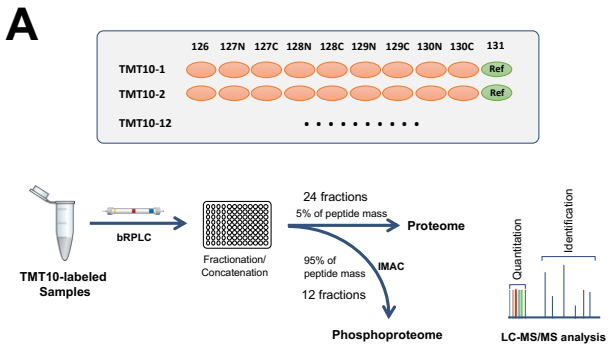
Cell Reports Medicine, Volume *1*

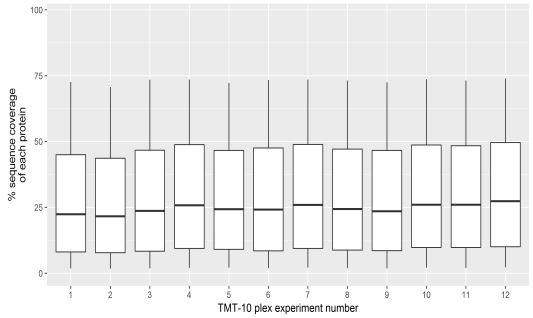
Supplemental Information

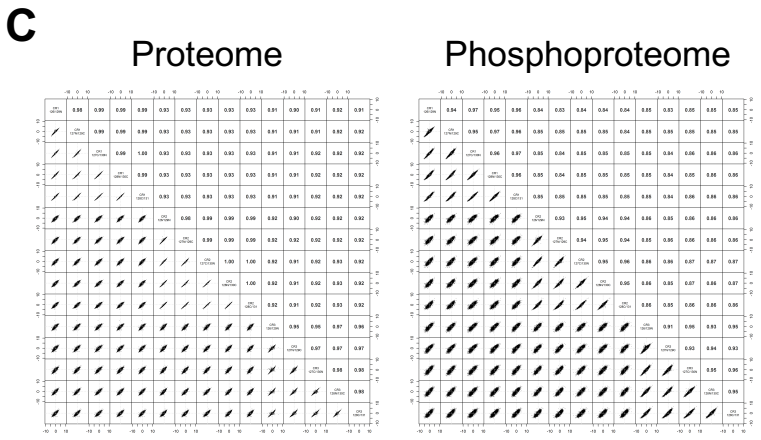
Proteogenomic Characterization of Ovarian HGSC Implicates Mitotic Kinases, Replication Stress in Observed Chromosomal Instability

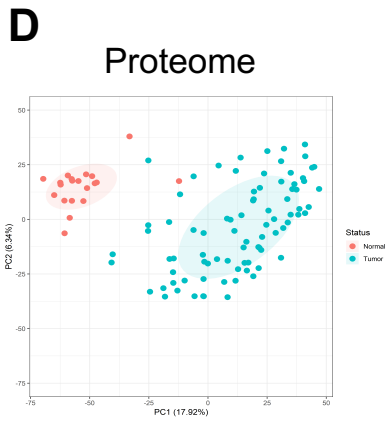
Jason E. McDermott, Osama A. Arshad, Vladislav A. Petyuk, Yi Fu, Marina A. Gritsenko, Therese R. Clauss, Ronald J. Moore, Athena A. Schepmoes, Rui Zhao, Matthew E. Monroe, Michael Schnaubelt, Chia-Feng Tsai, Samuel H. Payne, Chen Huang, Liang-Bo Wang, Steven Foltz, Matthew Wyczalkowski, Yige Wu, Ehwang Song, Molly A. Brewer, Mathangi Thiagarajan, Christopher R. Kinsinger, Ana I. Robles, Emily S. Boja, Henry Rodriguez, Daniel W. Chan, Bing Zhang, Zhen Zhang, Li Ding, Richard D. Smith, Tao Liu, Karin D. Rodland, and Clinical Proteomic Tumor Analysis Consortium

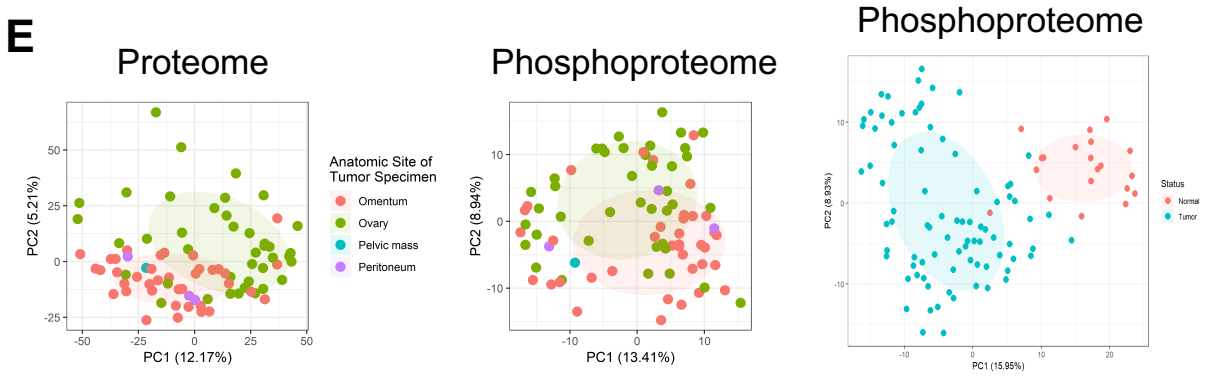
Supplemental Figures











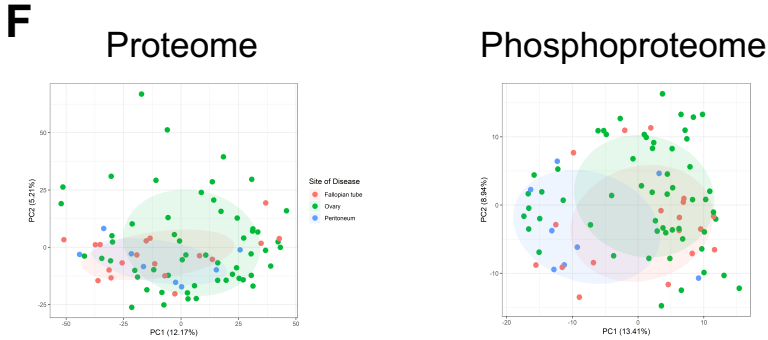


Figure S1. Integrated proteomics workflow and longitudinal mass-spectrometry data generation quality control. Related to STAR Methods and Figures 4 and 5. **(A)** Mass spectrometry (MS)-based TMT-10 based global proteome and phosphoproteome analysis workflow. The ovarian tumors and normal tissues were analyzed in twelve TMT 10-plex experiments, each with nine study samples and a common internal reference sample created by pooling all study samples (equal contribution). The TMT10-labeled samples were then fractionated, split (with 5% peptide mass analyzed directly for global proteome), and subjected to enrichment of phosphopeptides. Peptides were detected and quantified using information from the TMT-10 MS/MS spectra. **(B)** Distribution of sequence coverage of the identified proteins with tryptic peptides detected by MS/MS in each TMT-10 plex; whiskers show the 5–95 percentiles. **(C)** Robust and precise proteomics platforms. Longitudinal performance was tested by repeated proteome (left) and phosphoproteome (right) analysis of aliquots of the same patient-derived xenograft QC samples in standalone TMT-10 plexes, along with the ovarian cancer study samples; scatter plots and Pearson correlations comparing individual replicate measurements are shown. **(D)** PCA analysis of the proteome (left) and phosphoproteome (right) data readily separated the tumors from the normal FT tissue samples. (**E**) PCA analysis of the tumor-only proteome (left) and tumor-only phosphoproteome (right) data separate the site of the tumor sample for the proteome, but not well for the phosphoproteome. (**F**) PCA analysis of the tumor-only proteome (left) and tumor-only phosphoproteome (right) show some separation based on the site of disease.

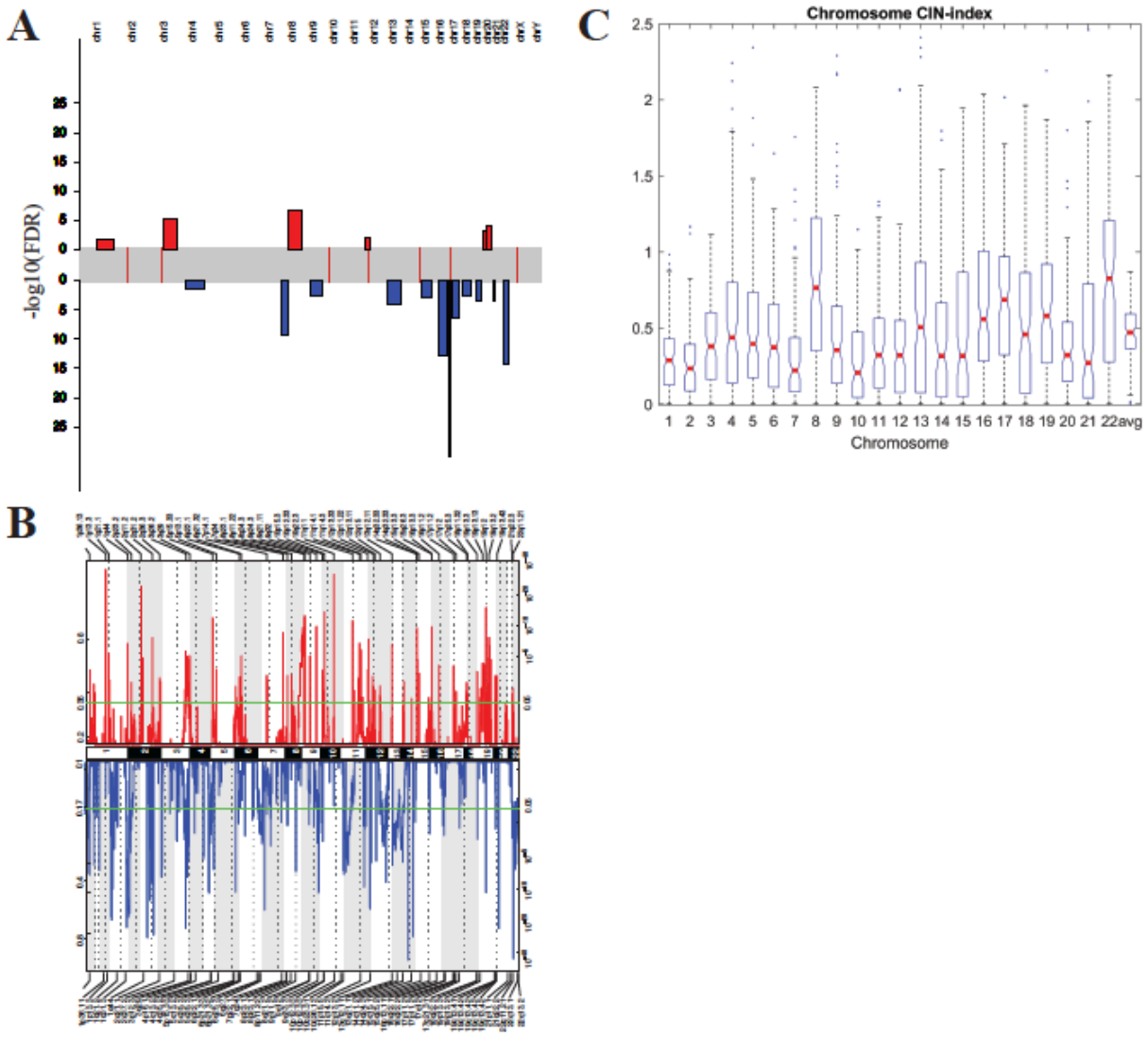


Figure S2. Somatic copy number aberrations of prospective HGSC samples. Related to Figures 1 and 2. (A) Arm-level events. Significant (FDR <0.05) arm-level events were identified by GISTIC. Red and blue bars represent amplifications and deletions respectively. **(B)** Focal events. GISTIC analysis of somatic copy number alterations identifies significantly recurring regions in the genome with focal copy number amplifications (red) and deletions (blue) across the tumor cohort. The x-axis represents the chromosomal locations with the dashed vertical lines indicating the centromeres. The y-axis shows the G-scores (left) signifying the magnitude and frequency of the aberrations, and the corresponding q-values (right) on a log scale with the dashed green line representing a q-value threshold of 0.05. The focal peaks are annotated at the top and bottom of the plot. **(C)** Box-plots of distributions of CIN index in tumor samples from

the current cohort calculated for each individual chromosome and in genome-wide average (the last boxplot). Chromosomes 22, 8, and 17 show the highest levels of median CIN index values.

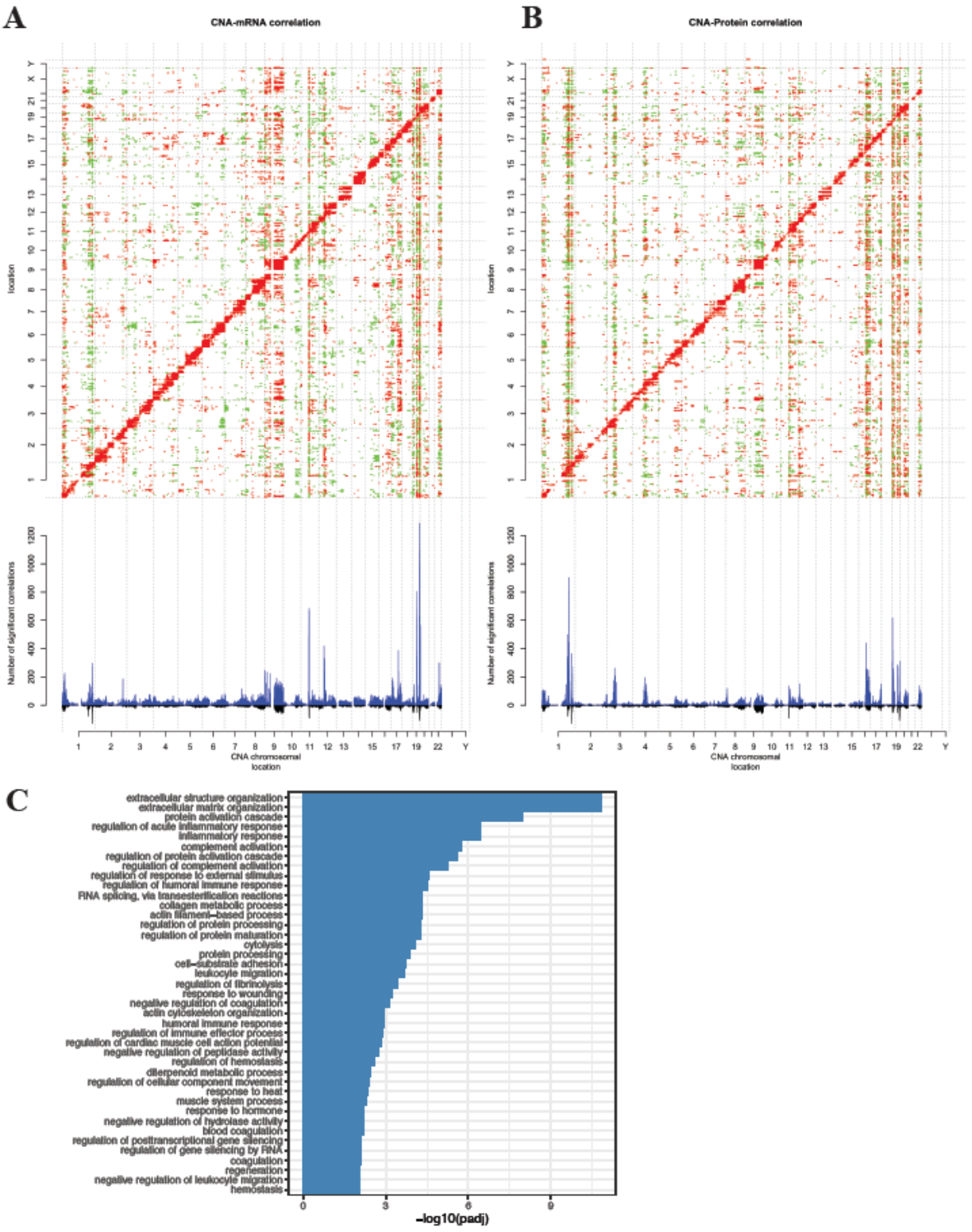


Figure S3. Proteogenomic characterization of prospective HGSC samples. Related to STAR Methods. The upper panels show the correlation of copy number abundance (x axes)

with RNA **(A)** and protein **(B)** levels (y axes). Genes are ordered by chromosomal locations on both axes. Significant correlations (Benjamini-Hochberg adjusted p-value <0.01, Spearman correlation coefficient) are indicated with positive correlations in red and negative correlations in green. Lower panels show the number of mRNAs **(A)** and proteins **(B)** significantly correlated with each CNA gene. Blue bars show the correlations specific to mRNA or protein and black bars represent the correlations common to both mRNA and protein. **(C)** Functional enrichment analysis of trans-affected proteins**.** Enriched GO biological processes associated with the set of proteins significantly correlated with the top trans-acting CNA hotspots on chromosomes 1, 16 and 19, i.e. the peaks identified in bottom panel in B representing aberrant loci significantly correlated with a large number of proteins, are shown. Log-transformed, FDR-corrected p-values are shown for the GO terms found to be most significantly enriched in the trans-affected proteins.

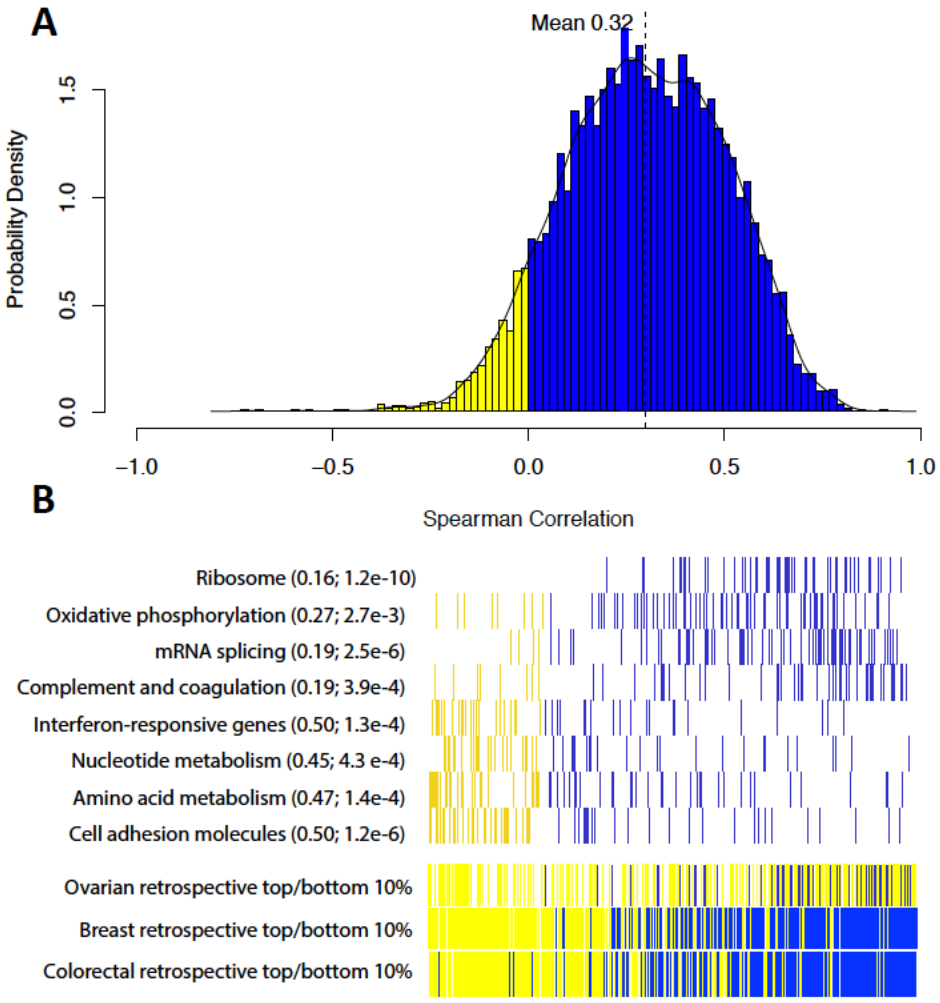


Figure S4. mRNA-protein correlation. Related to STAR Methods. (A) Distribution of the correlation of mRNA to cognate protein across all tumor samples using Pearson correlation. **(B)** Functional enrichment using gene set enrichment analysis shows a typical distribution of functional pathways enriched in the most highly correlated and least highly correlated fractions.

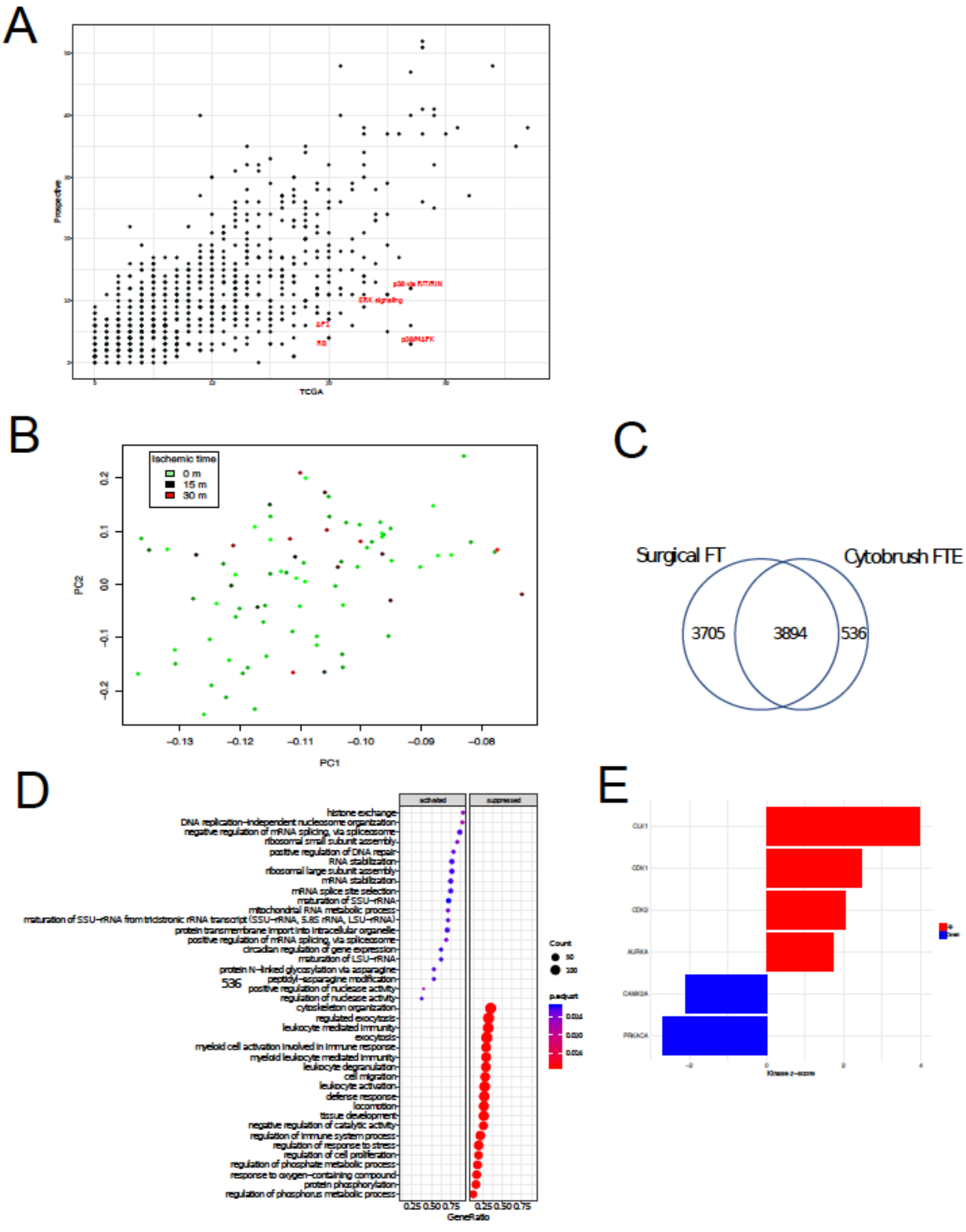


Figure S5. Prospective samples allow comparison with normal tissue, improved ischemia control. Related to Figures 3 and 4. (A) Pathway activation comparison of the current cohort

against the retrospective cohort. Scatterplot of the number of patients for which a pathway was found to be significantly enriched (Kolmogrov-Smirnov p-value < 0.05) in the retrospective (x axis) cohort against the prospective (y axis) cohort. (**B**) PCA of the phosphoproteome colored by ischemic time shows no separation. **(C)** Overlap of proteins between surgical FT and cytobrush FTE samples. **(D)** Gene-set enrichment analysis of proteins observed only in the cytobrush FTE preparations. **(E)** Kinase substrate enrichment analysis results after filtering out phosphosites that are known to be involved in warm ischemia.

Supplemental Tables

Table S1. Patient cohort clinical data. Related to STAR Methods.

Table S2. Proteomic and phosphoproteomic data. Related to STAR Methods.

Table S3. RNA-seq data. Related to STAR Methods.

Table S4. GISTIC analysis of arm-level and focal alterations. Related to STAR Methods.

Table S5. Proteome and phosphoproteome tumor vs normal differential analysis. Related to Figures 3 and 4.

Table S6. Tumor-vs-normal proteome gene-set enrichment analysis. Related to Figure 3.

Table S7. Phosphoproteome kinase substrate enrichment analysis. Related to Figure 4.

Table S8. Differential phosphosites known to be involved in kinase activity. Related to Figure 6.

Table S9. Analysis of covariates potentially affecting survival. Related to Figure 7.