Figure S1. Sample and variant quality control procedures, related to Figure 1

(A) Coverage distribution over WES target regions for the entire CPTAC cohort (*n* = 1,064). WES samples with coverage ≥20× were used for germline variant calling.

(B) Average WES coverage of 160 CPGs for the entire cohort. Data are represented as average coverage ± 1 standard deviation (mean ± SD).

(C) Number of exonic germline variants detected in normal samples from CPTAC WES data. Individuals are represented by dots, which are colored according to their genetic ancestry as predicted by our pipelines. Average number of variants per cancer type ± one standard deviation is also shown (mean ± SD).

(D) Boxplots representing the concordance of whole-exome-based variant calls with dbSnP (release 151), showing >99% concordance. Overall concordance for the entire cohort was 97.43%, with a TiTv ratio of 2.74.

(E) Overlap rate between variants called from WES and WGS datasets for seven cancer types.

(F) Principal-component analysis (PCA) plots showing WES and WGS-based genetic ancestry predictions for the CPTAC cohort. Ancestry predictions were obtained from WES data using a random forest classifier for all 1,064 individuals (STAR Methods). The 9 individuals of Slavic origin in the GBM, HNSCC, LSCC, PDAC, and UCEC cohorts which were misclassified as AMR in the WES-based predictions, but correctly classified as EUR in the WGS-based predictions are labeled (see STAR Methods).

(G) Histograms depicting the peptide-length distribution of both reference (top) and alternative (bottom) peptides detected in the proteome (left), phosphoproteome (middle), and acetylome (right).

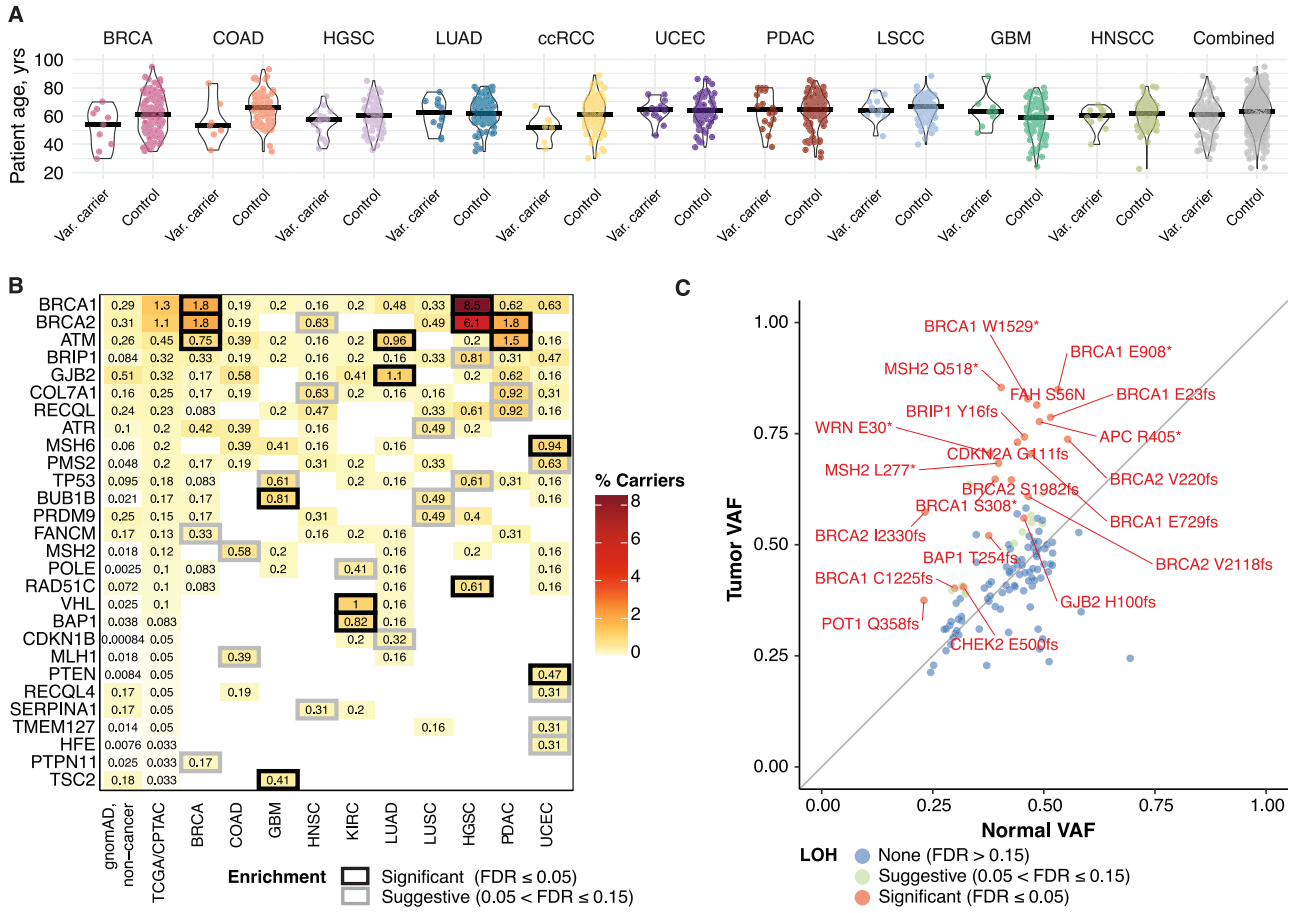


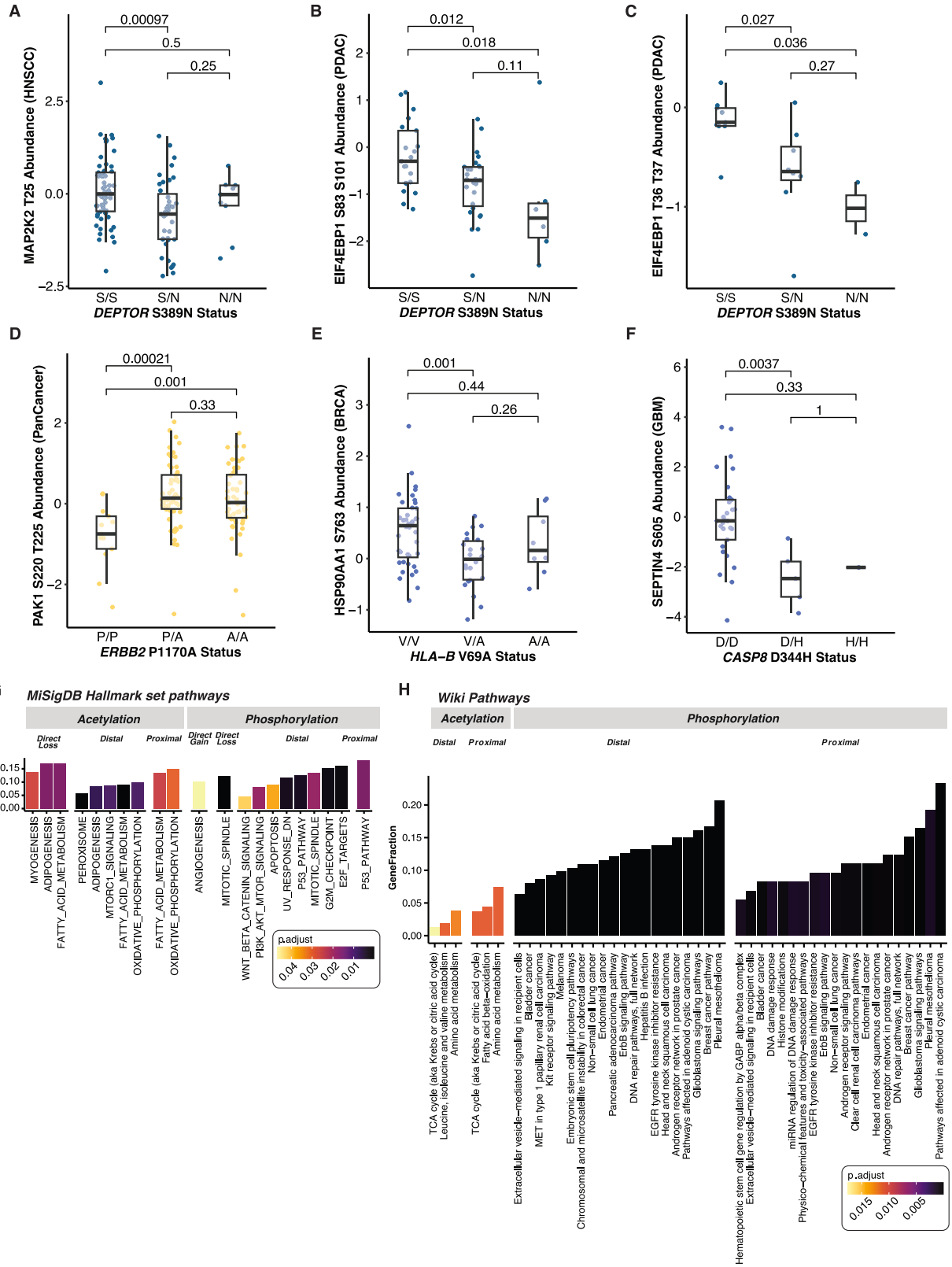
Figure S2. Impact of pathogenic rare variants, related to Figure 2

(A) Violin plots showing age at diagnosis distributions in carriers and non-carriers of rare pathogenic (P) and likely pathogenic (LP) germline variants across 10 cancer types.

(B) Heatmap showing fraction of samples that are carriers of P/LP rare variants across 10 cancer types in the combined sample set of CPTAC and TCGA cohorts. Cancer gene pairs with significant (FDR ≤ 0.05) and suggestive (FDR ≤ 0.15) enrichment with P/LP variants are indicated with black and gray outline respectively.

(legend continued on next page)

(C) Plot showing comparison of variant allele frequencies (VAFs) for P/LP in tumor and normal samples, highlighting variants undergoing LOH in the tumor. Each dot corresponds to one variant, with the diagonal line indicating equal tumor and normal VAFs (i.e., neutral selection). Green indicates suggestive LOH (FDR $\leq$ 0.15); red is significant LOH (FDR $\leq$ 0.05); and blue indicates events not statistically significant.
(D) Plot showing protein expression quantiles in NAT (x axis) and tumor (y axis) of the proteins from germline P/LP variant carriers.



(legend on next page)

Figure S3. Enrichment of biological pathways in genes harboring variants direct, proximal, and distal to PTM sites, related to Figure 3

(A–F) Significant effects of proximal and distal PTM variants on the phosphorylation levels of other proteins belonging to the same signaling pathway. Values above correspond to FDR from pairwise Wilcoxon tests between carrier groups.

(A–C) Boxplots showing phosphorylation abundance differences in members of the KEGG mTOR pathway (hsa04150) between carriers and non-carriers of the *DEPTOR* S389N germline variant in the (A) HNSCC and (B and C) PDAC cohorts.

(D) Boxplot showing pan-cancer phosphorylation level differences at the PAK1 S220/T225 sites between carriers and non-carriers of the *ERBB2* P1170A germline variant.

(E) Boxplot showing phosphorylation level differences at the HSP90AA1 S773 site between carriers and non-carriers of the *HLA-B* V69A germline variant in the BRCA cohort.

(F) Boxplot showing phosphorylation level differences at the SEPTIN4 S605 site between carriers and non-carriers of the *CASP8* D344H germline variant in the GBM cohort.

(G) Enrichment of biological pathways obtained from the MiSigDB Hallmark set.

(H) Enrichment of biological pathways obtained from WikiPathways.

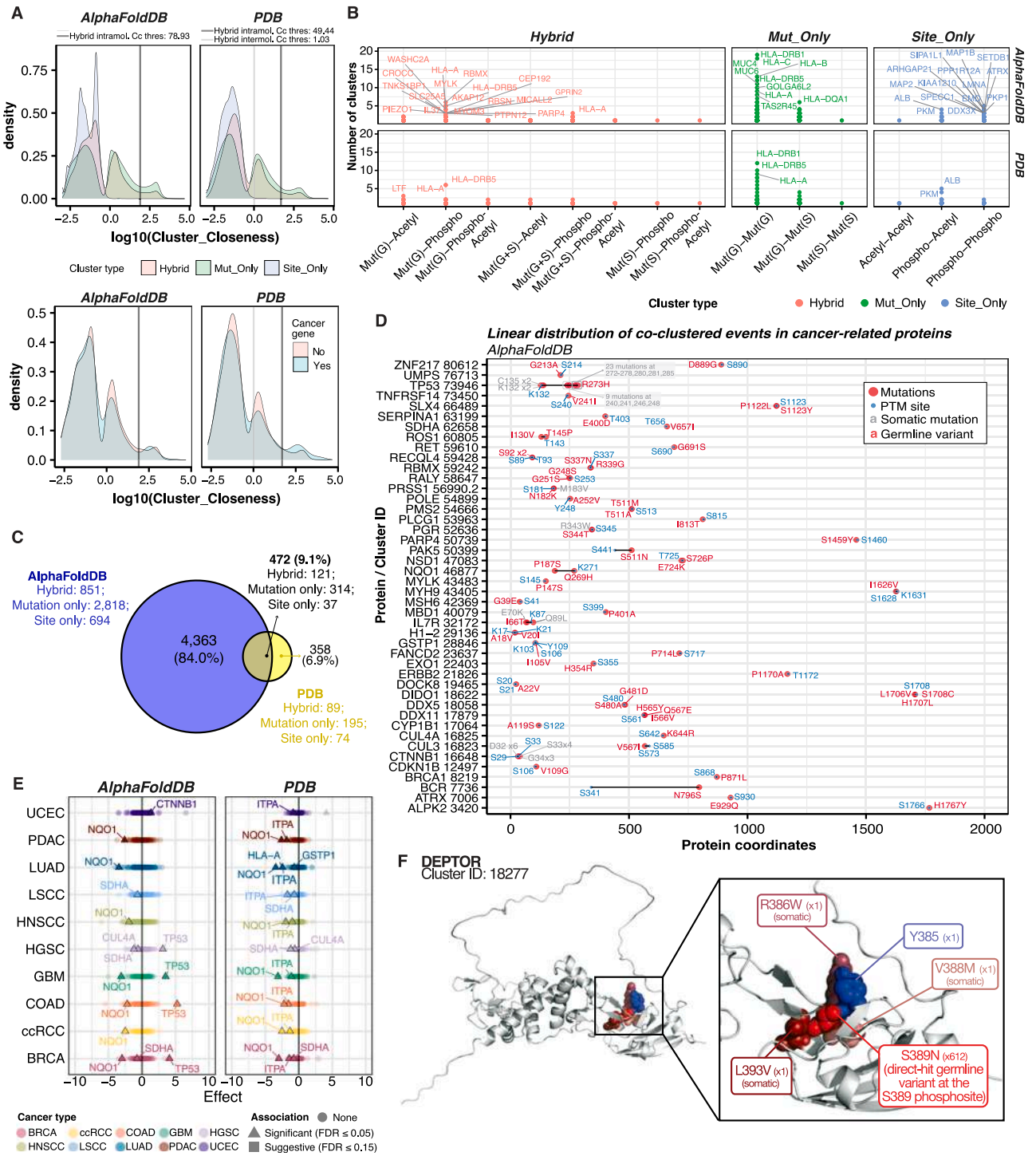


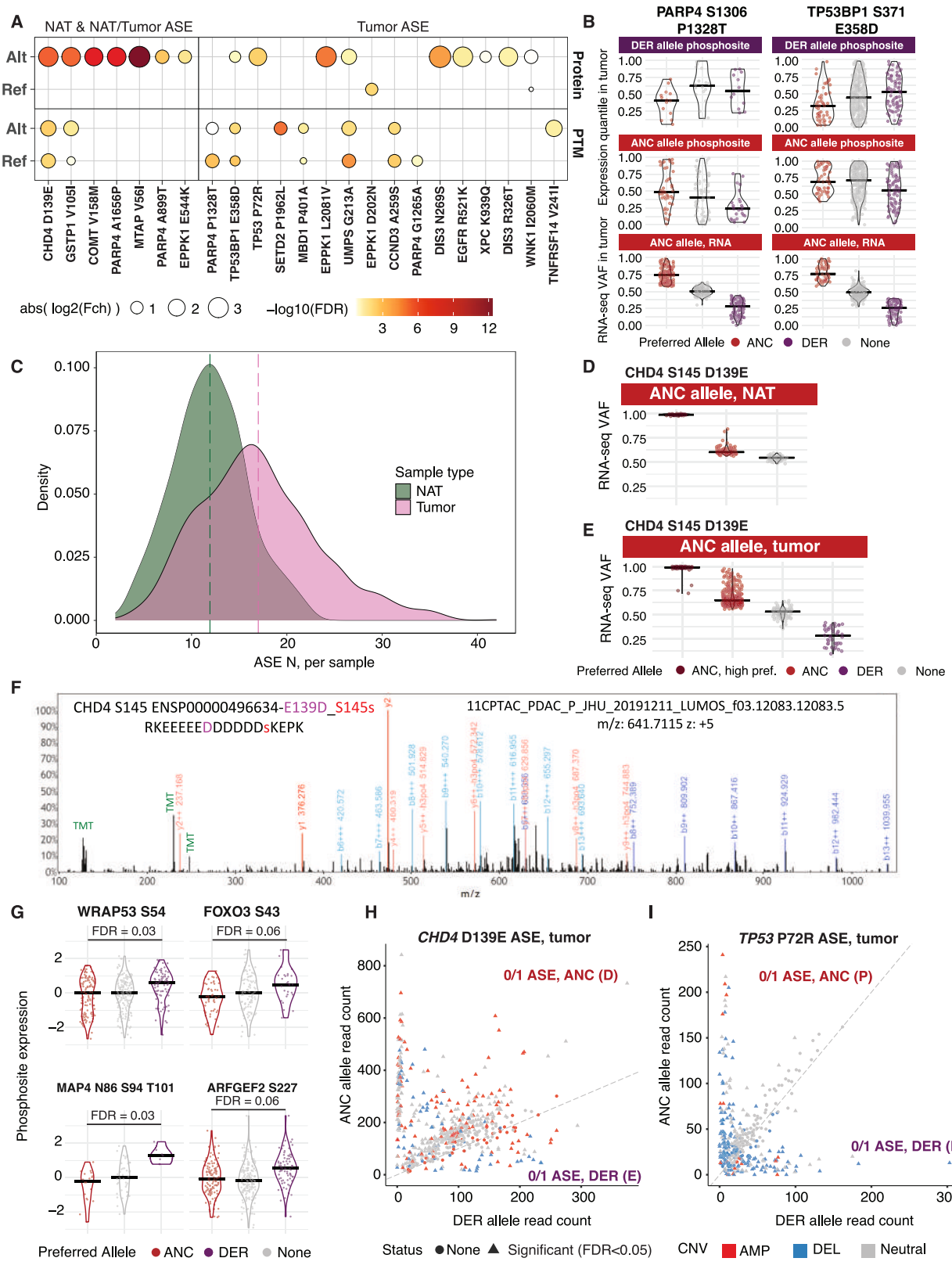
Figure S4. Spatial clustering analyses using HotPho, related to Figure 4

(A) (Top) The cluster closeness (Cc) score density distribution for hybrid, mutation-only, and site-only clusters detected with HotPho using structures from both AlphaFoldDB (AFDB) and RCSB PDB databases. The dark gray vertical line indicates the top 5% Cc threshold used to select intramolecular clusters, while the light gray vertical line indicates the top 20% Cc threshold used to select intermolecular clusters (only detected in PDB). (Bottom) The Cc density distribution for cancer-related genes (unique set from 160 CPGs, 299 driver genes, and 624 other cancer-related genes) vs. other genes.

(B) Number of hybrid, mutation-only, and site-only clusters detected using HotPho from both AFDB and PDB structures. Proteins with the highest number of clusters in each category are labeled. Cluster types are named based on the events they involve, with Mut (G) indicating the presence of a germline variant, Mut (S) indicating a somatic mutation, and Mut (G + S) indicating the presence of both.

(legend continued on next page)

(C) Venn diagram showing the number of clusters detected in both AFDB and PDB, as well as clusters unique to each dataset.
(D) Germline variants, somatic mutations, and PTM sites on the linear protein coordinate for the clusters with the highest Cc score in cancer-related proteins (excluding HLA-A and HLA-B due to the large number of mutations in these clusters that would be difficult to visualize). Germline variants are highlighted in red, while somatic mutations are depicted in gray and smaller font size for reference.
(E) Plot showing proteins with significantly lower or higher levels associated with co-clustered mutations in hybrid clusters in each cancer type. Dots represent a gene-cancer association, colors represent the different cancer types, and the different shapes indicate different levels of significance. Effect is calculated as the slope in the regression model.
(F) DEPTOR protein structure (AFDB: AF-Q8TB45-F1) depicting cluster 18,277 involving a putative PTM loss germline variant (S389N). Blue toned spheres represent phospho sites; red toned spheres represent somatic mutations and germline variants. Variant and site annotations are depicted, followed by the number of individuals with events at that position (x1, for example, to note one individual).



(legend on next page)

Figure S5. Proteogenomic characterization of ASE effects,related to Figure 5

(A) Dotplot showing ASE events with significant and suggestive impact (FDR < 0.2) on protein and/or PTM levels of the respective genes. The size corresponds to the absolute value of $\log_2$(fold change) between samples with preference for the ANC allele vs. samples with preference for the DER allele. The color corresponds to the FDR significance level of the observed differences between the two groups of samples. The comparisons were performed separately for phosphosites containing REF or ALT alleles, peptides containing ALT allele and proteins with the REF allele.

(B) Violin plots showing DER and ANC allele phosphopeptides expressions (first and second rows) and RNA-seq VAF of the ANC allele in tumor (third row) for the PARP4 S1306 P1328T (left) and TP53BP1 S371 E358D (right) variants.

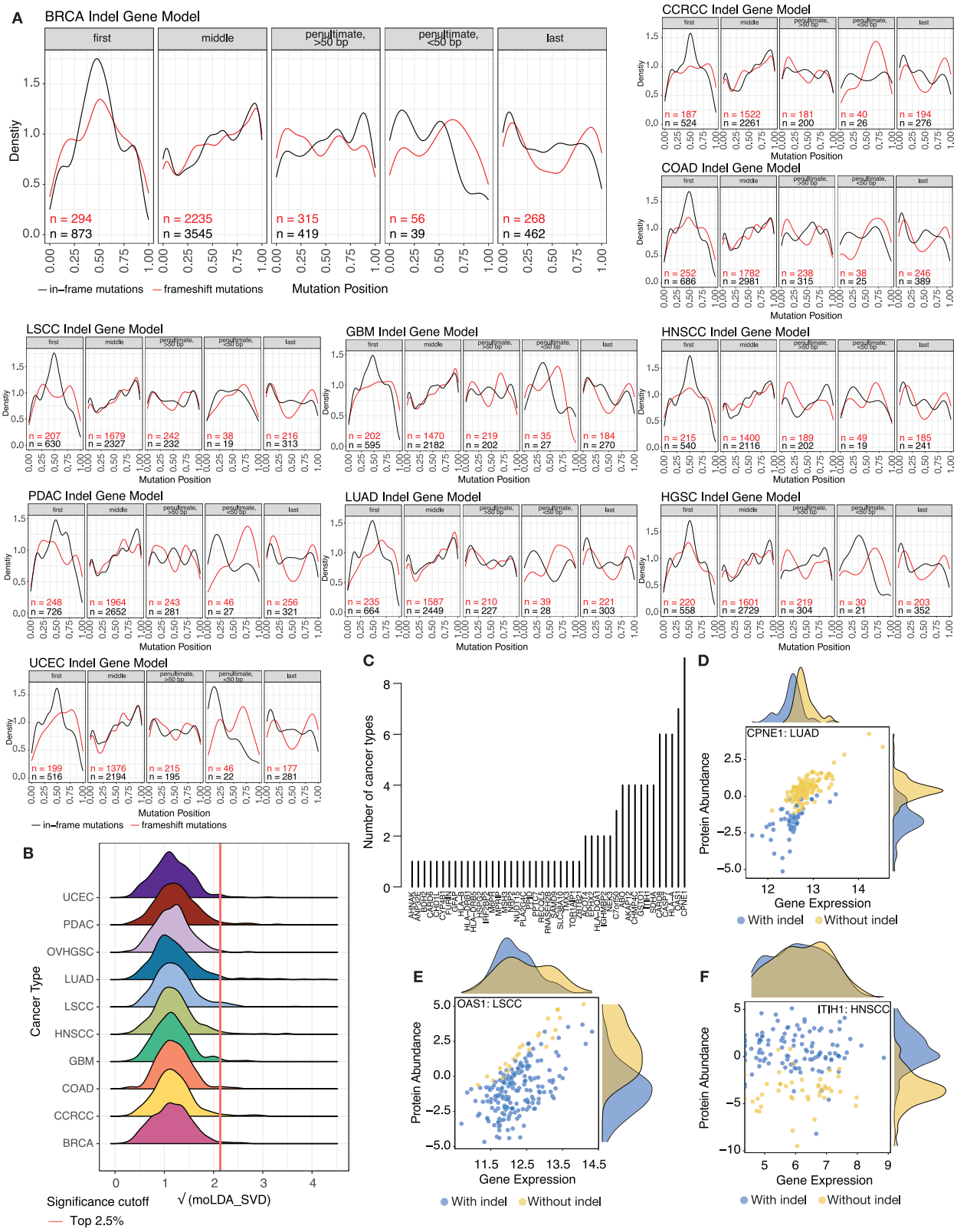
(C) Density plots showing the distribution of the number of ASEs per sample detected in tumor and NAT samples. Dashed lines correspond to the mean number of ASEs per sample in these two groups.

(D and E) Violin plots showing RNA-seq VAF of the ANC allele across three groups of HET samples in NAT (D) and tumor (E): with preference for the ANC or DER allele, or with no preference for either of the alleles.

(F) MS/MS spectrum of CHD4 phosphopeptide RKEEEEEDDDDDDsKEPK phosphorylated at S145 containing the germline variant E139D (ENSP00000496634-E139D_S145s), best scoring representative PSM of phosphosite detected in 10/25 PDAC TMT plexes.

(G) Violin plots showing the distribution of phosphosite expression across three groups of HET samples for the *TP53* P72R variant: with preference for the ANC allele (P72), preference for DER allele (R72), and without preference for either of those alleles. ASE events were detected using a two-sided binomial test with a null probablity of success 0.5 in a Bernoulli experiment. FDR adjusted *p* values are provided.

(H and I) Scatterplots showing read counts for ANC and DER allele for *CHD4* D139E (H), and *TP53* P72R (I) variants. Each dot corresponds to an individual sample, and color denotes if the sample has amplification (AMP) or deletion (DEL) of *CHD4* (H) or *TP53* (I).



(legend on next page)

Figure S6. EJ models for indels and moLDA summary, related to Figure 6

(A) Similar to the main Figure 6, EJ models are shown for each cancer type highlighting the first, middle, and last exons. Two penultimate regions within 50 bps of the last exon and greater than 50 bps away from the last exon are also displayed.

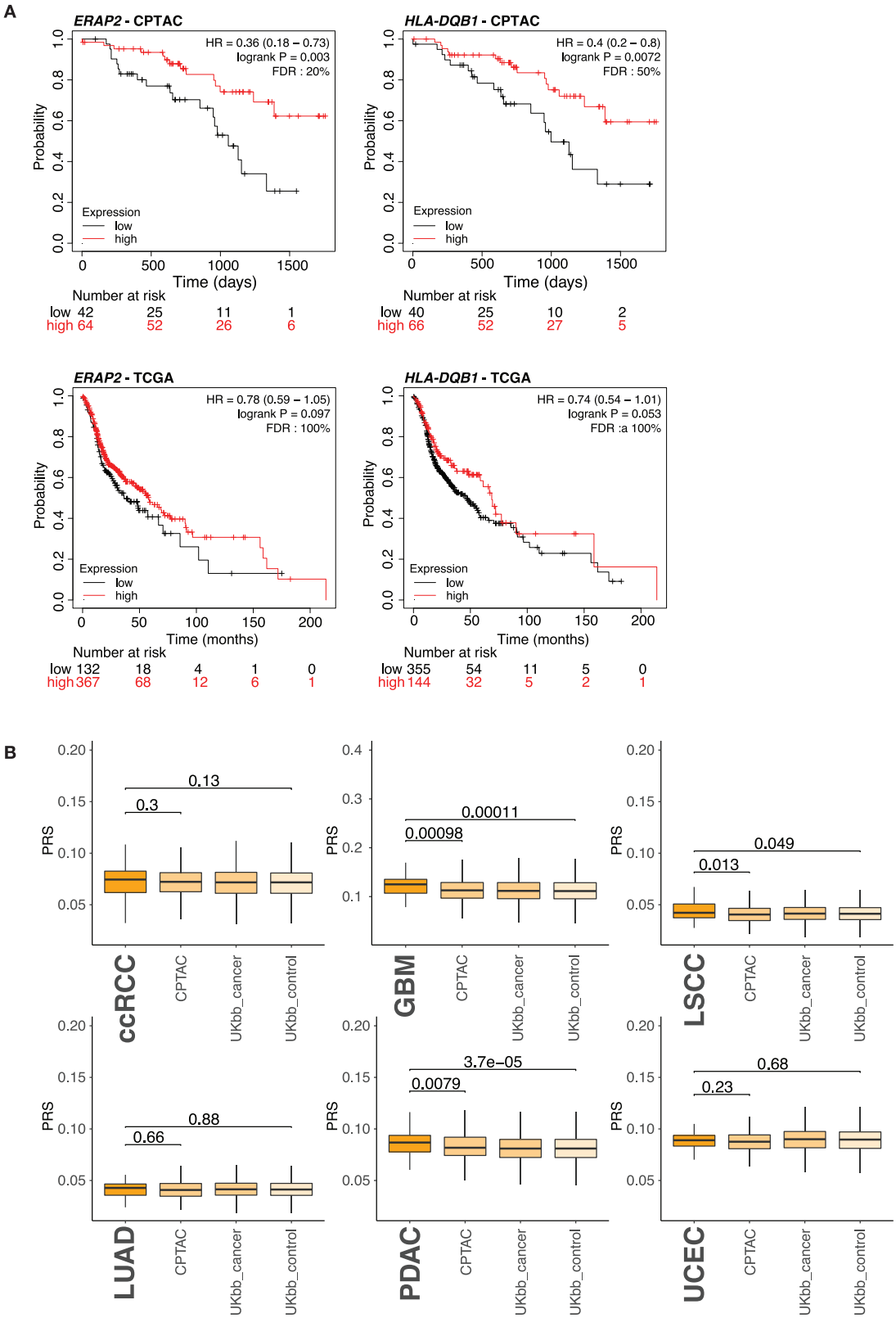
(B) Square-root normalized distribution of moLDA SVD scores for each cancer type. Vertical threshold is set at the PanCancer top 2.5% of scores.

(C) Genes that exceed the SVD threshold that fall into multiple cancer types.

(D) moLDA result for CPNE1 in LUAD. Log normalized RNA-seq gene expression on the x axis and protein abundance on the y axis. Each point represents a tumor from that cancer type.

(E) Same as (D) but for the *OAS1* gene in LSCC.

(F) Same as (D) but for *ITIH1* gene in HNSCC.



(legend on next page)

Figure S7. Survival plots and PRS distributions, related to Figure 7

(A) Kaplan-Meier survival plots based on ERAP2 and HLA-DQB1 expression and overall survival in CPTAC and TCGA HNSCC cohorts.

(B) Polygenic risk score (PRS) distributions. For 6 cancer types, we calculated the PRS in the CPTAC individuals using common risk variants discovered by the largest GWAS available in each specific cancer type. These values were compared with the distributions of PRSs in three other groups, namely (1) CPTAC individuals for the remaining cancer types (“CPTAC”), (2) UKBB individuals diagnosed with any cancer type (“Ukbb_cancer”, and (3) rest of UKBB individuals (“Ukbb_controls”). The *p* values for statistical significance for the comparisons against CPTAC and Ukbb_controls, respectively, are provided for each cancer type (t test).