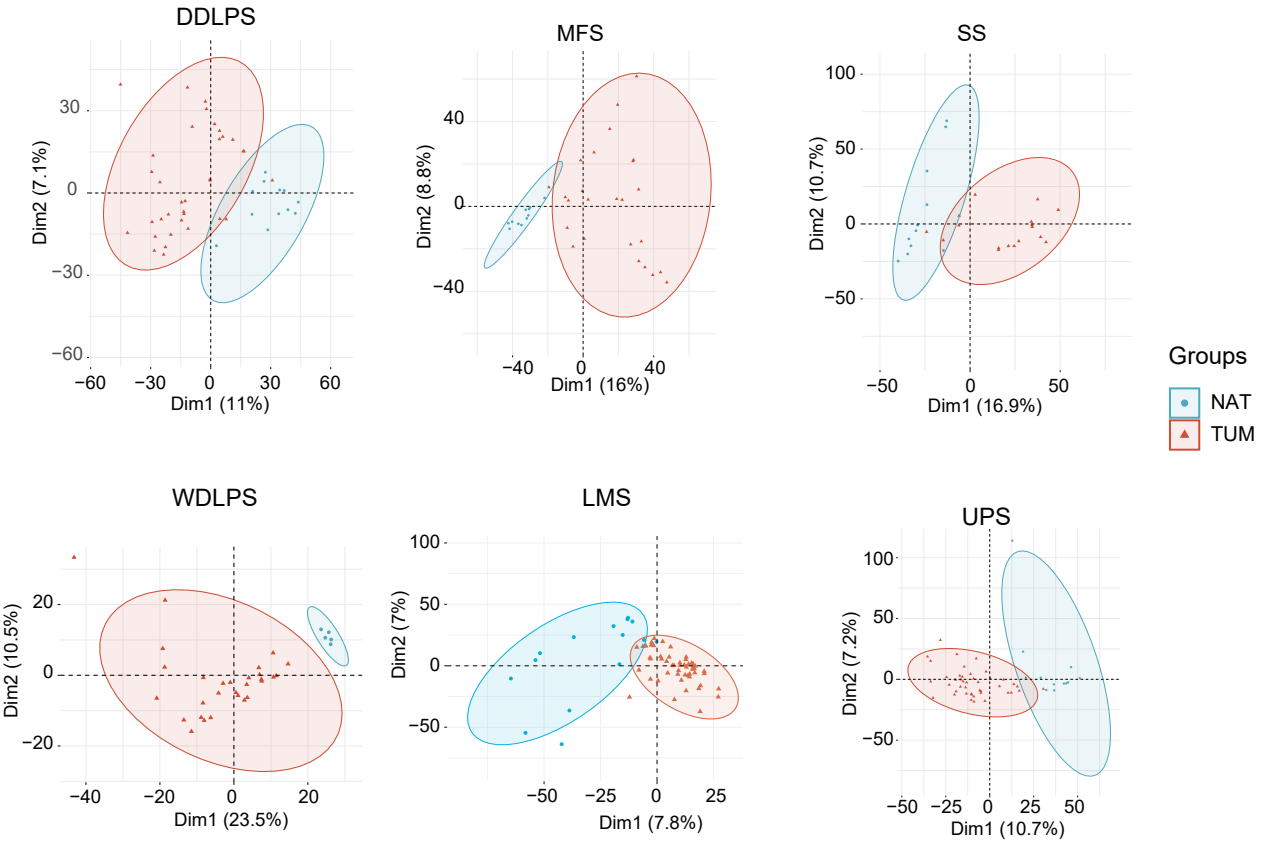
A



B

	Cluster 1	Cluster 2	Specificity
NAT	10	81	89%
T	154	118	56.60%
Purity	93.90%	59.30%	

C

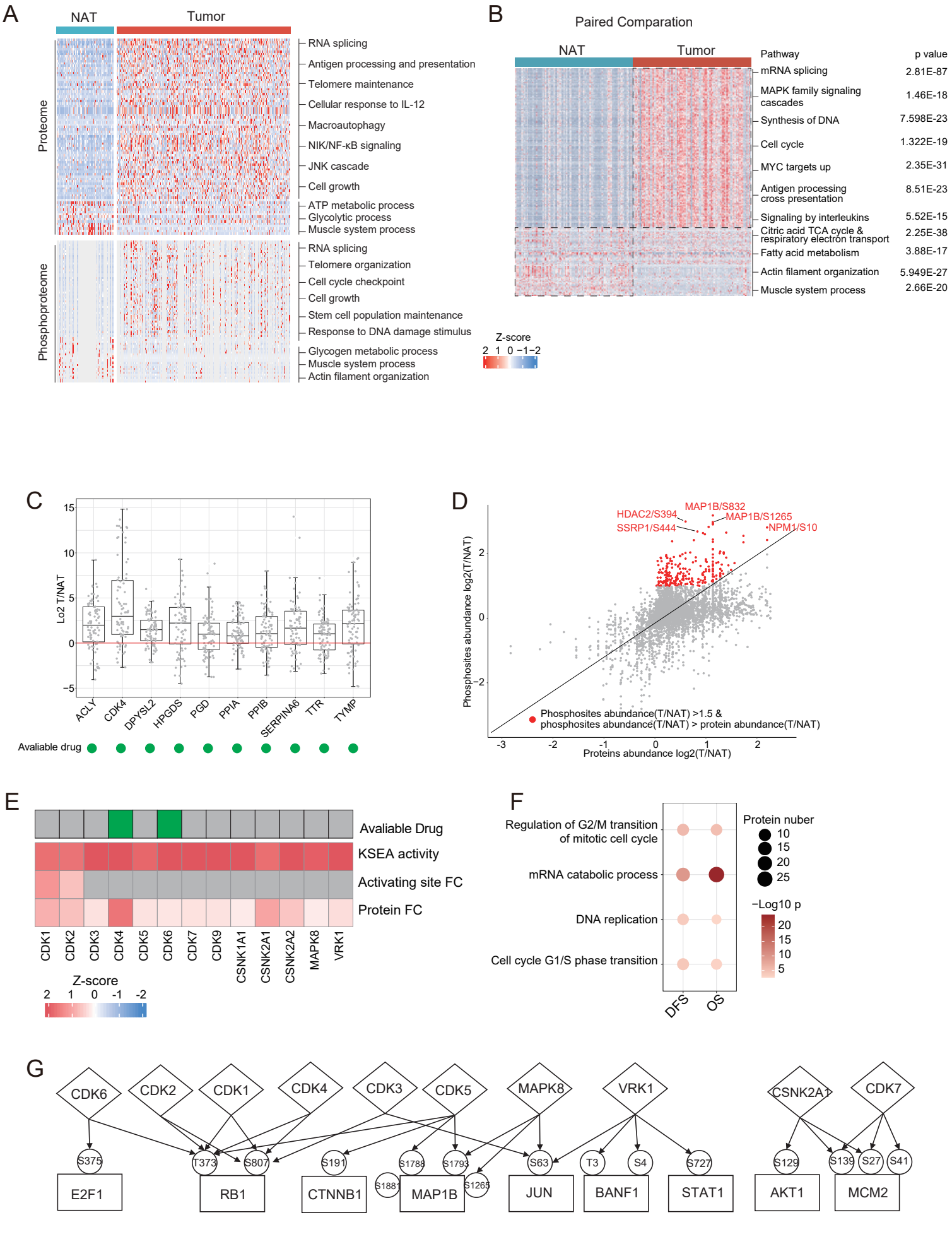


Supplementary figure 3. PCA analysis of NAT and tumor samples.

A. PCA for tumor and NAT samples based on proteomic data. The sample number for each group: NAT (n = 91) and Tumor (n = 272).
B. The table about unsupervised clustering results of NATs and tumor samples.
C. PCA plots illustrate separation levels between NAT and tumor samples in histological subtypes. The sample number for each group: DDLPS (35 tumors and 14 NATs), MFS (26 tumors and 12 NATs), SS (18 tumors and 11 NATs), WDLPS (36 tumors and 5 NATs), LMS (52 tumors and 15 NATs), and UPS (43 tumors and 12 NATs).

Source data are provided as a Source Data file.

Supplement Figure 4

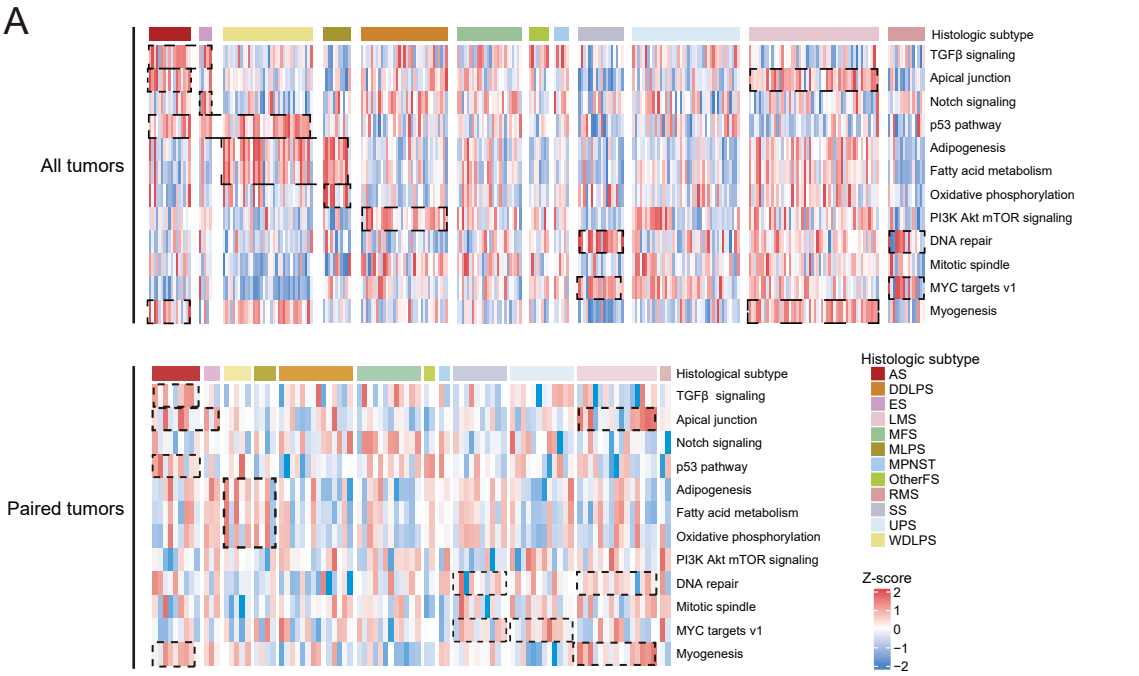


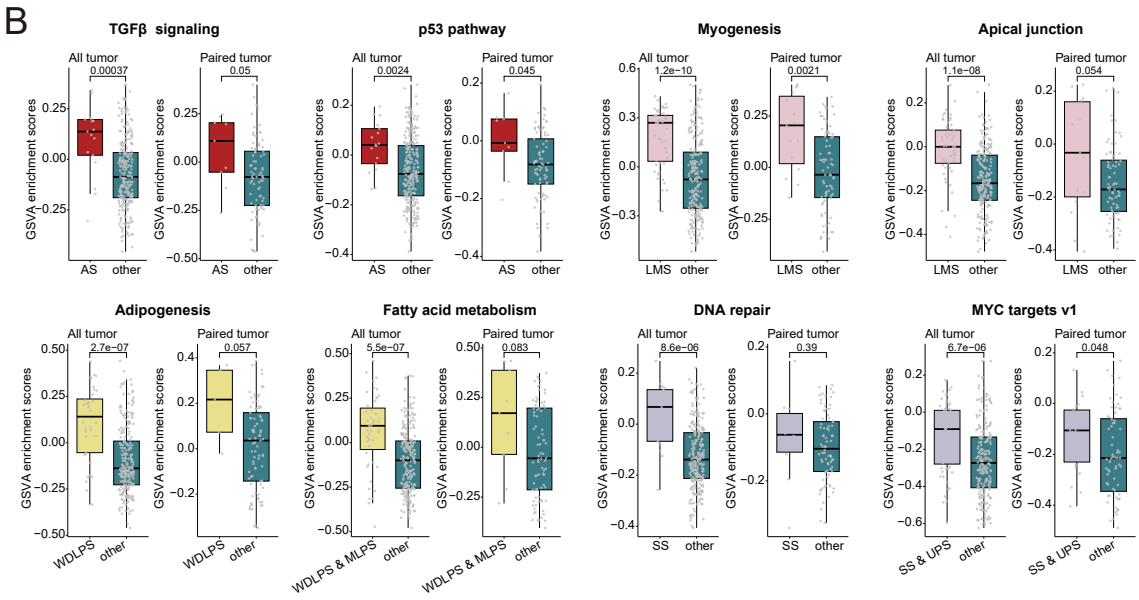
Supplementary figure 4. Proteomic features of STS compared with NATs

A. The heatmap presents the significantly differentially expressed proteins and phosphosites between tumors and NATs. The enriched pathways calculated by GO enrichment analysis are annotated in the right.
B. The heatmap presents the significant difference of enriched pathways between tumors and NATs through the pairwise comparison.
C. Boxplot presents the log2-transformed fold changes of tumors versus NATs of some proteins with available drugs. 91 samples with both tumor and NAT data are used for analysis. The middle bar represents the median and the box represents the interquartile range. Bars extend to 1.5× the interquartile range.
D. Comparison of abundance changes between phosphosites and their corresponding proteins.
E. Kinases with increased activity inferred from phosphorylation of its substrates (normalized enrichment score) or increased phosphorylation of its activating site.
F. The cell-cycle related pathways enriched with proteins or phosphoproteins associated with OS or DFS (Fisher's exact test).
G. Interaction network of kinases in figure E and some of their key substrates.

Source data are provided as a Source Data file.

Supplementary Figure 5



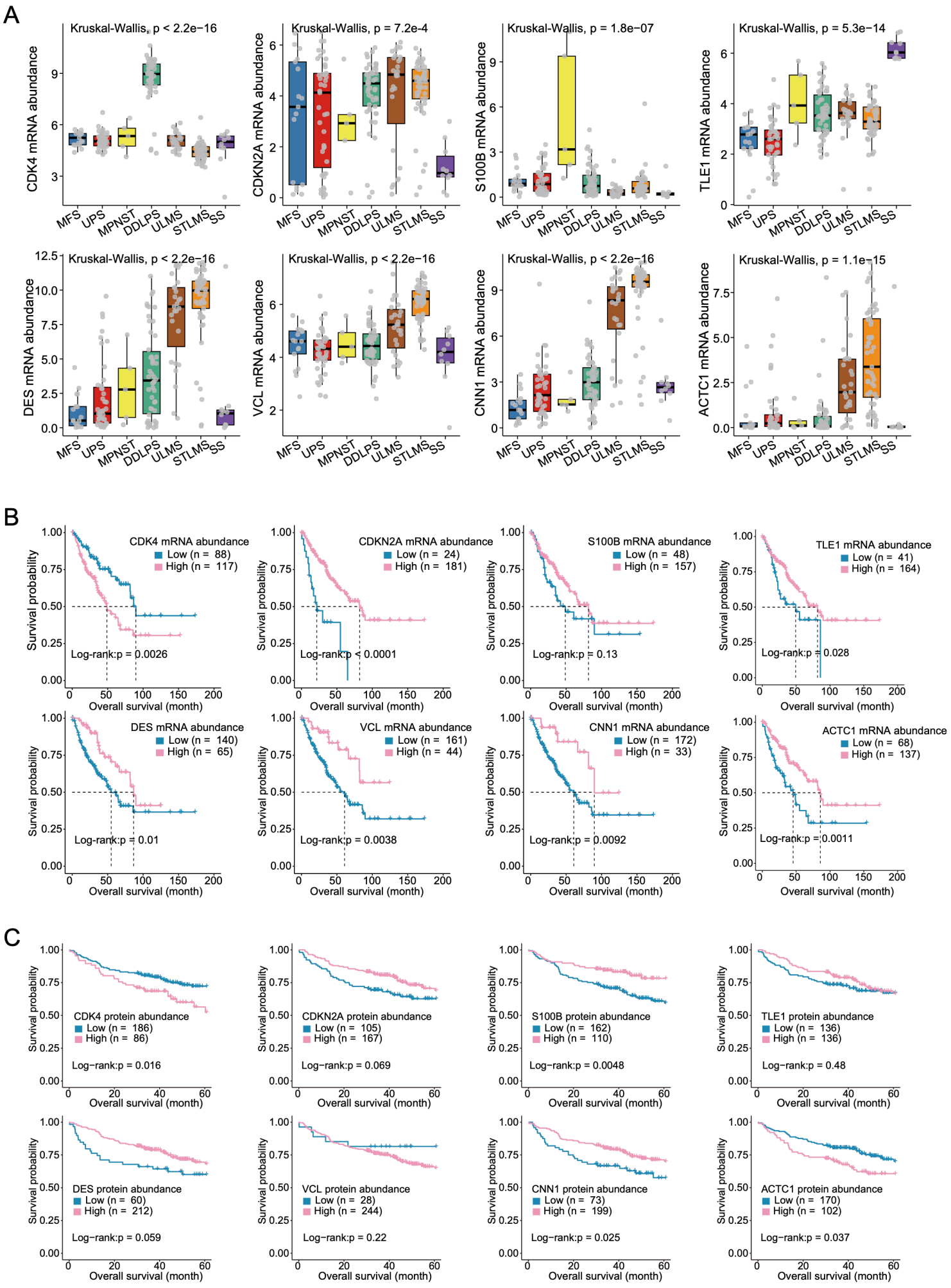


Supplementary figure 5. Proteomic features of different histological subtypes

A. Heatmaps illustrate enriched cancer hallmarks in STS histologic subtypes through non-pairwise (top) and pairwise (down) methods.
B. Boxplots presents the enriched pathways in specific histological subtypes processed through non-pairwise or pairwise method. The middle bar represents the median and the box represents the interquartile range. Bars extend to 1.5× the interquartile range. The two-sided student's t test is used for statistical comparison.

Source data are provided as a Source Data file.

Supplementary Figure 6



Supplementary figure 6. Survival analysis of some known diagnosis markers for STS pathologic subtypes

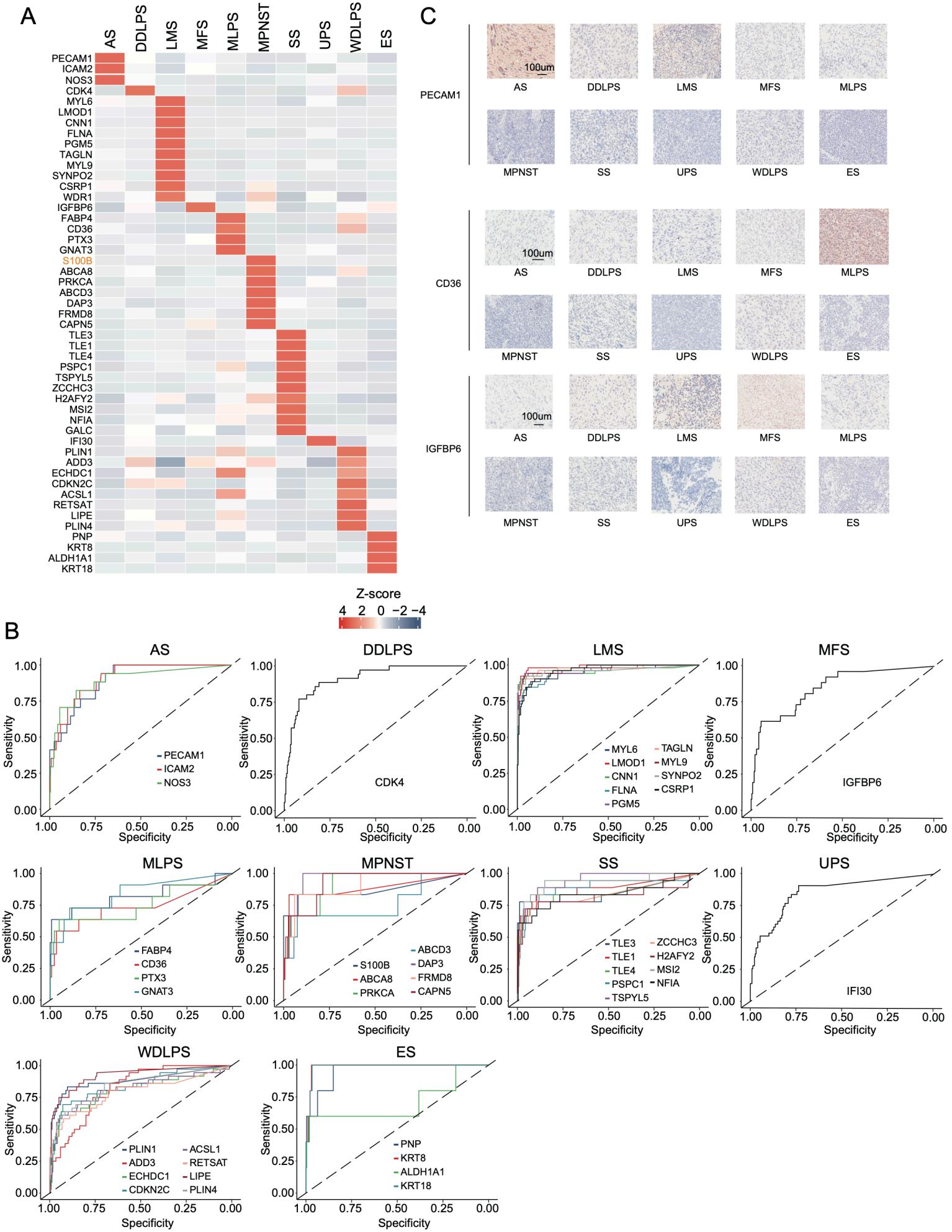
A. Boxplots presents transcriptional data of some known diagnosis markers for STS pathologic subtypes in the TCGA sarcoma cohort, including CDK4, CDKN2A, S100B, DES, VCL, CNN1. The sample number in each groups: DDLPS (n = 51), MFS (n = 17), MPNST (n = 5), SS (n = 10), STLMS (n = 53), ULMS (n = 27), and UPS (n = 44). The middle bar represents the median and the box represents the interquartile range. Bars extend to 1.5× the interquartile range. The Kruskal-Wallis's test is used for statistical analysis.

B. Kaplan-Meier curves for OS of tumor patients stratified by mRNA abundance of diagnosis markers. Data is from the TCGA sarcoma cohort (log-rank test).

C. Kaplan-Meier curves for OS of tumor patients stratified by protein abundance of diagnosis markers (log-rank test).

Source data are provided as a Source Data file.

Supplementary Figure 7



Supplementary figure 7. Candidates of prognostic markers

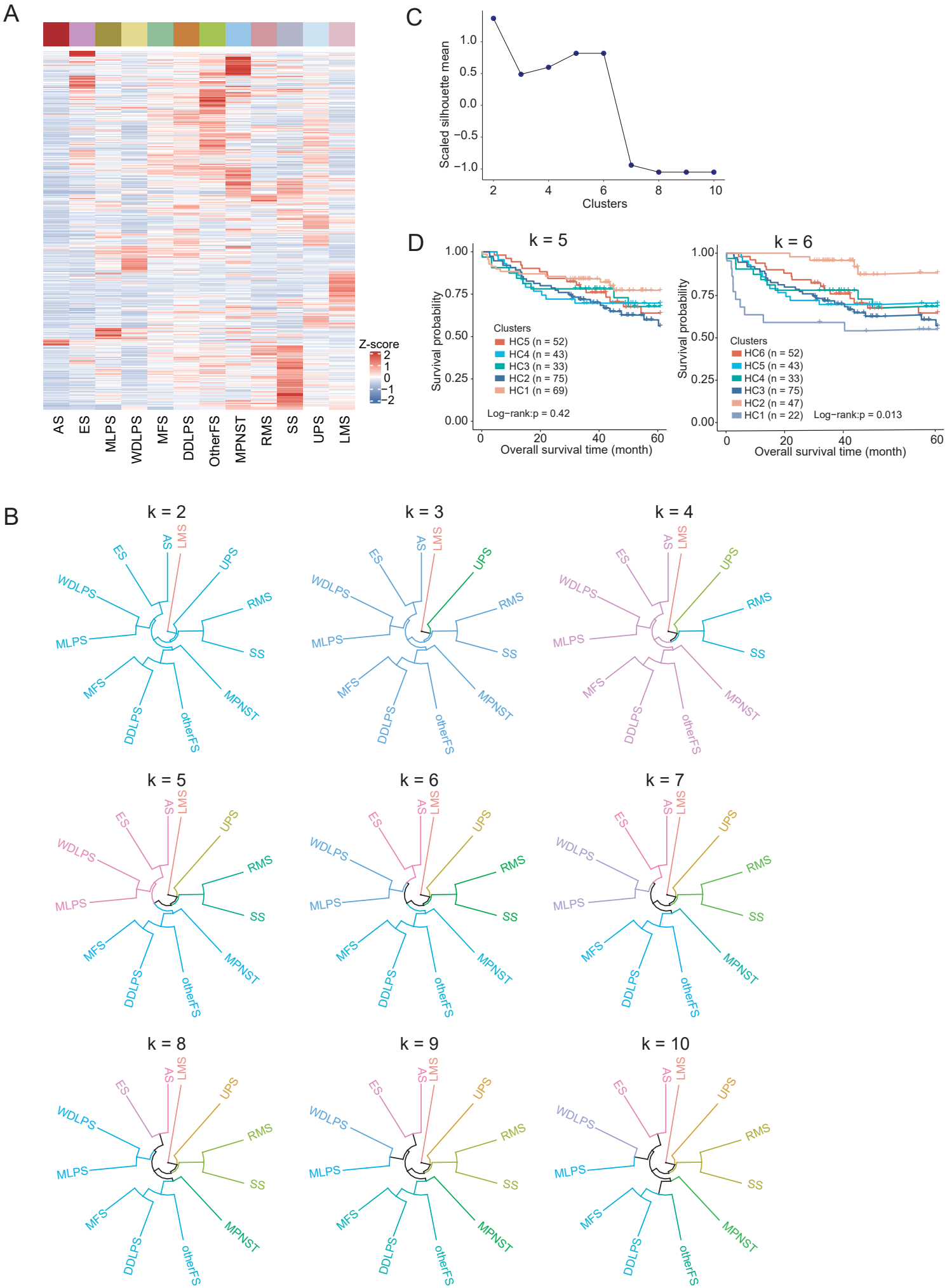
A. Heatmap illustrating the exclusive enrichment of potential prognosis markers in single STS histologic subtypes.

B. ROC curves showing the sensitivity and specificity of potential prognostic markers.

C. Immunohistochemistry of PECAM1, CDK36 and IGFBP6, Scale bar = 100 μm.

Source data are provided as a Source Data file.

Supplementary Figure 8

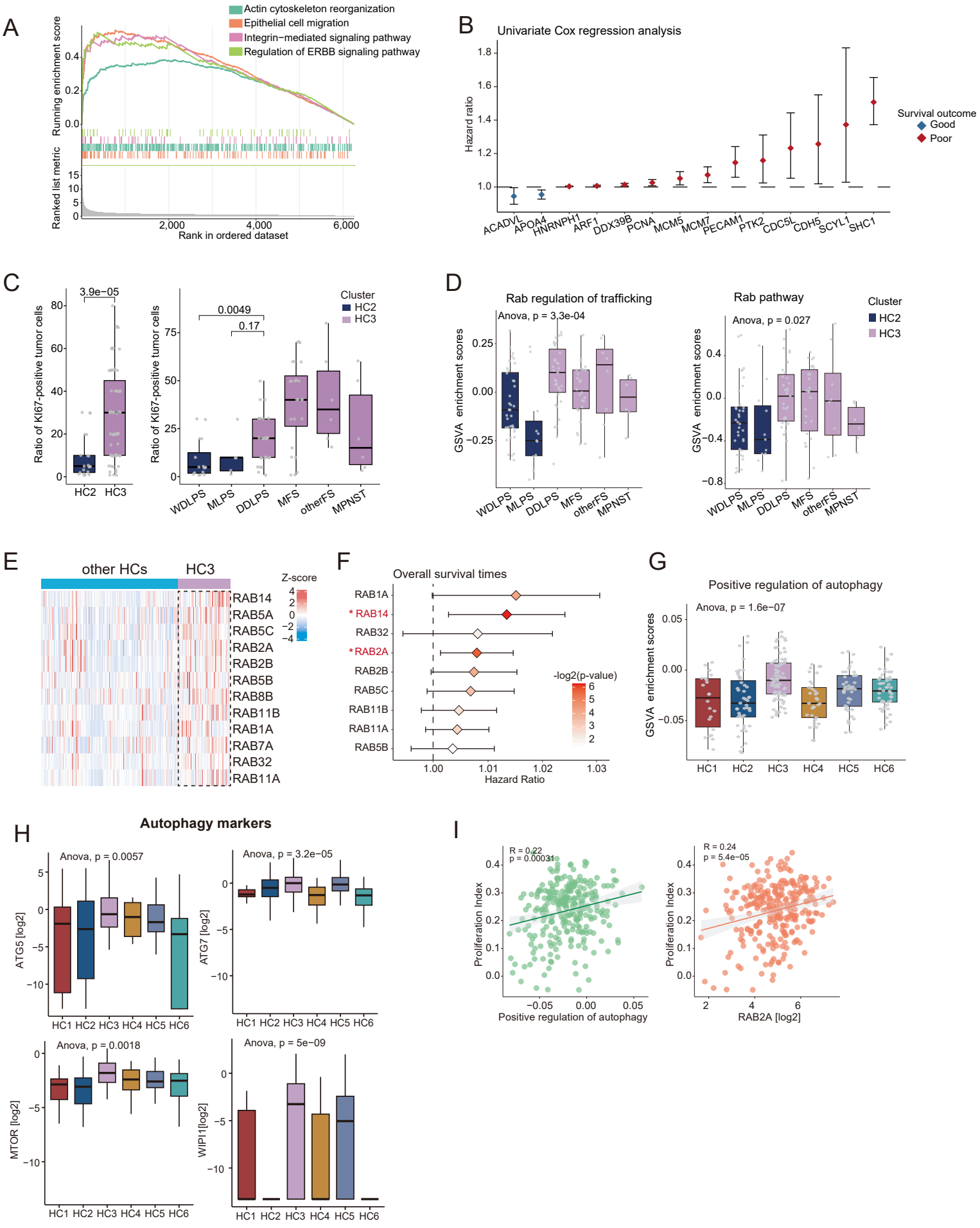


Supplementary figure 8. Processes of hierarchical clustering

A. Heatmap presents proteins used for hierarchical clusters.
B. The circled cluster dendrograms of sarcoma histological subtypes with cluster numbers from 2 to 10.
C. The scaled mean values of silhouette coefficients for different cluster numbers.
D. Kaplan-Meier curves for overall survival times when cluster number is 5 or 6. When cluster number is 6, different clusters have distinguished prognosis.

Source data are provided as a Source Data file.

Supplementary Figure 9

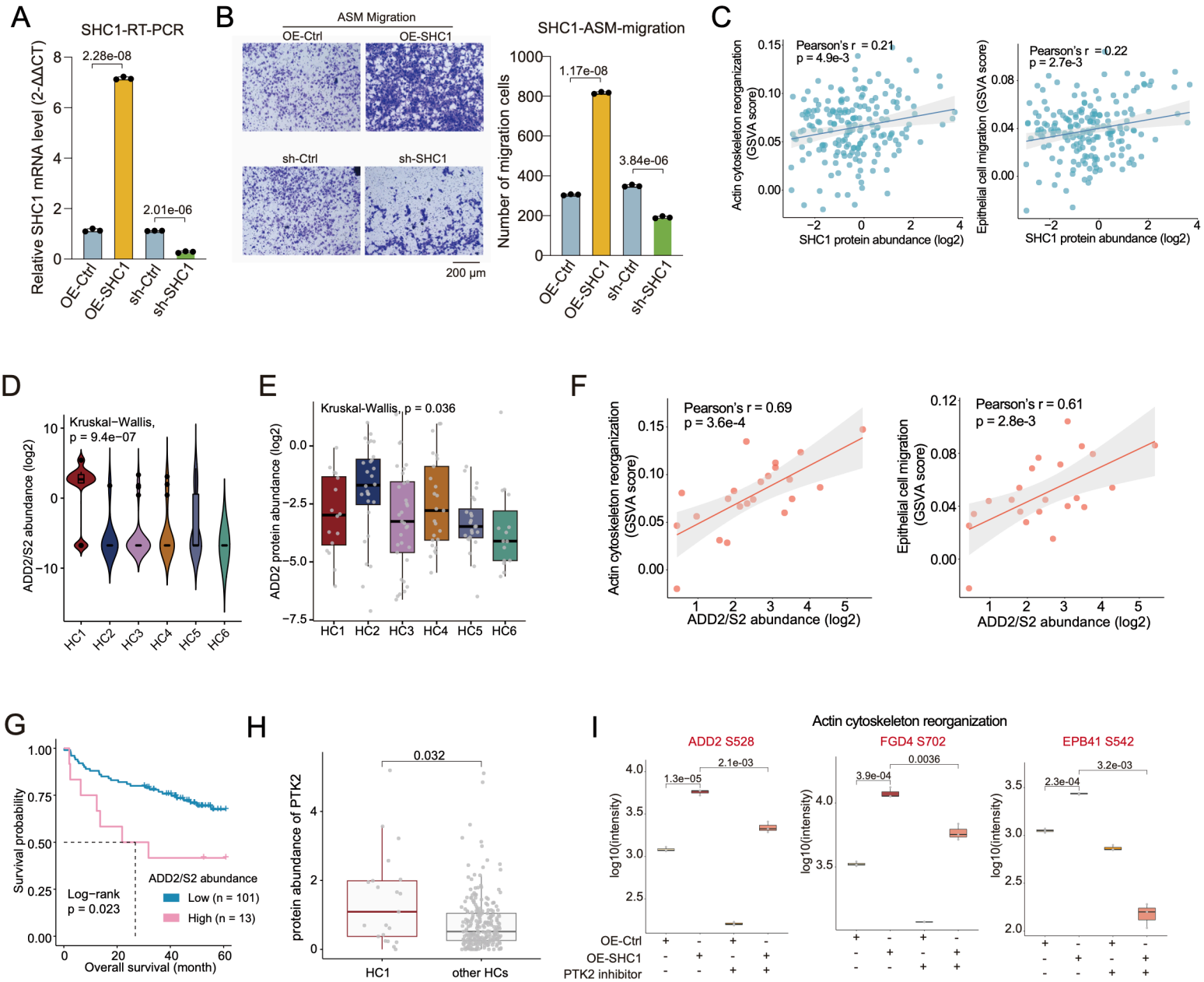


Supplementary figure 9. Characteristics of hierarchical clusters

A. GSEA plot for characteristic pathways of HC1 based on the rank of protein abundance.
B. Confidence intervals (95%) of hazard ratio coefficients for prognosis-related proteins based on multivariate Cox proportional hazard models.
C. Boxplots illustrates the ratio of KI67-positive tumor cells in HC2 and HC3 (left) and histological subtypes belonging to HC2/HC3 (right). The two-sided student's t test is used for statistical analysis.
D. Boxplots shows GSVA scores of Rab regulation of trafficking and Rab pathway in histological subtypes belonging to HC2/HC3. The ANOVA analysis is used for statistical analysis.
E. The heatmap presents Rab GTPases enriched in HC3.
F. The forest plot shows the hazard ratios of Rab GTPases enriched in HC3.
G. The boxplot presents the enrichment scores of autophagy in different hierarchical clusters.
H. Boxplots presents the abundances of autophagy markers in different hierarchical clusters. The ANOVA analysis is used for statistical analysis.
I. The scatter plot presents the positive correlation between proliferation index and autophagy pathway. 272 samples are used for the associated analysis. The Pearson's correlation is used for associated analysis. The error band represents the 95 confidence interval of the regression line.

For B and F, the dots represent the hazard ratio values and the bars represent the 95% confidence intervals. For C, D, G and H, the middle bar represents the median and the box represents the interquartile range. Bars extend to 1.5× the interquartile range. Source data are provided as a Source Data file.

Supplementary Figure 10



Supplementary figure 10. The mechanism of how SHC1 regulates cell migration

A. The expression of SHC1 in SHC1-OE-ASM, SHC1-KD-ASM (sh-SHC1), and the control group by RT-PCR (n = three biological repeats for each group). The two-sided student's t test is used for statistical analysis.

B. The effects of SHC1 on the migration of ASM cells were confirmed by the transwell assay. The bar plots (right panel) indicate counts of migrated ASM cells under different treatments. For each group, the migrated cells are counted three times in different views.

C. The scatter plot describes the correlation between the protein abundance of SHC1 (log2) and GSVA enrichment scores of "actin cytoskeleton reorganization" (left panel) or "positive regulation of epithelial cell migration" (right panel). 214 samples are used for the associated analysis.

D. The violin plot indicates the phosphorylation abundance of ADD2 Ser2 across six hierarchical clusters (Kruskal-Wallis's test).

E. The boxplot indicates the protein abundance of ADD across six hierarchical clusters (Kruskal-Wallis's test).

F. Scatter plots described the correlations between the phosphorylation abundance of ADD2 Ser2 and GSVA enrichment scores of "actin cytoskeleton reorganization" (left panel) or "positive regulation of epithelial cell migration" (right panel). 22 samples are used for associated analysis.

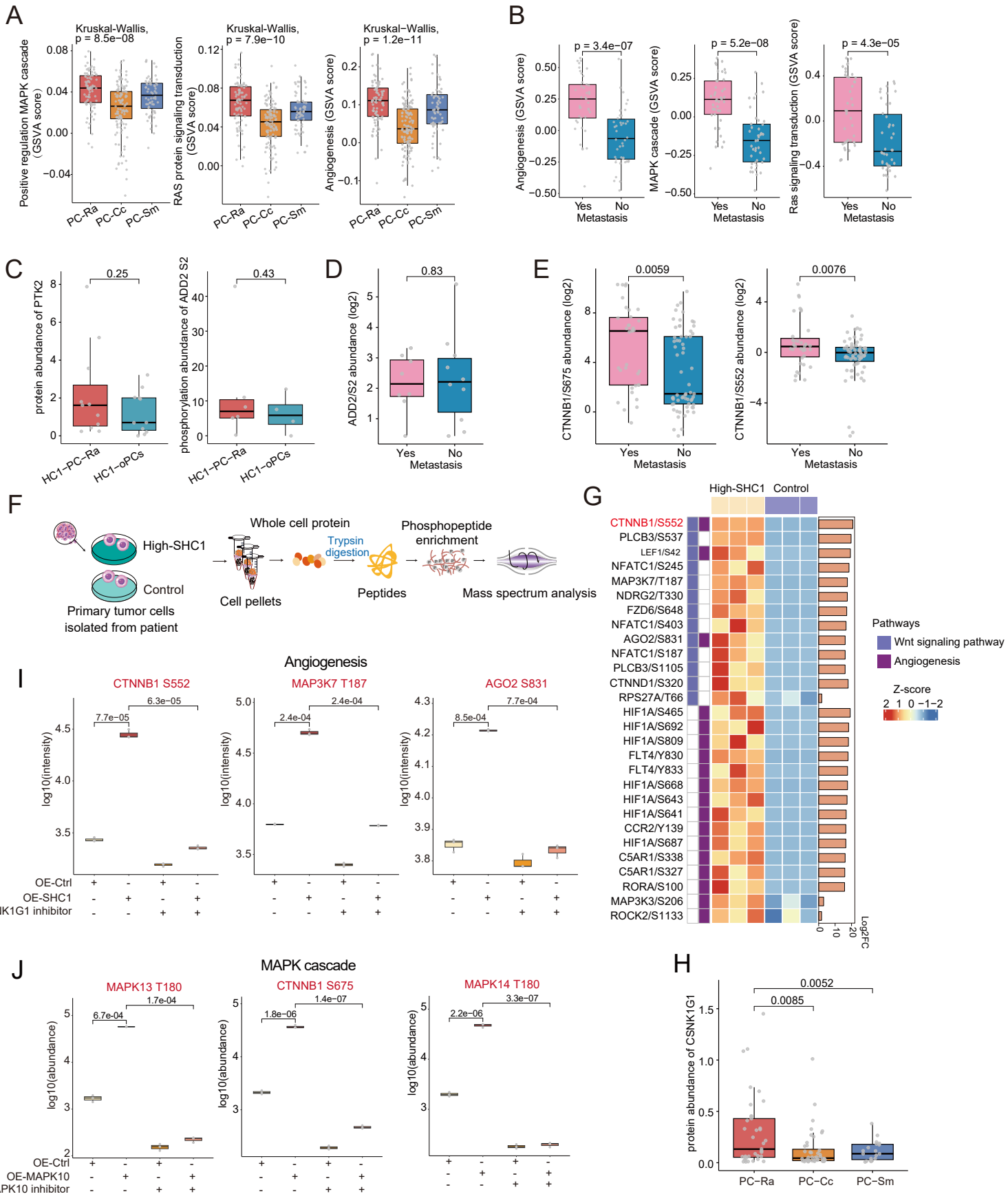
G. Kaplan-Meier curves for OS of patients stratified by phosphosites abundance of ADD2 Ser2 (low, n = 101; high, n = 13) (log-rank test).

H. The boxplot indicates the protein expression of PTK2 between HC1 and other HCs.

I. The boxplots indicate the phosphorylation intensity of phosphosites participating in actin cytoskeleton reorganization under different treatments

For A, B, H, and I, the two-sided student's t test is used for statistical analysis. For D and E, the Kruskal-Wallis's test is used for statistical analysis. For A and B, the data is presented as mean value +/- SE. For D, E, H, and I, the middle bar represents the median and the box represents the interquartile range. Bars extend to 1.5× the interquartile range. For C and F, the Pearson's correlation is used for statistical analysis and the error band represents the 95% confidence interval of the regression line. Source data are provided as a Source Data file.

Supplementary Figure 11



Supplementary figure 11. Molecular features of PC-Ra

A. Boxplots indicated GSVA scores of "positive regulation of MAPK cascade", "RAS protein signaling transduction", and "angiogenesis" across tumor samples stratified by proteomics clusters. The sample number for each group: PC-Ra (n = 86), PC-Cc (n = 122), and PC-Sm (n = 64). The Kruskal-Wallis's test is used for data analysis.

B. Boxplots indicated GSVA scores of the pathway, positive regulation of MAPK cascade, RAS protein signaling transduction, and angiogenesis among tumor samples with and without metastasis.

C. Boxplots illustrate the abundances of PTK2 and ADD2 S2 in HC1-PC-Ra and HC1-oPCs.

D. The boxplot shows the phosphosite abundance of ADD2 Ser2 in patients with and without metastasis (two-sided student's t test).

E. Boxplots shows the phosphosite abundance of CTNNB1 Ser675 (left) or Ser552 (right) in patients with and without metastasis (two-sided student's t test).

F. Flow chart showing IP-MS steps of primary patient-derived cancer cells (PDCs) isolated from patients with low (control) or high SHC1 abundance. The patients with low or high SHC1 abundance are detected by SHC1 abundance of proteomic data.

G. The heatmap presents the enrichment of phosphosites related to Wnt signaling pathway and angiogenesis in high-SHC1 PDCs. Fold changes of protein abundance are displayed on right side.

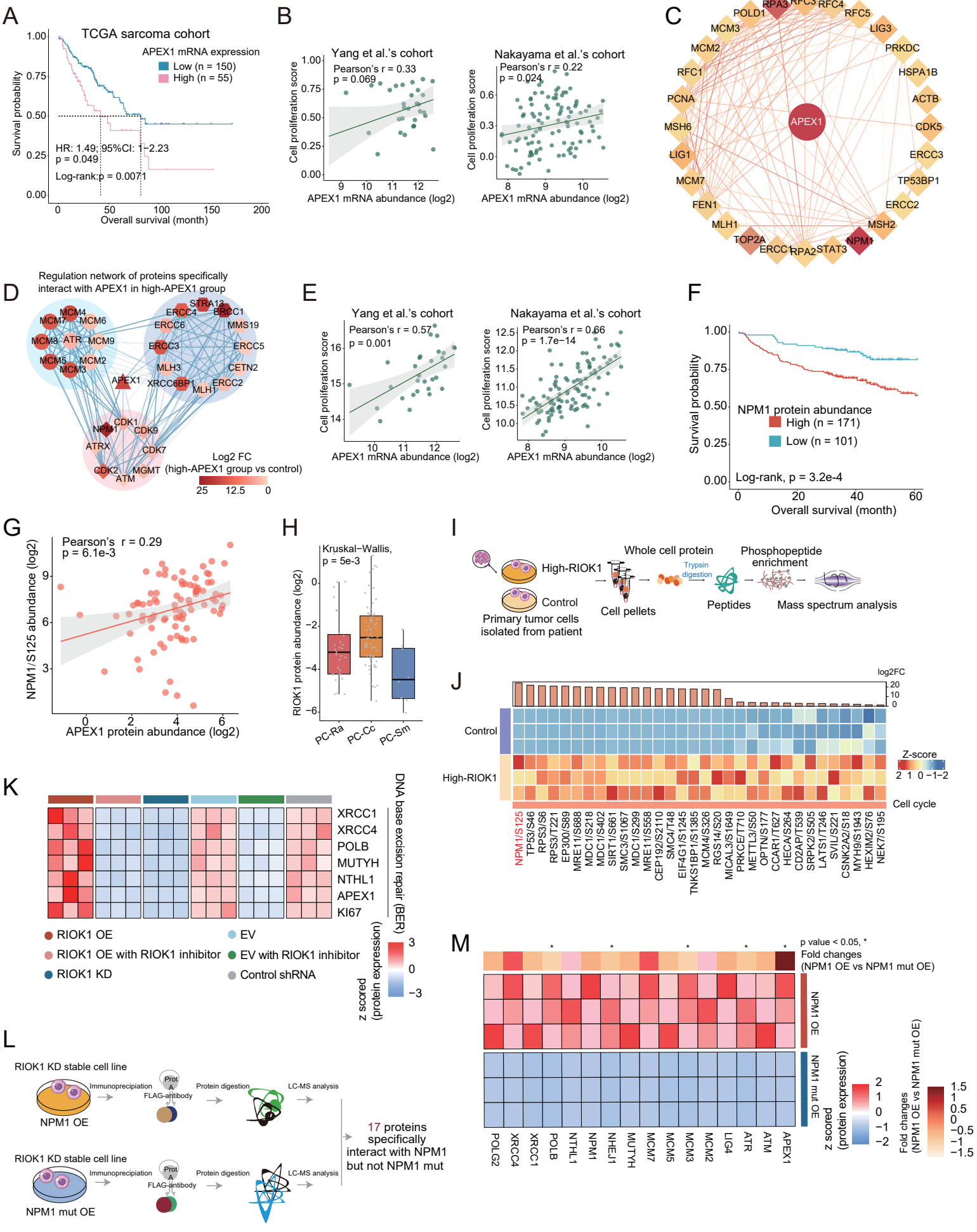
H. The boxplot indicates the expression of CSNK1G1 in different proteomic clusters.

I. The boxplots indicate the phosphorylation levels of CTNNB1 Ser552 and other phosphosites participating in angiogenesis under different treatments.

J. The boxplots indicated the phosphorylation levels of CTNNB1 Ser675 and other phosphosites participating in MAPK signaling cascade under different treatments.

For A, B, C, D, E, H, I, and J, the middle bar represents the median and the box represents the interquartile range. Bars extend to 1.5× the interquartile range. For B, C, D, E, H, I, and J, the two-sided student's t test is used for statistical analysis. For I and J, three biological repeats are used for each group. Source data are provided as a Source Data file.

Supplementary Figure 12

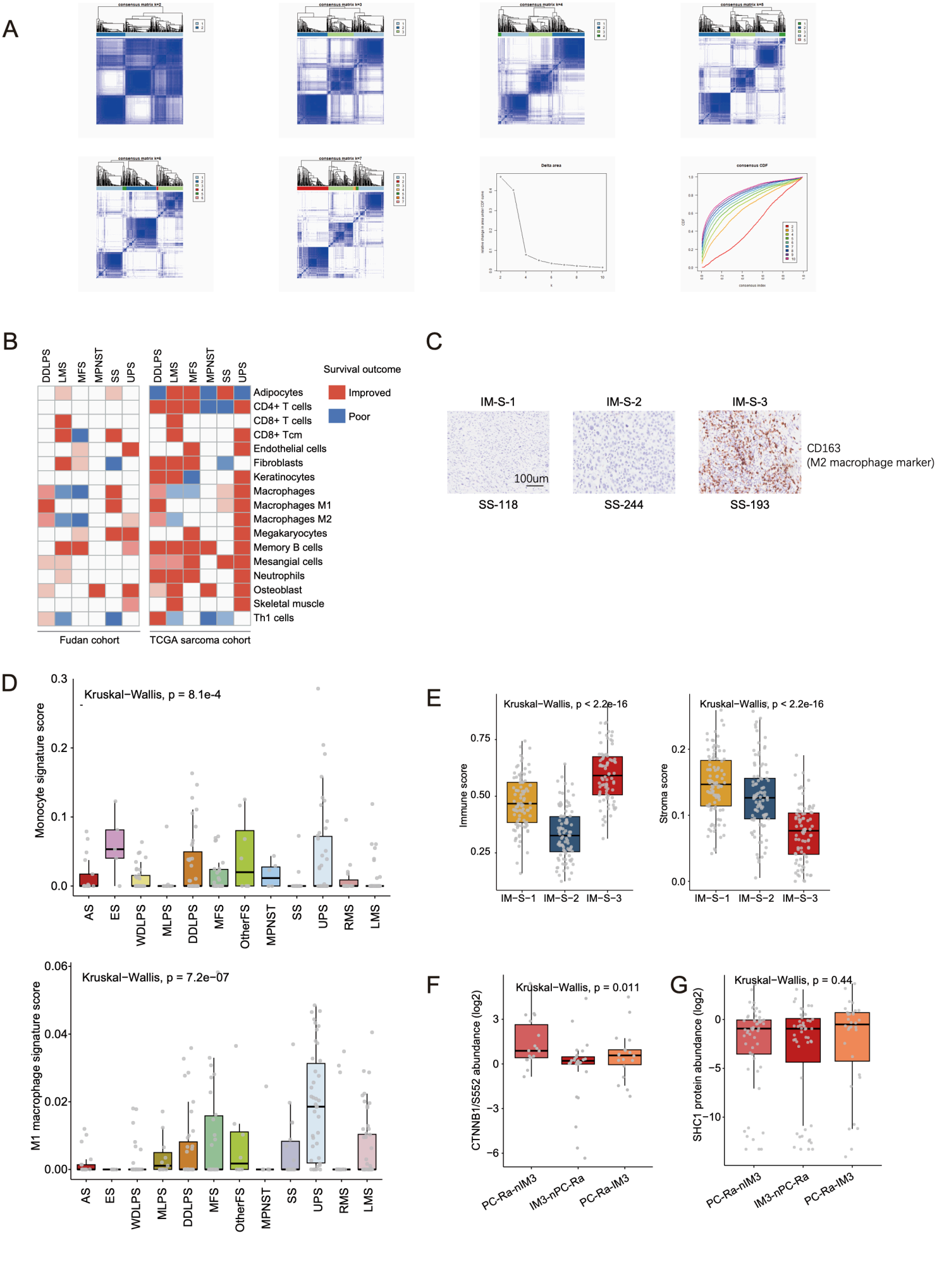


Supplementary figure 12. Molecular features of PC-Cc

A. Kaplan-Meier curves for OS of tumor samples stratified by protein abundance level of APEX1 from TCGA cohort (log-rank test and Cox proportional hazards model).
B. Scatterplot indicates the correlation between log2-transformed protein abundance of APEX1 and cell proliferation score in two public independent sarcoma cohorts.
C. The interaction network of APEX1 and related proteins. The interaction relationships are gotten from STING database and the protein nodes are colored by fold changes of PC-Cc versus other PCs.
D. Regulation network of APEX1 in high-APEX1 PDCs group. Protein nodes were colored by fold change values (high APEX1 group vs control). The relationships represented by blue lines were obtained from STING database.
E. Scatter plots indicated the correlation between log2-transformed protein abundance of NPM1 and cell proliferation score in two public independent sarcoma cohorts.
F. Kaplan-Meier curves for OS of samples stratified by NPM1 protein abundance levels (log-rank test).
G. The scatter plot indicated the correlation between abundance of APEX1 and NPM1 Ser125.
H. The boxplot indicates the RIOK1 protein abundance (lo2-transformed) across tumor samples stratified by proteomic clusters. The middle bar represents the median and the box represents the interquartile range. Bars extend to 1.5× the interquartile range. The Kruskal-Wallis' test is used for statistical analysis.
I. Flow chart showing IP-MS steps of PDCs isolated from patients with low (control) or high RIOK1 abundance. The patients with low or high RIOK1 abundance are detected by RIOK1 abundance of proteomic data.
J. The heatmap illustrates the most enriched phosphosites in high-RIOK1 PDCs (n = 3 biological repeats per group).
K. The heatmap reveals the expression patterns of DNA base excision proteins across the cells associated with various treatment (n = 3 biological repeats per group).
L. The schematic work flow of the IP-MS experiment for the NPM1.
M. The heatmap reveals the expression patterns of DNA base excision proteins across the NPM1-OE-RIOK1-KD-RKN, NPM1-mut-OE-RIOK1-KD-RKN (n = 3 biological repeats per group). The symbol, '*', represents that the p value < 0.05.

For B, E, and G, the Pearson's correlation is used for associated analysis. The error band represents the 95% confidence interval of the regression line. Source data are provided as a Source Data file.

Supplementary Figure 13



Supplementary figure 13. Immune signatures of histologic subtypes and immune clusters

A. Identification of immune clusters (K = 2–6) based on proteomic data (n = 272) by consensus clustering upon their abundance (**Methods**). Consensus matrix, consensus cumulative distribution function (CDF) plot, delta area (change in CDF area) plot, and tracking plot plots are shown.

B. Significant OS associations ($p < 0.05$) for high immune score by histologic subtypes in our cohort and TCGA sarcoma cohort. Log-rank test is used for survival analysis.

C. Immunohistochemistry of CD163, Scale bar = 100 μm

D. Boxplots indicate different infiltration scores of monocytes (top panel) and M1 macrophage (bottom panel) in STS histologic subtypes.

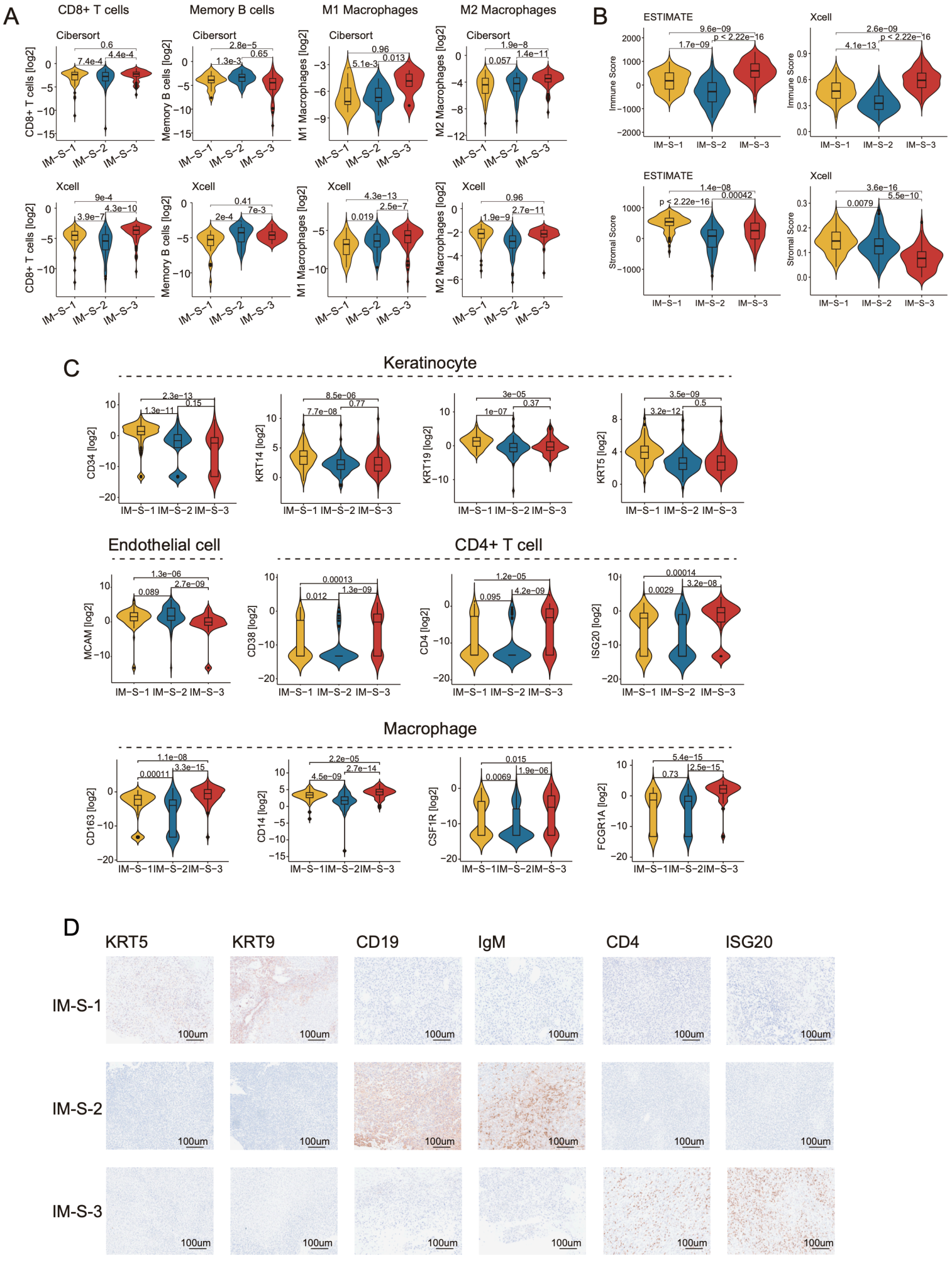
E. Boxplots indicate different immune and stroma scores in three immune clusters.

F. The boxplot indicates CTNNB1 Ser552 abundance (log2-transformed) across PC-Ra-nIM3, im3-nPC-Ra, and PC-Ra-IM3.

G. The boxplot indicates SHC1 protein abundance (log2-transformed) across PC-Ra-nIM3, im3-nPC-Ra, and PC-Ra-IM3.

For D, E, and F, the middle bar represents the median and the box represents the interquartile range. Bars extend to 1.5× the interquartile range. The Kruskal-Wallis's test is used for statistical analysis. Source data are provided as a Source Data file.

Supplementary Figure 14

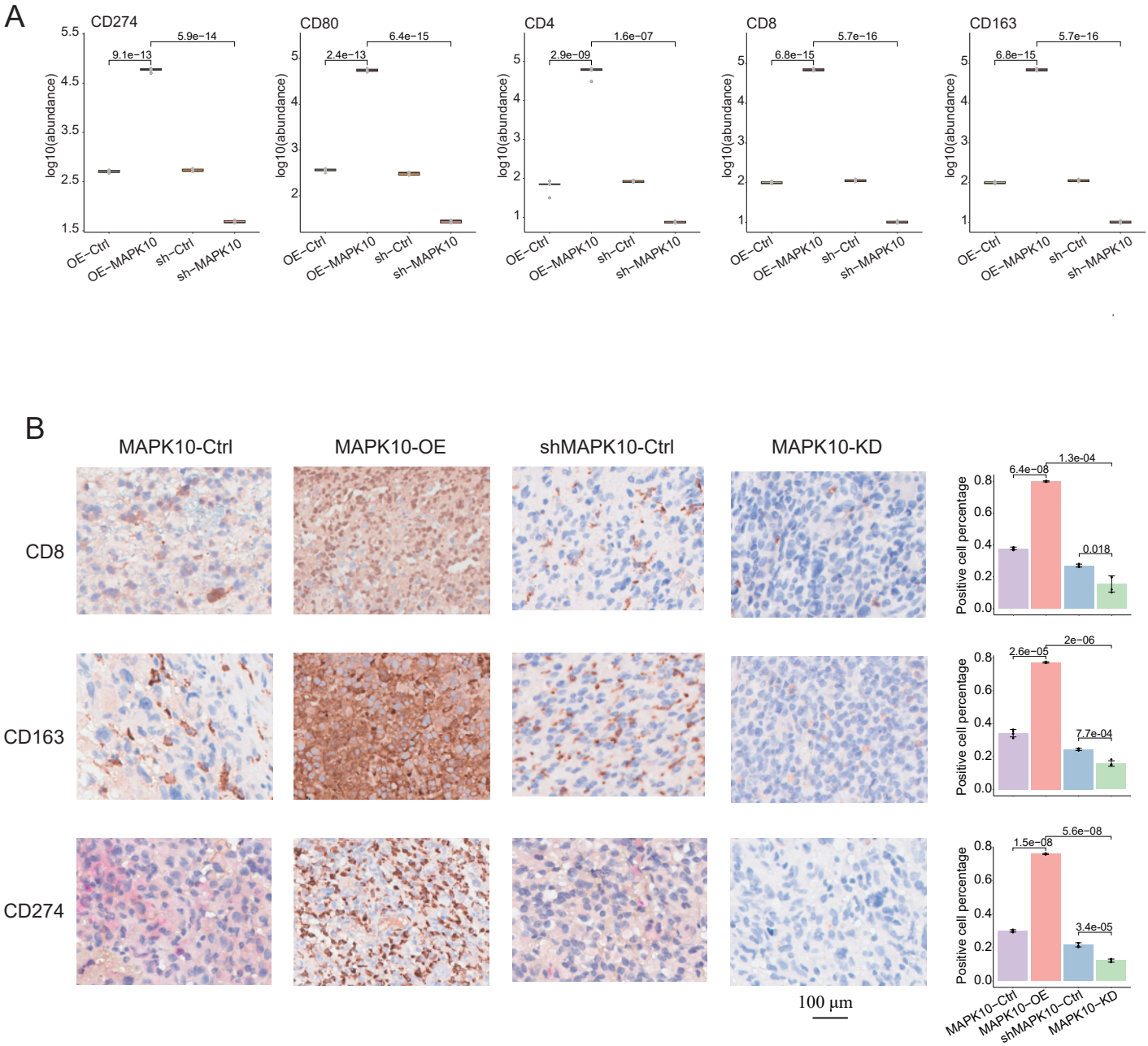


Supplementary figure 14. The immune cell signatures and cell markers in different immune clusters

A. Boxplots illustrate cell signature scores inferred by xCell and CIBERSORT algorithm among three immune clusters.
B. Boxplots illustrate the immune scores and stromal scores calculated by xCell and ESTIMATE algorithm among the three immune clusters.
C. Boxplots illustrate proteomic abundance of immune cell markers in immune clusters.
D. IHC staining presents expressions of immune and stromal cell markers in three immune clusters.

For A, B, and C, the middle bar represents the median and the box represents the interquartile range. Bars extend to 1.5× the interquartile range. The violin area represents the distribution density of the samples. The two-sided student's t test is used for statistical analysis. Source data are provided as a Source Data file.

Supplementary Figure 15



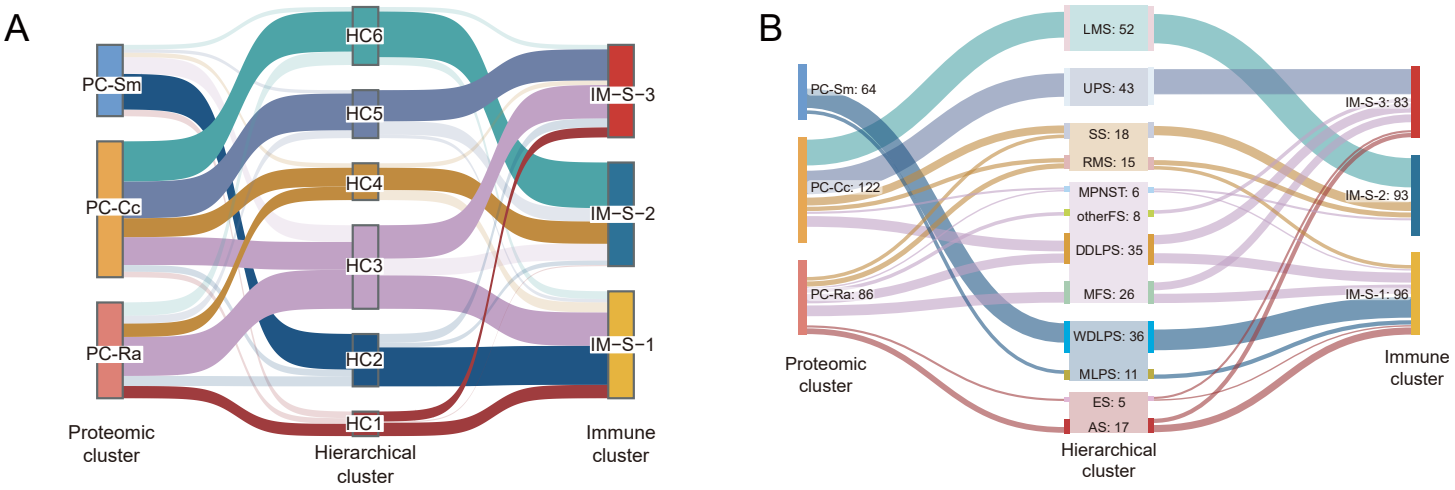
Supplementary figure 15. The impact of MAPK10 on immune infiltration in mouse xenograft model

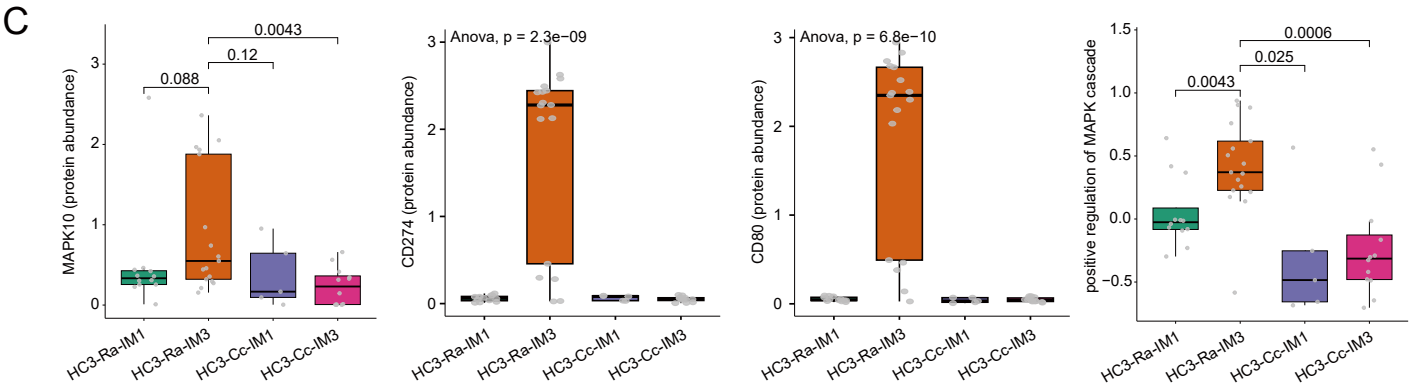
A. Boxplots illustrate the expressions of immune cell markers, including CD274, CD80, CD4, and CD8 in differently treated mouse xenograft models. Each group has three biological repeats. The middle bar represents the median and the box represents the interquartile range. Bars extend to 1.5× the interquartile range. The student's t test is used for statistical analysis.

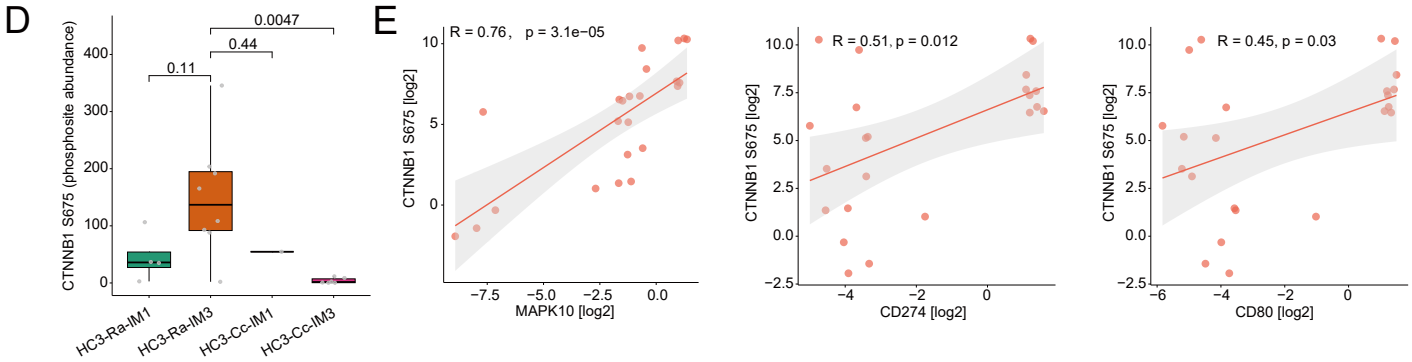
B. IHC images illustrate the expression of CD8, CD163, and CD274 in subcutaneous tumors of the C57/BL6J mice transplanted with SW872 sarcoma cell lines. Positive cell percentage is presented on the right. Each group has three biological repeats. The data is presented as mean value +/- SE. The two-sided student's t test is used for statistical analysis.

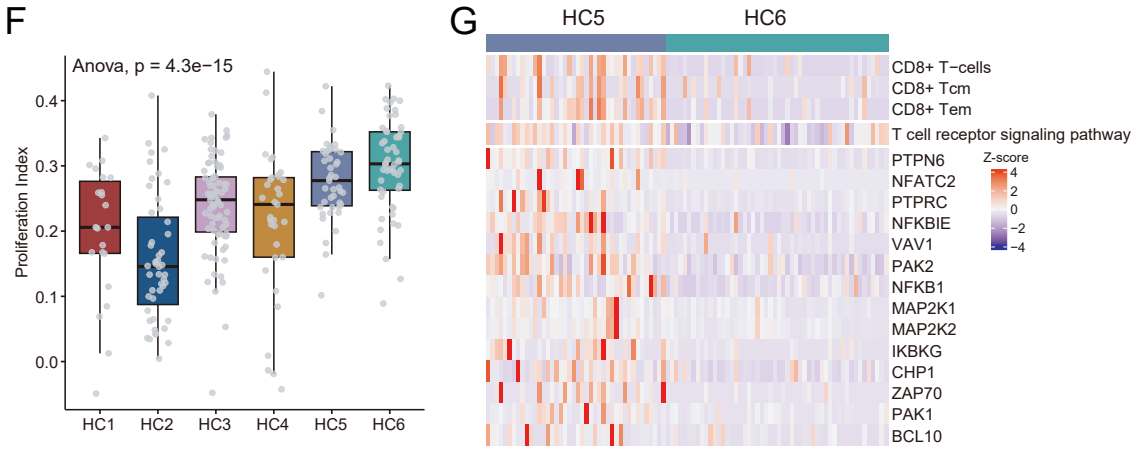
Source data are provided as a Source Data file.

Supplementary Figure 16









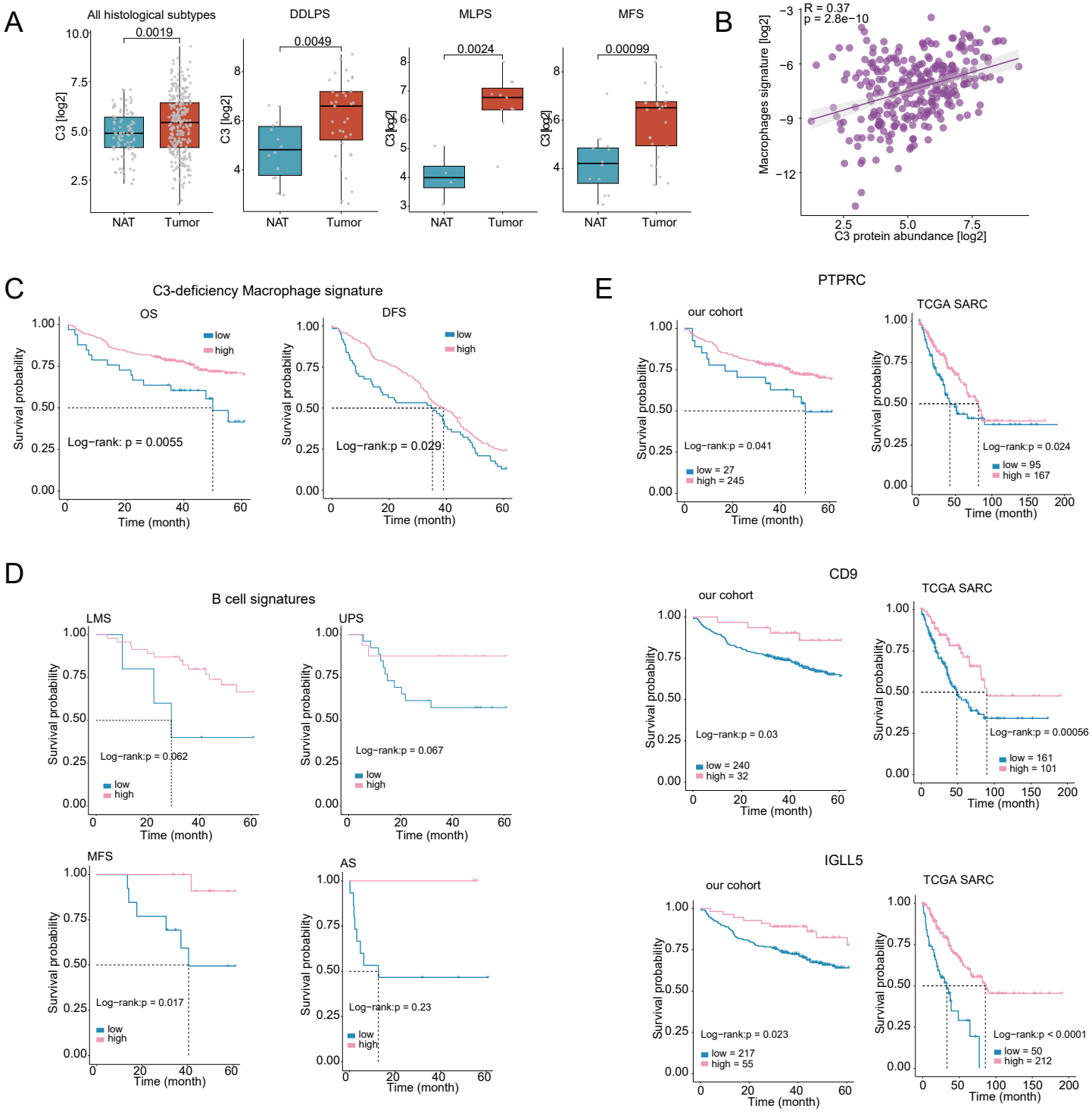
Supplementary figure 16. Integration analysis of clustering result from different levels

A. Sankey plot illustrating relationships between hierarchical clusters and proteomic clusters or immune clusters.
B. Sankey plot illustrating relationships between sarcoma histology subtypes and proteomic clusters or immune clusters.
C. Boxplots illustrate enrichment of protein abundance of MAPK10, the phosphorylation level of CTNNB1 Ser675, and the signaling pathway, positive regulation of MAPK cascade, in the HC3-Ra-IM3. The ANOVA and two-sided student's t test are used for statistical analysis.
D. The boxplot illustrates enrichment of CTNNB1 Ser675 in HC3-Ra-IM3. The two-sided student's t test is used for statistical analysis.
E. Scatter plots present positive correlations between phosphorylation level of CTNNB1 Ser675 and protein abundance of MAPK10, CD274, or CD80 in HC3 group. 23 samples are filtered and the Pearson's correlation is utilized for associated analysis. The error band represents the 95% confidence interval of the regression line.
F. The boxplot presents proliferation index in different hierarchical clusters. The ANOVA test is used for statistical analysis. The ANOVA analysis is used for statistical comparison.
G. The heatmap presents the enrichment of $CD8^+$ T cells, T cell receptor signaling pathway, and related proteins in HC5.

For C, D, and F, the middle bar represents the median and the box represents the interquartile range. Bars extend to 1.5× the interquartile range.

Source data are provided as a Source Data file.

Supplementary Figure 17



Supplementary figure 17. Integration analysis of clustering result from different levels

A. Boxplots illustrate the proteomic expression of C3 in NATs and tumors. Two-sided student's t test is used for data analysis.
B. The scatter plot presents a positive correlation between C3 protein abundance and macrophage signatures in pan-sarcoma. Pearson's correlation is used for associated analysis. The error band represents the 95% confidence interval of the regression line.
C. Kaplan-Meier curves for OS and DFS stratified by levels of C3-deficiency macrophage signatures in pan-sarcoma.
D. Kaplan-Meier curve for OS stratified by B cell signatures in LMS, UPS, MFS, and AS.
E. Kaplan-Meier curve for OS stratified by B cell markers. top: our cohort, stratified by proteomic abundance; bottom: TCGA SARC cohort, stratified by mRNA expression.

For C, D, and E, the log-rank test is utilized for survival analysis. Source data are provided as a Source Data file.